JMIR Serious Games

Journal Impact Factor (JIF) (2023): 3.8
Volume 13 (2025) ISSN 2291-9279 Editor-in-Chief: Gunther Eysenbach, MD, MPH

Contents

Original Papers

Serious Games Integrating Perceptual Learning and Stereopsis Training in Children With Amblyopia: Single-Arm Pre-Post Feasibility Study (e77402)
Yiwei Mo, Peng Chen, Manting Hou, Fang Cui, Chao Luo, Chao Deng, Jinkang Lin, Qi Wu. 13

Development of a Serious Game App (Digimenz) for Patients With Dementia: Prospective Pilot Study for Usability Testing in Inpatient Treatment and Long-Term Care (e69812)
Sören Brähmer, Benjamin Iffland, Stefan Kreisel, Martin Driessen, Eva Trompetter, Meret Schomburg, Max Toepper, Carolin Steuwe. 34

Age Differences in Flow Experience During Body Movement–Controlled Video Game Rehabilitation Tasks: Cross-Sectional Study (e76278)
Shiyi Tong, Lixuan Li, Yuankun Zhang, Weiyan Ren, Fang Pu. 51

Comparing Caregiver Perceptions of a Social Robot and Tablet for Serious Game Delivery in Dementia Care: Cross-Sectional Comparison Study (e76209)
Dorothy Bai, Kelvin Tan Cheng Kian, Po-Yin Chen, Yeh-Liang Hsu, Gong-Hong Lin. 61

Virtual Reality for Analgesia During Intrauterine Device Insertion: Randomized Controlled Trial (e72917)
Chloe Higgins, Claudia Zecena Morales, Judi Hocking, Kate Tyson, Cheryl Leung, Luke Larmour, Paul Leong, Beverley Vollenhoven. 73

Exploring the Immediate and Long-Term Effects of Immersive Virtual Reality on Behavioral and Psychological Symptoms of Dementia and Caregiver Burden: Longitudinal Observational Study (e73044)
Ling-Chun Huang, Ching-Fang Chien, Yuan-Han Yang. 83

A Virtual Reality Force Control Training System on Brain Activation: Functional Near-Infrared Spectroscopy (fNIRS) Study (e63874)
Luigi Gan, Chien-Ju Lin, Hsiao-Feng Chieh, Kai-Nan An, Fong-Chin Su. 95

Adolescents' and Parents' Perspectives on Using the MedSMARxT Families Intervention in Emergency Departments for Opioid Medication Safety Education: Mixed Methods Study (e68814)
Olufunmilola Abraham, Sara Nadi, Irene Hurst. 116

Factors Influencing Virtual Reality as a Distraction Tool for Venipuncture in Children: Observational Pilot Feasibility Study (e66656)
Chris Worth, Leyi Yang, Catherine Fullwood, Indraneel Banerjee. 131

Active Video Games Training for Older Adults: Comparative Study of User Experience, Workload, Pleasure, and Intensity (e67314)
Néva Béraud-Peigné, Alexandra Perrot, Pauline Maillot. 137

XSL•FO
RenderX

Physical Activity and Enjoyment in Active Virtual Reality Games in Youth: Comparative Analysis of Gorilla Tag and Beat Saber (e66593)
Brenden Boots, Daniel Berg, Easton Hewitt, Keith Naugle, Kelly Naugle. . . . 151

Comparative Efficacy of Video Games Versus Midazolam in Reducing Perioperative Anxiety in Pediatric Patients: Systematic Review and Meta-Analysis (e67007)
Ziyue Luo, Sisi Deng, Ruihao Zhou, Ling Ye, Tao Zhu, Guo Chen. . . . 164

Acceptability of and Willingness to Use Virtual Reality Exergames for Weight Loss Among Young Adults With Overweight or Obesity in China: Qualitative Study (e66998)
Yanya Chen, Bingsheng Guan, Yaqi Zhang, Suen Lee, Jia-yu Liu, Sicun Li, Ming Liu, Xiaoshen Zhang, Wai-kit Ming. . . . 177

Personal Protection Equipment Training as a Virtual Reality Game in Immersive Environments: Development Study and Pilot Randomized Controlled Trial (e69021)
Liang Zhou, Haoyang Liu, Mengjie Fan, Jiahao Liu, Xingyao Yu, Xintian Zhao, Shaoxing Zhang. . . . 187

Impact of a 3-Month Recall Using High-Fidelity Simulation or Screen-Based Simulation on Learning Retention During Neonatal Resuscitation Training for Residents in Anesthesia and Intensive Care: Randomized Controlled Trial (e57057)
Anne-Claire Louvel, Cécile Dopff, Gauthier Loron, Daphne Michelet. . . . 205

Perceptions Toward an Attentional Bias Modification Mobile Game Among Individuals With Low Socioeconomic Status Who Smoke: Qualitative Study (e59515)
Michael Wakeman, Lydia Tesfaye, Gunnar Baskin, Tim Gregory, Greg Gruse, Erin Leahy, Brandon Kendrick, Sherine El-Toukhy. . . . 216

Exergame (ExerG)-Based Physical-Cognitive Training for Rehabilitation in Adults With Motor and Balance Impairments: Usability Study (e66515)
Silvia Herren, Barbara Seebacher, Sarah Mildner, Yanick Riederer, Ulrike Pachmann, Nija Böckler, Stephan Niedecken, Sabrina Sgandurra, Leo Bonati, Isabella Hotz, Alexandra Schättin, Roman Jurt, Christian Brenneis, Katharina Lenfert, Frank Behrendt, Stefan Schmidlin, Lennart Nacke, Corina Schuster-Amft, Anna Martin-Niedecken. . . . 230

Transformative Gamified Binocular Therapy for Unilateral Amblyopia in Young Children: Pilot Prospective Efficacy and Safety Study (e63384)
Wenqing Zhu, Shuneng Gu, Jian Li, Jin Lin, Chanling Hu, Rui Liu. . . . 251

User Experience of a Semi-Immersive Musical Serious Game to Stimulate Cognitive Functions in Hospitalized Older Patients: Questionnaire Study (e57030)
Laurent Samson, Lena Carcreff, Frédéric Noublanche, Sophie Noublanche, Hélène Vermersch-Leiber, Cédric Annweiler. . . . 265

Effectiveness of a Virtual Reality Serious Video Game (The Secret Trail of Moon) for Emotional Regulation in Children With Attention-Deficit/Hyperactivity Disorder: Randomized Clinical Trial (e59124)
Marina Martin-Moratinos, Marcos Bella-Fernández, María Rodrigo-Yanguas, Carlos González-Tardón, Chao Li, Ping Wang, Ana Royuela, Pilar Lopez-Garcia, Hilario Blasco-Fontecilla. . . . 278

Gamified Web-Delivered Attentional Bias Modification Training for Adults With Chronic Pain: Randomized, Double-Blind, Placebo-Controlled Trial (e50635)
Julie Vermeir, Melanie White, Daniel Johnson, Geert Crombez, Dimitri Van Ryckeghem. . . . 301

Enhancing Immersion in Virtual Reality–Based Advanced Life Support Training: Randomized Controlled Trial (e68272)
Dilek Kitapcioglu, Mehmet Aksoy, Arun Ozkan, Tuba Usseli, Dilan Cabuk Colak, Tugrul Torun. . . . 327

Diaphragmatic Breathing Interfaces to Promote Relaxation for Mitigating Insomnia: Pilot Study (e67000)
Yi-Jen Lai, Hsiao-Yean Chiu, Ko-Chiu Wu, Chun-Wei Chang. . . . 339

XSL•FO
RenderX

Effects and Acceptability of a 1-Week Home-Based Virtual Reality Training for Supporting the Management of Stress and Anxiety: Randomized Pilot Trial (e50326)
Federica Pallavicini, Eleonora Orena, Lisa Arnoldi, Federica Achille, Stefano Stefanini, Maddalena Cassa, Alessandro Pepe, Guido Veronese, Luca Bernardelli, Francesca Sforza, Sara Fascendini, Carlo Defanti, Marco Gemma, Massimo Clerici, Giuseppe Riva, Fabrizia Mantovani. . . 3 5 8

Domain-Specific Evaluation of Exergame Metrics Among Older Adults With Mild Neurocognitive Disorder: Secondary Analysis of 2 Randomized Controlled Trials (e65878)
Wanda Kaiser, Eling de Bruin, Patrick Manser. . . . 376

The Positive Effect of Video-Game Play on College Students' Anxiety and Depression Symptoms During the COVID-19 Pandemic Shelter-in-Place Lockdowns: Mixed Methods Study (e58857)
Fengbin Hu, Zixue Tai, Jianping Liu. . . . 401

Nature Engagement Outcomes of Viewing Nature Through a 360° Video or a Tablet Screen: Randomized Trial (e63424)
Elena Brambilla, Karen Stendal, Vibeke Sundling, Giovanna Calogiuri. . . . 419

Effects of The Legend of Zelda: Breath of the Wild and Studio Ghibli Films on Young People's Sense of Exploration, Calm, Mastery and Skill, Purpose and Meaning, and Overall Happiness in Life: Exploratory Randomized Controlled Study (e76522)
Annisa Arigayota, Barbara Duffek, Congcong Hou, Andreas Eisingerich. . . . 436

Effects of a Board Game on Tic Management and Psychosocial Functioning in Adolescents With Tourette Syndrome: Randomized Controlled Trial (e76208)
Mei Lee, Huei Wang, Chung Lien, Zhi Chen. . . . 455

Virtual Reality–Based Assessment of Attention-Deficit/Hyperactivity Disorder and Comorbid Symptoms in Children: Framework Development and Standardization Study (e69146)
Harim Jeong, Minjoo Kang, Kennet Sorenson, Jacob Moore, Robert Blair, Ellen Leibenluft, Jeffrey Newcorn, Beth Krone, Singi Jeong, Donghee Kim, Soonjo Hwang. . . . 470

Usability Evaluation of a Virtual Reality Multisensory Sham-Feeding Device for Patients Undergoing Fasting Periods for Colorectal Cancer Surgery: Mixed Methods Study (e75641)
Xingzhu Yuan, Peiling Ye, Xinyue Liu, Yuehan Hu, Qin Xu, Wenyi Zhao, Ka Li. . . . 482

Performance and Perceptions of Health Care Professionals Using an Immersive Virtual Reality Tool for Home Care Training: Observational Feasibility and Acceptability Study (e75104)
José Mira, Javier Rios, Eva Gil-Hernández, Nida Abed, Vanesa Ribeiro-Neves, Clara Pérez-Esteve, Mercedes Guilabert-Mora, Almudena Arroyo, Maria Ballester. . . . 497

Efficacy of a Mobile Serious Game (SwaziYolo) for Increasing HIV Risk Perception: Randomized Controlled Trial (e70333)
Bhekumusa Lukhele, Mac Delay, Fortunate Shabalala, Mfundi Motsa, Alexander Kay, Christina El Saaidi, Masako Kihara, Masahiro Kihara, Katia Bruxvoort. . . . 512

Super Mario Bros. and Yoshi Games' Affordance of Childlike Wonder and Reduced Burnout Risk in Young Adults: In-Depth Mixed Methods Cross-Sectional Study (e84219)
Winze Tam, Congcong Hou, Andreas Eisingerich. . . . 527

Enhancing Equity in Schoolchildren's Basic Life Support Education in Brazil Through Serious Games: Cohort Study (e69252)
Uri Flato, Adriana Flato, Isabella Martins, Giuliana Simoes Nakano, Julia Romao, Manuela Nakano, Emilio dos Santos, Yuuki Villas Boas, Leonardo Medeiros, Vinicius Rossignoli, Gabriel Cunha, Rafael Batista, Pedro da Silva, Miguel Florentino, Amanda Rabelo, Thais Midega, Rogerio Passos, Helio Guimaraes, Karl Kern. . . . 545

XSL•FO
RenderX

A Serious Game (Health Unit in Focus) for Enhancing Undergraduate Education on Older Adults' Health: Design and Validation Study (e66289)
Yuri Barbalho, Calliandra Silva, Carla Fernandes, Raíza Trombini, Patrícia Melo, Aline Oliveira, Alayne Pereira, Alessandro Silva, Luciano Lima, Marina Stival, Diana Pinho, Silvana Funghetto. . . . 558

Effectiveness of a Web-Based Virtual Simulation to Train Nursing Students in Suicide Risk Assessment: Randomized Controlled Investigation (e69347)
Paul Roux, Yujiro Okuya, Cristina Morel, Mariane Soulès, Hugo Bottemanne, Eric Brunet-Gouet, Solène Frileux, Christine Passerieux, Nadia Younes, Jean Martin. . . . 570

Exploring the Impact of Home-Based Serious Smartphone Resuscitation Gaming on Stress Among Nursing Students Practicing Simulated Adult Basic Life Support: Randomized Waitlist Controlled Trial (e67623)
Nino Fija ko, Benjamin Abella, Špela Metli ar, Leon Kopitar, Robert Greif, Gregor Štiglic, Pavel Skok, Matej Strnad. . . . 590

Virtual Reality for Workplace Violence Training of Health Care Workers: Pilot Mixed Methods Usability Study (e70817)
Daniel Jackson, Thipkanok Wongphothiphan, John Luna, Tensing Maa, Mary Fristad, Yungui Huang, Brittany Schaffner, Jennifer Reese, Jason Wheeler, Brandon Abbott, Emre Sezgin. . . . 607

Effectiveness of Augmented Reality in the Teaching of Health University Students: Quasi-Experimental Study (e54312)
Rocío Martín-Valero, Alejandro Vega-Morales Sr, Francisco Martín-Vega, Manuel Rodriguez-Huguet, Maria Rodríguez-Martínez, Maria Vinolo-Gil. . . . 623

A Mixed Reality–Based Telesupervised Ultrasound Education Platform on 5G Network Compared to Direct Supervision: Prospective Randomized Pilot Trial (e63448)
Minha Kim, Meong Son, Suhyeon Moon, Won Cha, Ik Jo, Hee Yoon. . . . 635

Evaluating the Efficacy of a Serious Game to Deliver Health Education About Invasive Meningococcal Disease: Clustered Randomized Controlled Equivalence Trial (e60755)
Lauren Bloomfield, Julie Boston, Martin Masek, Lesley Andrew, Donna Barwood, Amanda Devine. . . . 649

Exploring Nursing Students' Experiences of Empathy and User Experiences in an Immersive Virtual Reality Simulation Game: Cross-Sectional Study (e62688)
Jaana-Maija Koivisto, Sanna Kämäräinen, Katri Mattsson, Satu Jumisko-Pyykkö, Riikka Ikonen, Elina Haavisto. . . . 664

Immersive Virtual Reality and AI (Generative Pretrained Transformer) to Enhance Student Preparedness for Objective Structured Clinical Examinations: Mixed Methods Study (e69428)
Shaniff Esmail, Brendan Concannon. . . . 680

Children's Improvement After Language and Rhythm Training With the Digital Medical Device Poppins for Dyslexia: Single-Arm Intervention Study (e76435)
Charline Grossard, Melanie Descamps, Hugues Pellerin, François Vonthron, David Cohen. . . . 700

Empowering Children With Down Syndrome by Enhancing Emergency Preparedness Through Serious Games: Quasi-Experimental Study With a Between-Group Design (e73690)
Samaa Shohieb, Suzan Bakhit, Ensaf Mohammed, Abdelghafar M Elhady. . . . 713

An Integrated Model for Circular Waste Management Using the Internet of Things, Semantic Web, and Gamification (Circonomy): Case Study in Indonesia (e66781)
Vitri Tundjungsari, Bambang Putranto, Muhamad Ulum, Nizirwan Anwar. . . . 733

XSL•FO
RenderX

Bridging Needs and Expectations of Individuals With Physical Disabilities and Community Services Stakeholders for the Cocreation of an Adapted Physical Activity Platform in Virtual Reality: Qualitative Study (e59704)
Aurélien Ramos, Maxence Boisvert, Elodie Traverse, Danielle Levac, Martin Lemay, Marika Demers, Martine Bordeleau, Sarah-Maude Ruest, Roxanne Périnet-Lacroix, Krista Best, Maxime Robert. . . . 750

Feasibility and Usability of an Artificial Intelligence—Powered Gamification Intervention for Enhancing Physical Activity Among College Students: Quasi-Experimental Study (e65498)
Yanan Gao, Jinxi Zhang, Zhonghui He, Zhixiong Zhou. . . . 762

A Serious Game to Study Reduced Field of View in Keyhole Surgery: Development and Experimental Study (e56269)
Phoebe Whitley, Connor Creasey, Matthew Clarkson, Stephen Thompson. . . . 776

Gamified mHealth System for Evaluating Upper Limb Motor Performance in Children: Cross-Sectional Feasibility Study (e57802)
Md Mia, Sheikh Ahamed, Samuel Nemanich. . . . 788

A Novel Gamified Exercise Program Incorporating Stampede Training for Enhancing Functional Fitness, Physical Activity Levels, and Quality of Life in Community-Dwelling Older Adults: Randomized Parallel Exploratory Trial (e73474)
Siao-Fen Huang, Ching-Wen Fan, Diana Nisrina, Tsai-Wei Lin, Wen Huang. . . . 804

FITLIGHT Training and Its Influence on Visual-Motor Reactions and Dribbling Speed in Female Basketball Players: Prospective Evaluation Study (e70519)
Ahmed Hassan. . . . 823

Insights on the User Experience and Feasibility of an Electromyography-Driven Exergame Combined With Blood Flow Restriction for Strength Training in Hospitalized Older Adults: Mixed Methods Randomized Controlled Feasibility Study (e69400)
Ruben Debeuf, Reinhard Claeys, Margo Berlanger, Myrthe Bunt, Aziz Debain, Daan De Vlieger, Matthias Eggermont, Mahyar Firouzi, Stefania Guida, Katarína Kostková, Siddhartha Lieten, Lubos Omelina, Silvia Zaccardi, Bart Jansen, Eva Swinnen, David Beckwée. . . . 837

Effects of a Session of Exergames and Traditional Games on Inhibitory Control in Children With Autism Spectrum Disorder: Randomized Controlled Crossover Trial (e65562)
Juliana Miranda, Rodrigo Browne, Weslley da Silva, João Rodrigues dos Santos, Carmen Campbell, Isabela Ramos. . . . 854

Effect of a Gamified Family-Based Exercise Intervention on Adherence to 24-Hour Movement Behavior Recommendations in Preschool Children: Single-Center Pragmatic Trial (e60185)
Gaizka Legarra-Gorgoñon, Yesenia García-Alonso, Robinson Ramírez-Vélez, Loreto Alonso-Martínez, Mikel Izquierdo, Alicia Alonso-Martínez. 8 6 9

Comparing In-Person, Standard Telehealth, and Remote Musculoskeletal Examination With a Novel Augmented Reality Exercise Game System: Pilot Cross-Sectional Comparison Study (e57443)
Richard Wu, Keerthana Chakka, Sara Belko, Ninad Khargonkar, Kevin Desai, Balakrishnan Prabhakaran, Thiru Annaswamy. . . . 883

Virtual Reality High-Intensity Interval Training Exergaming Compared to Traditional High-Intensity Circuit Training Among Medical Students: Pilot Crossover Study (e63461)
Pietro Merola, Marcos Cardoso, Gabriel Barreto, Matheus Chagas, Luana Farias Oliveira Saunders, Bryan Saunders, Danilo Cortozi Berton. 8 9 4

Tongue Muscle Training App for Middle-Aged and Older Adults Incorporating Flow-Based Gameplay: Design and Feasibility Pilot Study (e53045)
Kuan-Chu Su, Ko-Chiu Wu, Kuei-Ru Chou, Chia-Hsu Huang. . . . 905

XSL•FO
RenderX

The Usability and Effect of a Novel Intelligent Rehabilitation Exergame System on Quality of Life in Frail Older Adults: Prospective Cohort Study (e50669)
Chien-Hsiang Chang, Chun-Chun Wei, Wei-Chih Lien, Tai-Hua Yang, Bo Liu, Yimo Lin, Poh Tan, Yang-Cheng Lin. 919

Playful Antisedentary Interactions for Online Meeting Scenarios: A Research Through Design Approach (e62778)
Jiaqi Jiang, Shanghao Li, Xian Li, Yingxin Xu, Jian Zhao, Pengcheng An. 950

The Experience of and Needs for Exergames in Older Adults With Mild Cognitive Impairment: Qualitative Interview Study (e53631)
Xi Chen, Dian Jiang, Hongting Ning, Lina Wu, Yifei Chen, Chi Zhang, Ruotong Peng, Yishu Zhu, Hui Feng. 970

Preliminary Efficacy of a Gamified Mobile App for Promoting Self-Health Management Among Nurses in the Post-COVID Era: Single-Group Pre-Post Study (e66262)
Shao Hsu, Li Lu, Pei Chou. 985

Active Video Games to Improve Behavioral Intentions and Cognitive Function in Patients With Schizophrenia: Randomized Controlled Trial (e69116)
Huan-Hwa Chen, Ching-Ching Lin, Man-Ling Yu, Hsiu-Lan Wu, Hui-Chu Shen, Hsiu-Fen Hsieh. 1000

Exploring Mixed-Reality Exergames for Sports Rehabilitation: Design Insights and Evaluation Findings (e68431)
Michelle Haas, Larissa Wild, Leander Schneeberger, Eveline Graf, Anna Martin-Niedecken. 1014

Achieving Gameplay Independence in Virtual Reality Exergames for Individuals With Mild Intellectual Disabilities: Pilot Study (e71823)
Julia Ci y ska, Janusz Maciaszek. 1033

Effects of Balance-Based Exergame Training With Variable Difficulty on Balance and Spatiotemporal Gait Outcomes in Adults With Mild Cognitive Impairment: Randomized Controlled Trial (e74092)
Aruba Saeed, Imran Amjad, Imran Niazi, Abdullah Alzahrani, Muhammad Shafiq, Heidi Haavik. 1053

Effects of an Immersive Virtual Reality–Based Exercise Intervention on Psychological and Physiological Outcomes in College Students: Randomized Controlled Trial (e75777)
Wenxi Liu, Daniel McDonough, John Oginni, Jessh Mavoungou, Zan Gao. 1075

Effects of Competitive Exergaming–Based Esports on Older Adults in Hong Kong: Nonrandomized Controlled Pilot Study (e77655)
Ka Leung, Yuchen Shi. 1088

Immersive Virtual Reality–Assisted Therapy for Distressing Voices in Psychosis: Qualitative Study of Participants' and Therapists' Experiences in the Challenge Trial (e77920)
Mads Christensen, Matilde Rydborg, Rikke Jørgensen, Cecilie Nielsen, Jan Mainz, Imogen Bell, Neil Thomas, Lisa Smith, Lise Mariegaard, Thomas Ward, Merete Nordentoft, Louise Glenthøj, Ditte Vernal. 1103

Kangaroo Stimulation Game in Tracheostomized Intensive Care–Related Dysphagia: Interventional Feasibility Study (e60685)
Marjolein Jansen, Ingrid van Iperen, Anke Kroner, Raphael Hemler, Esther Dekker-Holverda, Peter Spronk. 1132

Improvement of Game Users' Depressive Symptoms via Behavioral Activation in a Massive Multiplayer Online Game: Randomized Controlled Trial (e73734)
Kenji Yokotani, Masanori Takano, Nobuhito Abe, Takahiro Kato. 1141

XSL•FO
RenderX

Perceptions of Patients With Stroke Regarding an Immersive Virtual Reality–Based Exercise System for Upper Limb Rehabilitation: Questionnaire and Interview Study (e49847)
Jiayin Chen, Calvin Or, Zhixian Li, Eric Yeung, Tianrong Chen. . . . 1160

Improving Pediatric Patients' Magnetic Resonance Imaging Experience With an In-Bore Solution: Design and Usability Study (e55720)
Annerieke Heuvelink, Privender Saini, Özgür Ta ar, Sanne Nauts. . . . 1180

Testing the Effectiveness of a Gamified Emotional Cognitive Bias Modification Task as an Intervention for Low Mood: Randomized Controlled Trial (e65103)
Rumeysa Kuruo lu, Angela Attwood, Ian Penton-Voak. . . . 1194

Long-Term Usage of Breeze, a Gamified Breathing Training App, and Its Effect on Momentary Relaxation in People With Cancer: Cohort Study (e70297)
Anja Bischof, Tobias Budig, Sonja Schläpfer, Yanick Lukic, Fabian Schneider, Prabhakaran Santhanam, Jürgen Barth, Claudia Witt, Tobias Kowatsch. . . . 1206

Utilizing Gamification, Artificial Intelligence, and mHealth for the Professional Development of Maternal Care Providers: Exploratory Pilot Cross-Sectional Study Assessing Providers' Satisfaction in Primary Health Care Centers in Lebanon (e53735)
Mohamad Alameddine, Nadine Sabra, Nour El Arnaout, Asmaa El Dakdouki, Mahmoud El Jaouni, Randa Hamadeh, Abed Shanaa, Shadi Saleh. . . . 1221

Gamified Adaptive Approach Bias Modification in Individuals With Methamphetamine Use History From Communities in Sichuan: Pilot Randomized Controlled Trial (e56978)
Danlin Shen, Jianping Jiao, Liqun Zhang, Yanru Liu, Xiang Liu, Yuanhui Li, Tianjiao Zhang, Dai Li, Wei Hao. . . . 1238

Considering Theory-Based Gamification in the Co-Design and Development of a Virtual Reality Cognitive Remediation Intervention for Depression (bWell-D): Mixed Methods Study (e59514)
Mark Hewko, Vincent Gagnon Shaigetz, Michael Smith, Elicia Kohlenberg, Pooria Ahmadi, Maria Hernandez Hernandez, Catherine Proulx, Anne Cabral, Melanie Segado, Trisha Chakrabarty, Nusrat Choudhury. . . . 1250

Research Hotspots and Trends of Virtual Reality Intervention for Stroke: Bibliometric Analysis (e65993)
Yixin Wei, Yuan Chen, Runting Ma, Yitong Qiu, Wei Su, Li Zhang, Qiang Gao. . . . 1270

Feasibility, Usability, and Effects of Leisure-Based Cognitive Training Using a Fully Immersive Virtual Reality System in Older Adults: Single-Arm Pretest-Posttest Pilot Study (e66673)
I-Ching Chuang, Xiao-Ting Huang, I-Chen Chen, Yih-Ru Wu, Ching-Yi Wu. . . . 1286

Evaluating the Effectiveness of Immersive Virtual Reality Rehabilitation Games With Enhanced Visual Training for Hand Motor Function Improvement Using Electromyography: Randomized Controlled Trial (e74314)
Faisal Amin, Asim Waris, Muhammad Khan, Muhammad Ijaz, Hammad Nazeer, Syed Gilani, Fawwaz Hazzazi, Umer Shah. . . . 1298

Influence of Virtual Reality Illusions on Balance Performance and Immersive User Experience in Young Adults: A Within-Subject Experimental Study (e70376)
R Abayasiri, Antonio Padilha Lanari Bo, Taylor Dick, Nilufar Baghaei. . . . 1321

The Impact of Active Augmented Reality Games on Physical Activity and Cognition Among Older Adults: Feasibility Study (e73221)
Amy Lu, Bhagyashree Parkar, Shiveshchandra Gupta, Sierra Hall, Dominika Pindus, Arthur Kramer. . . . 1342

Vision AI-Based Gamified Cognitive Prosthesis for Executive Function: Feasibility and Usability Study (e74157)
Co Siow, Yao-Hua Yang, Cheng-Jui Tsai, Wan-Wan Yang, Chaur-Jong Hu, Jia-Ying Sung, Jowy Tani. . . . 1359

XSL•FO
RenderX

A Resource-Efficient, High-Dose, Gamified Neurorehabilitation Program for Chronic Stroke at Home: Retrospective Real-World Analysis (e69335)
Spencer Arbuckle, Anna Knill, Michelle Chan-Cortés, Gabriela Rozanski, Anastasia Ford, Louis Derungs, John Krakauer, Naveed Ejaz, David Putrino, Jenna Tosto-Mancuso, Meret Branscheidt. . . . 1370

Exergaming System for Exercise-Based Cardiac Rehabilitation in Patients With Heart Failure: Development and Usability Assessment Study of a Device Prototype (e71385)
Carles Blasco-Peris, Juan Alcolea Garrido, Barbara Seguí, Rocio Zaragoza, Vicente Climent-Paya, Laura Fuertes-Kenneally, Agustín Manresa-Rocamora, Ana Sanz-Rocher, Sabina Baladzhaeva, José Sarabia. . . . 1385

Mixed Reality–Based Physical Therapy in Older Adults With Sarcopenia: Preliminary Randomized Controlled Trial (e76357)
Yeongsang An, Seunghwa Min, Chanhee Park. . . . 1399

Neurotechnology-Based, Intensive, Supplementary Upper-Extremity Training for Inpatients With Subacute Stroke: Feasibility Study (e56397)
Reut Binyamin-Netser, Shirley Handelzalts, Noy Goldhamer, Inbar Avni, Adi Tayer Yeshurun, Yogev Koren, Ofri Bibas Levy, Shilo Kramer, Simona Bar Haim, Lior Shmuelof. . . . 1411

Reward Feedback Mechanism in Virtual Reality Serious Games in Interventions for Children With Attention Deficits: Pre- and Posttest Experimental Control Group Study (e67338)
Hao Fang, Changqing Fang, Yan Che, Xinyuan Peng, Xiaofan Zhang, Di Lin. . . . 1423

A Novel Eye Tracking–Based Gamified Assessment of Contrast Sensitivity Function in Children: Prospective Development and Reliability Study (e81082)
Yunsi He, Yijing Zhuang, Lei Feng, Xuanyu Xu, Ying Deng, Moruomi Li, Yangfei Pang, Ying Yao, Wentong Yu, Zixuan Xu, Yusong Zhou, Yudan Zhong, Qiuying Li, Qingqing Ye, Junpeng Yuan, Yun Wen, Jinrong Li. . . . 1437

A Novel Approach Using Serious Game Data to Predict the WISC-V Processing Speed Index in Children With Attention-Deficit/Hyperactivity Disorder: Machine Learning Study (e73408)
Jun-Su Kim, Yoo Jeong, Seung-Jae Kim, Su Jun, Jin-Yeop Park, Hyang-Sook Hoe, Jeong-Heon Song. . . . 1448

Discriminative Power of the Serious Game Attention Slackline in Children and Adolescents With and Without Attention-Deficit/Hyperactivity Disorder: Validation Study (e65170)
Nicolás Ruiz-Robledillo, Ignacio Lucas, Rosario Ferrer-Cascales, Natalia Albaladejo-Blázquez, Javier Sanchis, Juan Trujillo. . . . 1464

Machine Learning Analysis of Engagement Behaviors in Older Adults With Dementia Playing Mobile Games: Exploratory Study (e54797)
Melika Torabgar, Mathieu Figeys, Shaniff Esmail, Eleni Stroulia, Adriana Ríos Rincón. . . . 1478

Digital Ergonomics of NavegApp, a Novel Serious Game for Spatial Cognition Assessment: Content Validity and Usability Study (e66167)
Juan Sanchez-Escudero, David Aguillon, Stella Valencia, Mauricio Garcia-Barrera, Daniel Aguirre-Acevedo, Natalia Trujillo. . . . 1492

Effects of Computerized Cognitive Training on Vesicular Acetylcholine Transporter Levels using [18F]Fluoroethoxybenzovesamicol Positron Emission Tomography in Healthy Older Adults: Results from the Improving Neurological Health in Aging via Neuroplasticity-based Computerized Exercise (INHANCE) Randomized Clinical Trial (e75161)
Mouna Attarha, Ana de Figueiredo Pelegrino, Lydia Ouellet, Paule-Joanne Toussaint, Sarah-Jane Grant, Thomas Van Vleet, Etienne de Villers-Sidani. . . . 1509

Effect of Gamification on Task Engagement During an Eye-Tracking Test Battery in 5-Year-Old Children Born Preterm: Observational Study (e77109)
Andrea De Gobbis, Marta Malavolta, Nathalie De Beukelaer, Els Ortibus, Aleksander Sadikov, Vida Groznik. . . . 1535

XSL•FO
RenderX

Measuring Stress and Perceptions for a Virtual Reality–Based Pericardiocentesis Procedure Simulation for Medical Training: Usability Study (e68515)
Alberto Rubio-López, Rodrigo García Carmona, Laura Zarandieta Román, Alejandro Rubio Navas, Angel González-Pinto, Pablo Cardinal-Fernández. 1552

Applying Augmented Reality to Convey Medical Knowledge on Osteoclasts to Users of a Serious Game: Vignette Experiment (e64751)
Jascha Grübel, Julia Chatain, Claudio Schmid, Violaine Fayolle, Fabio Zünd, Reinhard Gruber, Bernd Stadlinger. 1568

Development and Evaluation of a Case-Based Serious Game for Diagnosis and Treatment Planning in Orthodontic Education: Quasi-Experimental Study (e73956)
Rochaya Chintavalakorn, Chayuth Chanwanichkulchai, Napat Buranasing, Napatsaporn Parivisutt, Naruchol Patchasri, Kawin Sipiyaruk. . . .
1 5 8 6

Use of an Immersive Virtual Reality Application to Educate Medical Students in Patient Handover: Pilot Study (e73907)
Laura Hanke, Patrick Schwoerer, Florentine Huettl, Lukas Vradelis, Kai-Uwe Strelow, Christian Boedecker, Patrick Saalfeld, Vuthea Chheang, Holger Buggenhagen, Hauke Lang, Christian Hansen, Tobias Huber. 1597

ADRENALINE, a Learning Game to Improve Prescribing Skills in Undergraduate Medical Students: Descriptive Study (e66334)
Cecile Yelnik, Aurélie Daumas, Yanele Poteaux, Annie Standaert, Natacha Grimbert, Raphaël Favory, Pierre Ravaux, Marc Lambert, Katia Quelennec. 1609

Efficacy of the Web-Based Gamified Infection Control Training System on Practices for Health Care Workers in Residential Care Homes: Clustered Randomized Controlled Trial (e71593)
Angela Leung, Doris Leung, Terence Lau, Justina Liu, Teris Cheung, Daphne Cheung, Simon Lam, Eliza Wong, Mimi Tse, Alex Molassiotis. .
1 6 1 9

Effectiveness of a Gamified Mobile App in Enhancing Treatment Adherence for Children With Amblyopia: Explorative Study (e60309)
Bo Liu, Yisheng Fan, Meng Xu, Fangyuan Chang, Yue Shi, Zhao Liu. 1635

Development and Evaluation of a Monocular Camera–Based Mobile Exergame for at-Home Intervention in Individuals at High Risk of Type 2 Diabetes: Randomized Controlled Trial (e75823)
Jianan Zhao, Dian Zhu, Yanan Wang, Yaqin Xia, Zeshi Zhu, Jihong Yu. 1653

Becoming a Rapid Shooter in a Game Using Embodied Electrical Muscle Stimulation: Development and Usability Study (e69330)
Jihwan Kim, Mingyu Kang, Jejoong Kim, Kwanguk Kim. 1664

Smartwatch-Based Tailored Gamification and User Modeling for Motivating Physical Exercise: Experimental Study With the Maximum Difference Scaling Segmentation Method (e66793)
Jie Yao, Di Song, Tao Xiao, Jiali Zhao. 1710

Digital Catalysts for Noncommunicable Disease Prevention Serious Games and Gamified Applications: Framework Design Study (e69246)
Christoph Aigner, René Baranyi, Thomas Grechenig. 1734

Haptic-Driven Serious Card Games for Older Adults: User Preferences Study (e73135)
Xin Huang, Nazlena Ali, Shafrida Sahrani. 1753

A Gamified Assessment Tool for Antisocial Personality Traits (Antisocial Personality Traits Evidence-Centered Design Gamified): Randomized Controlled Trial (e70453)
Yaobin Tang, Yongze Xu, Qunli Zhou, Ran Bian. 1766

XSL•FO
RenderX

In-Game Need Satisfaction, Frustration, and Gaming Addiction Patterns Across Subgroups of Adolescents Through Structural Equation Modeling: Cross-Sectional and Instrument Validation Study of the Youth Gaming Experience Scales (e63612)
Amparo Luján-Barrera, Lydia Cervera-Ortiz, Mariano Chóliz. 1793

Prevalence, Sex Differences, and Predictors of Internet Gaming Disorder Among Impoverished Rural Adolescents: Cross-Sectional and Prospective Cohort Study (e83522)
Chenhan Wang, Yifan Li, Shuhong Lin, Qiuping Huang, Yongyan Shi, Guangxian Yang, Erjia Huang, Xicheng Deng, Jinwen Luo. 1810

Reciprocal Relationship Between Self-Control Belief and Gaming Disorder in Children and Adolescents: Longitudinal Survey Study (e59441)
Shimin Zhu, Di Qi. 1824

Effects of Progressive Aerobic Training on Executive-Reward Network Connectivity and Symptoms of Internet Gaming Disorder: Randomized Controlled Trial (e83597)
Shaoyu Cui, Xuefeng Xu, Xin Luo, Meiting Wei, Guang-Heng Dong. 1836

Reviews

Exploring Virtual Reality and Exercise Simulator Interventions in Patients With Attention Deficit Hyperactivity Disorder: Comprehensive Literature Review (e57297)
Gurdeep Sarai, Prem Jayaraman, Oren Tirosh, Nilmini Wickramasinghe. 935

Identifying Key Principles and Commonalities in Digital Serious Game Design Frameworks: Scoping Review (e54075)
Raluca Maxim, Joan Arnedo-Moreno. 1680

Gamification Integration in Technological Devices for Motor Rehabilitation in Parkinson Disease: Scoping Review (e69433)
Pere Bosch-Barceló, Oriol Martínez-Navarro, Maria Masbernat-Almenara, Carlos Tersa-Miralles, Anni Pakarinen, Helena Fernández-Lago. . .
1 8 7 0

Home-Based Virtual Reality Training for Enhanced Balance, Strength, and Mobility Among Older Adults With Frailty: Systematic Review and Meta-Analysis (e67146)
Hammad Alhasan, Elaf Alandijani, Lara Bahamdan, Ghofran Khudary, Yara Aburaya, Abdulaziz Awali, Mansour Alshehri. 1886

Effectiveness of Gamification Interventions to Improve Physical Activity and Sedentary Behavior in Children and Adolescents: Systematic Review and Meta-Analysis (e68151)
Min Wang, Jisheng Xu, Xulin Zhou, Xingchen Li, Yu Zheng. 1901

Use of Gamified Digital Tools in Daily Tasks of Health Care Workers: Scoping Review (e70480)
Binita Paudel, Hussein Al-Shehabi, Rita Dörner, Charbel El Bcheraoui, Tessa Lennemann, Andrea Bernasconi. 1924

Efficacy of Virtual Reality–Based Interventions on Cognitive Function in Patients With Neuropsychiatric Disorders: Systematic Review and Meta-Analysis of Randomized Controlled Trials (e67501)
Qiujing Du, Yuhan Wei, Yuexuan Ma, Changqing Liu, Shanshan Du, Qi Zhang, Xiaotong Gong, Jiaju Yang, Qijie Li, Ka Li. 1943

Efficacy of Motion-Sensing Game–Assisted Pulmonary Rehabilitation in Patients With Chronic Obstructive Pulmonary Disease: Systematic Review and Meta-Analysis of Randomized Controlled Trials (e69562)
Cuirong Hu, Xia Lan, Yanwen Luo, Shu Zhu, Guilan Cheng. 1959

XSL•FO
RenderX

Scrutinizing the Gateway Relationship Between Gaming and Gambling Disorder: Scoping Review With a Focus on the Southeast Asian Region (e59740)
Kristiana Siste, Daniel King, Enjeline Hanafi, Lee Sen, Adrian Adrian, Belinda Murtani. 1978

Pervasive Games for Sexual Health Promotion: Scoping Literature Review (e58912)
Claudio Rubio, Felipe Besoain. 1993

Effectiveness of Mobile Health–Based Gamification Interventions for Improving Physical Activity in Individuals With Cardiovascular Diseases: Systematic Review and Meta-Analysis of Randomized Controlled Trials (e64410)
Tianzhuo Yu, Monica Parry, Tianyue Yu, Linqi Xu, Yuejin Wu, Ting Zeng, Xin Leng, Qian Tong, Feng Li. 2013

Electrophysiological Insights in Exergaming—Electroencephalography Data Recording and Movement Artifact Detection: Systematic Review (e50992)
Carolina Rico-Olarte, Diego Lopez, Bjoern Eskofier, Linda Becker. 2030

Effectiveness of Serious Games as Digital Therapeutics for Enhancing the Abilities of Children With Attention-Deficit/Hyperactivity Disorder (ADHD): Systematic Literature Review (e60937)
Jing Lin, Woo-Rin Chang. 2043

The Impact of Gamified Interventions on the Management of Chronic Obstructive Pulmonary Disease: Systematic Literature Review (e69510)
Jinsong Chen, Tingzhong Yang, Qilian He, Mingli Pang, Ying Cao, Zheng Liu, Linfei Li, Hsing-I Liu, Christopher Bullen. 2062

Technology-Assisted Motor-Cognitive Training Among Older Adults: Rapid Systematic Review of Randomized Controlled Trials (e67250)
Yaqin Li, Yaqian Liu, Angela Leung, Jed Montayre. 2083

Serious Games for Preventing Musculoskeletal Disorders in Occupational Settings: Scoping Review (e66913)
Thomas Rivière, Thibaud Hulin, Farzan Sasangohar, Mohsen Zare. 2106

Corrigenda and Addendas

Correction: A Novel Approach Using Serious Game Data to Predict the WISC-V Processing Speed Index in Children With Attention-Deficit/Hyperactivity Disorder: Machine Learning Study (e88500)
Jun-Su Kim, Yoo Jeong, Seung-Jae Kim, Su Jun, Jin-Yeop Park, Hyang-Sook Hoe, Jeong-Heon Song. 1851

Correction: Exergame (ExerG)-Based Physical-Cognitive Training for Rehabilitation in Adults With Motor and Balance Impairments: Usability Study (e73405)
Silvia Herren, Barbara Seebacher, Sarah Mildner, Yanick Riederer, Ulrike Pachmann, Nija Böckler, Stephan Niedecken, Sabrina Alicia Sgandurra, Leo Bonati, Isabella Hotz, Alexandra Schättin, Roman Jurt, Christian Brenneis, Katharina Lenfert, Frank Behrendt, Stefan Schmidlin, Lennart Nacke, Corina Schuster-Amft, Anna Martin-Niedecken. 1853

Correction: A Mixed Reality–Based Telesupervised Ultrasound Education Platform on 5G Network Compared to Direct Supervision: Prospective Randomized Pilot Trial (e77586)
Minha Kim, Meong Hi Son, Suhyeon Moon, Won Cha, Ik Jo, Hee Yoon. 1855

Correction: Effects and Acceptability of a 1-Week Home-Based Virtual Reality Training for Supporting the Management of Stress and Anxiety: Randomized Pilot Trial (e77791)
Federica Pallavicini, Eleonora Orena, Lisa Arnoldi, Federica Achille, Stefano Stefanini, Maddalena Cassa, Alessandro Pepe, Guido Veronese, Luca Bernardelli, Francesca Sforza, Sara Fascendini, Carlo Defanti, Marco Gemma, Massimo Clerici, Giuseppe Riva, Fabrizia Mantovani. . .
1 8 5 7

XSL•FO
RenderX

Correction: Serious Game for the Nursing Assessment of Home-Dwelling Older Adults: Development and Validation Study (e81884)
Erica Busca, Silvia Caristia, Sara Bidone, Alessia Bolamperti, Sara Campagna, Arianna Cattaneo, Rosaria Lea, Doriana Montani, Antonio Scalogna, Francesca Piovesan, Erika Bassi, Alberto Dal Molin. . . . 1859

Correction: FITLIGHT Training and Its Influence on Visual-Motor Reactions and Dribbling Speed in Female Basketball Players: Prospective Evaluation Study (e82614)
Ahmad Hassan. . . . 1861

Letters to the Editor

Author's Reply: Refining Open-World Game and Nostalgic Film Interventions for Broader and More Reliable Therapeutic Impact (e83239)
Andreas Eisingerich, Annisa Arigayota, Barbara Duffek, Congcong Hou. . . . 1862

Refining Open-World Game and Nostalgic Film Interventions for Broader and More Reliable Therapeutic Impact (e82157)
Chenlei Sun, Ke Meng. . . . 1864

Authors' Reply: Is the Pinball Machine a Blind Spot in Serious Games Research? (e73034)
Luis Rodríguez Timaná, Javier Castillo García, Teodiano Bastos Filho, Alvaro Ocampo González, Nazly Hincapié Monsalve, Nicolas Valencia Jimenez. . . . 1866

Is the Pinball Machine a Blind Spot in Serious Games Research? (e72354)
Jens Eckardt. . . . 1868

Research Letter

Mobile Exergames to Reduce Sedentary Time in Patients With Heart Failure: What Do Health Care Professionals Expect? (e69126)
Leonie Klompstra, Anna Strömberg, Antoni Bayes-Genis, Maria Boldo, Beatriz González, Nuria Santaularia, Tiny Jaarsma. . . . 2124

XSL•FO
RenderX

Serious Games Integrating Perceptual Learning and Stereopsis Training in Children With Amblyopia: Single-Arm Pre-Post Feasibility Study

Yiwei Mo[1,2*], MA; Peng Chen[1*], MSc; Manting Hou[1]; Fang Cui[3], PhD; Chao Luo[2], PhD; Chao Deng[1]; Jinkang Lin[1]; Qi Wu[4], MA

[1]School of Digital Media, Shenzhen Polytechnic University, Room 321, Block A, Zhixingyuan, Liuxiandong Campus, 7098 Liuxian Avenue, Nanshan District, Shenzhen, Guangdong, China

[2]School of Visual Arts Design, Hubei Institute of Fine Arts, Wuhan, Hubei, China

[3]School of Psychology, Shenzhen University, Shenzhen, Guangdong, China

[4]School of Art and Humanities, Hubei Institute of Fine Arts, Wuhan, Hubei, China

[*]these authors contributed equally

Corresponding Author:
Peng Chen, MSc
School of Digital Media, Shenzhen Polytechnic University, Room 321, Block A, Zhixingyuan, Liuxiandong Campus, 7098 Liuxian Avenue, Nanshan District, Shenzhen, Guangdong, China

Abstract

Background: Amblyopia, a leading cause of preventable childhood blindness, often remains inadequately addressed by traditional treatment methods such as refractive correction and occlusion therapy, which can be noninteractive and lead to poor adherence.

Objective: This study aimed to design and evaluate the feasibility and preliminary efficacy of a serious game intervention that integrates perceptual learning and stereoscopic vision training for pediatric amblyopia and to explore its potential as a complementary or alternative approach to conventional treatments such as occlusion or atropine therapy.

Methods: We evaluated visual acuity, accommodative sensitivity, binocular accommodation, stereopsis, and compliance-related data in a cohort of children with amblyopia aged 7 to 12 years before and after a 3-month intervention. Participants engaged in visual training via a serious game, attending sessions 4 times weekly for 30 minutes each.

Results: Best-corrected visual acuity improved significantly from 0.42 (SD 0.16) to 0.37 (SD 0.18) logMAR, yielding a mean difference of 0.05 (95% CI 0.03 - 0.08; t_{24}=4.82; Cohen d=0.96; P<.001). Accommodative sensitivity showed marked enhancement following the intervention. In the amblyopic eye, mean values increased from 1.44 (SD 2.18) to 4.96 (SD 2.91) cycles per minute, corresponding to a mean change of −3.52 (95% CI −4.59 to −2.45; t_{24}=−6.81; Cohen d=−1.36; P<.001). Under binocular viewing, accommodative sensitivity improved from 1.52 (SD 2.49) to 5.08 (SD 2.71) cycles per minute (Δ=−3.56, representing the mean paired difference (baseline minus post), 95% CI −4.88 to −2.24; t_{24}=−5.56; Cohen d=−1.11; P<.001). Stereoacuity also improved significantly, decreasing from 780.0 (SD 613.6) to 448.8 (SD 472.2) arc sec (Δ=331.20, 95% CI 134.20 - 528.20; t_{24}=3.47; Cohen d=0.69; P=.002). The proportion of participants with normal Worth 4 dot responses increased from 76% (19/25) at baseline to 96% (24/25) after the intervention, and treatment adherence was high under supervised clinical conditions.

Conclusions: The integration of gamified elements into amblyopia treatment was associated with high adherence under supervised clinical conditions and with significant short-term improvements in visual function, suggesting a promising complementary approach to conventional therapies. This interactive approach effectively combines perceptual learning with stereopsis training, presenting a potential alternative to conventional therapies.

(JMIR Serious Games 2025;13:e77402) doi:10.2196/77402

KEYWORDS

amblyopia; serious games; perceptual learning; stereopsis training; visual function improvement

Introduction

Amblyopia is a cortical visual impairment clinically manifested as a unilateral or bilateral reduction in visual acuity (VA) that cannot be explained by structural abnormalities of the eye [1]. It has an estimated prevalence of approximately 3.4% in children [2] and poses a significant risk of lifelong visual loss [3]. The first-line treatment is refractive correction [4]; if this proves ineffective, occlusion therapy is commonly implemented [5]. Although occlusion is considered the standard treatment, it lacks

XSL•FO
RenderX

binocular visual stimulation and is often associated with poor adherence [6,7]. Atropine may serve as an alternative intervention, but it may cause systemic side effects [8].

Advancements in our understanding of neurophysiological mechanisms have led to perceptual learning (PL) based on visual stimuli garnering increasing attention and emerging as a therapeutic approach for amblyopia treatment [9,10]. PL refers to the neuroplastic process induced by specific training that leads to persistent functional optimization of perceptual systems. It relies on the neuroplasticity of the visual cortex [11], enabling the visual system to dynamically adapt to task demands and trigger neuroplastic changes that enhance visual performance [11,12]. Clinical studies have confirmed that this training paradigm can significantly improve multiple dimensions of visual function in patients with amblyopia, including positional acuity [11,13], stereopsis [13], contrast sensitivity, and VA [14], while inducing long-term functional modifications [15].

Existing neuroimaging studies suggest that PL may be accompanied by plastic changes within the visual cortex and associated networks, such as alterations in the activity of primary and secondary visual areas and remodeling of white matter tracts. Moreover, certain gamified training paradigms have been linked to structural and functional modifications in additional brain regions involved in spatial navigation, executive control, and motor coordination. It should be emphasized that this study did not include neuroimaging assessments, nor did it directly measure these neural mechanisms. Therefore, the mechanistic statements below are presented solely as general interpretations based on previous literature, intended to provide a theoretical context for the design of the training tasks, rather than as causal explanations of the findings in this study.

Neuroimaging studies have provided direct evidence for the neuroplastic mechanisms underlying PL. Using multimodal imaging techniques, Zhai et al [16] found that 30 days of PL intervention induced significant enhancement of neural activity in the primary visual cortex (Brodmann areas 17 - 19) and bilateral temporal lobes of patients with amblyopia, accompanied by structural remodeling of white matter tracts. Notably, training paradigms incorporating gamified elements have been shown to selectively increase gray matter volume in the right hippocampus (associated with spatial navigation), prefrontal cortex (executive control center), and cerebellum (motor coordination hub) [17]. These findings not only confirm the persistent neuroplasticity of the adult brain but also provide neuroanatomical validation for the theory that cognitive training enhances higher-order brain functions. Current advancements highlight the dual value of PL in visual system remodeling [18,19] and amblyopia therapeutics. Clinical evidence demonstrates that synergistic integration of PL with traditional occlusion therapy significantly accelerates visual rehabilitation in adolescent amblyopes [20].

Stereopsis, as the ability of the 2 eyes to work together to perceive depth and spatial position [21], is one of the important visual functions in amblyopia treatment. Its formation relies on the brain's processing of parallax—the image offset received by the left and right eyes due to spatial position differences. The parallax effect is the core physical basis of stereoscopic vision, and by controlling this parallax amount, the brain is able to merge the 2D images and reconstruct 3D spatial information [22,23]. In amblyopia, interocular suppression disrupts binocular fusion mechanisms, leading to stereoscopic deficits [24]. Clinically, children with amblyopia with impaired stereopsis face a 2.2-fold higher risk of persistent amblyopia compared with those with intact binocularity [25].

Recent years have seen increasing interest in binocular and gamified interventions. In adults, Vedamurthy et al [26] showed that a dichoptic action video game significantly improved VA, contrast sensitivity, and stereopsis, with retention after a no-contact period. In children, however, the results are more cautious. The Pediatric Eye Disease Investigator Group indicated in a multicenter randomized controlled trial (RCT) that the Falling Blocks binocular iPad game (provided by the research team) did not outperform spectacle correction alone after 8 weeks, largely due to poor adherence [27]. Nevertheless, promising findings continue to emerge. Bari et al [28] demonstrated in Italian patients that a home-based binocular rehabilitation using red-blue anaglyphic games significantly improved best-corrected visual acuity (BCVA) within 8 weeks. Bocqué et al [29] found that the updated FAVAS 2020 platform significantly increased adherence compared with its 2015 version. In China, Zhu et al conducted a prospective pilot study in preschool children, showing that gamified binocular therapy significantly improved both distance VA and stereoacuity after 12 weeks, with good safety [30].

Virtual reality (VR) technologies are also being tested. Hirota et al [31] evaluated a motion-based VR dichoptic training system in healthy adults and confirmed its safety, with minimal ocular or systemic side effects. Moreover, Li et al [32] showed that stereoscopic 3D video games enhance the precision, but not the accuracy, of depth perception, highlighting the mechanism through which disparity processing can be optimized. Together, these studies suggest that amblyopia management is shifting from "occlusion and penalization" toward "PL, gamification, and VR."

Finally, amblyopia prevention also requires innovations in screening and triage pathways. Morse and Oatts [33] emphasized that current screening technologies are inconsistent, and new strategies—such as artificial intelligence–assisted diagnostics and biomarker-based approaches—are urgently needed to optimize early detection and referral. This systemic perspective provides a rationale for integrating low-cost anaglyph glasses with touchscreen devices in community and home-based training settings.

Within this evolving landscape, particular attention has turned to children in late childhood. Although ages 7 - 12 years are traditionally viewed as approaching the tail end of the sensitive period for amblyopia treatment, converging evidence indicates that residual visual cortical plasticity can be harnessed beyond early childhood [34,35], especially when interventions incorporate binocular integration, contrast rebalancing, and PL-based tasks. Studies in older children, adolescents, and adults have reported meaningful, though sometimes attenuated, gains with binocular, game-based, or PL interventions [34,36]. At the same time, clinical trials in 7 - 12-year-olds indicate that

XSL•FO
RenderX

conventional patching or atropine can still yield improvements, but with lower overall success rates, residual deficits, and adherence challenges compared with younger children [37,38]. Taken together, these findings suggest that 7 - 12-year-olds constitute a clinically relevant target group in which engaging, technology-supported binocular training warrants systematic evaluation rather than being dismissed as beyond therapeutic reach.

Based on these advances, this study introduces a gamified serious-game framework integrating PL and stereopsis training, using red-blue anaglyph glasses and tablet devices to generate controlled binocular stimuli. The aim is to explore the therapeutic potential of this approach in children aged 7 - 12 years with amblyopia, focusing on improvements in VA, accommodative function, and stereopsis, while positioning the intervention as a potential complement to conventional therapies. Given these gaps in the current evidence, several key research questions remain.

First, can a gamified intervention that integrates PL principles with stereopsis training demonstrably improve multiple dimensions of visual function in children with amblyopia—including VA, accommodative sensitivity, binocular accommodative function, and stereopsis—within a clinically feasible time frame?

Second, without direct comparison to conventional therapies (such as occlusion or atropine), does this intervention demonstrate feasibility as a potential supplement or alternative to standard treatment (eg, improving compliance and potentially providing a more accessible, lower-cost digital pathway)?

Third, how might objective adherence data captured through digital platforms be leveraged to refine individualized treatment protocols and to inform the design of future RCTs?

By addressing these questions, this study aims to provide preliminary clinical evidence for the feasibility, safety, and efficacy of gamified binocular digital therapy and to lay a foundation for subsequent large-scale, multicenter investigations.

Methods

Ethical Considerations

This study was a single-group, pre-post feasibility design. A “no-treatment” or “delayed-treatment” control group was not included because, during ethical review and discussions with parents, a substantial proportion of guardians declined to have their children randomly assigned to a waiting or nonintervention group. In accordance with routine clinical practice and the recommendations of the ethics committee, establishing a waiting group or suspending intervention was considered inappropriate under the current condition of insufficient clinical equipoise. To balance acceptable risk and practical feasibility, the study adopted a single-group intervention approach to evaluate feasibility and identify preliminary efficacy signals, with the following prespecified commitments and risk control measures.

First, refractive correction was maintained and standardized throughout the study (standard care was not interrupted).

Second, rescue criteria were established; if adverse events, new-onset diplopia, or significant changes in ocular alignment, or substantial deterioration in visual function relative to baseline occurred (thresholds predefined in the protocol), the training was immediately discontinued, and conventional treatment was resumed.

Third, during the intervention, occlusion or atropine therapy was not initiated proactively. If clinically required, initiation or adjustment was made independently by the attending ophthalmologist and recorded as a protocol deviation, which would be included in sensitivity analyses.

Finally, statistical analyses were limited to within-group changes and correlations without making any claims of superiority or inferiority relative to occlusion or atropine treatment.

The research protocol adhered to the principles of the Declaration of Helsinki and was approved by the Human Research Ethics Committee of the Faculty of Psychology, Shenzhen University (SZU_PSY_2024_100). After being fully informed of the study objectives, procedures, potential risks, and alternative treatment options, written informed consent was obtained from the guardians, while children provided assent or consent appropriate to their age and level of understanding. Participants were free to withdraw from the study at any time without any impact on their clinical care. No monetary or material incentives were offered. Data collection followed the principle of minimum necessity, using deidentified coding with restricted access. No sensitive personal information was collected, and all data were used solely for research purposes.

Participants

Participant recruitment began in July 2024, with intervention initiated in September 2024 and continuing for 12 weeks. All training sessions and data collection were completed by December 2024. A total of 25 children (Table 1) were enrolled in the study, including 9 males, all recruited from Luohu People’s Hospital, Shenzhen. The mean age was 8.04 (SD 1.457) years. Among the participants, 22 had anisometropic amblyopia, and 3 had strabismic amblyopia. Based on treatment history, 12 children had never received any previous treatment, 8 had undergone occlusion therapy only, 3 had received other forms of therapy, and 2 had previously received both occlusion therapy and other treatments. Classification of amblyopia subtypes and treatment history documentation followed the Clinical Guidelines for the Diagnosis and Treatment of Amblyopia in Children (China, 2021).

Furthermore, 2 experienced optometrists conducted comprehensive eye examinations for all participants as per some inclusion and exclusion criteria (Textbox 1).

XSL•FO
RenderX

Table . Demographic and baseline characteristics of participants.

Patient	Age (years)	Sex	logAE[a]	logFE[b]	Acc_AE[c]	Acc_FE[d]	Acc_BE[e]	Supp[f]	Stereo[g]
1	7	Male	0.30	0.10	2	1	1	OD[h]	800
2	9	Male	0.52	0.10	0	0	0	normal	1200
3	7	Male	0.52	0.10	0	0	0	normal	1200
4	7	Female	0.52	0.10	1	2	0	normal	1200
5	9	Male	0.30	0.00	7	6	7	normal	100
6	9	Female	0.30	0.10	0	0	0	normal	600
7	12	Female	0.52	0.10	0	0	0	OS[i]	1200
8	7	Female	1.00	0.10	0	0	0	OS	1200
9	8	Female	0.30	0.10	0	0	4	normal	1200
10	9	Female	0.70	0.10	0	0	0	normal	1200
11	7	Male	0.30	0.05	5	4	2	normal	200
12	8	Male	0.40	0.05	0	0	8	normal	200
13	12	Female	0.40	0.10	2	2	0	diplopia	1200
14	7	Male	0.30	0.05	0	0	0	normal	40
15	9	Female	0.30	0.05	0	0	0	normal	100
16	8	Male	0.30	0	4	2	3	normal	60
17	7	Female	0.30	0.05	3	2	0	normal	200
18	7	Female	0.30	0.10	1	3	2	normal	100
19	7	Female	0.30	0.05	7	8	7	normal	100
20	7	Female	0.40	0.05	2	0	1	normal	200
21	8	Female	0.40	0.10	2	6	3	OD	2400
22	7	Female	0.40	0.10	0	0	0	OD	1200
23	7	Female	0.40	0.10	0	0	0	normal	1200
24	7	Female	0.52	0.10	0	1	0	normal	1200
25	9	Male	0.52	0.10	0	0	0	normal	1200

[a]logAE: visual acuity of the amblyopic eye.

[b]logFE: visual acuity of the fellow eye.

[c]Acc_AE: accommodative sensitivity of the amblyopic eye.

[d]Acc_FE: accommodative sensitivity of the fellow eye.

[e]Acc_BE: binocular accommodative sensitivity.

[f]Supp: suppression.

[g]Stereo: stereopsis.

[h]OD: oculus dexter.

[i]OS: oculus sinister.

XSL•FO
RenderX

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria

- Aged between 7 and 12 years (this relatively broad age range was chosen to reflect real-world clinical referrals of late-childhood amblyopia; however, developmental heterogeneity across 7 - 12 y was not powered for formal subgroup analysis and is considered in the interpretation of results).
- Best-corrected visual acuity in 1 eye, confirmed by cycloplegic refraction, was below 0.8, with an interocular visual acuity difference of ≥0.2 logMAR.
- Visual acuity remained stable without significant fluctuation over the past 3 months.
- No other ocular diseases or neurodevelopmental disorders were present.
- Consent to participate and provision of written informed consent.

Exclusion criteria

- Developmental delay or cognitive impairment that prevented compliance with experimental procedures.
- Presence of retinal disease or ocular media opacity (eg, cataract or corneal opacity).
- Uncorrectable vision problems.
- Ongoing atropine treatment during the study period or use of atropine within the past 3 months.
- Premature birth ≥8 weeks.
- Presence of severe concurrent ocular or systemic diseases.

Game Principles and Design

This study was conducted in the visual function laboratory of the ophthalmology department at Luohu People's Hospital, Shenzhen, Guangdong, China. Participants underwent visual function training at this location for a duration of 3 months.

While extensive research has been conducted on PL and stereoscopic vision training, few studies have integrated the design principles of these 2 approaches into a single game tailored for children.

In the design of serious games, the parallax effect generated by red-blue glasses (left red, right blue) is used for training. After wearing these special glasses, the left eye can only see the red image, and the right eye can only see the blue image. By controlling the size of the parallax, this method helps patients train and enhance their stereoscopic vision abilities. Using the physical light-splitting principle of the red-blue glasses, the left eye of the wearer receives only the red image, while the right eye sees only the blue image. The red and blue channels of each game element are extracted separately using Adobe Photoshop's channel extraction function, ensuring that each frame displays 2 independent images. These 2 images are then horizontally displaced to create different levels of parallax, thereby achieving the effect of stereopsis training. Furthermore, to accommodate users with varying levels of stereopsis ability, the game includes a feature to adjust the size of the parallax: when a user has weaker stereoscopic vision (larger random-dot stereograms [RDS] value), a larger parallax is provided to reduce the difficulty, while a smaller parallax is used to increase the challenge for users with stronger ability (smaller RDS value).

The primary purpose of these 2 games is to effectively train children's ability to discern objects at varying depth positions through carefully designed task mechanics that focus children's attention on depth perception. These games were designed and developed by the research team using the Mugeda platform (Beijing Lexiang Yunchuang Technology Co, Ltd) and are specifically configured to run on the Magic Eye (Shenzhen Vizhi Holographic Technology Co, Ltd) hardware system. Both "Panda Eats Bamboo" and "Fishing Master" (developed by our research team) have been certified as class II medical devices by the National Medical Products Administration of China (Registration xiangxiezhuzhun20232210101).

Game 1, "Panda Eats Bamboo" (Figure 1B), guides children's attention explicitly toward evaluating depth differences among bamboo sticks, using their simple and repetitive circular motion as the training stimulus. The straightforward regularity and clear depth differences among the bamboo sticks make this game an ideal tool for depth perception training (refer to Multimedia Appendix 1 for the operational demonstration video). The design emphasizes helping children concentrate on depth judgments rather than complex gameplay mechanics or rules. Children aged 7 - 12 years generally possess adequate visual capacity to engage with and benefit from this structured perceptual training.

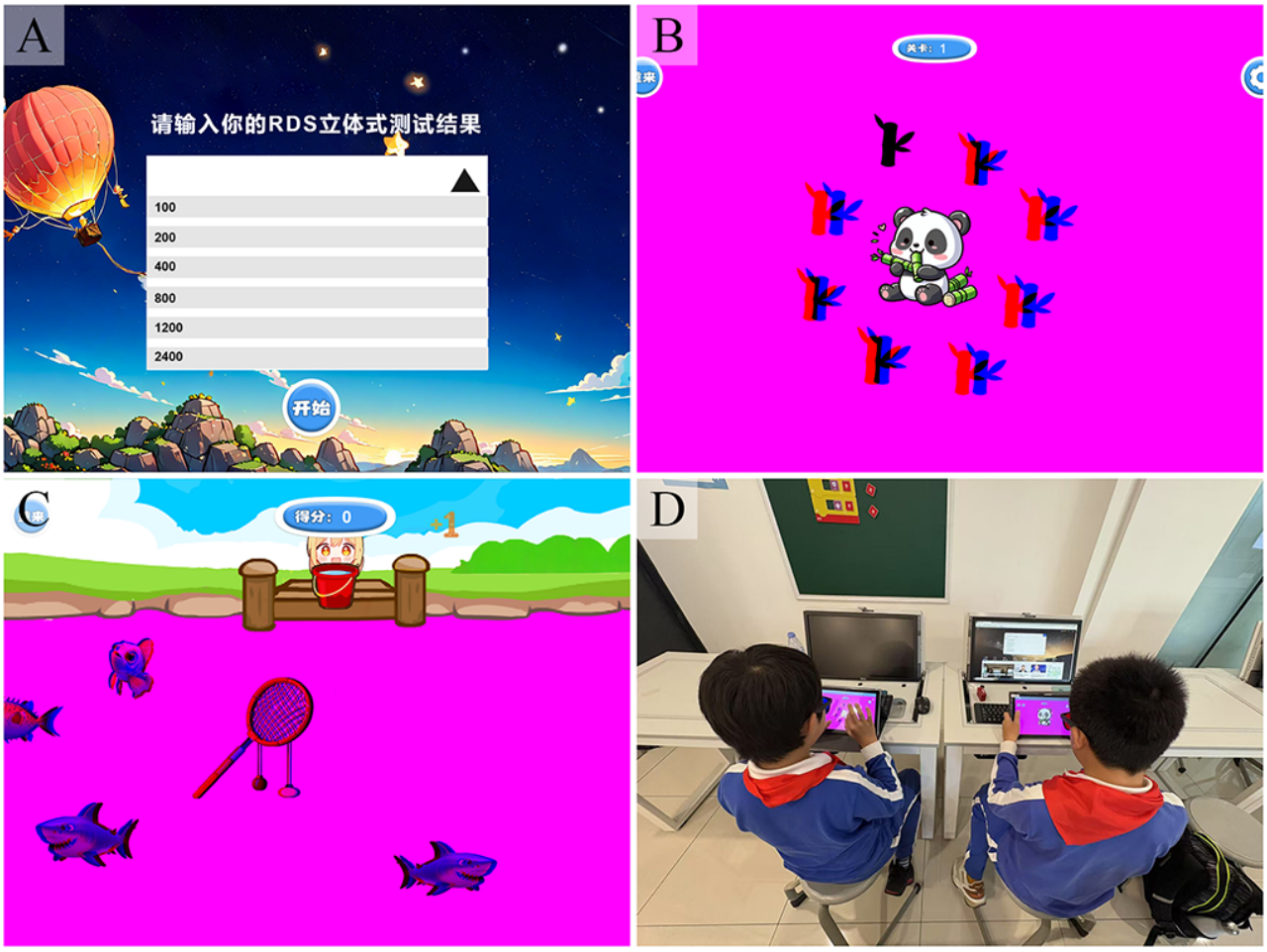
Game 2, "Fishing Master" (Figure 1C), achieves this by requiring players to concentrate on judging depth perception. It trains children's ability to distinguish objects at varying depths through a fishing mechanism. The dynamic characteristics of fish as moving targets enhance interactivity and engagement, effectively drawing children's attention to depth cues (refer to Multimedia Appendix 2 for the demonstration video). Players must accurately perceive and react to the fish's depth and location to successfully complete the catching action, which not only trains their spatial perception but also enhances their ability to discern changes in depth within dynamic scenarios.

The task design of this study was informed by the principles of PL and stereoscopic vision training, emphasizing adaptive task difficulty, enhancement of binocular disparity cues, and engagement with dynamic targets, with the goal of activating

XSL•FO
RenderX

plastic changes in the visual cortex. For example, when children repeatedly judge the depth of objects, neurons in the primary visual cortex and secondary visual area (V1 and V2) regions gradually enhance their ability to integrate and respond to binocular disparity signals. This training prompts neurons in the V1 area to readjust their receptive fields, allowing for more efficient encoding of an object's spatial position [39]. At the same time, dynamic tasks (such as tracking moving fish) activate the middle temporal visual area region in the brain's dorsal pathway, which specializes in processing depth information of moving objects. Repeated training strengthens the sensitivity of neurons in this area to motion disparity, thereby improving the ability to predict the depth of fast-moving objects [40]. Additionally, decision-making tasks (such as selecting the raised bamboo) enhance the collaboration between the prefrontal cortex and the cerebellum, helping children integrate higher-level attention control into their visual judgments [41]. The brain's white matter remodeling after training also supports functional improvement. For example, the corpus callosum fibers connecting the left and right visual areas become more efficient, promoting the integration of information from both eyes [42]. The task design mechanisms described above are based on theoretical frameworks proposed in previous literature; this study did not directly measure these neural pathways and therefore does not draw any causal or temporal conclusions regarding them.

Figure 1. (A) Initial interface: selecting stereo level [屏幕中文： "选择立体等级"→英文"Select stereo level"]; (B) Panda Eats Bamboo ["开始""下一关"→"Start""next"]; (C) Fishing Master ["投网""计分"→"cast""Score"]; (D) Training with the Magic Eye device.



Training Device: Magic Eye

Magic Eye is a standard tablet computer provided by Shenzhen Vizhi Holographic Technology Co, Ltd. The device is equipped with an octa-core processor and a naked-eye 3D LCD display and supports TF card and USB storage accessories, WLAN connectivity, and 3.5 mm stereo headphone output. Unlike conventional tablets, it features specialized stereoscopic display technology that allows interaction with games on the tablet. For amblyopia treatment, patients need to wear red and blue anaglyph glasses to achieve the stereoscopic effect necessary for therapeutic intervention. The choice of this platform was based on its wide availability and common use in local hospitals and clinics, rather than any commercial relationship, and no financial or commercial ties exist between the investigators and the supplier. Magic Eye consists of 4 modules. The perception training module enables interaction with games, facilitates game switching, and allows for the adjustment of parallax sizes. The eye movement detection module monitors user gaze behavior, detecting whether the user is focused on the screen and providing prompts if the user's attention deviates. The parameter control module provides unified management of user parameters with global accessibility, enabling reading and writing of fixed-format files and supporting remote updates or modifications of parameters by doctors or staff. The server management module logs detailed user training data, including

XSL•FO
RenderX

training dates, activities, duration, and other relevant information.

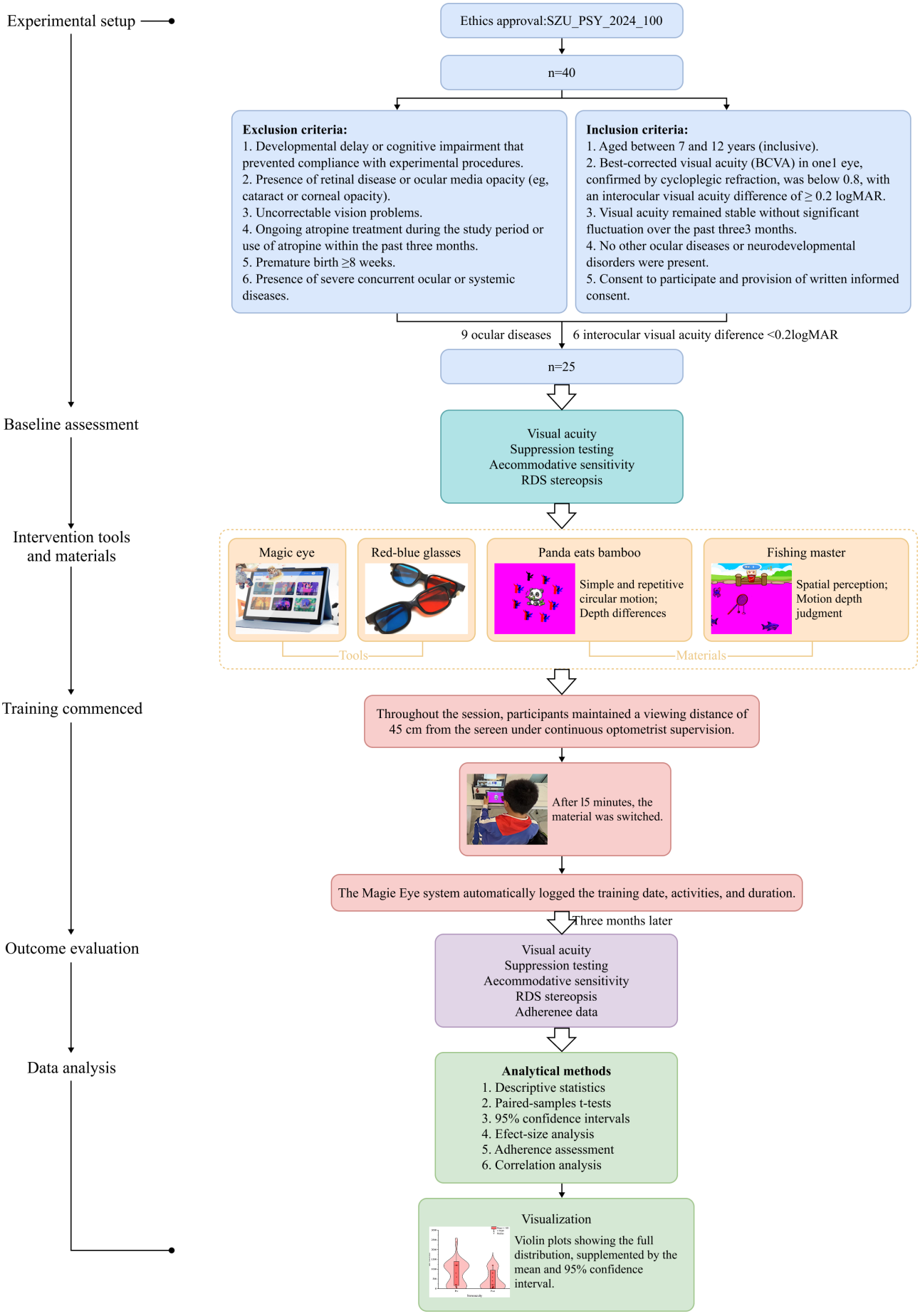
Experimental Procedure

Before the experiment, optometrists briefly introduced the use of Magic Eye to parents and participants. Visual training lasted 12 weeks, with 4 sessions per week, each session lasting 30 minutes. During each 30-minute session, the 2 games were played consecutively, with each game lasting 15 minutes. Training data, including training dates, activities, and duration, were collected through the Magic Eye system during each session. During training, participants were required to maintain a distance of approximately 45 cm from the device. All training was supervised by optometrists to ensure compliance with the protocol. A scheduled follow-up examination (including assessments of VA, accommodative function, stereopsis, and suppression) was arranged 6 months after the completion of training (Figure 2). This study reports only the immediate pre- and postintervention outcomes; follow-up data will be reported separately according to the preregistered protocol. To ensure that participants could independently complete the training, the serious games used in the experiment were briefly explained before the training sessions began. During each session, optometrists accompanied the participants, providing assistance and observing their performance when necessary.

XSL•FO
RenderX

Figure 2. Experimental procedure. This flowchart illustrates the experimental design and workflow of the study on amblyopia treatment using serious games. Participants were screened according to inclusion and exclusion criteria, underwent baseline visual function assessments, and then participated in the Magic Eye training program, which included multiple game-based visual tasks. The system automatically logged training data, and posttraining evaluations were conducted to assess visual outcomes and adherence. RDS: random-dot stereograms.

XSL•FO
RenderX



XSL•FO
RenderX

Outcome Measures

In binocular vision function assessment for children with amblyopia, various indicators play distinct roles in evaluating binocular coordination, accommodative ability, and stereoscopic vision.

Visual Function Data

VA was measured using the Tumbling E chart, which is suitable for children and patients unfamiliar with Roman letters. The Tumbling E chart consists of uppercase "E" shapes oriented in various directions [43]. Participants identified the orientation of the E as directed by the optometrist. VA measurements were converted to the LogMAR scale using the formula LogMAR=−log(VA) [44,45].

Suppression testing is a common clinical method for evaluating binocular vision function, primarily detecting monocular suppression or diplopia. The Worth 4 dot test is frequently used to assess simultaneous vision and fusion [46]. By identifying the number of dots seen, clinicians evaluate binocular vision function and disorders such as esotropia and exotropia [47].

Participants wore red-blue stereoscopic glasses and maintained a stable head position during the test. They first covered 1 eye and reported the number of dots seen, then uncovered the eye and reported the images perceived. Normal binocular vision participants see 4 dots; left-eye suppression patients see only 2 dots, indicating the left eye cannot coordinate with the right; right-eye suppression patients see 3 dots, suggesting a lack of coordination in the right eye. Patients with diplopia may see 5 dots, indicating poor coordination and possible esotropia or exotropia [47]. Suppression training, which involves repeated desuppression exercises, is critical for improving fusion ability and enhancing simultaneous and fusion vision [48].

Accommodative sensitivity testing assesses monocular and binocular accommodative function. Monocular sensitivity reflects the quality of accommodative function, while binocular sensitivity indicates the interaction between accommodation and convergence. Accommodative ability tests are used to evaluate this function [49]. This study followed the protocol outlined by García et al [50] to measure monocular and binocular accommodative ability.

Using ±2.00 D lens combinations, the speed at which participants could view targets (size 20/30) at 40 cm was measured in cycles per minute (CPM). The ±2 D flipper used in this study allowed adjustment of interpupillary distance from 44 mm to 89 mm and had a total weight of 40 g, making it lightweight and suitable for children to handle, thereby minimizing potential ergonomic effects on CPM scores. Participants reported the direction of targets circled in red. Correct responses resulted in lens flips, while incorrect responses were recorded as CPM within 1 minute. Nontested eyes were covered during monocular testing. The standard values for CPM are ≥11 for monocular testing and ≥8 for binocular testing. Values below these thresholds indicate accommodative insufficiency [51,52].

RDS stereoscopic vision tests assess binocular stereopsis, which represents a high level of binocular cooperation and is vital for children's spatial cognitive development. RDS is commonly used to evaluate stereoscopic vision in both pediatric and adult ophthalmology [53].

This study used the Randot Preschool Stereotest to measure stereoscopic acuity, with the minimum disparity at which correct shapes could be identified serving as the indicator [27]. The normal range for stereoscopic acuity is 40 - 60 arcseconds, with values exceeding 60 arcseconds considered abnormal [54].

Compliance Data

Adherence data were captured automatically via the Magic Eye platform. The system continuously logged session start and end times, effective training minutes (excluding pauses and background operation), training frequency, game type, level progression, and task difficulty; it also recorded interaction metrics (touch count, reaction time, and accuracy) and attention-quality indicators (gaze ratio and idle ratio). From these raw data, we derived the following adherence indices: completion rate (sessions completed÷prescribed sessions), duration rate (effective minutes÷prescribed total minutes), weekly adherence (proportion of weeks meeting the prescribed frequency), and quality-adjusted adherence (effective minutes weighted by attention-quality factors and normalized to percentage). All sessions were screened by predefined quality control rules: sessions with effective time <5 minutes or >10 minutes of continuous idle were labeled invalid and excluded.

Statistical Analysis

For a 2-tailed paired design (α=.05, 1−β=0.80) with n=25, the minimum detectable within-participants effect is Cohen d_z≈0.58, a measure of the standardized effect size for paired samples (noncentral-t; normal approximation gives 0.56). Observed accommodative effects (Cohen d_z≈1.1 - 1.4) exceed this threshold, whereas BCVA and stereoacuity show small-to-medium signals that warrant confirmation in controlled trials.

Data Analysis

Data analysis was performed using IBM SPSS software. Descriptive statistics and paired t-tests were used to examine within-group differences, while CIs and effect sizes were calculated to quantify the magnitude and reliability of these differences. Results were visualized using violin plots. Furthermore, Pearson correlation analysis was conducted to explore the relationships between improvements in amblyopic eye VA and measures of accommodative function, stereopsis, and compliance, thereby revealing the statistical associations among these variables. Analyses were restricted to baseline and immediate postintervention outcomes; prespecified durability end points at 6 months were not analyzed in this manuscript.

Results

Overview

The preintervention and postintervention changes in the primary outcomes are summarized in Table 2, followed by detailed statistical analyses and graphical representations for each outcome in subsequent sections.

XSL•FO
RenderX

Table . Postintervention changes in primary outcomes (for accommodative sensitivity, negative Δ values reflect postintervention increases (improvement), as differences were computed as pre-post).

Parameter	Preintervention, mean (SD)	Postintervention, mean (SD)	Cohen *d* (95% CI)	*t* test (*df*)	*P* value
Log_AE[a]	0.42 (0.16)	0.37 (0.18)	0.96 (0.03 to 0.08)	4.82 (24)	<.001
Acc_AE[b] (CPM[c])	1.44 (2.18)	4.96 (2.91)	−1.36 (−4.59 to −2.45)	−6.81 (24)	<.001
Acc_FE[d] (CPM)	1.48 (2.28)	3.76 (2.35)	−1.00 (−3.22 to −1.33)	−4.99 (24)	<.001
Acc_BE[e] (CPM)	1.52 (2.49)	5.08 (2.71)	−1.11 (−4.88 to −2.24)	−5.56 (24)	<.001
Stereo[f] (arc sec)	780.00 (613.57)	448.80 (472.23)	0.69 (134.24 to 528.16)	3.47 (24)	.002
Supp[g] (n, % normal)	19 (76%)	24 (96%)	—[h]	—	—

[a]Log_AE: visual acuity of the amblyopic eye.

[b]Acc_AE: accommodative sensitivity of the amblyopic eye.

[c]CPM: cycles per minute.

[d]Acc_FE: accommodative sensitivity of the fellow eye.

[e]Acc_BE: binocular accommodative sensitivity.

[f]Stereo: stereopsis.

[g]Supp: suppression.

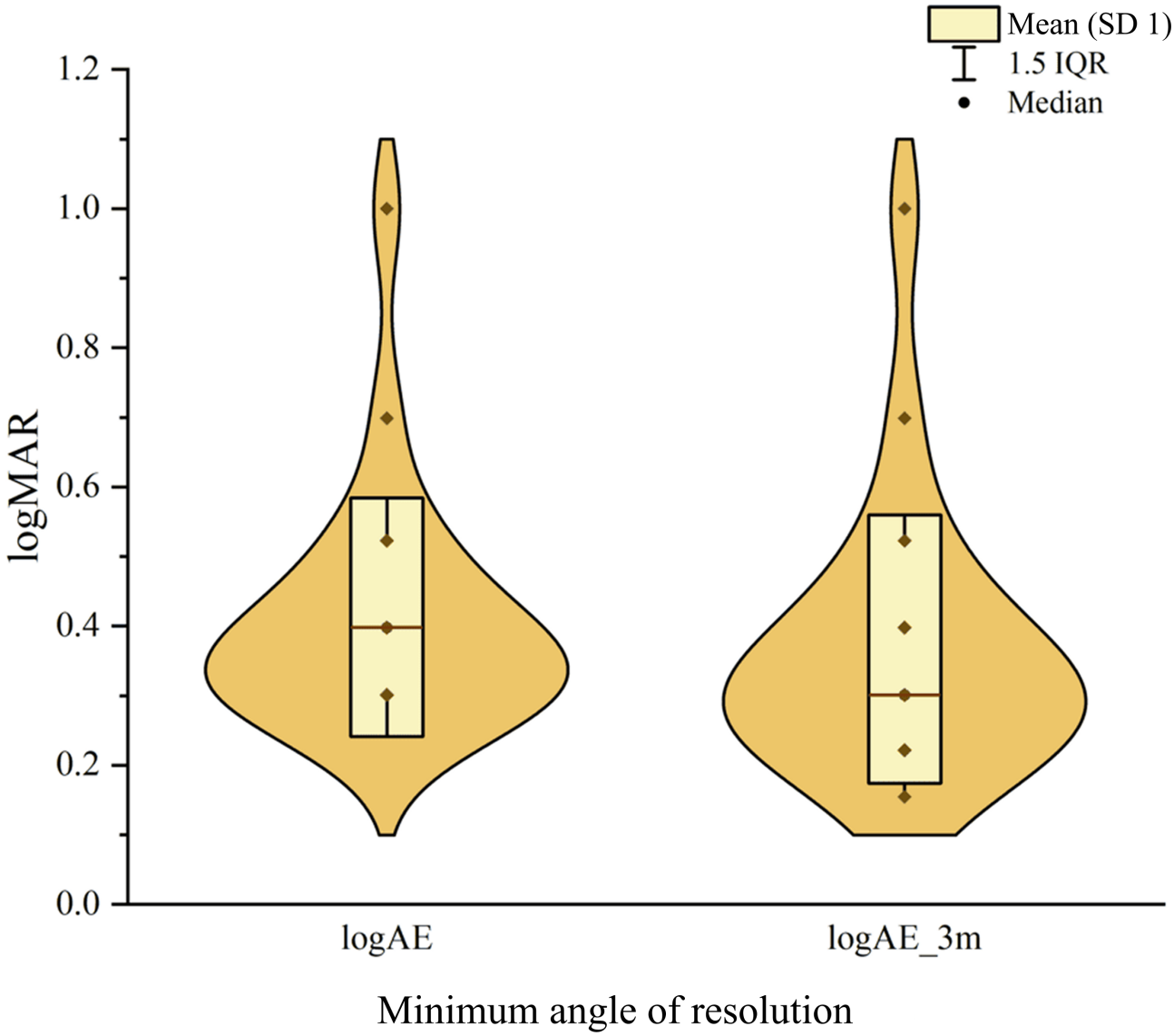
[h]Not applicable.

Changes in VA

Following the intervention, the amblyopic eye showed a significant improvement in logMAR VA (Figure 3). The mean VA of the amblyopic eye improved from 0.42 (SD 0.16) at baseline to 0.37 (SD 0.18) after the intervention, with a mean difference of 0.05. A paired-sample *t*-test revealed a t_{24} value of 4.82 (P<.001) and a 95% CI of 0.03 - 0.08, indicating a highly significant difference. Effect size analysis yielded a Cohen *d* of 0.96, suggesting a large improvement with clear clinical relevance.

XSL•FO
RenderX

Figure 3. Improvement in visual acuity of children with amblyopia before and after treatment. logAE: visual acuity of the amblyopic eye.

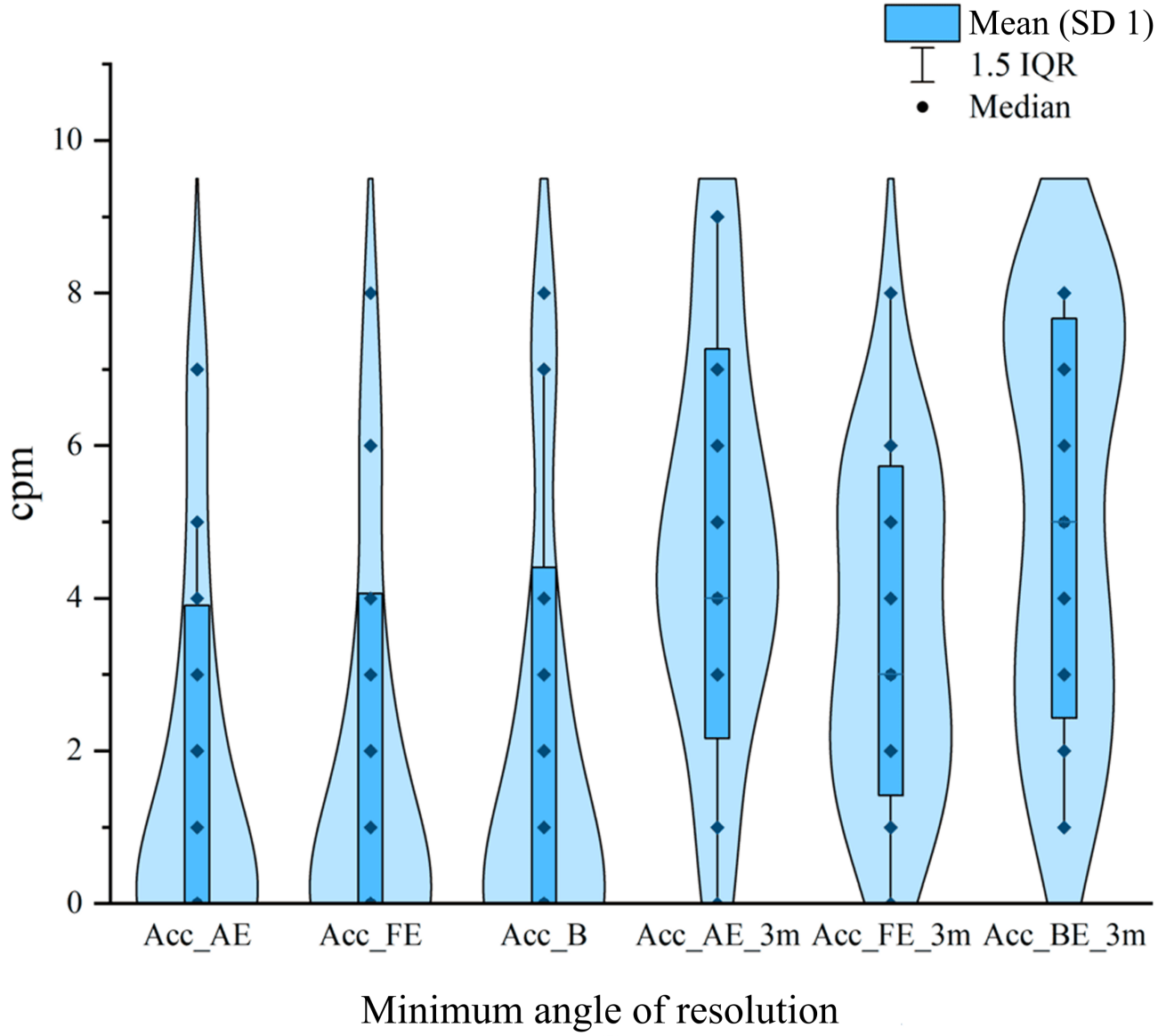


Accommodative Sensitivity

Following the intervention, both monocular and binocular accommodative abilities improved significantly (Figure 4). For the amblyopic eye, the mean accommodative sensitivity increased markedly from 1.44 (SD 2.18) before the intervention to 4.96 (SD 2.91) after the intervention, with a mean difference of –3.52. A paired *t*-test revealed a t_{24} value of –6.81 (P<.001) and a 95% CI of –4.59 to –2.45. The effect size (Cohen d=–1.36) indicated a large and clinically meaningful improvement in accommodative function of the amblyopic eye.

XSL•FO
RenderX

Figure 4. Improvement in binocular accommodative function of children with amblyopia before and after treatment. Acc_AE: accommodative sensitivity of the amblyopic eye; Acc_BE: binocular accommodative sensitivity; Acc_FE: accommodative sensitivity of the fellow eye; cpm: cycles per minute.



In the fellow eye, accommodative sensitivity also showed a significant enhancement, increasing from 1.48 (SD 2.28) to 3.76 (SD 2.35), with a mean difference of –2.28. The paired *t*-test yielded a t_{24} value of –4.99 (*P*<.001) and a 95% CI of –3.22 to –1.34, corresponding to a Cohen *d* of –1.00, suggesting a moderate-to-large facilitative effect on the fellow eye's accommodation.

Furthermore, binocular accommodative sensitivity increased significantly from 1.52 (SD 2.49) to 5.08 (SD 2.71), with a mean difference of –3.56. The 95% CI ranged from –4.88 to –2.24, and the paired *t*-test yielded t_{24}=–5.56 (*P*<.001). The effect size (Cohen *d*=–1.11) also reflected a substantial and clinically relevant improvement.

Overall, the intervention led to marked enhancements in both monocular and binocular accommodative abilities; improvement was greater in the amblyopic eye than in the fellow eye. These findings suggest that serious game–based visual training can effectively strengthen accommodative function in children and promote the recovery of binocular coordination.

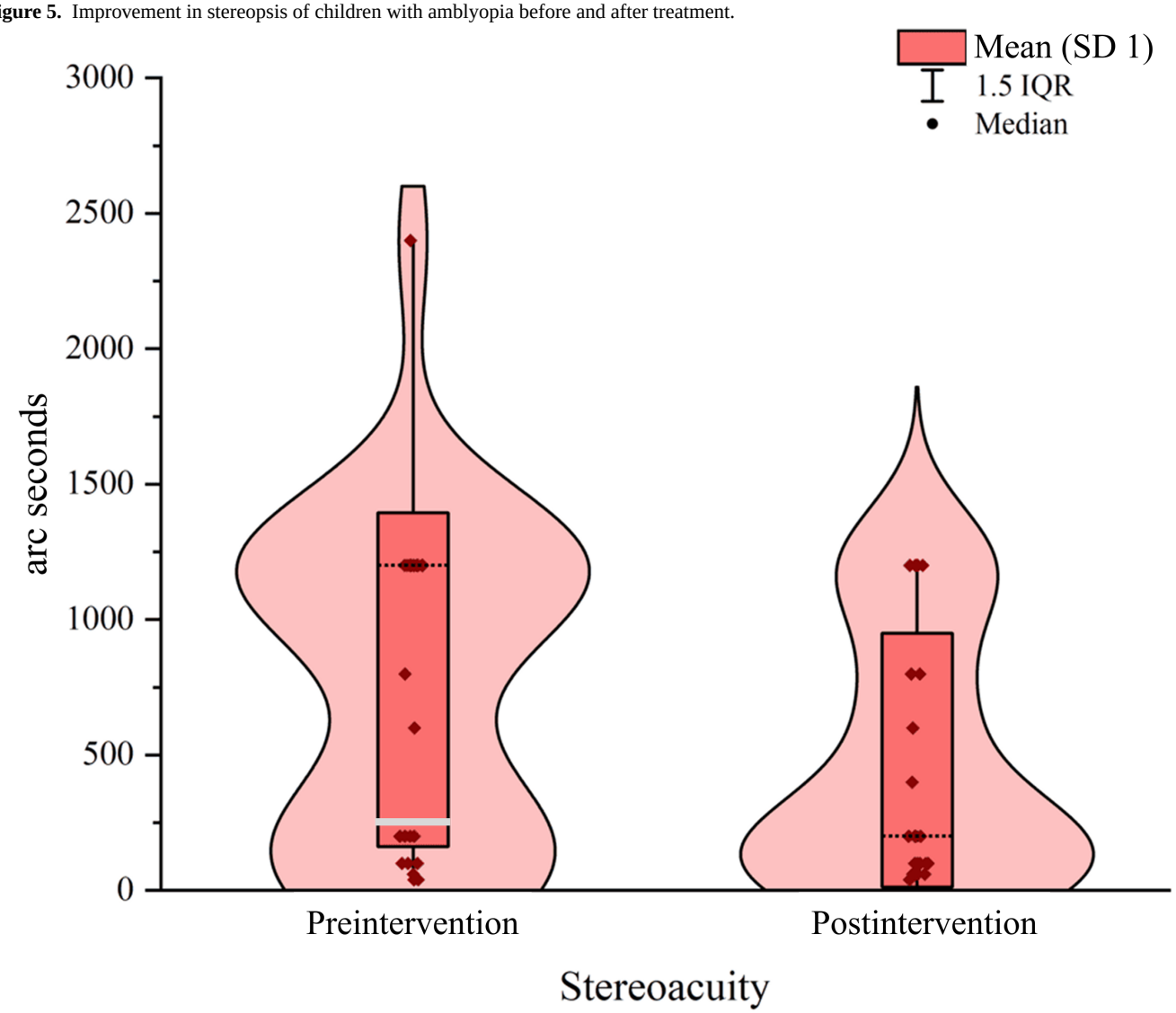
Stereoacuity

The intervention led to a significant improvement in stereoacuity, both statistically and clinically (Figure 5). Preintervention data indicated a mean stereoacuity of 780.00 (SD 613.57), which decreased significantly to 448.80 (SD 472.23) postintervention. This resulted in a mean difference of 331.20, with a 95% CI of 134.24-528.16. A paired *t*-test yielded a t_{24} value of 3.47 (*P*=.002), and Cohen *d* was 0.69, indicating a medium-to-large effect size. This demonstrates that the intervention had reliable statistical significance and certain clinical utility in improving participants' stereoacuity.

Supporting these statistical findings, the violin plot illustrated a significant postintervention shift toward lower stereoacuity values, with the median and primary distribution range moving downward. Data points became more concentrated in the range representing better stereoacuity levels, reflecting consistent improvements in participants' stereoacuity following the intervention.

XSL•FO
RenderX

Figure 5. Improvement in stereopsis of children with amblyopia before and after treatment.



Suppression

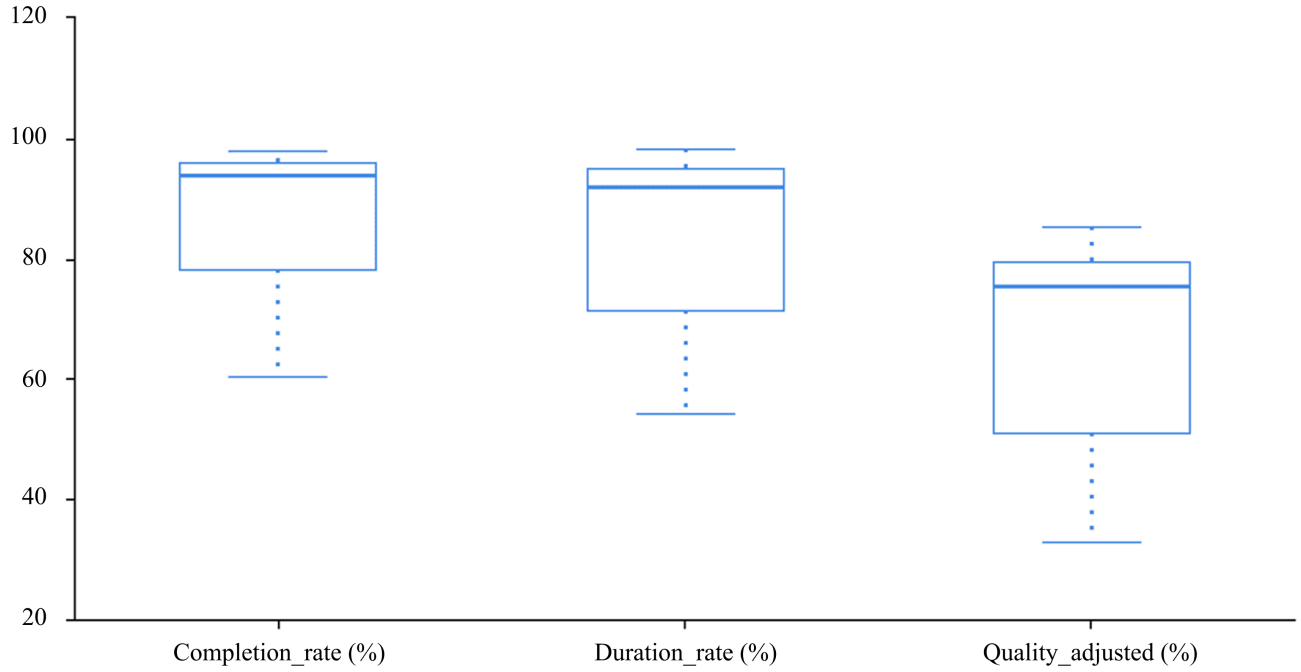
Before the intervention, 19 of 25 participants (76%) showed normal Worth 4 dot responses. The remaining 6 participants (24%) had abnormal findings, including monocular suppression or diplopia. After the intervention, 24 of 25 participants (96%) demonstrated normal Worth 4 dot responses, with only 1 participant showing residual monocular suppression.

Compliance Assessment

A total of 25 participants completed the compliance and usage behavior assessments (Figure 6). Among the scheduled 48 training sessions, the mean number of completed sessions was 42.32 (SD 5.60). The cumulative effective training duration was 1229 (SD 201.07) minutes, out of a total planned duration of 1440 minutes. The mean weekly compliance rate was 83.04% (SD 12.68%). The mean fixation ratio was 0.85 (SD 0.10), and the mean inactivity ratio was 0.10 (SD 0.06). The three compliance indicators were as follows: (1) completion rate (mean 88.16%, SD 11.67%), (2) duration rate (mean 85.41%, SD 13.95%), and (3) quality adjusted (mean 68.57%, SD 16.4%).

XSL•FO
RenderX

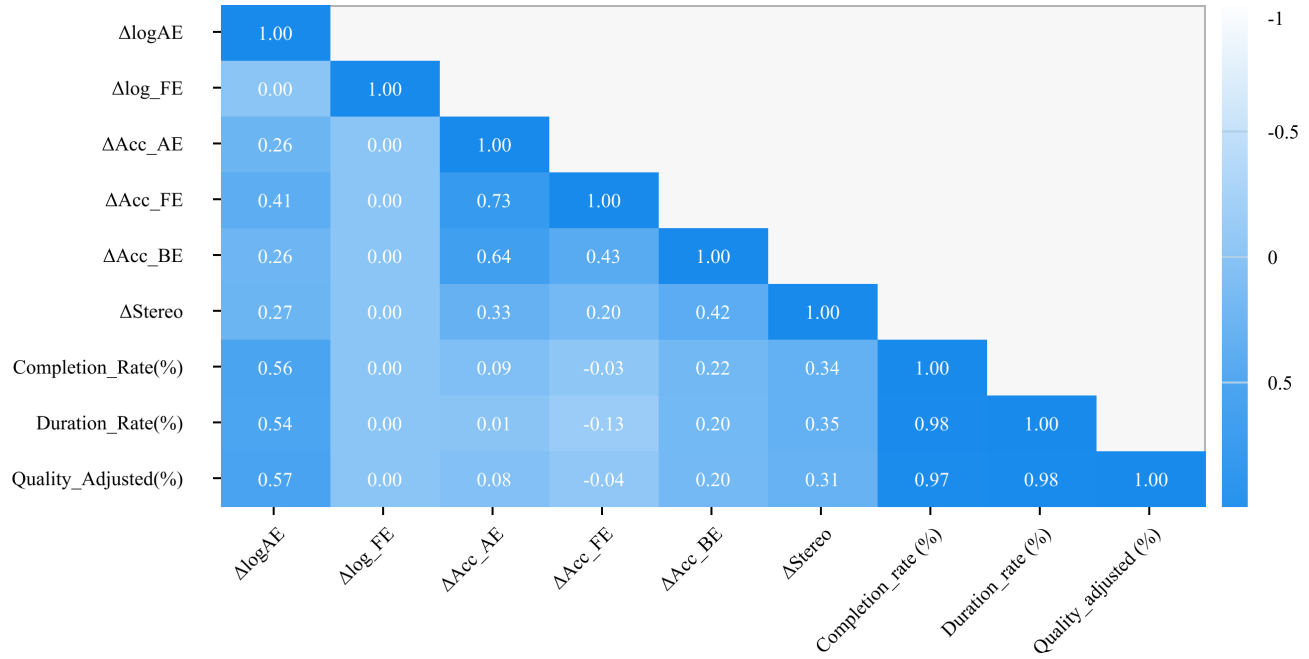
Figure 6. Overview of adherence metrics (completion, duration, and quality-adjusted).



Correlation Analysis

Improvement in VA of the amblyopic eye was significantly positively correlated with the change in accommodative function of the fellow eye (r=0.41, P=.04), session completion rate (r=0.56, P=.004), effective training duration rate (r=0.54, P=.006), and quality-adjusted compliance (r=0.57, P=.003) (Figure 7). No significant correlations were observed between amblyopic eye visual improvement and changes in fellow-eye VA, amblyopic-eye accommodation, binocular accommodation, or stereopsis (all P>.05).

Figure 7. Pearson correlation heatmap. Acc_AE: accommodative sensitivity of the amblyopic eye; Acc_BE: binocular accommodative sensitivity; Acc_FE: accommodative sensitivity of the fellow eye; logAE: visual acuity of the amblyopic eye; logFE: visual acuity of the fellow eye.



The compliance indicators were highly intercorrelated: session completion rate was strongly correlated with both effective training duration rate (r=0.98, P<.001) and quality-adjusted compliance (r=0.97, P<.001); similarly, effective training duration rate was highly correlated with quality-adjusted compliance (r=0.99, P<.001).

Regarding functional indices, changes in fellow-eye accommodation were positively correlated with both amblyopic-eye accommodation (r=0.73, P<.001) and binocular accommodation (r=0.43, P=.03). Additionally, changes in stereopsis were positively correlated with changes in binocular accommodation (r=0.42, P=.04). Given the number of pairwise correlations examined, these analyses are exploratory; no formal adjustment for multiple comparisons was applied, and marginal

XSL•FO
RenderX

associations (particularly the correlation between BCVA improvement and fellow-eye accommodative sensitivity, P=.04) should be interpreted with caution.

Discussion

Principal Findings

In this single-arm feasibility study, 12-week serious-game training integrating PL and stereopsis tasks produced (1) a modest but statistically significant BCVA gain in the amblyopic eye (mean 0.05, SD 0.056 logMAR); (2) large within-participants improvements in accommodative sensitivity for the amblyopic, fellow, and binocular conditions (Cohen d_z≈1.36, 1.00, and 1.11, respectively); (3) a medium improvement in stereoacuity (Cohen d_z≈0.69); and (4) an increase in normal Worth 4 dot responses from 19/25 (76%) at baseline to 24/25 (96%) postintervention, with 5 of 6 participants with baseline suppression or diplopia showing normalization. Device-logged metrics indicated high adherence under supervised clinical conditions, and exploratory correlations suggested that greater adherence and changes in accommodative function were associated with VA gains. Given the single-group design, small sample, absence of multiple-comparison adjustment for the correlation matrix, and the heterogeneous 7‐12-year age range, these findings should be regarded as hypothesis-generating. The preliminary improvements observed in this late-childhood cohort are consistent with the presence of residual plasticity but require confirmation in age-stratified RCTs before firm conclusions can be drawn.

Comparison With Previous Work

This study systematically evaluated the efficacy of gamified interventions in improving visual functions among children with amblyopia, validating the clinical applicability of integrating PL and stereopsis training strategies. The findings demonstrated that, compared with traditional treatments, the gamified intervention not only effectively enhanced basic VA but also improved accommodative function and showed notable efficacy in the restoration of stereopsis.

Compared with baseline, BCVA in the amblyopic eye improved from 0.42 (SD 0.16) to 0.37 (SD 0.18) logMAR (mean change=0.05 logMAR, t_{24}=4.82, P<.001), corresponding to an improvement of approximately 0.5 lines on a logMAR chart. Because this study used a single-arm, pre-post feasibility design with clinic-based, optometrist-supervised sessions, the observed within-participants improvement should be interpreted cautiously. Nonspecific factors, such as practice effects on visual testing, natural maturation in children, and the Hawthorne effect associated with supervised participation, may have contributed to the change. Therefore, direct comparisons with RCTs are not appropriate; randomized, controlled studies with adequate sample sizes are required to establish comparative efficacy.

It is noteworthy that To et al [55] achieved a VA improvement of 0.17 logMAR in adult patients, which closely aligns with the results of this study. However, their research implemented a high-intensity training protocol (1‐2 h daily), which might be less feasible for children. In contrast, this study used shorter sessions through gamified design, limiting training to 30 minutes per session to accommodate children's attention spans and ensure higher adherence. This approach not only enhanced children's engagement but also maintained treatment effectiveness, suggesting that shorter, frequent sessions may be preferable for pediatric populations.

Moreover, the accommodation function, an important yet often overlooked factor in visual rehabilitation for amblyopia, significantly improved in response to gamified interventions, as indicated by the ±2.00 D flipper test results. The mean accommodative sensitivity increased markedly, with 6 of 25 participants (24%) achieving normal levels (>8 CPM), representing a substantial improvement compared with baseline. Accommodative function is likely due to the dynamic focusing tasks embedded within the game mechanics, requiring rapid identification of variable virtual elements, thus effectively training the responsiveness of the ciliary muscles. Unlike previous studies, such as those by Herbison et al [56] and To et al [55], which did not specifically quantify accommodative sensitivity due to their focus on dichoptic stimuli alone, this study fills this critical gap and highlights the necessity of including dynamic focusing mechanisms in amblyopia training programs.

Regarding stereopsis, the median stereoacuity improved significantly from 780.00 arc seconds to 448.80 arc seconds (t_{24}=3.47, P=.002), and the RDS test pass rate increased from 2 of 25 participants (8%) to 6 (24%). Such consistent improvements were not observed in previous studies; for example, the Frisby stereoacuity test used by Herbison et al [56] showed no statistically significant enhancement, likely due to the majority of their participants having strabismic amblyopia without targeted fusion training. Similarly, To et al [55] reported improved stereopsis using polarized dichoptic Tetris, but their improvements did not surpass the achievements observed in this study's targeted, dynamic, and repetitive training approach.

Despite promising outcomes, this study's generalizability is limited by its single-center, small-sample design (n=25). Future research should incorporate advanced technologies, such as VR, and adopt multicenter RCT designs to further validate long-term efficacy. Additionally, exploring advanced display technologies, like VR, and conducting multicenter RCTs could further verify the long-term effectiveness of this approach. Future research might also quantify neuroplasticity maintenance thresholds (eg, monthly BCVA decay <0.02 logMAR) through multicenter longitudinal studies and establish personalized retraining recommendation systems based on artificial intelligence prediction models. To accommodate cognitive differences across age groups, developing age-specific game content libraries could be beneficial. Moreover, the simplicity of the current games introduces a potential learning curve effect, where initial performance gains may partly reflect increased familiarity with the tasks rather than true visual improvement. In addition, outcome measures were collected only at baseline and posttraining, without interim assessments. This limits the ability to characterize the trajectory of visual improvement or disentangle transient learning effects from long-term treatment gains. Future studies should incorporate intermediate testing sessions to better capture the temporal dynamics of therapeutic

XSL•FO
RenderX

response. Ultimately, these efforts aim to transition amblyopia management toward a comprehensive "technology-physiology-behavior" multidimensional intervention paradigm.

This study demonstrated that improvements in amblyopic eye VA were moderately and positively correlated with compliance-related indicators, including session completion rate, effective training duration rate, and quality-adjusted compliance—all of which reached statistical significance. This finding suggests that compliance is a key driver of short-term visual gains. The result is consistent with previous multicenter studies emphasizing that the efficacy of home-based binocular training is often constrained by adherence. However, large RCTs have not observed a clear dose–response relationship in actual data—training duration, or the magnitude of interocular contrast adjustment was not significantly associated with VA improvement [11]. These results indicate that good compliance is a necessary condition for benefit, but a "more-is-better" linear dose effect has not been substantiated.

In contrast, no significant correlation was found between improvement in amblyopic eye VA and changes in stereopsis. This finding aligns with several pediatric studies reporting that improvements in VA and stereopsis during short-term binocular training may occur asynchronously. Considering the correlation pattern observed in this study (nonsignificant association between VA and stereopsis changes), our results are consistent with previous findings. However, because this study did not include multiple timepoint analyses, the temporal sequence of these effects could not be determined, and the underlying mechanisms require further validation through prospective longitudinal data. Notably, stereopsis improvement correlated only with changes in binocular accommodation, not with VA improvement, suggesting that binocular coordination mechanisms—such as binocular accommodation and fusion—may play a more direct role in stereopsis recovery, rather than being mediated by monocular VA enhancement.

A moderate correlation was identified between improvement in BCVA and changes in fellow-eye accommodative sensitivity (r=0.41, P=.04), whereas no significant association was found with amblyopic-eye accommodative changes (P>.05). This pattern appears counterintuitive, as improvements in amblyopic-eye VA would typically be expected to relate to changes within the same eye. One possible interpretation is that binocular coordination or interocular facilitation mechanisms might manifest initially through compensatory adjustments in the fellow eye. However, given the small and clinically heterogeneous sample, this correlation may also represent a statistical artifact or reflect measurement variability. Accordingly, the result should be interpreted with caution, and future studies with larger samples and objective physiological assessments are needed to clarify the underlying mechanisms.

Overall, the correlation patterns observed in this study support the following interpretation: in home-based gamified binocular training, compliance serves as the critical threshold determining whether measurable improvement occurs, yet evidence for a linear dose–response effect remains insufficient. The restoration of stereopsis appears to rely more on binocular coordination and suppression-rebalancing mechanisms than on simple monocular VA enhancement. Unlike large-scale RCTs that reported no overall superiority or dose–response effect, our findings reveal individual-level positive associations detectable by statistical modeling—highlighting a practical direction for optimizing therapeutic efficacy through improved engagement and training quality.

All training sessions in this feasibility study were conducted under optometrist supervision in the clinic to ensure compliance and protocol fidelity. Therefore, the high adherence observed likely reflects the benefits of supervised conditions and may not directly translate to unsupervised home environments. Although the Magic Eye platform is designed to support remote logging and could potentially enable home-based deployment in the future, this feasibility phase did not evaluate such functionality. Future studies should explicitly test unsupervised, home-based implementation to determine whether similar adherence and efficacy can be maintained outside the clinic.

The current framework provides preliminary evidence that a gamified binocular training system can be feasibly implemented in clinical settings. With further validation and appropriate safety and monitoring mechanisms, this approach could be extended to establish a scalable "home training–community screening–hospital evaluation" network, potentially reducing clinical workload and improving accessibility in resource-limited regions. However, claims regarding reduced hospital visits or real-world cost-effectiveness remain hypothetical and require confirmation in prospective trials incorporating economic evaluation and unsupervised home-based use.

Limitations

This study has 3 primary limitations. First, as a single-center, single-arm, pre-post feasibility study with a small and clinically heterogeneous sample (n=25), this work cannot exclude practice or learning effects as alternative explanations for the observed improvements, thereby limiting causal inference. The limited sample size also reduced statistical power and precluded subgroup analyses, which may have obscured potential differences in treatment responsiveness across amblyopia subtypes and treatment histories. To mitigate bias, standard refractive correction was maintained throughout, and analyses were restricted to within-group changes and correlations without making superiority claims over patching or atropine. Future research will use multicenter randomized or stratified controlled trials with appropriate comparator arms (patching, atropine, nongamified, or sham conditions) and independent data monitoring to strengthen causal inference. Second, outcome assessments were limited to baseline and the 12-week end point, which precludes characterization of trajectories and durability and prevents establishing the temporal sequence of short-term learning versus longer-term retention; a 6-month follow-up has been prespecified, and subsequent studies will add interim assessments at weeks 2, 4, and 8 and extend follow-up to ≥6 - 12 months. Third, the intervention bundles multiple components (binocular rebalancing, disparity-based training, and gamification for motivation), making it difficult to isolate the specific active ingredients; in this report, gamification is conservatively treated as an adherence-enhancing strategy rather

XSL•FO
RenderX

than an independent therapeutic mechanism. Future studies will use factorial or multiarm designs (eg, gamified vs nongamified interactive, stereoscopic vs nonstereoscopic, and monocular vs binocular) to disentangle mechanisms and test both independent and interactive effects. In addition, the inclusion of children aged 7‐12 years introduces developmental heterogeneity in residual visual plasticity and treatment responsiveness. The sample size in this study was underpowered for age-stratified analyses; consequently, the reported effects should be interpreted as averaged signals across a late-childhood cohort rather than uniform efficacy at each age, and future trials should incorporate prespecified age strata.

Conclusions

This study integrated PL with stereoscopic perception training within a serious game intervention administered to children aged 7‐12 years with amblyopia over 12 weeks (4 sessions per wk, 30 min each). Within a clinically feasible time frame, it achieved significant improvements in multidimensional visual functions: VA in the amblyopic eye showed modest yet statistically significant improvement, accommodation agility markedly increased, stereopsis demonstrated moderate optimization, and overall suppression status improved, with normal Worth 4 dot responses increasing from 19/25 to 24/25. Platform-based logging demonstrated good adherence and effective training duration under fully supervised in-clinic conditions, indicating that the protocol is feasible and monitorable in a controlled clinical setting. Whether similar adherence can be maintained in unsupervised home environments remains unknown and should be explicitly evaluated in future studies. Potential benefits related to affordability and resource efficiency remain hypothetical and require confirmation through dedicated economic evaluations and comparative studies. Collectively, serious game-based training offers a promising complementary therapeutic pathway for pediatric amblyopia. Objective compliance data also provides an evidence base for individualized prescription refinement and future multicenter RCTs, informing dose setting and methodological optimization.

Acknowledgments

We thank the participating children and their families, the ophthalmology staff and optometrists at Luohu People's Hospital for coordination and assessments, and the research assistants for data handling and figure preparation. Technical support for the Magic Eye device was provided by Shenzhen Vizhi Holographic Technology Co, Ltd, which had no role in the design, conduct, analysis, or reporting of this study. We used ChatGPT-4o (OpenAI) for limited grammar checking; the authors are fully responsible for the final content.

Funding

This work was supported by the Research Project of the Guangdong Provincial Department of Education (grant 2024CJPT002). The funder had no role in study design, data collection, analysis, interpretation, or manuscript preparation.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: YM, PC, FC, CL
Data curation: CD, JL, QW
Formal analysis: CL
Funding acquisition: PC
Investigation: MH, QW
Methodology: MH, CL
Software: CD, JL
Supervision: PC
Writing – original draft: YM, MH
Writing – review & editing: FC, CL, PC

Conflicts of Interest

None declared.

Multimedia Appendix 1
Video of "Panda Eats Bamboo" operational demonstration.
[MP4 File, 9215 KB - games_v13i1e77402_app1.mp4]

Multimedia Appendix 2
Video of "Fishing Master" operational demonstration.

XSL•FO
RenderX

[MP4 File, 10424 KB - games_v13i1e77402_app2.mp4]

References

1. Attebo K, Mitchell P, Cumming R, Smith W, Jolly N, Sparkes R. Prevalence and causes of amblyopia in an adult population. Ophthalmology 1998 Jan;105(1):154-159. [doi: 10.1016/s0161-6420(98)91862-0] [Medline: 9442792]
2. Mostafaie A, Ghojazadeh M, Hosseinifard H, et al. A systematic review of amblyopia prevalence among the children of the world. Rom J Ophthalmol 2020;64(4):342-355. [doi: 10.22336/rjo.2020.56] [Medline: 33367172]
3. Rahi JS, Logan S, Timms C, Russell-Eggitt I, Taylor D. Risk, causes, and outcomes of visual impairment after loss of vision in the non-amblyopic eye: a population-based study. The Lancet 2002 Aug;360(9333):597-602. [doi: 10.1016/S0140-6736(02)09782-9]
4. McKee SP, Levi DM, Movshon JA. The pattern of visual deficits in amblyopia. J Vis 2003;3(5):380-405. [doi: 10.1167/3.5.5] [Medline: 12875634]
5. Awan M, Proudlock FA, Gottlob I. A randomized controlled trial of unilateral strabismic and mixed amblyopia using occlusion dose monitors to record compliance. Invest Ophthalmol Vis Sci 2005 Apr;46(4):1435-1439. [doi: 10.1167/iovs.04-0971] [Medline: 15790912]
6. Wallace MP, Stewart CE, Moseley MJ, et al. Compliance with occlusion therapy for childhood amblyopia. Invest Ophthalmol Vis Sci 2013 Sep 17;54(9):6158-6166. [doi: 10.1167/iovs.13-11861] [Medline: 23882695]
7. Bhola R, Keech RV, Kutschke P, Pfeifer W, Scott WE. Recurrence of amblyopia after occlusion therapy. Ophthalmology 2006 Nov;113(11):2097-2100. [doi: 10.1016/j.ophtha.2006.04.034] [Medline: 17074568]
8. A randomized trial of atropine vs patching for treatment of moderate amblyopia in children. Arch Ophthalmol 2002 Mar 1;120(3):268. [doi: 10.1001/archopht.120.3.268]
9. Li J, Spiegel DP, Hess RF, et al. Dichoptic training improves contrast sensitivity in adults with amblyopia. Vision Res 2015 Sep;114:161-172. [doi: 10.1016/j.visres.2015.01.017] [Medline: 25676883]
10. Papageorgiou E, Asproudis I, Maconachie G, Tsironi EE, Gottlob I. The treatment of amblyopia: current practice and emerging trends. Graefes Arch Clin Exp Ophthalmol 2019 Jun;257(6):1061-1078. [doi: 10.1007/s00417-019-04254-w]
11. Levi DM, Li RW. Improving the performance of the amblyopic visual system. Philos Trans R Soc Lond B Biol Sci 2009 Feb 12;364(1515):399-407. [doi: 10.1098/rstb.2008.0203] [Medline: 19008199]
12. Huang CB, Zhou Y, Lu ZL. Broad bandwidth of perceptual learning in the visual system of adults with anisometropic amblyopia. Proc Natl Acad Sci U S A 2008 Mar 11;105(10):4068-4073. [doi: 10.1073/pnas.0800824105] [Medline: 18316716]
13. Ding J, Levi DM. Recovery of stereopsis through perceptual learning in human adults with abnormal binocular vision. Proc Natl Acad Sci U S A 2011 Sep 13;108(37):E733-E741. [doi: 10.1073/pnas.1105183108] [Medline: 21896742]
14. Polat U, Ma-Naim T, Spierer A. Treatment of children with amblyopia by perceptual learning. Vision Res 2009 Oct;49(21):2599-2603. [doi: 10.1016/j.visres.2009.07.008] [Medline: 19622368]
15. Zhou Y, Huang C, Xu P, et al. Perceptual learning improves contrast sensitivity and visual acuity in adults with anisometropic amblyopia. Vision Res 2006 Mar;46(5):739-750. [doi: 10.1016/j.visres.2005.07.031] [Medline: 16153674]
16. Zhai J, Chen M, Liu L, et al. Perceptual learning treatment in patients with anisometropic amblyopia: a neuroimaging study. Br J Ophthalmol 2013 Nov;97(11):1420-1424. [doi: 10.1136/bjophthalmol-2013-303778] [Medline: 24037607]
17. Dye MWG, Green CS, Bavelier D. The development of attention skills in action video game players. Neuropsychologia 2009 Jul;47(8-9):1780-1789. [doi: 10.1016/j.neuropsychologia.2009.02.002] [Medline: 19428410]
18. Schoups A, Vogels R, Qian N, Orban G. Practising orientation identification improves orientation coding in V1 neurons. Nature New Biol 2001 Aug 2;412(6846):549-553. [doi: 10.1038/35087601] [Medline: 11484056]
19. Gilbert CD, Sigman M, Crist RE. The neural basis of perceptual learning. Neuron 2001 Sep 13;31(5):681-697. [doi: 10.1016/s0896-6273(01)00424-x] [Medline: 11567610]
20. Hess RF, Thompson B, Black JM, et al. An iPod treatment of amblyopia: an updated binocular approach. Optometry 2012 Feb 15;83(2):87-94. [doi: 10.1016/j.optm.2011.08.013] [Medline: 23231369]
21. Qian N. Binocular disparity and the perception of depth. Neuron 1997 Mar;18(3):359-368. [doi: 10.1016/s0896-6273(00)81238-6] [Medline: 9115731]
22. Poggio GF, Poggio T. The analysis of stereopsis. Annu Rev Neurosci 1984;7:379-412. [doi: 10.1146/annurev.ne.07.030184.002115] [Medline: 6370081]
23. Gonzalez F, Perez R. Neural mechanisms underlying stereoscopic vision. Prog Neurobiol 1998 Jun;55(3):191-224. [doi: 10.1016/s0301-0082(98)00012-4] [Medline: 9643554]
24. Levi DM. Rethinking amblyopia 2020. Vision Res 2020 Nov;176:118-129. [doi: 10.1016/j.visres.2020.07.014] [Medline: 32866759]
25. Birch EE. Amblyopia and binocular vision. Prog Retin Eye Res 2013 Mar;33:67-84. [doi: 10.1016/j.preteyeres.2012.11.001] [Medline: 23201436]
26. Vedamurthy I, Nahum M, Huang SJ, et al. A dichoptic custom-made action video game as a treatment for adult amblyopia. Vision Res 2015 Sep;114:173-187. [doi: 10.1016/j.visres.2015.04.008] [Medline: 25917239]

XSL•FO
RenderX

27. Holmes JM, Manh VM, Lazar EL, et al. Effect of a binocular iPad game vs part-time patching in children aged 5 to 12 years with amblyopia: a randomized clinical trial. JAMA Ophthalmol 2016 Dec 1;134(12):1391-1400. [doi: 10.1001/jamaophthalmol.2016.4262] [Medline: 27812703]
28. Bari G, D'Ambrosio A, Petrizzelli F, Laborante A. Amblyopia rehabilitation: a preliminary study on the efficacy of an alternative therapeutic method within Italian patients. Adv Rehabil Sci Pract 2024;13:27536351241297249. [doi: 10.1177/27536351241297249] [Medline: 39574834]
29. Bocqué C, Wang J, Rickmann A, Julich-Haertel H, Kaempf U, Januschowski K. Gamification to support adherence to a therapeutic ambylopia treatment for children: retrospective study using a focal ambient visual acuity stimulation game. JMIR Pediatr Parent 2023 Feb 1;6:e32282. [doi: 10.2196/32282] [Medline: 36724007]
30. Zhu W, Gu S, Li J, Lin J, Hu C, Liu R. Transformative gamified binocular therapy for unilateral amblyopia in young children: pilot prospective efficacy and safety study. JMIR Serious Games 2025 Jan 16;13:e63384. [doi: 10.2196/63384] [Medline: 39819474]
31. Hirota M, Okumura Y, Nagino K, et al. Safety evaluation in healthy adults of motion-based virtual reality dichoptic training for pediatric patients with amblyopia: prospective intervention study. JMIR Form Res 2025 Jun 17;9:e69801. [doi: 10.2196/69801] [Medline: 40526901]
32. Li RW, Li BZ, Chat SW, Levi DM. Playing stereoscopic video games enhances the precision but not the accuracy of depth perception. Sci Rep 2024 Dec 30;14(1):31709. [doi: 10.1038/s41598-024-82194-0] [Medline: 39738325]
33. Morse J, Oatts JT. Amblyopia screening: the current state and opportunities for optimization. Expert Rev Med Devices 2025 Jan;22(1):1-4. [doi: 10.1080/17434440.2024.2449490] [Medline: 39760640]
34. Levi DM, Li RW. Perceptual learning as a potential treatment for amblyopia: a mini-review. Vision Res 2009 Oct;49(21):2535-2549. [doi: 10.1016/j.visres.2009.02.010] [Medline: 19250947]
35. Astle AT, Webb BS, McGraw P, Optics P. Can perceptual learning be used to treat amblyopia beyond the critical period of visual development? Ophthalmic Physiologic Optic 2011 Nov;31(6):564-573. [doi: 10.1111/j.1475-1313.2011.00873.x]
36. Pediatric Eye Disease Investigator Group, Holmes JM, Manny RE, et al. A randomized trial of binocular Dig Rush game treatment for amblyopia in children aged 7 to 12 years. Ophthalmology 2019 Mar;126(3):456-466. [doi: 10.1016/j.ophtha.2018.10.032] [Medline: 30352226]
37. Scheiman MM, Hertle RW, Beck RW, et al. Randomized trial of treatment of amblyopia in children aged 7 to 17 years. Arch Ophthalmol 2005 Apr;123(4):437-447. [doi: 10.1001/archopht.123.4.437] [Medline: 15824215]
38. Wang J. Compliance and patching and atropine amblyopia treatments. Vision Res 2015 Sep;114:31-40. [doi: 10.1016/j.visres.2015.02.012] [Medline: 25743080]
39. HUBEL DH, WIESEL TN. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J Physiol 1962 Jan;160(1):106-154. [doi: 10.1113/jphysiol.1962.sp006837] [Medline: 14449617]
40. Born RT, Bradley DC. Structure and function of visual area MT. Annu Rev Neurosci 2005;28:157-189. [doi: 10.1146/annurev.neuro.26.041002.131052] [Medline: 16022593]
41. Andersen RA, Buneo CA. Intentional maps in posterior parietal cortex. Annu Rev Neurosci 2002;25:189-220. [doi: 10.1146/annurev.neuro.25.112701.142922] [Medline: 12052908]
42. Sagi Y, Tavor I, Hofstetter S, Tzur-Moryosef S, Blumenfeld-Katzir T, Assaf Y. Learning in the fast lane: new insights into neuroplasticity. Neuron 2012 Mar 22;73(6):1195-1203. [doi: 10.1016/j.neuron.2012.01.025] [Medline: 22445346]
43. Lai YH, Wang HZ, Hsu HT. Development of visual acuity in preschool children as measured with Landolt C and Tumbling E charts. J AAPOS 2011 Jun;15(3):251-255. [doi: 10.1016/j.jaapos.2011.03.010] [Medline: 21777798]
44. Schulze-Bonsel K, Feltgen N, Burau H, Hansen L, Bach M. Visual acuities "hand motion" and "counting fingers" can be quantified with the Freiburg visual acuity test. Invest Ophthalmol Vis Sci 2006 Mar;47(3):1236-1240. [doi: 10.1167/iovs.05-0981] [Medline: 16505064]
45. Holladay JT. Proper method for calculating average visual acuity. J Refract Surg 1997;13(4):388-391. [doi: 10.3928/1081-597X-19970701-16] [Medline: 9268940]
46. Brooks SE, Johnson D, Fischer N. Anisometropia and binocularity. Ophthalmology 1996 Jul;103(7):1139-1143. [doi: 10.1016/s0161-6420(96)30555-1] [Medline: 8684806]
47. Jeon HS, Choi HY. Binocular function test. In: Primary Eye Examination 2019:71-82. [doi: 10.1007/978-981-10-6940-6_6]
48. Hess RF, Thompson B. Amblyopia and the binocular approach to its therapy. Vision Res 2015 Sep;114(4-16):4-16. [doi: 10.1016/j.visres.2015.02.009] [Medline: 25906685]
49. Radhakrishnan H, Allen PM, Charman WN. Dynamics of accommodative facility in myopes. Invest Ophthalmol Vis Sci 2007 Sep;48(9):4375-4382. [doi: 10.1167/iovs.07-0269] [Medline: 17724230]
50. Garcı´a Á, Cacho P, Lara F, Megı´as R. The relation between accommodative facility and general binocular dysfunction. Ophthalmic Physiologic Optic 2000 Mar;20(2):98-104. [doi: 10.1046/j.1475-1313.2000.00478.x]
51. McKENZIE KM, Kerr SR, Rouse MW, DeLAND PN. Study of accommodative facility testing reliability. Optom Vis Sci 1987 Mar;64(3):186-194. [doi: 10.1097/00006324-198703000-00005]
52. Scheiman M, Cotter S, Kulp MT, et al. Treatment of accommodative dysfunction in children: results from a randomized clinical trial. Optom Vis Sci 2011 Nov;88(11):1343-1352. [doi: 10.1097/OPX.0b013e31822f4d7c] [Medline: 21873922]

XSL•FO
RenderX

53. Frisby JP. Random-dot stereograms for clinical assessment of stereopsis in children. Dev Med Child Neurol 1975 Dec;17(6):802-806. [doi: 10.1111/j.1469-8749.1975.tb04709.x] [Medline: 1205003]
54. Birch E, Williams C, Drover J, et al. Randot Preschool Stereoacuity Test: normative data and validity. J AAPOS 2008 Feb;12(1):23-26. [doi: 10.1016/j.jaapos.2007.06.003] [Medline: 17720573]
55. To L, Thompson B, Blum JR, Maehara G, Hess RF, Cooperstock JR. A game platform for treatment of amblyopia. IEEE Trans Neural Syst Rehabil Eng 2011 Jun;19(3):280-289. [doi: 10.1109/TNSRE.2011.2115255] [Medline: 21335317]
56. Herbison N, MacKeith D, Vivian A, et al. Randomised controlled trial of video clips and interactive games to improve vision in children with amblyopia using the I-BiT system. Br J Ophthalmol 2016 Nov;100(11):1511-1516. [doi: 10.1136/bjophthalmol-2015-307798] [Medline: 26951772]

Abbreviations

BCVA: best-corrected visual acuity
CPM: cycles per minute
PL: perceptual learning
RCT: randomized controlled trial
RDS: random-dot stereograms
VA: visual acuity
VR: virtual reality

Edited by A Coristine; submitted 13.05.25; peer-reviewed by E Oliveira, L Quevedo-Junyent, T Lima; accepted 11.11.25; published 17.12.25.

Please cite as:
Mo Y, Chen P, Hou M, Cui F, Luo C, Deng C, Lin J, Wu Q
Serious Games Integrating Perceptual Learning and Stereopsis Training in Children With Amblyopia: Single-Arm Pre-Post Feasibility Study
JMIR Serious Games 2025;13:e77402
URL: https://games.jmir.org/2025/1/e77402
doi:10.2196/77402

© Yiwei Mo, Peng Chen, Manting Hou, Fang Cui, Chao Luo, Chao Deng, Jinkang Lin, Qi Wu. Originally published in JMIR Serious Games (https://games.jmir.org), 17.12.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Development of a Serious Game App (Digimenz) for Patients With Dementia: Prospective Pilot Study for Usability Testing in Inpatient Treatment and Long-Term Care

Sören Freerik Brähmer[1,2], MSc; Benjamin Iffland[1], Dr rer nat; Stefan Kreisel[2], Dr Med; Martin Driessen[2], Prof Dr Med; Eva M Trompetter[2], MSc; Meret Schomburg[2], MSc; Max Toepper[2*], Dr rer nat; Carolin Steuwe[2*], Dr rer nat

[1]Clinical Psychology and Psychotherapy Working Group, Department of Psychology, Bielefeld University, Universitätsstr 25, Bielefeld, Germany

[2]Department of Psychiatry and Psychotherapy, University Medical Center OWL, Evangelisches Klinikum Bethel, Bielefeld University, Bielefeld, Germany

[*]these authors contributed equally

Corresponding Author:
Sören Freerik Brähmer, MSc
Clinical Psychology and Psychotherapy Working Group, Department of Psychology, Bielefeld University, Universitätsstr 25, Bielefeld, Germany

Abstract

Background: In the face of an increasing treatment need among people with dementia, effective and efficient interventions with a focus on quality of life need to be established. In this context, serious games have received increasing attention. However, there is a lack of apps specifically designed for people with dementia.

Objective: In this prospective pilot study, we examined the usability of a newly developed serious game app ("Digimenz").

Methods: A total of 43 people with cognitive impairment and mild to severe dementia completed the repeated-measures study procedure. Participants were recruited from an inpatient geriatric psychiatric ward and a long-term care facility. Participants were asked to complete 4 conditions in randomized order, including playing 3 different serious games (experimental conditions) and reading a newspaper (control condition). Each condition was completed once, and the total duration was 60 to 90 minutes per participant. Data on app usability were collected through self-ratings and observation after each condition. We tested for differences in usability among the conditions and the recruitment sites, and analyzed the relation of usability to cognitive capacity.

Results: The serious games were accepted in both settings (long-term care: 30/30, 100% interested; psychiatric ward: 31/41, 76% interested), although study completion was lower in the psychiatric subsample (15/41, 37%) than in the long-term care subsample (28/30, 93%). Global usability was rated good (System Usability Scale global mean score: 79). More severely impaired patients had more pronounced difficulties in learning how to play the games ($\rho_{MMSE,\ Learnability}$=−0.61, 95% CI −0.78 to −0.36; *P*<.001) and playing them alone ($\rho_{MMSE,\ Support}$=−0.49, 95% CI −0.69 to −0.19; *P*<.001). Nevertheless, playing the games was associated with a more positive mood (likelihood ratio χ^2_3=25.09; *P*<.001), independent of the level of cognitive functioning (likelihood ratio χ^2_1=0.64; *P*=.42). All games were played with a moderate error rate (0.19 - 0.49).

Conclusions: Our results indicated a positive association between serious game usage and well-being in patients with dementia, given adequate support. This is a valuable addition to the understanding of serious game usage in dementia care. Although challenging, user-centered development of serious games with people who are severely impaired by dementia is an important research target. Limitations like low data quality and a simplified design are inherent in this study population. Nevertheless, we demonstrated how usability testing in this target group is possible through careful definition and operationalization. The inclusion of different data sources, different recruitment sites, and different levels of cognitive impairment increased the generalizability of the findings. To accommodate severely impaired patients, future developments should incorporate a broader range of difficulties and adaptations to group settings.

Trial Registration: German Clinical Trials Register DRKS00031363; https://drks.de/search/en/trial/DRKS00031363/details

(JMIR Serious Games 2025;13:e69812) doi:10.2196/69812

KEYWORDS

dementia; inpatient treatment; long-term care; nonpharmaceutical treatment; eHealth; serious games; pilot study; usability

XSL•FO
RenderX

Introduction

Interventions in Dementia Care

The number of people living with dementia has been rising globally from 20.2 million in 1990 to more than 50 million in 2019 and is estimated to exceed 150 million in 2050 [1,2]. The rising prevalence leads to increasing numbers of patients with dementia who receive inpatient treatment or live in inpatient care facilities. It is important to provide suitable interventions for these patients. Possible intervention targets include improved quality of life, delayed cognitive decline, and successful treatment of comorbid disorders such as depression [3]. Given the broad range of dementia types and courses, these targets are not easy to address [4-7]. To satisfy patient needs, interventions are required to account for the various types of impairments and allow for individual tailoring. Accordingly, the Alzheimer Society of Canada [8] has put forward a person-centered approach as the best practice for dementia care, and dementia-friendly environments are receiving increasing attention [9,10]. In fact, nonpharmaceutical treatment (NPT) has become the recommended treatment for the behavioral and psychological symptoms of dementia [11]. NPT covers a wide range of approaches, including exercise and motor rehabilitation, cognitive rehabilitation, and psychological therapy, as well as information technology and assistive technology [12].

Unfortunately, the need for a person-centered and individually tailored approach puts a high burden on health care systems and private caregivers. Particularly in later disease stages, dementia causes larger health care challenges than other diseases [13]. This often conflicts with the already high costs, the lack of staff, and the short duration of inpatient treatment [14-16]. A possible approach to address these difficulties could be the use of serious games (SGs).

SGs in the Treatment of Dementia

SGs have received increasing attention as a means to apply NPT in dementia treatment [17,18]. SGs are defined as games that are designed to not only entertain but also enhance learning, for example, in the area of psychoeducation or cognitive training [19]. There is a small but growing body of SGs being specifically targeted at people with dementia [20-26]. To accommodate the differential needs of this target group, different author groups propose development guidelines for SGs [27,28]. However, the number of studies involving SGs that meet these guidelines [20,29] or engage in systematic usability testing of SGs [30] is limited. Additionally, more severely impaired target groups are seldom thought of as potential SG users [18]. While unsupervised SG usage seems inherently less likely for people with severe dementia, high user satisfaction (ie, positive feelings toward SG usage) has been found to be independent of cognitive impairment under the premise of adequate app design [25,31]. Given their potential benefits as a delivery method for NPT, we argue that it is a valuable aim to develop SGs specifically for people with dementia, with consideration of severely impaired patients. Since the development of medical apps often tends to pay not enough attention to evidence-based standards [32], further studies are needed to establish utility estimates of SGs for patients with dementia to enhance clinical credibility [12].

Usability of SGs for Patients With Dementia

This paper presents a pilot study of the usability of a newly developed SG (“Digimenz”) for older patients, promoting participative development. Usability in the context of dementia care has previously been operationalized by the ISO 9241 model [30,33]. It defines usability in terms of effectiveness, efficiency, and user experience. The ability to learn the handling of software is also a common part of different usability definitions [34]. While it is missing in ISO 9241, it appears to be a crucial factor for patients with dementia who have limited learning capacities by definition. Therefore, we have followed Abran et al [35] who propose to augment ISO 9241 by the concept of learnability and use this as an integrative model of usability (Table 1).

Table . Definition of usability.

Construct	Definition
Effectiveness	Achieve the given aims with precision and completeness
Efficiency	Resources needed in relation to the results achieved
User experience	Perceptions, emotions, and attitudes related to software use, for example, user satisfaction or interest
Learnability	Time needed to learn the handling of the software

Rationale of the Study

The purpose of this study was to generate a feedback loop for the iterative and evidence-based development of an SG app by providing scientific testing and making the results available for further development of this and other SG apps. This will add to a diverse, person-centered, and resource-saving body of intervention options for patients with dementia.

To ensure that the full spectrum of cognitive impairment is covered when examining the usability of the SG app, we recruited participants from two settings: (1) a long-term care residential facility (where people live with, on average, a lesser degree of cognitive impairment) and (2) a geriatric psychiatric inpatient ward (where the degree of deficits is expected to be more severe). We examined whether the Digimenz app is usable across different degrees of impairment while controlling for the effects of the recruitment site.

We were particularly interested in whether the SG shows sufficient usability in terms of global usability, effectiveness, efficiency, user experience, and learnability. We hypothesized

XSL•FO
RenderX

that less impaired patients would report higher overall usability of the SG app and show higher effectiveness, efficiency, and learnability, but would have an equally satisfying user experience as patients with a higher degree of impairment. Moreover, usability parameters are expected to be similar in the psychiatric context and in a nursing home.

Methods

Participants

Following the adapted CONSORT (Consolidated Standards of Reporting Trials) statement for pilot trials [36], the sample size was adapted to the rationale of usability testing. To our knowledge, there is no evidence regarding the effect size for the comparison of the usability parameters of SGs among users with different levels of cognitive impairment. We therefore chose a sample size comparable to that in other usability trials involving people with cognitive impairment [30] in order to allow reliable conclusions about SG usability and inform future large-scale intervention studies. The final sample included 43 participants.

Subsample 1 (n=28) was recruited from a nursing home for older adults specialized in the management of cognitive impairment. Residents are referred to this kind of long-term care facility when care in their original homes is no longer adequate or possible. Subsample 2 (n=15) was recruited from 2 geriatric-psychiatric wards specialized in dementia management in a psychiatric hospital. Patients are referred to this kind of psychiatric ward in case of acute exacerbation of mental impairments or comorbid mental disorders.

The inclusion criterion for all participants was documented cognitive deficits (mild cognitive impairment [MCI] or dementia) diagnosed at admission to the respective institution by a specialist in the field according to the criteria by McKhann et al [37] and Albert et al [38]. Further, participants were included only if they had sufficient German language proficiency. Since we wanted to represent a broad range of cognitive impairments, we tuned the inclusion criteria for sensitivity and not specificity. Therefore, a specific etiology of cognitive impairment was not relevant for inclusion. Moreover, we did not define age limits, numerical cutoffs for cognitive tests, or comorbid mental disorders as exclusion criteria.

Ethical Considerations

The study protocol complies with the Declaration of Helsinki and received ethical approval from the Bielefeld University Ethics Review Board (EUB-2022 - 265 s). Potential participants were approached to examine their explicit preferences regarding study participation. The purpose and methodology of the study were explained in simple terms to be accessible for those who were nominally more severely cognitively impaired. If there was any sign of rejection, no further recruitment was attempted. If a potential participant expressed interest in participation, informed consent was acquired. To this end, an information sheet and consent form were read and signed by the participant, if mental capacity allowed it. Otherwise, informed consent was obtained from a legal representative or relative. Study data were deidentified and saved on the secure servers of our institution. To guarantee the anonymity of participating patients, participant-related data will not be made available. No compensation was provided.

Study Design

In this prospective study, we used a repeated-measures design. Each participant was asked to complete 4 conditions in randomized order. Three of these conditions involved playing 3 different games from the SG app (“Digimenz”). The fourth condition included reading a newspaper serving as a naturalistic control condition. All conditions consisted of 2 runs each (5 min), with one supported by the conducting researcher and the other performed with minimal support. Prior to the experiment and after each individual condition, the participants answered a set of questions and the conducting researcher gave external ratings. Additional unstructured qualitative feedback was registered. Global cognitive functioning was assessed in a separate session. The total duration for each participant was about 60 - 90 minutes. To minimize researcher bias while retaining a high data quality, data collection was performed by 2 research assistants who were instructed and supervised by the principal investigators (MT, EMT, and SK), who have extensive experience in dementia research. The study has been registered in the German Clinical Trials Register (DRKS00031363).

Intervention

Digimenz App Overview

The Digimenz app is a commercial tablet app linked to research through a European Union–funded project (IT4Anxiety). The project’s objective is to connect end users of e-mental health apps, scientists, and industry professionals, providing a platform for codevelopment, testing, and implementation of innovative treatment options. The Digimenz app has been developed in close cooperation with people affected by dementia and their caregivers. They were involved in an iterative, user-centered process of concept development, prototype testing, feedback collection, and software development until deriving the SG app tested in this study. Three games have been incorporated in the Digimenz app, all of which require psychomotor integration: recognizing objects and moving them by drag and drop (“Drag&Drop”), recognizing distorted pictures (“Object Recognition”), and sorting letters into words (“Spelling”). The difficulty of the games is adjusted automatically within 4 levels per game (higher levels entail higher difficulty). The level can change dynamically during a gaming session. It is determined through a running score, which is calculated based on the user’s current response time and error rate. The score is updated once per second and after each user input. Dynamic difficulty adjustment is a central recommendation for making SGs usable for patients in different stages of cognitive decline [27]. Further guidelines from Eichhorn et al [28] were considered regarding the design and timing of assignments, help, and feedback. The motor requirements for the games include dragging and tapping. Therefore, hand-eye coordination is required, but no complex, multistep, or whole-body movements are involved. The different SGs relate to different cognitive subfunctions: all games require semantic memory, working memory, and executive functioning. The Drag&Drop game additionally requires motor functioning and visuospatial functioning. The Spelling game requires lexical

XSL•FO
RenderX

functioning. Therefore, we did not compute mean values of usability measures across the games but analyzed usability for the individual games separately.

Drag&Drop Game

In this SG, animals must be sorted in a stable. The screen presents a choice of animals (targets and distractors) on a meadow, next to an empty stable (Figure 1). An assignment is given (eg, “Put the lamb in the stable”), and the users must choose the correct animal, place their finger on it, and place it in the stable via drag and drop. A higher difficulty is achieved by presenting more distractors and by giving more abstract assignments (eg, “Put all animals with hooves in the stable”). If an incorrect animal is placed in the stable, it disappears from the screen, and a red cross is presented (counted as an error). If a correct animal is placed in the stable, a green checkmark is presented, and the animal remains in the stable. The task cannot be skipped but is solved by moving all animals into the stable.

Figure 1. Example screen of the Drag&Drop game in the Digimenz app.



Object Recognition Game

In this game, a picture of an animal is presented, but it is covered by multiple tiles. The tiles disappear slowly, making it easier to recognize the picture until the full picture can be seen. Next to the picture, multiple animal names are presented as buttons (one target and one or multiple distractors). The assignment is to tap on the animal name that matches the picture. Difficulty is varied by changing the number of distractors and by altering the speed at which the tiles disappear. If a wrong name is chosen, it turns red to distinguish it from the remaining choices (counted as an error). The task cannot be skipped, but the distractors disappear after some time, such that only the right answer remains.

Spelling Game

In the third game, a picture of an animal is presented. In the lower part of the screen, the letters needed to spell the name of the animal are presented in randomized order. The task is to tap on the letters in the correct order to spell the name of the animal. If a wrong letter is chosen, it blinks in red (counted as an error). Difficulty is influenced by providing longer animal names. The task cannot be skipped but must be solved by tapping the letters presented on the screen in random order until discovering the right order by chance.

Control Condition

To control whether changes in mood would be attributable to SG usage as opposed to interaction with the researcher or study participation, we added an illustrated newspaper as a naturalistic control condition. It can be used in one-to-one interaction with varying amounts of support and is thus comparable to the SGs from the Digimenz app. Depending on the usage style (looking at pictures, reading headlines or short texts, and reading longer texts), it can accommodate different degrees of cognitive impairment.

Measures

To test the applicability of the SGs, we recorded reasons for nonparticipation in the recruitment phase and dropout thereafter. To test usability and its relation to cognitive functioning, as well as differences between the subsamples, we operationalized global usability and the individual usability constructs as follows:

XSL•FO
RenderX

- Global usability: The System Usability Scale (SUS) [39] (total score 0 - 100) was assessed after completing all 4 conditions. We used a simplified version for people with cognitive impairment [40]. SUS scores can be classified by comparing them to the existing literature. Since no specific norms for SGs exist, we used the universal norms defined by Sauro and Lewis [41], who stated that a SUS score of 68 denotes average usability (school grade C) and good usability can be assumed for SUS scores higher than 74 (school grade B).
- Effectiveness: We used the error rate in the second run of each SG to operationalize their correct usage. Multiple errors could be made per task, and therefore, the error rate can be greater than 1. We also collected data on the highest level reached by each participant to allow better interpretation of error rates.
- Efficiency: We used the amount of support needed in the minimal-support run (second run of each condition; 0 ["no support needed"] to 5 ["much support needed"]) as indicated by the researcher to operationalize the efficiency of the intervention. Scores were based on a scoring guideline that required information about the type, intensity, and frequency of support.
- User experience: At baseline and after each condition, the investigator and the participant rated the participant's mood (positive or negative affect) using a visual analog mood scale (VAMS) [42]. The researcher provided their own rating before asking the participant for their rating.
- Learnability: The item "Playing the game alone is too difficult for me" was answered by the participant on a 5-point Likert scale after each condition. The researcher answered the item "The game is too difficult for the participant" on the same scale after each condition. In addition, the proportion of participants needing support (yes or no) was compared between the first and second runs of each condition.

Efficiency, user experience, and learnability were compared between the SGs and the control condition (newspaper). Sufficient usability in the different settings was assumed in the case of noninferiority of these parameters when comparing the SGs to the control condition. A "good" global usability rating (SUS scores >74) was the second indicator of sufficient usability. Error rate cannot be used in this way since it is not defined for the control condition. In addition to these quantitative measures of usability, we registered any additional feedback on usability issues given by the participants during SG usage and while rating the scales on usability. Global cognitive functioning was assessed with the Mini-Mental State Examination (MMSE) [43]. Demographic information (age and sex) and any existing diagnoses were registered from the medical record.

Quantitative Data Analysis

We present descriptive statistics of the rates of nonuse and dropout as operationalization of the applicability of the SGs. We modeled the usability measures according to the conditions and the degree of cognitive impairment (MMSE) and their interaction, which we hypothesized to be the most relevant predictors of usability. We controlled for the effects of the recruitment site (subsample), given the differences between the 2 sites. We used mixed effect models (linear mixed models for continuous outcome variables and cumulative link mixed models [CLMMs] for ordinal outcome variables) with random intercepts to account for the repeated-measures design. These models have the advantage of using all available data at each measurement, precluding the need for listwise deletion (ie, a person does not have to be deleted if missing 1 measurement). Nagelkerke pseudo R^2 is reported as an effect size measure by comparing the complete model to the model without the respective predictor and its interaction terms. To quantify differences between conditions, we computed Hedges g using the within-subject variance [44] and the adjustment for small samples: $g = M_{condition\,a} - M_{condition\,b} / SD_{diff} * 2(1 - r_{condition\,a,\,condition\,b}) * (1 - 3/4n - 5)$ [45]. Values of g >0.2 may be interpreted as small, g>0.5 as medium, and g>0.8 as large effect sizes [44]. All analyses were performed in R version 4.4.1 (R Foundation for Statistical Computing). Linear mixed models were fitted with the lme4 package [46], and CLMMs were fitted with the ordinal package [47]. The significance level was set to α=.05 for all inferential tests.

Qualitative Analysis of Feedback

Qualitative feedback data on usability issues were subjected to a thematic content analysis following Kuckartz and Rädiker [48]. This is a multistep, iterative procedure. The steps are as follows:

1. Initial work with the data: Feedback and field notes were reviewed, annotated, and summarized (authors SFB and MS).
2. Deductive development of main categories: We predefined categories for the 3 SGs to capture any feedback specific to one of the SGs (authors SFB and MT).
3. Coding of data with main categories: The complete qualitative dataset was coded as far as possible with the predefined categories (author SFB).
4. Inductive development of further categories: Based on the first coding process, we defined further categories to account for themes not covered by the predefined categories (authors SFB and MT).
5. Recoding of data with all categories: The complete qualitative dataset was coded again, using both deductive and inductive categories (author SFB).
6. Interpretation and presentation: The results were summarized into a concise report (authors SFB, MT, and CS).

Steps 4 and 5 were repeated several times to fine-tune the categories.

Results

Sample Characteristics and Acceptability of the Intervention

In subsample 1 (nursing home), of 30 eligible individuals, 30 (100%) consented to participate and 28 (93%) completed the study procedure. In subsample 2 (psychiatric ward), of 41 patients eligible for participation, 21 (51%) consented to participate and 15 (37%) completed the study procedure. Table

XSL•FO
RenderX

2 details the characteristics of both samples. Figure 2 provides an overview of the reasons for nonparticipation and dropout. In subsample 1, all people approached were interested in the SGs, and the reasons for dropout were not specific to the games themselves. In subsample 2, 10 people (24%) denied participation, which was specifically related to SG usage (7 [17%] were disinterested, and 3 [7%] were frustrated by the SGs), while 31 people (76%) were generally interested in the SGs. The odds ratio for dropout between informed consent and study completion was 5.6 (P=.05), that is, the odds of dropping out of the study were 5.6 times higher in the psychiatric ward subsample than in the nursing home subsample.

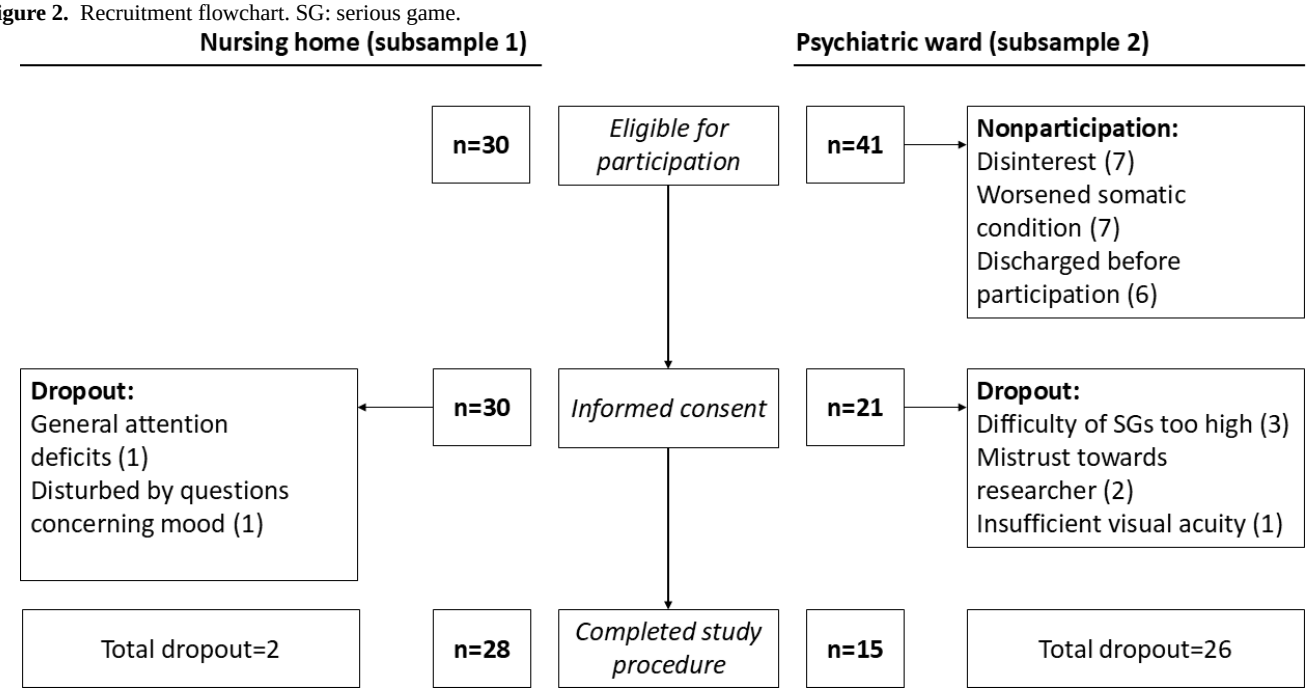
Table . Sample characteristics.

Characteristic	Nursing home (subsample 1) (N=28)	Psychiatric ward (subsample 2) (N=15)
Age (years)		
Mean (SD)	85.6 (7.3)	80.2 (7.5)
Range	57-97	68-94
Cognitive functioning (MMSE[a] score)		
Mean (SD)	20.1 (6.4)	14.7 (6.1)
Range	4-28	4-22
Sex (female), n (%)	22 (79)	9 (60)
Formally diagnosed with dementia (yes), n (%)	11 (39)	15 (100)
Main diagnoses, n (%)		
Dementia due to Alzheimer disease with late onset	1 (4)	2 (13)
Unspecified dementia, unspecified severity	10 (36)	6 (40)
Multicause dementia due to Alzheimer disease and vascular disease	—[b]	4 (27)
Vascular dementia, subcortical	—	2 (13)
Delirium superimposed on dementia	—	1 (7)

[a]MMSE: Mini-Mental State Examination; measured on a scale from 0 - 30.

[b]Not applicable.

Figure 2. Recruitment flowchart. SG: serious game.



XSL•FO
RenderX

Global Usability and Usability Dimensions

To compute the total SUS score, missing values were replaced by the scale mean, as recommended by Lewis [49]. In the psychiatric subsample, the SUS total score was 71.5 (SD 24.33), while in the nursing home subsample, the SUS total score was 85.10 (SD 12.67). Less cognitive impairment was associated with better usability ratings in both subsamples, with no significant difference in the correlation coefficients between the subsamples (total sample: Spearman ρ=0.31; *P*=.03). The results for the individual usability dimensions as defined in Table 1 are presented for the total sample in Table 3. Differences between the subsamples were tested in the following multilevel analyses.

Table . Descriptive statistics for usability dimensions (total sample).

Usability dimension and operationalization	Condition								MMSE[a,b]		
	Newspaper		Drag&Drop		Object Recognition		Spelling		*r*[c] (95% CI)	ρ[d] (95% CI)	*P* value
	Mean (SD)	*g* [e]	Mean (SD)	*g*	Mean (SD)	*g*	Mean (SD)	*g*			
Effectiveness											
Error rate	—[f]	—	0.35 (0.40)	—	0.19 (0.25)	−0.43	0.49 (0.43)	0.38	−0.14 (−1.00 to 0.13)	—	.20
Efficiency											
Amount of support in run 2 (0-5)	1.36 (2.10)	—	1.43 (1.82)	0.04	1.16 (1.62)	−0.11	1.79 (1.90)	0.21	—	−0.49 (−0.69 to −0.19)	<.001
User experience											
Mood (self-rated; 1-5)	3.36 (0.90)	—	4.00 (0.75)	0.78	3.85 (0.85)	0.55	3.78 (0.86)	0.46	—	0.07 (−0.31 to 0.41)	.67
Mood (externally rated; 1-5)	3.44 (0.63)	—	3.85 (0.69)	0.60	3.85 (0.85)	0.61	3.74 (0.73)	0.46	—	0.12 (−0.18 to 0.41)	.45
Learnability											
Too difficult (self-rated; 1-5)	1.98 (1.35)	—	1.51 (0.98)	−0.38	1.48 (0.99)	−0.41	1.71 (1.04)	−0.21	—	−0.15 (−0.44 to 0.17)	.18
Too difficult (externally rated; 1-5)	2.42 (1.53)	—	2.24 (1.56)	−0.12	1.98 (1.32)	−0.31	2.37 (1.41)	−0.04	—	−0.61 (−0.78 to −0.36)	<.001
Need for support (% yes in run 1)	33 (48)	—	65 (48)	0.65	53 (50)	0.40	74 (44)	0.87	−0.61 (−1.00 to −0.42)	—	<.001
Need for support (% yes in run 2)	32 (47)	—	44 (50)	0.23	40 (49)	0.14	58 (50)	0.51	−0.48 (−1.00 to −0.25)	—	<.001

[a]MMSE: Mini-Mental State Examination.

[b]Correlation coefficients denote the correlation of the respective variable with the MMSE score across the 3 serious game conditions.

[c]Pearson *r*.

[d]Spearman ρ.

[e]For error rate, each condition is compared to the Drag&Drop condition; for all other variables, each condition is compared to the newspaper condition. Moreover, *g*>0.2 indicates a small effect, *g*>0.5 indicates a medium effect, and *g*>0.8 indicates a large effect.

[f]Not applicable.

Effectiveness

No significant effects of MMSE ($F_{1,36.50}$=0.10; *P*=.75; DR^2=0.09) and subsample ($F_{1,36.45}$=0.16; *P*=.69; DR^2<0.001) on error rate were found in the multilevel analysis. Among the SGs, there were significant differences in error rates ($F_{2,74.28}$=9.43; *P*<.001; DR^2=0.31; Object Recognition game: −0.10; *P*=.09; Spelling game: +0.15; *P*=.01; see Table 3 for

XSL•FO
RenderX

effect sizes). The interaction of MMSE and condition was not significant ($F_{2,74.02}$=0.13; P=.15; DR^2=0.09). The Nagelkerke R^2 for the complete model compared to the null model was R^2=0.37. Regarding the difficulty adjustment, Table 4 shows a descriptive analysis of the difficulty levels. Overall, the highest (ie, most difficult) level was seldom reached, and participants with a higher MMSE reached higher levels of difficulty.

Table . Participants per difficulty level and association with the MMSE[a].

Condition (game)	Level				ρ[b]	*P* value
	1[c], n (%)	2[c], n (%)	3[c], n (%)	4[c], n (%)		
Drag&Drop	21 (49)	9 (21)	10 (23)	3 (7)	0.58	<.001
Object Recognition	16 (37)	15 (35)	8 (19)	4 (9)	0.47	.002
Spelling	28 (65)	14 (33)	1 (2)	—[d]	0.46	.003

[a]MMSE: Mini-Mental State Examination.

[b]Spearman ρ showing the association between the MMSE and the maximum level reached.

[c]Number and proportion of participants who reached no higher than this level.

[d]Not applicable.

Efficiency

Participants with a higher MMSE needed significantly less support than those with a lower MMSE (Table 5; Figure 3). The correlation analysis presented in Table 3 suggests that this effect was of medium size ($\rho_{MMSE,\ Amount\ Support}$=−0.49). For the Spelling game, significantly more support was needed than for the newspaper condition (+1.27; P=.02). This was not the case for the Drag&Drop game (+0.17; P=.77) and the Object Recognition game (−0.09; P=.88) (see Table 3 for effect sizes). No significant differences between the subsamples were found in the multilevel analysis. The Nagelkerke R^2 for the complete model compared to the null model was R^2=0.16.

Table . ANOVA for the amount of support (cumulative link mixed model)[a].

Predictor	LR[b] chi-square (*df*)	*P* value	ΔR^2
Condition	10.32 (3)	.02	0.07
Subsample	0.90 (1)	.35	0.006
MMSE[c]	13.60 (1)	<.001	0.09
Condition*MMSE	1.56 (3)	.67	0.01

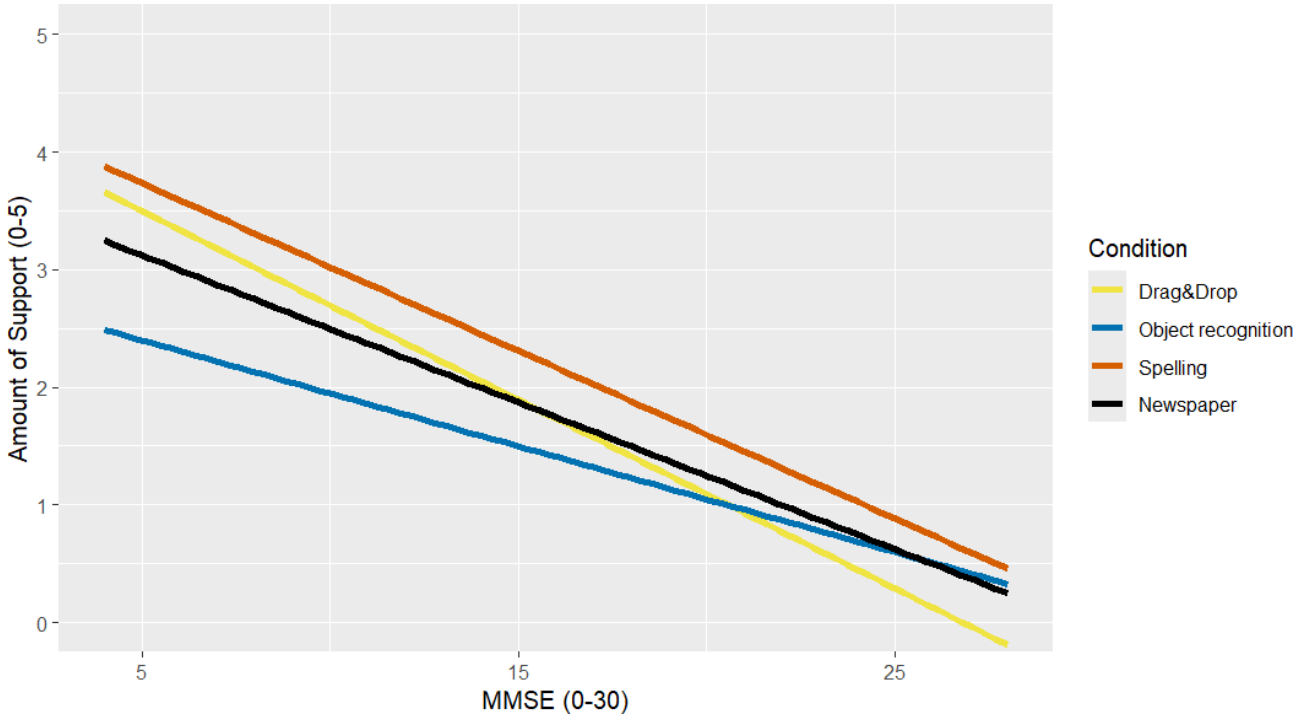
[a]Model formula: amount of support ~ 1 + condition + sample + MMSE + condition*MMSE + (1|subject).

[b]LR: likelihood ratio.

[c]MMSE: Mini-Mental State Examination.

XSL•FO
RenderX

Figure 3. Association between support needed and Mini-Mental State Examination (MMSE) scores in all experimental conditions.



User Experience

The multilevel analysis showed that self-reported mood was higher in all 3 game conditions than in the newspaper condition (Drag&Drop: +2.55; *P*<.001; Object Recognition: +2.17; *P*<.001; Spelling: +1.56; *P*=.004; see Table 3 for effect sizes), when correcting for baseline mood (Table 6). The MMSE was not significantly correlated with self-reported mood (Table 6; Figure 4). There were no significant differences between the 2 subsamples. The Nagelkerke R^2 for the complete model compared to the null model was R^2=0.24. The external rating of mood showed a similar pattern with higher mood ratings in all 3 game conditions compared to the newspaper condition (Drag&Drop: +1.65; *P*<.001; Object Recognition: +1.28; *P*=.007; Spelling: +1.08; *P*=.02; see Table 3 for effect sizes). The overall correlation of self-reported and externally rated mood was *r*=0.64 (95% CI 0.43-0.79; *P*<.001).

Table . ANOVA for self-reported mood (cumulative link mixed model)[a].

Predictor	LR[b] chi-square (*df*)	*P* value	ΔR^2
Condition	25.09 (3)	<.001	0.17
Baseline mood	9.76 (1)	.002	0.07
Subsample	0.70 (1)	.40	0.005
MMSE[c]	0.64 (1)	.42	0.005
Condition*MMSE	0.81 (3)	.85	0.006

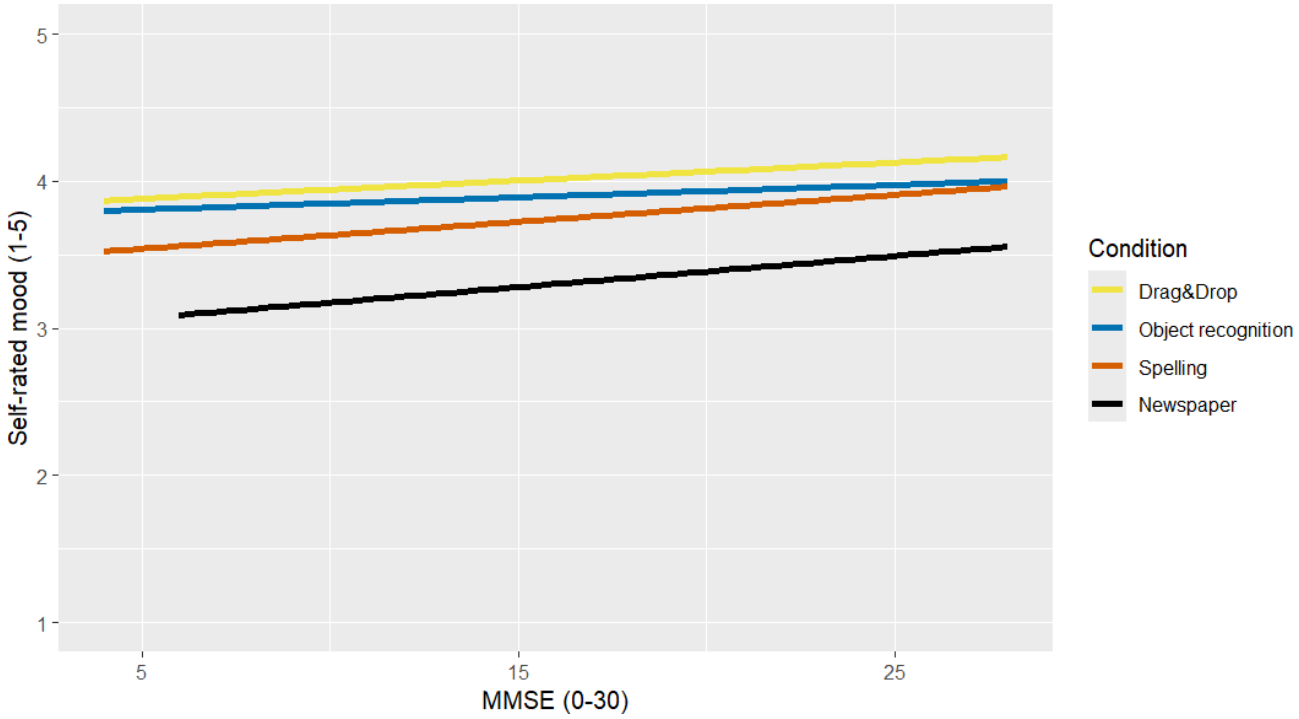
[a]Model formula: self-rated mood ~ 1 + condition + sample + MMSE + condition*MMSE + baseline mood + (1|subject).

[b]LR: likelihood ratio.

[c]MMSE: Mini-Mental State Examination.

XSL•FO
RenderX

Figure 4. Association between self-rated mood and Mini-Mental State Examination (MMSE) scores in all experimental conditions.



Learnability

According to the multilevel analysis (Table 7), self-rated difficulty was independent of the MMSE (estimate=−0.06; *P*=.54). Compared to the newspaper condition, it was significantly lower for the Drag&Drop game (−2.01; *P*=.002) and Object Recognition game (−1.76; *P*=.005), but not for the Spelling game (−0.63; *P*=.54; see Table 3 for effect sizes). The 2 subsamples did not differ significantly in their difficulty ratings. The Nagelkerke R^2 for the complete model compared to the null model was R^2=0.14. The external rating of difficulty, in contrast, showed a strong negative association with MMSE (estimate=−0.22; *P*<.001) (ie, higher MMSE scores were associated with lower external ratings of difficulty). The overall correlation of self-reported and externally rated difficulty was ρ=0.52 (95% CI 0.25-0.72; *P*<.001).

Table . ANOVA for self-rated difficulty (cumulative link mixed model)[a].

Predictor	LR[b] chi-square (*df*)	*P* value	ΔR^2
Condition	12.05 (3)	.007	0.08
Subsample	2.27 (1)	.13	0.02
MMSE[c]	0.67 (1)	.41	0.005
Condition*MMSE	3.39 (3)	.29	0.03

[a]Model formula: self-rated difficulty ~ 1 + condition + sample + MMSE + condition*MMSE + (1|subject).

[b]LR: likelihood ratio.

[c]MMSE: Mini-Mental State Examination.

Our second indicator of learnability was the proportion of participants needing support in the full-support run versus the minimal-support run (Multimedia Appendix 1).

In the full-support run, more participants needed support in the 3 game conditions than the newspaper condition (Drag&Drop: +0.28; *P*<.001; Object Recognition: +0.19; *P*=.02; Spelling: +0.41; *P*<.001; see Table 3 for effect sizes; $F_{3,275.18}$=10.93; *P*<.001; DR^2=0.14). From the full-support run to the minimal-support run, the rate of support decreased significantly ($F_{1,274.92}$=9.86; *P*=.002; DR^2=0.05). Accordingly, there were no significant differences in support needed between the game conditions and the newspaper condition in the second run (Drag&Drop: −0.20; *P*=.09; Object Recognition: −0.11; *P*=.34; Spelling: −0.16; *P*=.17; see Table 3 for effect sizes). The interaction effect of run and condition was not significant ($F_{3,274.9}$=1.09; *P*=.36; DR^2=0.02). Participants with a higher MMSE score were less likely to need support at all ($F_{1,37.89}$=16.76; *P*<.001; DR^2=0.07). There were no significant differences between the 2 subsamples ($F_{1,37.88}$=0.004; *P*=.95; DR^2 <0.001). The Nagelkerke R^2 for the complete model compared to the null model was R^2=0.25.

XSL•FO
RenderX

Qualitative Feedback Data

The results of the thematic content analysis [48] of usability issues mentioned during SG usage and data collection are presented in Table 8. We chose to categorize usability issues according to the source they were attributed to. Besides usability issues specific to the SGs, participants also mentioned problems with the handling of the tablet and personal impairment that hindered tablet or SG usage.

Table . Thematic analysis of usability issues (verbal feedback).

Main category and subcategory	Usability issues
Usability issues attributed to personal impairment	
Sensory and motor problems	• Impaired vision (n=2; MMSE[a] scores of 16 and 26) • Motor problems due to Parkinson disease (n=1; MMSE score of 17)
Cognitive problems	• Concentration deficits (n=4; MMSE scores of 14, 15, 17, and 27) • Tired (n=1; MMSE score of 18)
Usability issues attributed to tablet usage	• Difficulties handling the tablet (eg, difficult to use the touchscreen) (n=3; MMSE scores of 4, 14, and 22) • Eyes hurt after looking at the tablet for a while (n=1; MMSE score of 26)
Usability issues attributed to serious games	
All serious games	• Reinforcement from the app not strong enough (n=1; MMSE score of 22) • Computer-animated voice hard to understand (n=1; MMSE score of 16) • Serious games too easy/repetitive (n=4; MMSE scores of 18, 24, 27, and 27)
Drag&Drop	• Too many stimuli, cannot focus attention (n=1; MMSE score of 19) • Difficulties distinguishing animals with similar shape (eg, cat and dog) (n=4; MMSE scores of 4, 14, 17, and 19) • Motion exhaustive for the arm (n=1; MMSE score of 15) • Instructions not understood (n=1; MMSE score of 18)
Object Recognition	• Animals not recognized (n=2; MMSE scores of 9 and 9)
Spelling	• Animals not recognized (n=4; MMSE scores of 4, 20, 22, and 26) • Individual animals not known (n=1; MMSE score of 21) • Difficulties with spelling (n=2; MMSE scores of 22 and 25)

[a]MMSE: Mini-Mental State Examination.

Discussion

Principal Findings

This study aimed to evaluate the usability of SGs for older adults with cognitive impairment, ranging from MCI to mild and severe dementia. We hypothesized that the SGs would be applicable for a nursing home as well as a psychiatric ward, that the usability would be sufficient, and that it would be differentially influenced by the level of cognitive impairment.

Most people approached were open to testing the SGs; however, a considerably smaller proportion managed to complete the study procedure in the psychiatric ward subsample, given greater impairment there. The global usability was rated good (SUS), and the indicators for individual usability dimensions showed that using the SGs was possible with a moderate error rate and support. Moreover, the support rate decreased from the first to the second usage, implying a learning process, and the gaming experience was associated with positive mood. While the amount of support needed and the ability to learn the handling of the

XSL•FO
RenderX

games were related to cognitive impairment, the error rate and mood state after playing were not.

Comparison With Prior Work

Applicability

The SGs can be considered applicable in both recruitment settings, with the restriction that study participation and openness to the SGs would be lower in the hospital context, owing most likely to the higher proportion of severely impaired participants [50]. One aspect is cognitive impairment, and importantly, nonparticipation, study dropout, and usability issues within the psychiatric subsample were attributed by several participants to somatic conditions and sensory and motor impairments. Moreover, the time it took to acquire informed consent from legal advisors sometimes precluded participation before discharge or transfer to another ward. While this is partly due to research requirements, quick patient turnover may generally limit the applicability of NPTs in routine care. Only 7 patients denied participation due to disinterest. To understand their specific reasons for nonparticipation, their internal experiences would need to be investigated. This, however, is difficult as more severe impairments are inversely related to the capacity of objectifiable communication. Moreover, the validity of questionnaire data collected from the patients is uncertain due to limited understanding of the items, and proxy ratings may be subject to bias [51]. It is therefore not surprising that most samples in SG usability studies are composed mainly of persons with MCI or mild dementia (MD) [30]. These user groups are mostly found to be using mobile devices and SGs successfully [18,52,53], which is in line with the high participation rate in our nursing home subsample (less affected). Encouragingly though, we did not find any significant differences in usability between the recruitment sites. This result is in line with our hypothesis of sufficient and similar usability in both subsamples.

Nevertheless, outside of standardized study procedures, people with MCI or dementia report extensive difficulties and reservations concerning mobile devices [54,55]. Therefore, to raise the acceptability of SGs and aid their uptake into routine care, motivational strategies may be helpful not only within the SGs but also to initiate usage [28]. Possible strategies may include social support or social referencing [56]. For example, future applications of the Digimenz SGs might benefit from adaptation to a group setting. This would facilitate the resource-efficient allocation of professional support and mutual support among participants [57,58]. Social interaction itself is also associated with increased well-being and reduced cognitive decline [59,60]. At the same time, caregivers would need to be trained to determine an appropriate level of support in individual and group settings. Another challenge may be the availability of devices. As in this case, SGs should be developed with an emphasis on compatibility across devices and operating systems. This reduces the need for specialized equipment and lowers the entry threshold for providers and users.

Usability

Regarding the SGs themselves, the global usability rating showed a positive correlation with the level of cognitive functioning (MMSE). This was reflected in difficulties in understanding the SGs and playing them without support (learnability and efficiency), which are related to cognitive impairment. These findings are in line with our hypotheses. Beyond global cognitive functioning, qualitative feedback data suggested several usability issues specific to the SGs. In the Drag&Drop game, several participants mentioned having difficulties distinguishing the animals, potentially hindering effective usage. In the Spelling game, the combined requirement of recognizing animals and spelling their names seemed to be especially demanding. Accordingly, quantitative estimates of effectiveness, efficiency, and learnability were least favorable for the Spelling game. The Object Recognition game, in contrast, was least demanding, according to quantitative and qualitative findings. For all 3 SGs, the support rate decreased from the first to the second run, which may indicate a learning process. Häikiö et al [61] showed that older users could significantly improve their interaction with a touch-based device after just 1 contact. Evidently, there remained a large proportion of participants who needed support in the second run, although not significantly more than in the newspaper condition. The qualitative analysis further suggested that several usability issues arose from factors not specific to the SGs, such as sensory and motor impairment, as well as general problems with handling the tablet. Although the SGs used here have already taken into account many of the usability recommendations given by Bouchard et al [27], they will need further adjustments to accommodate such a diverse user group as ours. Easier as well as more difficult levels would be needed. Improvements to in-game help and the user interface may include clearer contrast, better distinguishable shapes, visual clues, adjustable text and button sizes, broader use of multisensory feedback, and a refined voice guide.

The difficulties in using the SGs do not seem to be associated with the enjoyment taken from playing (user experience), since mood was consistently higher in the 3 game conditions than in the newspaper condition, irrespective of the cognitive status. This is again in line with our hypotheses. However, the study design does not allow for causal interpretation. Moreover, the differences in support given during the full-support run between the SGs and the newspaper condition may be a confounding factor. High levels of fun or enjoyment were also found in other studies evaluating SGs [20,31,62,63]. In some studies, healthy controls found the games more enjoyable than people with dementia or MCI [25,64]. Two studies assessing SGs specifically designed for people with dementia reported less enjoyment among healthy or less impaired participants, since they found the games too easy [20,63]. Here, 4 participants explicitly stated that they found the SGs too easy, most of whom had relatively little cognitive impairment. Contrasting to our hypothesis, we found no significant relation between the error rate and the level of cognitive functioning. This is probably due to the adaptive difficulty of the games and the support at hand. Less impaired participants reached more difficult levels of the games, while more severely impaired patients received more support. This may have equalized the error rates and the self-ratings of difficulty across levels of cognitive impairment.

XSL•FO
RenderX

Strengths

It is a strength of this study that we chose theory-based operationalization of usability, enabling us to draw differentiated conclusions regarding the usability of the SGs and the relation to cognitive functioning. By using different data sources (external and internal ratings, qualitative data, and usage data from the SGs), we can provide comprehensive insights into the usage behavior and user experience. The diverse nature of our sample in terms of cognitive impairment and care settings allows us to set our results in the context of previous studies and generalize our conclusions to more severely impaired people who are often excluded from research.

Limitations

The inclusive design comes with the drawback of limited formal data quality and the necessity to limit this pilot study to a more constrained procedure. First, we did not include repeated gaming sessions. Therefore, we cannot provide information on the potential long-term effects of the SGs. Second, self-reported data may be inaccurate given limited introspective abilities and difficulties understanding questions, potentially reducing the reliability of the results. We attempted to overcome this limitation by adding observational data and usage data. However, there is some potential for researcher bias in observational data since the assessment of state variables (eg, mood) in patients with dementia generally lacks reliability [51]. We took measures to ensure high-quality observational data by providing scoring guidelines and supervising the ratings done by investigators. Third, the number of study completers in the psychiatric ward was rather small in absolute terms, although it was average for this highly vulnerable population. Therefore, inferential tests must be interpreted with caution. Fourth, the control condition differed from the experimental conditions in more than one way (medium and content). For future research endeavors, it would be preferable to add analog versions of the SGs or a digital version of the newspaper to isolate the effects related to these attributes.

Future Directions

Based on the promising findings from this pilot trial, it would be worthwhile to pursue a more extensive study design with a larger sample size, more recruitment sites, and repeated or longer gaming sessions, as well as repeated assessments over a longer period of time. Future studies should systematically register potential confounding factors that may influence usability at baseline (ie, sensory and motor impairment, and familiarity with technology) and incorporate the perspectives of caregivers and clinical staff. Future studies in the field would also benefit from establishing standards for quantifying the difficulty of SGs to allow the comparison of performance parameters within and across studies. This would help to evaluate the exact support requirements adapted to different profiles of impairment. Studying “people with dementia” without further specification is not sufficient for this aim, as has already been noted by Eichhorn et al [28].

Conclusions

From the quantitative results presented herein, we can conclude that users with a wide variety of cognitive impairments were able to adequately use the SGs under the premise of adequate support. They reported good global usability, reported positive mood after playing, and showed a moderate error rate. Since a decrease in support could be seen from the first to the second usage of each SG, we can conclude that the games were sufficiently learnable.

The qualitative results underlined the importance of a user-centered design, considering cognitive as well as somatosensory impairments. In this regard, the adaptive difficulty levels represented an important step toward personalized care. Less impaired patients reached higher levels, contributing to the broad applicability of the SGs. Nevertheless, additional levels at both the low and high ends of difficulty would have been worthwhile. No significant differences in usability parameters were found between recruitment settings. SGs may therefore be suitable for not only people with MCI or MD but also those with severe cognitive impairment who are treated in an inpatient psychiatric setting. However, this requires overcoming practical restraints like quick patient turnover.

Acknowledgments

This study was initiated as part of the IT4Anxiety project, hosted at our facility by MD. We thank the team from Digimenz for providing the Digimenz app and the necessary hardware for testing. We acknowledge the commitment of Ms Carla Seitz, who conducted a large portion of the testing sessions as part of her master’s thesis. Further, we want to thank all participants, caregivers, nurses, and doctors involved in the study for their participation and their support during recruitment. We acknowledge support for publication costs from the Open Access Publication Fund of Bielefeld University and the Deutsche Forschungsgemeinschaft (DFG). This work was supported by EU Interreg North West Europe NWE 983.

The authors attest that there was no use of generative artificial intelligence technology in the generation of text, figures, or other informational content in this manuscript.

Data Availability

Participant-related data will not be made available in order to guarantee the anonymity of participating patients. Requests to access aggregated data can be directed to SFB (soeren.braehmer@uni-bielefeld.de) or CS (carolin.steuwe@evkb.de).

XSL•FO
RenderX

Authors' Contributions

Conceptualization: SFB, BI, SK, MT, CS
Data curation: SFB
Formal analysis: SFB, BI, EMT, MS, MT, CS
Funding acquisition: MD, CS
Investigation: SFB, MS
Methodology: SFB, BI, EMT, MT, CS
Project administration: MT, CS
Resources: BI, SK, MD
Supervision: SK, MD, EMT
Validation: BI, SK, EMT, MT, CS
Visualization: SFB
Writing – original draft: SFB
Writing – review & editing: SFB, BI, SK, MD, EMT, MS, MT, CS

Conflicts of Interest

None declared.

Multimedia Appendix 1
Need for support (yes or no) in the full-support run versus the minimal-support run.
[PNG File, 12 KB - games_v13i1e69812_app1.png]

References

1. GBD 2016 Dementia Collaborators. Global, regional, and national burden of Alzheimer's disease and other dementias, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet Neurol 2019 Jan;18(1):88-106. [doi: 10.1016/S1474-4422(18)30403-4] [Medline: 30497964]
2. GBD 2019 Dementia Forecasting Collaborators. Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: an analysis for the Global Burden of Disease Study 2019. Lancet Public Health 2022 Feb;7(2):e105-e125. [doi: 10.1016/S2468-2667(21)00249-8] [Medline: 34998485]
3. Bahar-Fuchs A, Clare L, Woods B. Cognitive training and cognitive rehabilitation for mild to moderate Alzheimer's disease and vascular dementia. Cochrane Database Syst Rev 2013 Jun 5;2013(6):CD003260. [doi: 10.1002/14651858.CD003260.pub2] [Medline: 23740535]
4. Bang J, Spina S, Miller BL. Frontotemporal dementia. Lancet 2015 Oct 24;386(10004):1672-1682. [doi: 10.1016/S0140-6736(15)00461-4] [Medline: 26595641]
5. O'Brien JT, Thomas A. Vascular dementia. Lancet 2015 Oct 24;386(10004):1698-1706. [doi: 10.1016/S0140-6736(15)00463-8] [Medline: 26595643]
6. Scheltens P, Blennow K, Breteler MMB, et al. Alzheimer's disease. Lancet 2016 Jul 30;388(10043):505-517. [doi: 10.1016/S0140-6736(15)01124-1] [Medline: 26921134]
7. Walker Z, Possin KL, Boeve BF, Aarsland D. Lewy body dementias. Lancet 2015 Oct 24;386(10004):1683-1697. [doi: 10.1016/S0140-6736(15)00462-6] [Medline: 26595642]
8. Guidelines for care: person-centred care of people with dementia living in care homes: framework. : Alzheimer Society of Canada; 2011 URL: https://alzheimer.ca/sites/default/files/documents/Guidelines-for-Care_Alzheimer-Society-Canada.pdf [accessed 2025-10-10]
9. Waller S, Masterson A. Designing dementia-friendly hospital environments. Future Hosp J 2015 Feb;2(1):63-68. [doi: 10.7861/futurehosp.2-1-63] [Medline: 31098081]
10. Calkins MP. From research to application: supportive and therapeutic environments for people living with dementia. Gerontologist 2018 Jan 18;58(suppl_1):S114-S128. [doi: 10.1093/geront/gnx146] [Medline: 29361065]
11. Kenigsberg PA, Aquino JP, Bérard A, et al. Dementia beyond 2025: knowledge and uncertainties. Dementia (London) 2016 Jan;15(1):6-21. [doi: 10.1177/1471301215574785] [Medline: 25740575]
12. Zucchella C, Sinforiani E, Tamburin S, et al. The multidisciplinary approach to Alzheimer's disease and dementia. a narrative review of non-pharmacological treatment. Front Neurol 2018;9(1058):1058. [doi: 10.3389/fneur.2018.01058] [Medline: 30619031]
13. Kelley AS, McGarry K, Gorges R, Skinner JS. The burden of health care costs for patients with dementia in the last 5 years of life. Ann Intern Med 2015 Nov 17;163(10):729-736. [doi: 10.7326/M15-0381] [Medline: 26502320]
14. Stensland M, Watson PR, Grazier KL. An examination of costs, charges, and payments for inpatient psychiatric treatment in community hospitals. Psychiatr Serv 2012 Jul;63(7):666-671. [doi: 10.1176/appi.ps.201100402] [Medline: 22588167]
15. Glick ID, Sharfstein SS, Schwartz HI. Inpatient psychiatric care in the 21st century: the need for reform. Psychiatr Serv 2011 Feb;62(2):206-209. [doi: 10.1176/ps.62.2.pss6202_0206] [Medline: 21285100]

XSL•FO
RenderX

16. Goulter N, Kavanagh DJ, Gardner G. What keeps nurses busy in the mental health setting? J Psychiatr Ment Health Nurs 2015 Aug;22(6):449-456. [doi: 10.1111/jpm.12173] [Medline: 25939246]
17. McCallum S, Boletsis C. Dementia games: a literature review of dementia-related serious games. In: Ma M, Oliveira MF, Petersen S, Hauge JB, editors. Serious Games Development and Applications SGDA 2013 Lecture Notes in Computer Science: Springer; 2013, Vol. 8101:15-27. [doi: 10.1007/978-3-642-40790-1_2]
18. Ning H, Li R, Ye X, Zhang Y, Liu L. A review on serious games for dementia care in ageing societies. IEEE J Transl Eng Health Med 2020;8:1400411. [doi: 10.1109/JTEHM.2020.2998055] [Medline: 32537264]
19. Michael DR, Chen SL. Serious Games: Games That Educate, Train, and Inform: Muska & Lipman/Premier-Trade; 2005.
20. Manera V, Petit PD, Derreumaux A, et al. "Kitchen and cooking," a serious game for mild cognitive impairment and Alzheimer's disease: a pilot study. Front Aging Neurosci 2015;7:24. [doi: 10.3389/fnagi.2015.00024] [Medline: 25852542]
21. Boletsis C, McCallum S. Augmented reality cubes for cognitive gaming: preliminary usability and game experience testing. International Journal of Serious Games 2016;3(1). [doi: 10.17083/ijsg.v3i1.106]
22. Tong T, Chignell M, Tierney MC, Lee J. A serious game for clinical assessment of cognitive status: validation study. JMIR Serious Games 2016 May 27;4(1):e7. [doi: 10.2196/games.5006] [Medline: 27234145]
23. Rings S, Steinicke F, Picker T, Prasuhn C. Memory journalist: creating virtual reality exergames for the treatment of older adults with dementia. Presented at: 2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW); Mar 22-26, 2020; Atlanta, GA, USA. [doi: 10.1109/VRW50115.2020.00194]
24. Mader S, Dupire J, Guardiola E, Netkin S. Conception de jeux thérapeutiques pour seniors: l'exemple du village aux oiseaux [Article in French]. ReadkonG. URL: https://fr.readkong.com/page/conception-de-jeux-therapeutiques-pour-seniors-l-exemple-1786764 [accessed 2025-10-10]
25. Ben-Sadoun G, Sacco G, Manera V, et al. Physical and cognitive stimulation using an exergame in subjects with normal aging, mild and moderate cognitive impairment. J Alzheimers Dis 2016 Jun 30;53(4):1299-1314. [doi: 10.3233/JAD-160268] [Medline: 27372645]
26. Benveniste S, Jouvelot P, Pin B, Péquignot R. The MINWii project: renarcissization of patients suffering from Alzheimer's disease through video game-based music therapy. Entertain Comput 2012 Dec;3(4):111-120. [doi: 10.1016/j.entcom.2011.12.004]
27. Bouchard B, Imbeault F, Bouzouane A, Menelas BAJ. Developing serious games specifically adapted to people suffering from Alzheimer. In: Ma M, Oliveira MF, Hauge JB, Duin H, Thoben KD, editors. Serious Games Development and Applications SGDA 2012 Lecture Notes in Computer Science: Springer; 2012, Vol. 7528:243-254. [doi: 10.1007/978-3-642-33687-4_21]
28. Eichhorn C, Plecher DA, Lurz M, et al. Combining motivating strategies with design concepts for mobile apps to increase usability for the elderly and Alzheimer patients. In: Gao Q, Zhou J, editors. Human Aspects of IT for the Aged Population. Healthy and Active Aging. HCII 2020. Lecture Notes in Computer Science: Springer; 2020, Vol. 12208:47-66. [doi: 10.1007/978-3-030-50249-2_4]
29. Eichhorn C, Plecher DA, Klinker G, et al. Innovative game concepts for Alzheimer patients. In: Zhou J, Salvendy G, editors. Human Aspects of IT for the Aged Population Applications in Health, Assistance, and Entertainment ITAP 2018 Lecture Notes in Computer Science: Springer; 2018, Vol. 10927:526-545. [doi: 10.1007/978-3-319-92037-5_37]
30. Contreras-Somoza LM, Irazoki E, Toribio-Guzmán JM, et al. Usability and user experience of cognitive intervention technologies for elderly people with MCI or dementia: a systematic review. Front Psychol 2021;12:636116. [doi: 10.3389/fpsyg.2021.636116] [Medline: 33967901]
31. Gonzáles-Palau F, Franco M, Toribio J, Losada R, Parra E, Bamidis P. Designing a computer-based rehabilitation solution for older adults: the importance of testing usability. PsychNology Journal 2013;11(2):119-136 [FREE Full text]
32. Gao C, Zhou L, Liu Z, Wang H, Bowers B. Mobile application for diabetes self-management in China: do they fit for older adults? Int J Med Inform 2017 May;101:68-74. [doi: 10.1016/j.ijmedinf.2017.02.005] [Medline: 28347449]
33. ISO 9241-11:2018: ergonomics of human-system interaction part 11: usability: definitions and concepts. International Organization for Standardization. 2018. URL: https://www.iso.org/standard/63500.html [accessed 2024-01-08]
34. Dubey SK, Rana A. Analytical roadmap to usability definitions and decompositions. International Journal of Engineering Science and Technology 2010;2(9):4723-4729 [FREE Full text]
35. Abran A, Khelifi A, Suryn W, Seffah A. Usability meanings and interpretations in ISO standards. Software Quality Journal 2003 Nov;11(4):325-338. [doi: 10.1023/A:1025869312943]
36. Eldridge SM, Chan CL, Campbell MJ, et al. CONSORT 2010 statement: extension to randomised pilot and feasibility trials. BMJ 2016 Oct 24;355:i5239. [doi: 10.1136/bmj.i5239] [Medline: 27777223]
37. McKhann GM, Knopman DS, Chertkow H, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011 May;7(3):263-269. [doi: 10.1016/j.jalz.2011.03.005] [Medline: 21514250]
38. Albert MS, DeKosky ST, Dickson D, et al. The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011 May;7(3):270-279. [doi: 10.1016/j.jalz.2011.03.008] [Medline: 21514249]

XSL•FO
RenderX

39. Brooke J. SUS: a “quick and dirty” usability scale. In: Jordan PW, Thomas B, Weerdmeester BA, McClelland IL, editors. Usability Evaluation In Industry: CRC Press; 1996:189-194.
40. Holden RJ. A simplified system usability scale (SUS) for cognitively impaired and older adults. Proceedings of the International Symposium on Human Factors and Ergonomics in Health Care 2020 Sep;9(1):180-182. [doi: 10.1177/2327857920091021]
41. Sauro J, Lewis JR. Quantifying the User Experience: Practical Statistics for User Research, 2nd edition: Elsevier; 2016.
42. Temple RO, Stern RA, Latham J, Ruffolo JS, Arruda JE, Tremont G. Assessment of mood state in dementia by use of the visual analog mood scales (VAMS). Am J Geriatr Psychiatry 2004 Sep;12(5):527-530. [doi: 10.1097/00019442-200409000-00012]
43. Folstein MF, Folstein SE, McHugh PR. “Mini-mental state”. A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975 Nov;12(3):189-198. [doi: 10.1016/0022-3956(75)90026-6] [Medline: 1202204]
44. Cohen J. Statistical Power Analysis for the Behavioral Sciences, 2nd edition: Routledge; 2013. [doi: 10.4324/9780203771587]
45. Cheung MWL. Computing effect sizes for meta‐analysis. In: Cheung MWL, editor. Meta‐Analysis: A Structural Equation Modeling Approach: John Wiley & Sons; 2015:48-80. [doi: 10.1002/9781118957813.ch3]
46. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Soft 2015;67(1):1-48. [doi: 10.18637/jss.v067.i01]
47. Ordinal: regression models for ordinal data. R Project. URL: https://cran.r-project.org/web/packages/ordinal/index.html [accessed 2025-10-10]
48. Kuckartz U, Rädiker S. Qualitative Inhaltsanalyse Methoden, Praxis, Umsetzung Mit Software Und Künstlicher Intelligenz [Book in German]: Beltz Juventa; 2024.
49. Lewis JR. The system usability scale: past, present, and future. International Journal of Human–Computer Interaction 2018 Jul 3;34(7):577-590. [doi: 10.1080/10447318.2018.1455307]
50. Field B, Mountain G, Burgess J, et al. Recruiting hard to reach populations to studies: breaking the silence: an example from a study that recruited people with dementia. BMJ Open 2019 Nov 19;9(11):e030829. [doi: 10.1136/bmjopen-2019-030829] [Medline: 31748295]
51. Lepore M, Shuman S, Wiener J, Gould E. Challenges in involving people with dementia as study participants in research on care and services. Assistant Secretary for Planning and Evaluation - US Department of Health and Human Services. 2017. URL: http://aspe.hhs.gov/sites/default/files/private/pdf/256696/session%25205%2520background.pdf [accessed 2025-10-10]
52. Tong T, Chan JH, Chignell M. Serious games for dementia. Presented at: 26th International Conference on World Wide Web Companion; Apr 3-7, 2017; Perth, Australia. [doi: 10.1145/3041021.3054930]
53. Zmily A, Mowafi Y, Mashal E. Study of the usability of spaced retrieval exercise using mobile devices for Alzheimer’s disease rehabilitation. JMIR Mhealth Uhealth 2014 Aug 14;2(3):e31. [doi: 10.2196/mhealth.3136] [Medline: 25124077]
54. Nygård L, Starkhammar S. The use of everyday technology by people with dementia living alone: mapping out the difficulties. Aging Ment Health 2007 Mar;11(2):144-155. [doi: 10.1080/13607860600844168] [Medline: 17453547]
55. Massimi M, Baecker RM. An empirical study of seniors’ perceptions of mobile phones as memory aids. TAGlab. URL: https://ron.taglab.ca/papers/B16.pdf [accessed 2025-10-10]
56. Sama PR, Eapen ZJ, Weinfurt KP, Shah BR, Schulman KA. An evaluation of mobile health application tools. JMIR Mhealth Uhealth 2014 May 1;2(2):e19. [doi: 10.2196/mhealth.3088] [Medline: 25099179]
57. Tobiasson H. Game over or play it again and again…participatory design approach within special housing. Digitala Vetenskapliga Arkivet. 2010. URL: https://www.diva-portal.org/smash/get/diva2:291681/FULLTEXT02.pdf [accessed 2025-10-10]
58. McLaughlin A, Gandy M, Allaire J, Whitlock L. Putting fun into video games for older adults. Ergonomics in Design: The Quarterly of Human Factors Applications 2012 Apr;20(2):13-22. [doi: 10.1177/1064804611435654]
59. Kang H. Correlates of social engagement in nursing home residents with dementia. Asian Nurs Res (Korean Soc Nurs Sci) 2012 Jun;6(2):75-81. [doi: 10.1016/j.anr.2012.05.006] [Medline: 25030831]
60. Barnes LL, Mendes de Leon CF, Wilson RS, Bienias JL, Evans DA. Social resources and cognitive decline in a population of older African Americans and whites. Neurology (ECronicon) 2004 Dec 28;63(12):2322-2326. [doi: 10.1212/01.wnl.0000147473.04043.b3] [Medline: 15623694]
61. Häikiö J, Wallin A, Isomursu M, Ailisto H, Matinmikko T, Huomo T. Touch-based user interface for elderly users. Presented at: 9th International Conference on Human Computer Interaction with Mobile Devices and Services; Sep 9-12, 2007; Singapore. [doi: 10.1145/1377999.1378021]
62. Boulay M, Benveniste S, Boespflug S, Jouvelot P, Rigaud AS. A pilot usability study of MINWii, a music therapy game for demented patients. Technol Health Care 2011;19(4):233-246. [doi: 10.3233/THC-2011-0628] [Medline: 21849735]
63. Tziraki C, Berenbaum R, Gross D, Abikhzer J, Ben-David BM. Designing serious computer games for people with moderate and advanced dementia: interdisciplinary theory-driven pilot study. JMIR Serious Games 2017 Jul 31;5(3):e16. [doi: 10.2196/games.6514] [Medline: 28760730]
64. Garcia-Sanjuan F, Jaen J, Nacher V. Tangibot: A tangible-mediated robot to support cognitive games for ageing people—a usability study. Pervasive Mob Comput 2017 Jan;34:91-105. [doi: 10.1016/j.pmcj.2016.08.007]

XSL•FO
RenderX

Abbreviations

CLMM: cumulative link mixed model
MCI: mild cognitive impairment
MD: mild dementia
MMSE: Mini-Mental State Examination
NPT: nonpharmaceutical treatment
SG: serious game
SUS: System Usability Scale

Edited by A Coristine; submitted 09.12.24; peer-reviewed by M Munikumar, P Nohl-Deryk; revised version received 29.08.25; accepted 29.08.25; published 27.10.25.

Please cite as:
Brähmer SF, Iffland B, Kreisel S, Driessen M, Trompetter EM, Schomburg M, Toepper M, Steuwe C
Development of a Serious Game App (Digimenz) for Patients With Dementia: Prospective Pilot Study for Usability Testing in Inpatient Treatment and Long-Term Care
JMIR Serious Games 2025;13:e69812
URL: https://games.jmir.org/2025/1/e69812
doi:10.2196/69812

© Sören Freerik Brähmer, Benjamin Iffland, Stefan Kreisel, Martin Driessen, Eva M Trompetter, Meret Schomburg, Max Toepper, Carolin Steuwe. Originally published in JMIR Serious Games (https://games.jmir.org), 27.10.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Age Differences in Flow Experience During Body Movement–Controlled Video Game Rehabilitation Tasks: Cross-Sectional Study

Shiyi Tong[1], BEng; Lixuan Li[2], MEng; Yuankun Zhang[1], BEng; Weiyan Ren[3], PhD; Fang Pu[1], PhD

[1]School of Biological Science and Medical Engineering, Beihang University, 37 Xueyuan Road, Haidian District, Beijing, China

[2]Department of Medical Innovation and Research, Chinese PLA General Hospital, Beijing, China

[3]School of Engineering Medicine, Beihang University, Beijing, China

Corresponding Author:
Fang Pu, PhD
School of Biological Science and Medical Engineering, Beihang University, 37 Xueyuan Road, Haidian District, Beijing, China

Abstract

Background: Body movement–controlled video games (BMCVGs) are increasingly adopted in rehabilitation because they combine physical training with interactive engagement. Flow experience, a critical factor for enhancing user engagement and training efficacy, exhibits age-related differences that are essential for designing age-appropriate rehabilitation tasks. However, current BMCVG rehabilitation tasks often overlook these age-related differences in subjective experience, leading to insufficient engagement among older adults.

Objective: This study aimed to explore differences in flow experience between younger and older adults when performing the same BMCVG rehabilitation task and to provide empirical evidence for designing personalized and age-appropriate programs.

Methods: A total of 40 participants were recruited, including 21 older adults (mean age 63.00, SD 6.64 y; n=10, 48% male participants) and 19 younger adults (mean age 24.68, SD 1.16 y; n=9, 47% male participants). Participants performed the "Space Pop" task in *Kinect Adventures*, simulating limb coordination training. Flow experience was assessed using the Chinese version of the Flow State Scale-2, which measures 9 dimensions of flow. Group differences were analyzed using the nonparametric Wilcoxon rank-sum tests, and effect sizes (Cohen *d*) were calculated via bootstrap estimation.

Results: Older adults exhibited significantly lower overall flow experience than younger adults (W=339.5; *P*<.001; Cohen *d*=1.45; η^2=0.37). Significant differences were also found in the dimensions of "challenge-skill balance" (W=339; *P*<.001); "clear goals" (W=271; *P*=.04); "sense of control" (W=389.5; *P*<.001); and "loss of self-consciousness" (W=268; *P*=.048). The largest effect was observed in the "sense of control" dimension (Cohen *d*=3.22; η^2=0.74), indicating it was the most significantly impacted by age. Other dimensions (eg, concentration and time transformation) showed no significant differences.

Conclusions: Age plays a significant role in shaping flow experiences during BMCVG rehabilitation tasks. Older adults' reduced flow may be attributed to declines in cognitive processing speed, motor control, and self-efficacy, which particularly impair their sense of control and goal clarity. Tailoring designs through strategies such as dynamic difficulty adjustment, clearer goal cues, and reduced motor demands is crucial. These adaptations can enhance older adults' sense of control and immersion, promoting active participation and ultimately improving rehabilitation outcomes.

(JMIR Serious Games 2025;13:e76278) doi:10.2196/76278

KEYWORDS

flow; rehabilitation; video games; exergames; task design; age differences

Introduction

Role of Flow Experience in Body Movement–Controlled Video Games for Rehabilitation

"Flow" is a state of optimal experience where individuals are fully engaged in challenging activities, experiencing high focus, control, and enjoyment [1]. Due to its positive effects, flow has been widely studied in sports, music, gaming, learning, and other fields [2-5], enhancing performance, engagement, and task persistence.

In the context of body movement–controlled video games (BMCVGs) applied to rehabilitation, the characteristics of flow are particularly important. BMCVGs are a genre of video gaming where users control the game through their own body movements [6,7]. These games, which also belong to exergames, are increasingly being explored for their potential in rehabilitation, such as fall prevention and improving balance

XSL•FO
RenderX

and mobility [8,9]. However, despite advancements in BMCVG hardware, rehabilitation task design often fails to integrate mechanisms that sustain engagement, leading to poor subjective experiences [10]. Current tasks often use a "one-size-fits-all" design, ignoring age-related differences in subjective experiences caused by older adults' cognitive and motor ability declines, which affects training compliance. Flow theory provides a framework for assessing and optimizing user experiences, making it a useful tool for the design of BMCVG tasks intended for rehabilitation [11].

Theoretical Basis for Age-Related Flow Differences

Flow is influenced by individual differences, including personality traits, gender, motivation, and mental health [2,12-15]. Research has shown that video game characteristics and player identification are important precursors to the flow state [6,16-18]. Age, a critical factor in psychological experiences, may influence flow in BMCVGs. Murman [19] noted that cognitive abilities (eg, information processing speed and executive function) decline with age, and Gazzaley and Nobre [20] emphasized older adults' difficulties in attention allocation and maintenance. Boot et al [21] found that older adults have less video game experience and different preferences from younger individuals. These cognitive, attentional, and preference changes may impact their flow experience in BMCVG tasks. Older adults may face challenges in achieving flow due to cognitive and motor declines. However, empirical research on age differences in the flow experience within BMCVG rehabilitation remains limited, especially for typical rehabilitation tasks such as limb coordination training.

Research Objectives and Hypotheses

This study compared 9D flow experiences between young and older adults in a BMCVG task that simulates activities often targeted in rehabilitation to test the hypothesis that age differences affect flow experience, aiming to (1) identify key dimensions of age-related flow experience differences and (2) provide theoretical evidence for adapting task difficulty and optimizing interfaces for older adults, promoting better design of BMCVG tasks for rehabilitation.

Methods

Participants

In total, 40 participants were recruited and divided into an older adult group (21 participants; mean age 63.00, SD 6.64 y; n=10, 48% male participants) and a young adult group (19 participants; mean age 24.68, SD 1.16 y; n=9, 47% male participants). Older adults were recruited from the community of Beihang University, while younger adults were students at Beihang University. Data collection took place between October and December 2024.

The inclusion criteria were as follows: (1) normal or corrected-to-normal vision, (2) no neuromuscular diseases, (3) no previous experience with Kinect, and (4) the ability to complete questionnaires independently. The exclusion criteria were as follows: (1) a current or historical medical recommendation against strenuous exercise and (2) any chronic illnesses that could impact physical or cognitive performance.

Ethical Considerations

The study was approved by the Science and Ethics Committee of the School of Engineering Medicine, Beihang University (BM20230154). All participants were briefed on the study purposes and procedures, and they all gave written informed consent before participation. All collected data were deidentified by assigning unique participant ID codes in place of names. Each study participant was thanked and received 150 RMB (US $21) as a token of gratitude.

Experimental Tools

The experimental equipment included an Xbox 360 (Microsoft Corp) gaming console, a Kinect v1 sensor (Microsoft Corp), and a 28-inch 4K display (Samsung U28E590D). The task selected for this study was "Space Pop" from Kinect Adventures (Microsoft Studios, Microsoft Corp), a BMCVG chosen because it involves skills pertinent to rehabilitation goals. The rehabilitation task selected was "Space Pop" from Kinect Adventures, which required participants to control a virtual character to pop bubbles through limb movements (forward, backward, left, and right) and arm actions (flapping and dropping). Real-time scores and ranks (none, bronze, silver, gold, and platinum) were displayed, with the task goal of popping as many bubbles as possible to simulate limb coordination and reaction speed training in rehabilitation.

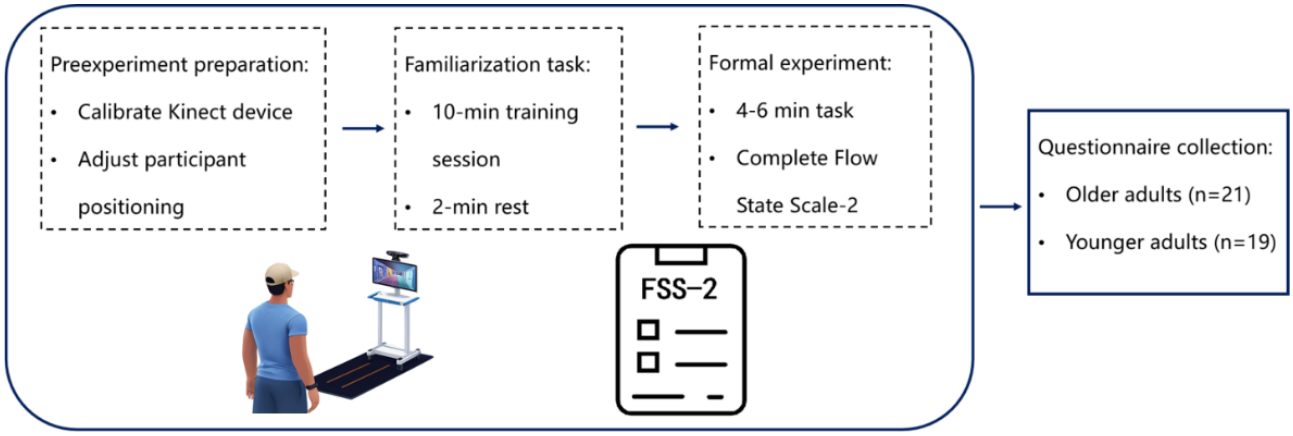
The Flow State Scale-2 (FSS-2; Multimedia Appendix 1) [22] was adapted into the Chinese version (CFSS-2) through translation and back-translation. The scale consists of 9 dimensions (36 items). Participants completed a paper version of the questionnaire using a 5-point Likert scale (1=strongly disagree and 5=strongly agree). Instructions emphasized that there were no right or wrong answers.

Experimental Procedure

Before the experiment, the Kinect sensor was calibrated for optimal recognition, and participant positioning was adjusted to the recommended distance (2 - 3 m from the display) within the Kinect's effective range. Participants completed a 10-minute training session (repeating the task twice) to ensure familiarity with the task operations, followed by a 2-minute rest to restore physical stamina. The formal task lasted 4 - 6 minutes. After task completion, participants immediately completed the CFSS-2 questionnaire based on their experience during the task. The experiment was terminated immediately if any participant reported physical discomfort. The experimental procedure is outlined in Figure 1.

XSL•FO
RenderX

Figure 1. Experimental procedure for the body movement–controlled video game rehabilitation study using Kinect "space pop" task (Beihang University, Beijing, October-December 2024; 21/40, 53% older adults and 19/40, 47% younger adults).



Statistical Analysis

For each questionnaire, the mean score of the 4 items in each dimension was calculated as the dimension score and the mean score of all items in the questionnaire was calculated as the participant's flow experience score during the task.

Using R (version 4.3.3; R Foundation for Statistical Computing), reliability and validity tests of the CFSS-2 were conducted by calculating the Cronbach α coefficient and item factor loadings. Then, normality tests (Shapiro-Wilk test) and homogeneity of variance tests (Levene test) were performed for the older and younger groups. If data met both normality and homogeneity of variance criteria, independent sample 2-tailed *t* tests were used to analyze significant differences in overall flow experience and its dimensions between the 2 groups. If either criterion was not met, the nonparametric Wilcoxon rank-sum test was applied to examine significant differences in flow experience, with a significance level set at α=.05. Furthermore, due to the small sample size of this study, the corrected Cohen *d* (Hedges *g*) was used to calculate the effect size, and the 95% CI was estimated via the bootstrap method to enhance the robustness of the results. Posterior statistical power was calculated using G*Power (version 3.1.9.7; Heinrich-Heine-Universität Düsseldorf).

Results

Participants took an average of 1.2 (SD 0.1) minutes to complete the questionnaire after the task. A total of 40 questionnaires were collected (100% response rate). Table 1 shows the demographic characteristics of the participants. Each participant completed all experimental tasks. The older and younger groups did not differ significantly in terms of sex distribution (*P*>.99), while, as expected, there was a significant difference in age between the 2 groups (*P*<.001).

Table . Participants' demographic characteristics (N=40).

Characteristics	Older adults (n=21)	Younger adults (n=19)	*P* value
Age (y), mean (SD)	63.00 (6.64)	24.68 (1.16)	<.001
Sex, n (%)			>.99
Male	10 (48)	9 (47)	
Female	11 (52)	10 (53)	

The Cronbach α coefficient for the scale ranged from .83 to .98 (total scale α=.86), with item factor loadings between 0.44 and 0.98 (mean factor loading=0.87), indicating satisfactory internal consistency and cross-cultural, cross-domain applicability. Because the data did not conform to normal distribution or homogeneity of variance, we used the Wilcoxon rank-sum test (the nonparametric equivalent of the independent sample *t* test) for between-group comparisons (α=.05). The analysis revealed significant differences between the older and younger groups in overall flow experience and several dimensions (Table 2). A radar chart was plotted to further analyze the differences in the mean flow scores between the 2 groups (Figure 2), showing that younger adults scored higher than older adults in most dimensions (except for "time transformation" and "autotelic experience").

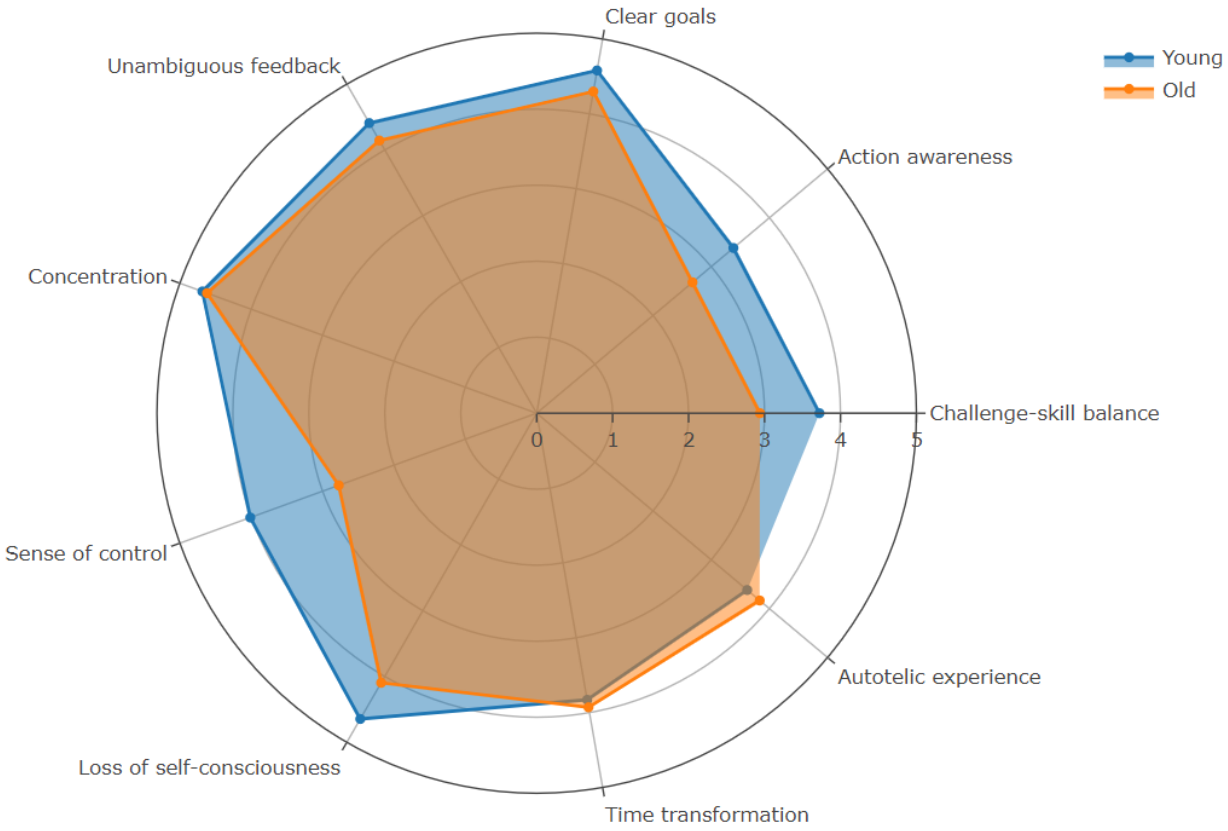
XSL•FO
RenderX

Table . Age-related differences in flow experience dimensions (N=40; nonparametric Wilcoxon rank-sum test).

Dimensions	W value	*P* value	Cohen *d*	η^2
Challenge-skill balance	339	<.001[a]	1.28	0.31
Action awareness	267	.07	0.73	0.13
Clear goals	271	.04[a]	0.65	0.10
Unambiguous feedback	256	.09	0.69	0.12
Concentration	182.5	.62	0.12	<0.01
Sense of control	389.5	<.001[a]	3.22	0.74
Loss of self-consciousness	268	.048[a]	0.57	0.08
Time transformation	224	.48	-0.12	<0.01
Autotelic experience	181.5	.63	-0.25	0.02
Total	339.5	<.001[a]	1.45	0.37

[a]*P*<.05.

Figure 2. Radar chart of mean flow experience scores across 9 dimensions in older and younger adults.



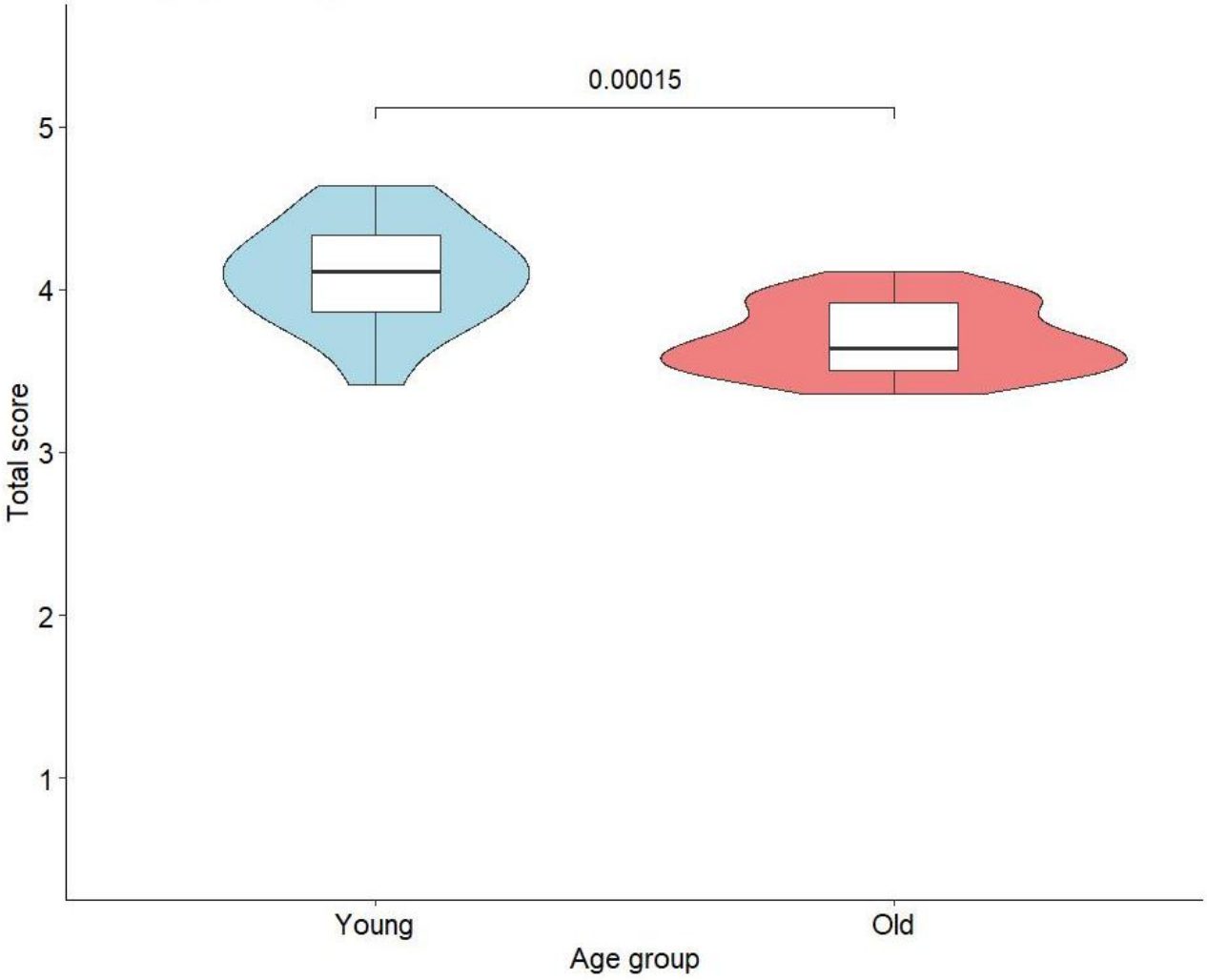
For overall flow experience, there was a significant difference between the older and younger groups (W=339.5; *P*<.001), indicating that older and younger adults obtain distinct flow experiences in the same BMCVG task, with the flow distribution among older adults being more concentrated (Figure 3). The effect size was large (Cohen *d*=1.45; η^2=0.37), suggesting a substantial impact of age on overall flow experience. The corresponding statistical power reached 0.98, which also strengthened the reliability of the significant results.

Regarding different flow dimensions (Figure 4), significant age differences were found in 4 dimensions: "challenge-skill balance" (W=339; *P*<.001; Cohen *d*=1.28; η^2=0.31); "clear goals" (W=271; *P*=.04; Cohen *d*=0.65; η^2=0.10); "sense of control" (W=389.5; *P*<.001; Cohen *d*=3.22; η^2=0.74); and "loss of self-consciousness" (W=268; *P*=.048; Cohen *d*=0.57; η^2=0.08). Among these, the "sense of control" dimension had the largest effect size (η^2=0.74). Other dimensions such as "action awareness" (*P*=.07; Cohen *d*=0.73; η^2=0.13) and

XSL•FO
RenderX

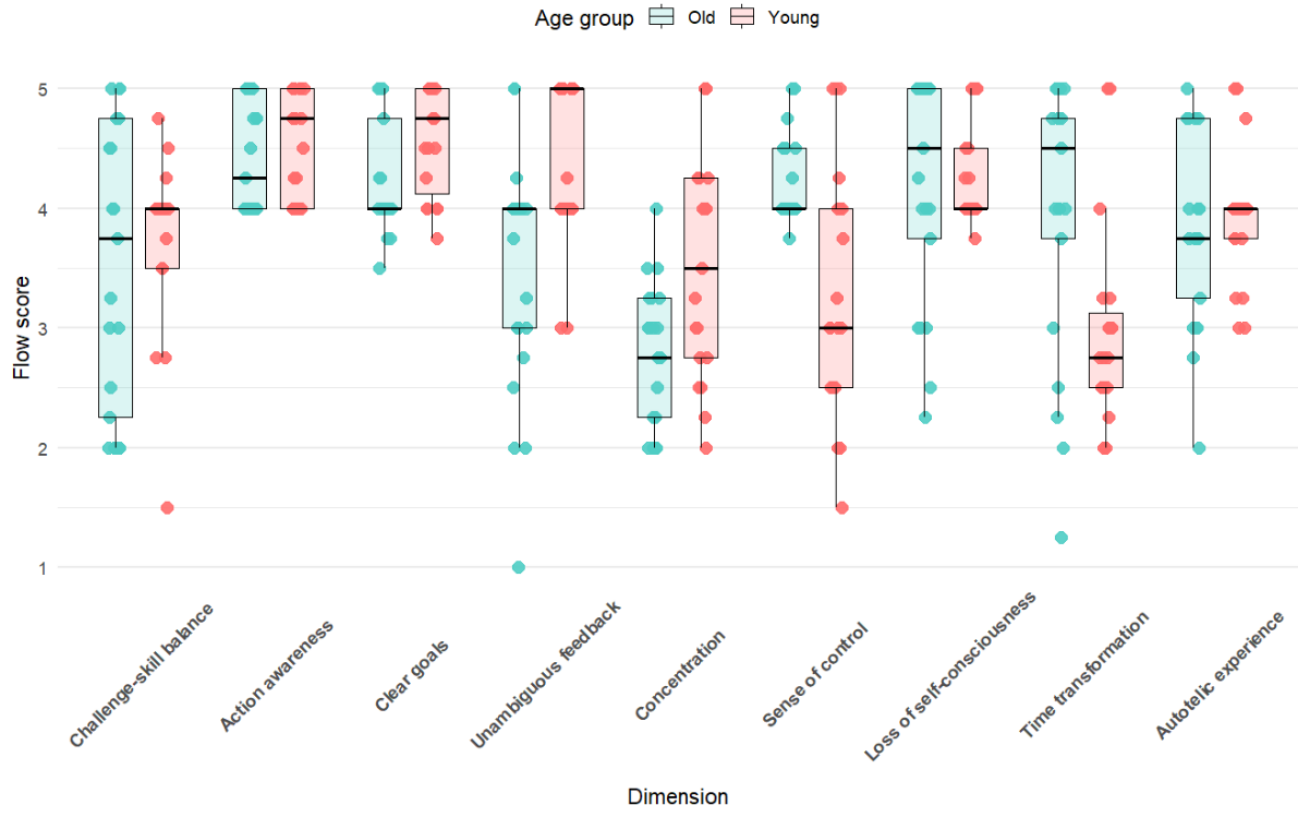
"unambiguous feedback" (P=.09; Cohen d=0.69; η^2=0.12) had P values close to the significance threshold, with effect size indicators suggesting a possible weak influence, but they did not reach statistical significance. No significant age differences were observed in "concentration" (P=.62; Cohen d=0.12; η^2<0.01), "time transformation" (P=.48; Cohen d=−0.12; η^2<0.01), and "autotelic experience" (P=.63; Cohen d=−0.25 η^2=0.02), all of which had small effect sizes.

Figure 3. Violin plot of overall flow experience scores in older versus younger adults.



XSL•FO
RenderX

Figure 4. Comparison of flow experience scores across 9 dimensions between older and younger adults.



Discussion

Principal Findings

This study explored the impact of age on flow experience during BMCVG rehabilitation tasks. The results demonstrated that older adults scored significantly lower than younger adults in terms of overall flow experience and several core dimensions. Specifically, statistically significant differences were observed between the 2 groups in 4 dimensions: "challenge-skill balance" (P<.001; η^2=0.31); "clear goals" (P=.04; η^2=0.10); "sense of control" (P<.001; η^2=0.74); and "loss of self-consciousness" (P=.048; η^2=0.08). Among these dimensions, "sense of control" had the largest effect size, indicating that age differences had the most prominent impact on this dimension. These findings provide direct empirical evidence that age significantly shapes flow experience in BMCVG rehabilitation tasks.

Comparison to Previous Work

The perceived match between task challenge and personal skill among older adults was significantly lower than that among younger adults (P<.001; η^2=0.31), meaning older adults perceived task challenges as being either too high or too low for their own skills, resulting in a poor match between challenge and skill. This may be related to age-related cognitive decline. Meta-analyses have shown that physical activity levels have a protective effect on cognitive function [23-25]. Older adults generally experience a degradation of brain function due to reduced physical activity [26]. Especially when task challenges increase, the activation level of brain regions in older adults is equivalent to or lower than that in younger adults, with poorer behavioral performance [27]. Additionally, the complexity of motor tasks (such as those requiring high coordination) places more demanding cognitive requirements on older adults, leading to more obvious cognitive decline. We speculate that cognitive differences between older and younger adults when performing the same task may contribute to the differences in the "challenge-skill balance" dimension.

In the "clear goals" dimension, older adults scored significantly lower than younger adults (P=.04; η^2=0.10), indicating they had more difficulty clearly perceiving task goals during the task. This is likely related to the decline in cognitive abilities and task comprehension in older adults. With age, individuals experience slower information processing and reduced attention maintenance, making it harder for them to quickly understand and internalize task goals. Research has shown that older adults process information more slowly, which may lead to poor performance in tasks requiring rapid target recognition and response [28]. Furthermore, older adults' perception of task goals may be more influenced by situational factors, such as the complexity of visual information and the clarity of cues, which may further affect their goal perception in BMCVG tasks. Studies have shown that in visual search tasks, if the target presentation time is short or there are many interfering factors, older adults may have difficulty clearly perceiving the target [29].

Older adults had a significantly lower sense of control over the task than younger adults (P<.001; η^2=0.74), meaning they were less likely to feel a sense of mastery over the task, and the difference in this dimension was the most significant. This result supports the findings of Hibbs et al [30], who found that older adults had a weaker sense of control than younger adults in low-screen virtual reality conditions, and this study further

XSL•FO
RenderX

confirms the significance of this result. This may be related to the objective physical and psychological conditions of older adults. On the one hand, due to the decline in motor abilities, older adults may feel limited control over tasks that require rapid response or fine motor control, affecting their sense of task mastery. Research has shown that older adults exhibit reduced accuracy, increased variability, and prolonged execution time when performing bimanual motor tasks, failing to achieve expected performance [31]. On the other hand, older adults may experience more anxiety or lack confidence in task completion when faced with new tasks [32]. This decline in self-efficacy may affect their sense of control over tasks. Especially when faced with new or complex tasks, older adults may doubt their abilities, leading to a lower sense of control. As people enter old age, physical functions and activity levels gradually decline, and psychological characteristics also change [33]. Due to reduced sensory abilities, slower neuromotor speed, and memory decline, older adults become more cautious. We therefore propose that the decline in physical abilities and self-efficacy is the primary reason for the lower sense of control observed in older adults.

The "loss of self-consciousness" dimension showed that older adults scored significantly lower than younger adults (P=.048; η^2=0.08), indicating they had more difficulty fully immersing themselves and experiencing self-forgetfulness in BMCVG tasks. This is consistent with the research by Boot et al [34] on age-related differences in immersion in virtual environments, which pointed out that many older adults have less experience and proficiency with emerging technologies, which may become a barrier to their integration into new virtual environments. In addition, research has shown that there are differences in the sense of immersion in BMCVGs among different age groups, which may be related to motor performance and subjective experience [30]. We hypothesize that inattention or memory loss in older adults may lead to poor motor performance and concerns about task failure or mistakes, making it difficult for them to fully engage in the BMCVG experience while maintaining a high level of self-awareness.

In summary, older adults' lower scores in these dimensions may stem from declines in cognitive abilities, objective physical strength, and self-efficacy, while younger adults adapt more easily, maintaining immersion and task mastery. These findings highlight the need for age-sensitive rehabilitation task design. Designers should optimize task difficulty, provide clearer goals, and reduce motor demands, especially to enhance older adults' sense of control and immersion.

It is worth noting that in this study, dimensions such as "action awareness" (P=.07; η^2=0.13); "unambiguous feedback" (P=.09; η^2=0.12); "concentration" (P=.62; η^2<0.01); "time transformation" (P=.48; η^2<0.01); and "autotelic experience" (P=.63; η^2=0.02) did not show significant age differences. Among them, the first two were near the significance threshold (.05), and the effect sizes were close to medium levels. Although they did not reach significance, increasing the sample size in the future may further reveal potential differences. The low effect sizes in "concentration," "time transformation," and "autotelic experience" suggest that these flow characteristics may be more universal. The consistency in the "concentration" dimension may be related to the intuitiveness of the task design, where the task's low cognitive load and situational attractiveness allow older adults to maintain high focus. The lack of differences in "time transformation" may be due to the short task duration (average 5 min), where differences in time perception between older and younger adults may not have been fully reflected. The nonsignificant difference in "autotelic experience" is consistent with previous research findings, reflecting the universality of the task's interestingness design. Türksoy et al [13] pointed out that similar task interestingness can provide similar intrinsic motivation experiences. Therefore, the low effect sizes (η^2<0.02) in "concentration," "time transformation," and "autotelic experience" are more likely to reflect true commonalities between age groups. For the "action awareness" (P=.07) and "unambiguous feedback" (P=.09) dimensions close to the significance threshold, their medium effect sizes (η^2=0.12 - 0.13) suggest that a larger sample size may be needed to confirm potential differences in these dimensions.

Limitations

This study has several limitations. First, the use of a single BMCVG task limits the external validity of our findings. The task primarily involves limb coordination and reaction speed, and the observed age-related differences in flow might not be generalizable to other types of rehabilitation tasks, such as those focused on cognitive training, fine motor skills, or balance control. The task-dependent nature of flow experience means that different game genres could yield different results. Second, our sample was confined to younger and older adults, excluding a middle-aged cohort. This binary comparison prevents a full understanding of how flow experience evolves across the lifespan. Incorporating a middle-aged group would provide a more complete picture of the developmental continuum of flow in BMCVG environments. Furthermore, the recruitment of older adults from the Beihang University community and younger adults from the university's student body poses another potential limitation. This specific sample may introduce a bias toward individuals with relatively higher education, cognitive function, and socioeconomic status, potentially affecting the generalizability of our results to a broader population with more diverse backgrounds. Third, while the gender distribution within our groups was relatively balanced, this study did not analyze potential gender-based differences in flow experience. Given that some research suggests gender can influence gaming preferences and experiences, this remains an area for future investigation. Finally, although the post hoc power analysis indicated adequate power for the primary outcome, such analyses depend on observed effect sizes and should not substitute for a priori sample size planning.

Future Directions

To build upon our findings, future research should address these limitations. A key priority is to expand the variety of BMCVG rehabilitation tasks investigated. Studies could compare tasks with different cognitive and motor demands to determine whether the age-related differences in flow observed here are consistent across various rehabilitation scenarios. In addition, although the Xbox Kinect platform provided a validated and

XSL•FO
RenderX

accessible tool for this study, future research should examine whether similar age-related flow differences are observed with newer motion-sensing technologies (eg, Kinect V2, depth cameras, and virtual reality systems). Such investigations would help establish the generalizability of our findings across different rehabilitation platforms and ensure technological relevance in evolving clinical and community settings. Furthermore, future studies should include a middle-aged demographic (eg, people aged between 30 and 55 y) to map the trajectory of flow experience across a broader age spectrum, moving beyond a simple young-versus-old comparison. This would allow for a more nuanced understanding of age-related changes. Finally, considering a wider range of individual differences, such as gender, personality traits, and educational level, could help tailor BMCVG rehabilitation designs to meet personalized user needs more effectively.

Conclusions

This cross-sectional study (N=40) revealed a significant impact of age on flow experience in BMCVG rehabilitation tasks. Older adults exhibited lower flow levels than younger adults, especially in “challenge-skill balance,” “clear goals,” “sense of control,” and “loss of self-consciousness.” The largest effect was observed in the “sense of control” dimension (Cohen *d*=3.22). These findings indicate that age-related differences in cognitive and motor functions affect subjective experiences during BMCVG rehabilitation tasks. Designing BMCVG rehabilitation tasks for older adults should emphasize optimization of task difficulty, explicit goal cues, and reduced motor demands, thereby improving engagement and outcomes.

Acknowledgments

The authors would like to thank all study participants for the time and effort they dedicated to participating in the research. For assistance with grammatical and spelling checks during the writing process, the authors used DeepSeek and Doubao. This study was funded by the National Key Research and Development Program of China (grant 2023YFC3603700); the Beijing Natural Science Foundation, China (grant L244083); the Fundamental Research Funds for the Central Universities, China (grant YWF-23-YGQB-042); and the Chongqing Technology Innovation and Application Development Key Project (CSTB2024TIAD-KPX0035).

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

ST was responsible for methodology, investigation, formal analysis, and visualization, as well as writing the original draft of the manuscript. LL was responsible for methodology and investigation and data collection. YZ was responsible for data curation and reviewing and editing the manuscript. WR was responsible for supervision and reviewing and editing the manuscript. FP was responsible for conceptualization, resources, supervision, as well as reviewing and editing the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Flow State Scale-2 questionnaire.
[PDF File, 150 KB - games_v13i1e76278_app1.pdf]

References

1. Csikszentmihalyi M. Beyond Boredom & Anxiety - Experiencing Flow in Work & Play: Jossey-Bass Publishers; 2000.
2. Paula R, Castanheira LF, da Silva OA, Ferreira RW, Gaia JW, Pires DA. Perception of the flow dimensions and influence of the variable sex in school handball athletes: a cross-sectional study. Retos 2024;51:1110-1115. [doi: 10.47197/retos.v51.99336]
3. Li J, Xu Y, Wang X, et al. Flow and performance: a quantitative study of elicitation modeling in a piano performance perspective. Front Psychol 2024 Jul 15;15:1386831. [doi: 10.3389/fpsyg.2024.1386831] [Medline: 39077214]
4. Paz GO, Freisztav MC. Flow state as a performance measure in Esports. Interdisciplinaria 2024;41(1):14-15. [doi: 10.16888/interd.2024.41.1.14]
5. Kahyaoğlu Erdoğmuş Y, Kurt AA. Digital game-based learning: pedagogical agent and feedback types on achievement, flow experience, and cognitive load. Educ Inf Technol 2024 Jul;29:12943-12968. [doi: 10.1007/s10639-023-12368-2]
6. Thin AG, Hansen L, McEachen D. Flow experience and mood states while playing body movement-controlled video games. Games Cult 2011;6(5):414-428. [doi: 10.1177/1555412011402677]

XSL•FO
RenderX

7. Zhang G. Creating enhanced user experience through persona and interactive design: a case of designing a motion sensing game. In: Marcus A, Wang W, editors. Design, User Experience, and Usability. Design Philosophy and Theory: Springer; 2019:382-394. [doi: 10.1007/978-3-030-23570-3_28]
8. Bronner S, Pinsker R, Naik R, Noah JA. Physiological and psychophysiological responses to an exer-game training protocol. J Sci Med Sport 2016 Mar;19(3):267-271. [doi: 10.1016/j.jsams.2015.03.003] [Medline: 25824058]
9. Alghamdi M, Vallis LA, Leat SJ. The association between visual attention and body movement-controlled video games, balance and mobility in older adults. BMC Geriatr 2021 Jun 30;21(1):405. [doi: 10.1186/s12877-021-02358-9] [Medline: 34193074]
10. Cao S. Virtual reality applications in rehabilitation. In: Kurosu M, editor. Human-Computer Interaction Theory, Design, Development and Practice: Springer; 2016:3-10. [doi: 10.1007/978-3-319-39510-4_1]
11. Suh A, Prophet J. The state of immersive technology research: a literature analysis. Comput Human Behav 2018 Sep;86:77-90. [doi: 10.1016/j.chb.2018.04.019]
12. Ullén F, de Manzano Ö, Almeida R, et al. Proneness for psychological flow in everyday life: associations with personality and intelligence. Pers Individ Dif 2012 Jan;52(2):167-172. [doi: 10.1016/j.paid.2011.10.003]
13. Türksoy A, Altıncı EE, Üster U. Relationship between motivation and dispositional flow state on football players participating in the U13-U16 football leagues. Procedia Soc Behav Sci 2015 May;185:301-306. [doi: 10.1016/j.sbspro.2015.03.423]
14. Burke RJ, Koyuncu M, Fiksenbaum L. Flow, work satisfactions and psychological well-being among nurses in Turkish hospitals. Arch Nurs Pract Care 2016 Feb;2(1):010-017. [doi: 10.17352/anpc.000007]
15. Asakawa K. Flow experience, culture, and well-being: how do autotelic Japanese college students feel, behave, and think in their daily lives? J Happiness Stud 2010;11:205-223. [doi: 10.1007/s10902-008-9132-3]
16. Hull DC, Williams GA, Griffiths MD. Video game characteristics, happiness and flow as predictors of addiction among video game players: a pilot study. J Behav Addict 2013 Sep;2(3):145-152. [doi: 10.1556/JBA.2.2013.005] [Medline: 25215196]
17. Laffan DA, Greaney J, Barton H, Kaye LK. The relationships between the structural video game characteristics, video game engagement and happiness among individuals who play video games. Comput Human Behav 2016 Dec;65:544-549. [doi: 10.1016/j.chb.2016.09.004]
18. Soutter AR, Hitchens M. The relationship between character identification and flow state within video games. Comput Human Behav 2016 Feb;55:1030-1038. [doi: 10.1016/j.chb.2015.11.012]
19. Murman DL. The impact of age on cognition. Semin Hear 2015 Aug;36(3):111-121. [doi: 10.1055/s-0035-1555115] [Medline: 27516712]
20. Gazzaley A, Nobre AC. Top-down modulation: bridging selective attention and working memory. Trends Cogn Sci 2012 Feb;16(2):129-135. [doi: 10.1016/j.tics.2011.11.014]
21. Boot WR, Andringa R, Harrell ER, Dieciuc MA, Roque NA. Older adults and video gaming for leisure: lessons from the Center for Research and Education on Aging and Technology Enhancement (CREATE). Gerontechnology 2020 Jul;19(2):138-146. [doi: 10.4017/gt.2020.19.2.006.00]
22. Jackson SA, Eklund RC. Assessing flow in physical activity: the Flow State Scale-2 and Dispositional Flow Scale-2. J Sport Exerc Psychol 2002;24(2):133-150. [doi: 10.1123/jsep.24.2.133]
23. Colcombe S, Kramer AF. Fitness effects on the cognitive function of older adults: a meta-analytic study. Psychol Sci 2003 Mar;14(2):125-130. [doi: 10.1111/1467-9280.t01-1-01430] [Medline: 12661673]
24. Kramer AF, Erickson KI. Effects of physical activity on cognition, well-being, and brain: human interventions. Alzheimers Dement 2007 Apr;3(2 Suppl):S45-S51. [doi: 10.1016/j.jalz.2007.01.008] [Medline: 19595974]
25. Sofi F, Valecchi D, Bacci D, et al. Physical activity and risk of cognitive decline: a meta-analysis of prospective studies. J Intern Med 2011 Jan;269(1):107-117. [doi: 10.1111/j.1365-2796.2010.02281.x] [Medline: 20831630]
26. Ingold M, Tulliani N, Chan CC, Liu KP. Cognitive function of older adults engaging in physical activity. BMC Geriatr 2020 Jul 2;20(1):229. [doi: 10.1186/s12877-020-01620-w] [Medline: 32616014]
27. Reuter-Lorenz PA, Cappell KA. Neurocognitive aging and the compensation hypothesis. Curr Dir Psychol Sci 2008;17(3):177-182. [doi: 10.1111/j.1467-8721.2008.00570.x]
28. Verhaeghen P. The Elements of Cognitive Aging: Meta-Analyses of Age-Related Differences in Processing Speed and Their Consequences: Oxford University Press; 2013. [doi: 10.1093/acprof:oso/9780195368697.001.0001]
29. Liu H, Liu H, Li F, Han B, Wang C. Effect of cognitive control on attentional processing of emotional information among older adults: evidence from an eye-tracking study. Front Aging Neurosci 2021 Apr 29;13:644379. [doi: 10.3389/fnagi.2021.644379] [Medline: 33994995]
30. Hibbs A, Tempest G, Hettinga F, Barry G. Impact of virtual reality immersion on exercise performance and perceptions in young, middle-aged and older adults. PLoS One 2024 Oct 30;19(10):e0307683. [doi: 10.1371/journal.pone.0307683] [Medline: 39475865]
31. Krehbiel LM, Kang N, Cauraugh JH. Age-related differences in bimanual movements: a systematic review and meta-analysis. Exp Gerontol 2017 Nov;98:199-206. [doi: 10.1016/j.exger.2017.09.001] [Medline: 28890358]
32. An J, Zhu X, Wan K, et al. Older adults' self-perception, technology anxiety, and intention to use digital public services. BMC Public Health 2024 Dec 19;24(1):3533. [doi: 10.1186/s12889-024-21088-2] [Medline: 39702092]

XSL•FO
RenderX

33. Butler RN, Lewis MI. Aging & Mental Health: Positive Psychosocial Approaches: C. V. Mosby; 1973.
34. Boot WR, Dilanchian A, Andringa R. Exploring older adults' perceptions of presence and immersion in diverse virtual environments. Innov Aging 2019 Nov;3(Supplement_1):S239-S240. [doi: 10.1093/geroni/igz038.895]

Abbreviations

BMCVG: body movement–controlled video game
CFSS-2: Chinese version of the Flow State Scale-2
FSS-2: Flow State Scale-2

Edited by A Coristine; submitted 20.04.25; peer-reviewed by D Ajayi, J Ambros-Antemate, V Wattanasoontorn; revised version received 09.10.25; accepted 09.10.25; published 29.10.25.

Please cite as:
Tong S, Li L, Zhang Y, Ren W, Pu F
Age Differences in Flow Experience During Body Movement–Controlled Video Game Rehabilitation Tasks: Cross-Sectional Study
JMIR Serious Games 2025;13:e76278
URL: https://games.jmir.org/2025/1/e76278
doi:10.2196/76278

© Shiyi Tong, Lixuan Li, Yuankun Zhang, Weiyan Ren, Fang Pu. Originally published in JMIR Serious Games (https://games.jmir.org), 29.10.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Comparing Caregiver Perceptions of a Social Robot and Tablet for Serious Game Delivery in Dementia Care: Cross-Sectional Comparison Study

Dorothy Bai[1], PhD; Kelvin Tan Cheng Kian[2], PhD; Po-Yin Chen[3], PhD; Yeh-Liang Hsu[4], PhD; Gong-Hong Lin[5], PhD

[1]School of Gerontology and Long-Term Care, College of Nursing, Taipei Medical University, Taipei, Taiwan

[2]Minor in Applied Ageing Studies programme, S R Nathan School of Human Development, Singapore University of Social Sciences, Singapore

[3]Department of Physical Therapy and Assistive Technology, National Yang Ming Chiao Tung University, Taipei, Taiwan

[4]Gerontechnology Research Center, Yuan Ze University, Taoyuan, Taiwan

[5]International PhD Program in Gerontology and Long-Term Care, College of Nursing, Taipei Medical University, No. 250, Wuxing Street, Xinyi District, Taipei, Taiwan

Corresponding Author:
Gong-Hong Lin, PhD
International PhD Program in Gerontology and Long-Term Care, College of Nursing, Taipei Medical University, No. 250, Wuxing Street, Xinyi District, Taipei, Taiwan

Abstract

Background: Social robots integrated with serious games hold promise as innovative nonpharmacological strategies in dementia care. However, limited studies have adopted quantitative, platform-level comparisons from the perspective of formal caregivers, who are key stakeholders in technology implementation in dementia care settings.

Objective: This study aimed to evaluate the feasibility, usability, and overall user experience of a serious game–based interaction model delivered via a screen-equipped social robot, compared to a tablet-based version of the same model, from the perspective of formal dementia caregivers.

Methods: A cross-sectional comparative study was conducted with 120 formal dementia caregivers. Each caregiver individually interacted with both a screen-equipped social robot and a touchscreen tablet, delivering identical serious game content incorporating cognitive exercises, music therapy, and reminiscence. The robot featured multimodal interaction capabilities, including voice, gestures, movement, and facial expression display, while the tablet relied on standard touchscreen functions. Caregivers evaluated both platforms using the User Experience Questionnaire (UEQ), System Usability Scale (SUS), and a customized Technology Acceptance Model (TAM). Group comparisons were performed using *t* tests, with post hoc Benjamini-Hochberg correction applied to control for multiple comparisons.

Results: Caregivers generally favored the social robot over the tablet. The robot received higher total UEQ scores (mean 1.29, SD 1.14, vs mean 0.99, SD 1.08; *P*=.004), particularly in enjoyment (*P*=.002), friendliness (*P*=.006), clarity (*P*=.002), organization (*P*=.02), interest (*P*=.01), and innovation (*P*=.002). In the SUS, caregivers rated the robot higher for quick learning (mean 2.71, SD 0.79 vs mean 2.44, SD 0.81; *P*=.002), while overall SUS scores were comparable. TAM results indicated higher total scores for the robot (mean 4.03, SD 0.47 vs mean 3.67, SD 0.58; *P*=.002), with stronger ratings in perceived usefulness (*P*=.002), ease of use (*P*=.002), attitudes (*P*=.002), and behavioral intentions (*P*=.002). All *P* values are from 2-tailed *t* tests and were adjusted using the Benjamini–Hochberg procedure.

Conclusions: The social robot used in this study was perceived by formal dementia caregivers as providing a more favorable user experience and eliciting a stronger intention to use compared to a tablet-based platform. These findings support the feasibility of social robots as a platform for delivering technology-supported activities in dementia care and provide a foundation for future research on their implementation and outcomes in dementia care.

(JMIR Serious Games 2025;13:e76209) doi:10.2196/76209

KEYWORDS

social robot; serious games; dementia; formal caregivers; user experience; usability; technology acceptance

XSL•FO
RenderX

Introduction

Nonpharmacological interventions have emerged as essential complementary treatments in dementia care, addressing the limitations and side effects associated with pharmacological options [1-3]. Traditional nonpharmacological approaches, however, often encounter practical constraints such as environmental limitations, low engagement from individuals with dementia, and significant reliance on caregiver involvement, increasing caregiver burden and reducing sustainability [3,4]. Advances in smart technologies, including touchscreen devices and software applications, artificial intelligence, and sensor-enabled devices, have enabled new approaches to dementia interventions [5-7]. Among these, serious games offer structured digital activities that aim to enhance cognitive functioning and emotional well-being [8].

Serious games embedded within technological platforms, such as tablets and social robots, hold considerable potential for dementia care by offering engaging and stimulating activities that promote cognitive engagement and social interaction [9]. Social robots, in particular, have shown promise due to their capacity to provide companionship, cognitive stimulation, and emotional support, significantly benefiting individuals living with dementia [9-11]. Integrating serious games into social robots can enhance these benefits by delivering structured, engaging, and personalized interventions [12,13].

Formal caregivers play a pivotal role in adopting and implementing these technological interventions. However, caregivers frequently face significant challenges, including high workload, stress, burnout, and limited resources for meaningful engagement activities [14]. Social robots incorporating serious games can mitigate these challenges by reducing caregiver burden, facilitating engaging interactions, and promoting more efficient and effective caregiving practices. Formal caregivers' attitudes, usability perceptions, and acceptance levels toward these technologies are therefore critical to successfully integrating social robots with serious games into dementia care settings [14,15].

Despite growing interest in technology-enhanced dementia interventions, research directly comparing caregiver experiences using social robots versus traditional tablets for delivering serious games remains limited. Existing studies often examine these technologies independently rather than comparatively, and many are qualitative in nature, resulting in a significant research gap regarding the relative benefits, usability, and acceptance of social robots compared to more conventional technological devices [16,17]. Moreover, limited studies have adopted quantitative, platform-level comparisons from the perspective of formal caregivers, who are key stakeholders in technology implementation in dementia care settings [18]. This highlights the need for structured evaluations to inform real-world decision-making.

To address this gap, the current study aimed to evaluate the feasibility, usability, and overall user experience of a serious game–based interaction model delivered via a social robot. Specifically, we compared caregivers' experiences using a screen-equipped social robot with those using a tablet-based version of the same model in dementia care settings. The findings are anticipated to provide practical insights for implementing serious games through different technological platforms and inform future developments in technology-enhanced dementia care.

Methods

Study Design and Participants

This study used a cross-sectional comparative design. A total of 120 formal caregivers for individuals living with dementia were recruited from various older adult care institutions. Inclusion criteria were (1) age ≥20 years, (2) currently working as a formal caregiver for persons with dementia, (3) ability to communicate in Mandarin or Taiwanese, and (4) willingness to participate voluntarily with informed consent. Caregivers who provided care for individuals with dementia fewer than 2 days per week were excluded.

Serious Game Delivery Platforms and Interaction Design

A screen-equipped social robot with a child-like appearance and multimodal interaction capabilities was selected as the primary platform (Figure 1). The robot is approximately 32 cm tall, weighs 2.5 kg, and includes a screen capable of displaying facial expressions and multimedia content. It features 5 LED lights, 12 degrees of freedom in joint movement, and 5 tactile sensing areas on its head, cheeks, and hands. It is equipped with wireless connectivity, voice recognition, facial and image detection, and an infrared sensor for fall prevention. Mobility is enabled through integrated wheels for dynamic interaction.

The adaptive interaction model embedded in the robot was designed to deliver serious game–based content using established nonpharmacological strategies, including music therapy, reminiscence, and physical activity. While the content was developed primarily for people with dementia, it was intended to be used with caregiver support, especially during initial exposure or setup. The robot was programmed to sing, dance, and perform nostalgic songs to encourage engagement. A suite of cognitive training games was developed to address cognitive domains, including memory, language, spatial awareness, mathematical ability, judgment, abstract thinking, and attention. Each game included 3 difficulty levels and supported both single- and dual-player modes. During the caregiver sessions, both single- and dual-player modes were used. The dual-player mode was available only on the social robot, which featured tactile sensors on its arms and cheeks to detect which participant touched first and thereby determine who could respond to the question.

XSL•FO
RenderX

Figure 1. The social robot used in this study to deliver serious games for dementia care.



The robot also supported voice command interaction, allowing users to initiate actions verbally and improving interaction efficiency and engagement. Caregivers explored different difficulty levels based on their interest and available time and were asked to experience all 3 levels in at least one game to familiarize themselves with the difficulty structure. In addition, after each game or question, the robot provided immediate encouraging feedback through facial expressions, movements, and speech to reinforce participation. The same content was implemented on a touchscreen tablet to allow direct comparison of user experience across platforms. Each caregiver interacted with both devices individually for approximately 15 - 20 minutes per platform. A summary of the intervention structure and platform-specific features is provided in Table 1.

XSL•FO
RenderX

Table . Platform-specific features of the social robot and tablet.

Feature	Social robot	Tablet
Primary interaction mode	Touchscreen, voice, gesture, facial expressions	Touchscreen only
Screen size	7-inch LCD touchscreen	10.5-inch touchscreen
Movement or mobility	Yes; uses wheels and joint articulation for movement	None
Additional features	Expressive gestures, motivational verbal feedback, facial recognition	Standard tablet audio/visual feedback
Platform system	Android-based (version 9)	Android-based (version 13)
Content delivered	Cognitive games, music therapy, reminiscence activities	Same as the social robot
Session length per platform	Approximately 15 - 20 minutes per platform	Same as the social robot
Game difficulty levels	3 levels, manual selection	Same as the social robot
Player modes	Single-player, dual-player	Single-player
Voice command interaction	Yes; voice commands to initiate actions or respond during games	None
Emotional feedback	Yes; immediate encouraging feedback using expressions, movement, and speech after each action	Limited visual or auditory responses on screen only.

While the serious game content was similar across both platforms, the robot offered additional interaction modalities, including facial expressions, voice commands, physical gestures, and emotional feedback, and also supported a dual-player mode. In contrast, the tablet relied on standard touchscreen input and basic audio-visual feedback. Both the robot and tablet operated on Android-based systems. The software used on both platforms was developed in-house for this study and is not publicly available; however, a demonstration video can be provided upon request for academic or research purposes.

Data Collection

During this program, only formal caregivers participated; people with dementia were not involved in this evaluation. Caregivers were educated about the social robot and the interaction model and were then given individual, hands-on time to interact with the robot. Afterward, caregivers were educated about the tablet-based version of the same model. We acknowledge that this order may introduce potential order effects, which are discussed further in the limitations section. Caregivers completed all questionnaires based solely on their own user experience with each device. To compare the user experience of the social robot with that of a tablet used in dementia care, the study used 3 widely accepted questionnaires, including the User Experience Questionnaire (UEQ), the System Usability Scale (SUS), and a customized Technology Acceptance Model (TAM) in addition to demographic data such as age, gender, educational level, and duration and frequency of dementia care (Multimedia Appendix 1).

The UEQ, a 12-item tool with a 7-point Likert scale, measures users' feelings and attitudes toward a product [19]. The SUS, a 10-item questionnaire developed by John Brooke in 1986, assesses usability with responses ranging from "strongly disagree" to "strongly agree," and converts scores to a 0 - 100 scale to gauge system user-friendliness [20]. The TAM, designed by Fred Davis in 1989, evaluates perceived usefulness and ease of use, using a Likert scale to understand users' beliefs, attitudes, and intentions toward technology adoption [21]. A customized version of the TAM, which includes aspects specific to dementia care to make it relevant to the context of this study, was approved by 3 experts. In this study, TAM items were designed to reflect caregivers' subjective perceptions, including beliefs about usefulness in dementia care. While not objective clinical assessments, these perceptions are important factors influencing caregivers' willingness to adopt and use the system. These questionnaires provided a comprehensive assessment of formal caregivers' experiences with both the social robot and the tablet used in dementia care.

Data Analysis

Descriptive statistics (mean, SD, frequency, and percentage) were used to summarize demographic and questionnaire data. Group comparisons between the social robot and tablet experiences were conducted using *t* tests for continuous variables. To control for multiple comparisons across UEQ, SUS, and TAM subscale scores, a post hoc Benjamini-Hochberg correction was applied to the resulting *P* values, controlling the false discovery rate at 5% [22]. All analyses were performed using Stata Statistical Software (Release 28; StataCorp LLC) [23], with a significance level set at $P<.05$ and statistical power of 80%.

Ethical Considerations

This study was approved by the Taipei Medical University Joint Institutional Review Board (approval N202211039) and conducted in accordance with the Declaration of Helsinki and applicable institutional and national guidelines. Written informed consent was obtained from all participants before data collection; participation was voluntary and participants could withdraw at any time without penalty. All collected data were deidentified before analysis and stored on secure, access-controlled servers

XSL•FO
RenderX

at Taipei Medical University, with access limited to the study team; hard-copy documents were kept in locked cabinets within secure offices. Data will be retained for 7 years per institutional policy and destroyed thereafter. Participants did not receive compensation for their participation. No identifiable images or personal information are included in this manuscript. No data from care recipients were collected.

Results

Table 2 presents the demographic and caregiving characteristics of the 120 formal caregivers. The majority were female (103/120, 85.8%) with a mean age of 45.1 (SD 9.3) years. Most participants were care attendants (77/120, 64.2%), followed by nurses (27/120, 22.5%) and other roles (16/120, 13.3%). Educational attainment was mostly at the undergraduate level (85/120, 70.8%), with 11 (9.2%) participants holding postgraduate degrees. Nearly half (53/120, 44.2%) reported having a family member or friend with dementia. In terms of caregiving experience, 50 (41.7%) participants had between 3-<10 years, with an average of 4.3 (SD 1.32) caregiving days per week. The mean perceived care burden score was 5.10 (SD 1.93).

Table . Demographic data of formal dementia caregivers.

Characteristic	Values
Age (years), mean (SD)	45.1 (12.2)
Sex, n/N (%)	
Male	17/120 (14.2)
Female	103/120 (85.8)
Job position, n/N (%)	
Nurse	27/120 (22.5)
Care attendant	77/120 (64.2)
Other	16/120 (13.3)
Educational level, n/N (%)	
High school and below	24/120 (20.0)
Undergraduate	85/120 (70.8)
Postgraduate	11/120 (9.2)
Any family/friends with dementia, n/N (%)	
Yes	53/120 (44.2)
No	67/120 (55.8)
Years of dementia caregiving, n/N (%)	
Less than 1 year	26/120 (21.7)
1 to less than 3 years	32/120 (26.7)
3 to less than 10 years	50/120 (41.7)
10 years or more	12/120 (10.0)
Weekly care days for dementia (days), mean (SD)	4.30 (1.32)
Care pressure (points), mean (SD)	5.10 (1.93)

As shown in Table 3, caregivers reported significantly higher user experience ratings for the social robot than the tablet across several domains. In the “attractiveness” domain, the robot outperformed the tablet in enjoyment (mean 2, SD 1.43 vs mean 1.43, SD 1.53; *P*=.002) and friendliness (mean 1.43, SD 1.53 vs mean 1.08, SD 1.47; *P*=.006), resulting in a higher total attractiveness score (mean 1.63, SD 1.28 vs mean 1.21, SD 1.78; *P*=.002). In the “perspicuity” domain, the robot was rated as more understandable (mean 1.66, SD 1.51 vs mean 1.28, SD 1.43; *P*=.01) and clearer (mean 1.01, SD 1.61 vs mean 0.64, SD 1.51; *P*=.002), with a higher average perspicuity score (mean 1.06, SD 1.20 vs mean 0.62, SD 1.32; *P*=.002).

XSL•FO
RenderX

Table . Comparison of User Experience Questionnaire (UEQ) scores between the social robot and tablet among formal dementia caregivers (N=120).

Item	Social robot, mean (SD)	Tablet, mean (SD)	Raw *P* value	BH-adjusted *P* value[a]
Attractiveness				
Enjoyable	2 (1.43)	1.43 (1.53)	<.001	.002
Good	1.42 (1.45)	1.00 (1.56)	.07	.10
Pleasing	0.61 (1.90)	0.68 (1.70)	<.001	.002
Pleasant	0.88 (1.68)	0.67 (1.62)	.04	.05
Attractive	1.15 (1.75)	1.02 (1.61)	.001	.002
Friendly	1.43 (1.53)	1.08 (1.47)	.003	.006
Average score	1.63 (1.28)	1.21 (1.78)	<.001	.002
Perspicuity				
Understandable	1.66 (1.51)	1.28 (1.43)	.007	.01
Easy to learn	1.36 (1.40)	1.22 (1.41)	.22	.27
Easy	0.42 (1.49)	0.86 (1.50)	.008	.01
Clear	1.01 (1.61)	0.64 (1.51)	<.001	.002
Average score	1.06 (1.20)	0.62 (1.32)	.001	.002
Efficiency				
Fast	1.42 (1.53)	1.18 (1.41)	.003	.006
Efficient	1.48 (1.64)	1.25 (1.49)	.56	.63
Practical	0.90 (1.51)	0.44 (1.75)	.95	.98
Organized	1.69 (1.61)	1.19 (1.61)	.02	.02
Average score	0.96 (1.14)	0.97 (1.16)	.99	>.99
Dependability				
Predictable	1.69 (1.58)	0.88 (1.52)	.26	.32
Supportive	1.39 (1.62)	1.09 (1.37)	.06	.09
Secure	1.13 (1.53)	1.08 (1.48)	.75	.81
Meets expectations	1.42 (1.63)	0.93 (1.54)	.19	.24
Average score	1.27 (1.20)	1.12 (1.22)	.15	.20
Stimulation				
Valuable	1.18 (1.49)	0.99 (1.43)	.39	.46
Exciting	1.18 (1.54)	1.09 (1.44)	.01	.02
Interesting	1.05 (1.52)	0.38 (1.78)	.008	.01
Motivating	1.19 (1.57)	1.18 (1.50)	.003	.006
Average score	1.41 (1.32)	1.08 (1.25)	.005	.009
Novelty				
Creative	1.08 (1.46)	0.73 (1.57)	.72	.79
Inventive	1.61 (1.64)	1.09 (1.52)	.02	.03
Leading edge	1.64 (1.52)	1.18 (1.44)	<.001	.002
Innovative	1.61 (1.56)	1.04 (1.46)	<.001	.002
Average score	1.23 (1.34)	0.81 (1.19)	<.001	.002
Average for all items	1.29 (1.14)	0.99 (1.08)	.002	.004

[a]*P* value adjusted using the Benjamini-Hochberg correction for multiple comparisons.

In terms of “efficiency,” caregivers perceived the robot as faster (mean 1.42, SD 1.53 vs mean 1.18, SD 1.41; *P*=.006) and more organized (mean 1.69, SD 1.61 vs mean 1.19, SD 1.61; *P*=.02). Caregivers also rated the robot higher in “stimulation,”

XSL•FO
RenderX

particularly for excitement (mean 1.18, SD 1.54 vs mean 1.09, SD 1.44; *P*=.02) and interest (mean 1.05, SD 1.52 vs mean 0.38, SD 1.78; *P*=.01). In the "novelty" domain, the robot was seen as more inventive (mean 1.61, SD 1.64 vs mean 1.09, SD 1.52; *P*=.03) and cutting-edge (mean 1.64, SD 1.52 vs mean 1.18, SD 1.44; *P*=.002). Overall, the average UEQ score favored the robot (mean 1.29, SD 1.14 vs mean 0.99, SD 1.08; *P*=.004). These differences remained statistically significant after Benjamini-Hochberg correction for multiple comparisons.

As shown in Table 4, caregivers rated the robot significantly higher than the tablet for quick learning (mean 2.71, SD 0.79 vs mean 2.44, SD 0.81; *P*=.002), with statistical significance maintained after Benjamini-Hochberg correction. The total SUS scores were comparable between platforms.

Table . Comparison of System Usability Scale scores between the social robot and tablet among formal dementia caregivers (N=120).

Item	Social robot, mean (SD)	Tablet, mean (SD)	Raw *P* value	BH-adjusted *P* value[a]
Frequent use likelihood	2.77 (0.83)	2.84 (0.85)	.44	.06
Perceived simplicity	2.22 (0.96)	2.17 (1.01)	.56	.63
Ease of use	2.83 (0.84)	2.74 (0.86)	.39	.46
Independence	2.08 (1.22)	2.06 (1.06)	.80	.86
Function integration	2.69 (0.81)	2.71 (0.76)	.86	.90
Consistency	2.21 (1.01)	2.09 (0.90)	.12	.16
Quick learning	2.71 (0.79)	2.44 (0.81)	.001	.002
Smooth operation	2.50 (0.96)	2.33 (1.01)	.04	.06
Confidence in use	2.90 (0.88)	2.90 (0.77)	>.99	>.99
Minimal learning requirement	2.35 (1.07)	2.22 (1.08)	.10	.13
Total	63.13 (14.01)	61.23 (14.70)	.12	.16

[a]*P* value adjusted using the Benjamini-Hochberg correction for multiple comparisons.

As summarized in Table 5, caregivers perceived the robot as significantly more useful. It was rated higher for its ability to decelerate cognitive decline (mean 4.15, SD 0.60 vs mean 3.64, SD 0.87; *P*=.002), enhance caregiving effectiveness (mean 3.86, SD 0.70 vs mean 3.49, SD 0.78; *P*=.002), and for overall perceived usefulness (mean 3.99, SD 0.53 vs mean 3.53, SD 0.69; *P*=.002). The robot was also rated higher in perceived ease of use, particularly for interface clarity (mean 4.10, SD 0.60 vs mean 3.84, SD 0.75; *P*=.002) and general ease of use (mean 3.86, SD 0.79 vs mean 3.57, SD 0.81; *P*=.002), resulting in a higher average ease-of-use score (mean 3.95, SD 0.55 vs mean 3.70, SD 0.56; *P*=.002).

XSL•FO
RenderX

Table . Comparison of caregiver-reported scores on the Technology Acceptance Model between the social robot and tablet (N=120).

Item	Social robot, mean (SD)	Tablet, mean (SD)	Raw *P* value	BH-adjusted *P* value[a]
Perceived usefulness				
Decelerating cognitive decline in people with dementia	4.15 (0.60)	3.64 (0.87)	<.001	.002
Alleviating psychological symptoms for people with dementia	4.03 (0.73)	3.57 (0.81)	<.001	.002
Enhancing social willingness for people with dementia	3.99 (0.76)	3.51 (0.92)	<.001	.002
Enhancing caregiving effectiveness	3.86 (0.70)	3.49 (0.78)	<.001	.002
Improving caregiving performance	3.90 (0.74)	3.41 (0.89)	<.001	.002
Reducing caregiving stress	3.92 (0.78)	3.49 (0.90)	<.001	.002
Improving caregiving simplification	4.06 (0.69)	3.63 (0.83)	<.001	.002
Average score	3.99 (0.53)	3.53 (0.69)	<.001	.002
Perceived ease-of-use				
Interface clarity	4.10 (0.60)	3.84 (0.75)	.001	.002
Sound clarity	4.15 (0.62)	3.88 (0.71)	<.001	.002
Understandability	3.84 (0.83)	3.51 (0.85)	<.001	.002
Overall quality rating	4 (0.69)	3.78 (0.74)	.002	.004
Ease of use	3.86 (0.79)	3.57 (0.81)	.001	.002
Self-resolution of operational issues	3.73 (0.84)	3.58 (0.81)	.06	.08
Ease of becoming familiar and proficient	3.95 (0.72)	3.72 (0.77)	.002	.004
Average score	3.95 (0.55)	3.70 (0.56)	<.001	.002
Attitudes				
User enjoyment	4.29 (0.65)	3.89 (0.77)	<.001	.002
Overall positive view	4.30 (0.67)	3.88 (0.86)	<.001	.002
Average score	4.30 (0.57)	3.89 (0.75)	<.001	.002
Intentions				
Intention to reuse	4.20 (0.71)	3.91 (0.84)	.001	.002
Intention to maintain usage	3.79 (0.87)	3.46 (0.93)	<.001	.002
Intention to recommend to others	4.19 (0.68)	3.82 (0.91)	<.001	.002
Intention to frequently use	4.25 (0.63)	3.83 (0.84)	<.001	.002
Average score	4.11 (0.60)	3.75 (0.74)	<.001	.002
Average for all items	4.03 (0.47)	3.67 (0.58)	<.001	.002

[a]*P* value adjusted using the Benjamini-Hochberg correction for multiple comparisons.

Attitudinally, caregivers expressed more enjoyment (mean 4.29, SD 0.65 vs mean 3.89, SD 0.77; *P*=.002) and a more favorable overall impression (mean 4.30, SD 0.67 vs mean 3.88, SD 0.86; *P*=.002) of the robot. Intentions to reuse (mean 4.20, SD 0.71 vs mean 3.91, SD 0.84; *P*=.002) and recommend (mean 4.19, SD 0.68 vs mean 3.82, SD 0.91; *P*=.002) were also significantly higher. Statistical significance for these differences was confirmed after Benjamini-Hochberg correction for multiple comparisons.

XSL•FO
RenderX

Discussion

Principal Findings

This study evaluated and compared the acceptance and user experience of a screen-equipped social robot and tablet for serious game delivery in dementia care from the perspectives of formal caregivers. While the serious game content was similar across both platforms, the robot incorporated multimodal interaction capabilities, including voice, gestures, movement, and facial expression display, while the tablet relied on standard touchscreen functions. Our findings show that formal caregivers generally favored the robot over the tablet, rating it significantly higher in attractiveness, clarity, stimulation, novelty, quick learning, and perceived usefulness. These findings suggest that social robots may serve as a feasible and engaging option for delivering technology-supported activities in dementia care settings.

Comparison to Previous Work

Recent studies have shown growing acceptance of social robots among older adults, including those with dementia [9,24-26]. In addition to enhancing the emotional well-being of people with dementia, social robots have also been suggested to help reduce caregivers' burden by supporting engagement and interaction in care settings [27]. While the perspectives of people with dementia are central, caregivers' perceptions are also important, as they are key stakeholders in selecting and supporting technologies in dementia care [18]. However, limited studies, particularly quantitative ones, have examined how formal caregivers perceive such tools. Direct comparisons between platforms like social robots and tablets are also rare. By examining caregiver experiences across both platforms, this study offers practical insights for delivering technology-supported activities in dementia care. The following sections discuss our findings in relation to previous work using the UEQ, SUS, and TAM frameworks.

In this study, the social robot outperformed the tablet in most dimensions of the UEQ, particularly in attractiveness, stimulation, and novelty. These findings echo previous research showing that social robots can increase user engagement and motivation among individuals with dementia and their caregivers [14,28,29]. The robot's physical embodiment, expressiveness, and personalized interaction may contribute to this superior experience, highlighting the importance of multisensory design in serious game delivery for dementia care.

Although total SUS scores between the robot and tablet were similar, the robot was rated significantly higher in "quick learning." This suggests that while both platforms were usable, the robot provided a more intuitive and seamless experience for caregivers. Such advantages are especially relevant in dementia care, where caregivers benefit from technologies that are easy to learn and integrate into daily routines. These findings are consistent with prior studies reporting generally positive usability ratings for social robots, particularly in contexts requiring minimal training and smooth task execution [16,18].

The significantly higher scores for the social robot across almost all domains of TAM further confirm its perceived utility. According to the Theory of Planned Behavior, perceived behavioral control and positive attitudes are critical to forming behavioral intentions, which influence actual technology adoption [30]. The strong behavioral intentions to reuse and recommend the social robot point toward promising long-term integration into real-world care environments [14]. While caregivers are not positioned to evaluate clinical outcomes, their perceptions, such as whether the system may help slow cognitive decline, are relevant indicators of motivational factors and acceptance in real-world use [31].

Limitations

This study has several limitations. First, this study evaluated the intervention from the perspective of formal caregivers only; people with dementia were not involved. While caregivers are key stakeholders in technology implementation and offer valuable insights into feasibility and supported use, this limits generalizability to actual end users. Caregivers differ from people with dementia in demographic characteristics, cognitive ability, and technology familiarity, which may affect engagement with the system [14,18,31]. Previous studies showed that people with dementia often respond positively to social robots, though individual preferences and cognitive challenges can affect their engagement [9]. Future research should include direct input from people with dementia to validate usability and inform user-centered design.

Second, this study relied solely on quantitative self-report assessments without behavioral measures or follow-up data, limiting conclusions about real-world usability and adoption. Future studies should include observational or usage data and longitudinal follow-up to assess sustained engagement and outcomes. Third, all participants interacted with the social robot before the tablet, which may have introduced potential order effects. Although this sequence ensured standardized exposure, future studies should consider counterbalanced designs to reduce potential bias. Finally, this study involved multiple statistical comparisons, which may increase the risk of type I error. To address this, a post hoc Benjamini-Hochberg correction was applied, and key results remained significant after adjustment [22].

Future Directions

While caregivers provide important insights into technology implementation, future research should directly involve people with dementia to assess usability, engagement, and inform user-centered improvements. Studies should incorporate behavioral observations, randomized or counterbalanced designs, and longer-term follow-up to evaluate sustained use and outcomes. Given the social robot's multimodal and emotionally expressive features, future research should examine how specific elements of robot interaction influence engagement and effects. The refined interaction model may support the design of adaptive tools tailored to diverse cognitive and care needs. Finally, studies should compare platforms such as social robots and tablets in terms of their impact on health outcomes in people with dementia and caregiver burden.

XSL•FO
RenderX

Conclusions

Serious games offer a promising approach to enhancing nonpharmacological dementia care, and this study demonstrates that, from the perspective of formal caregivers, delivering serious games through a screen-equipped social robot offers a more engaging, user-friendly, and acceptable experience compared with a tablet-based version. The robot's multimodal interaction capabilities contributed to higher ratings in user experience, perceived usefulness, and behavioral intentions to use. While these results do not allow conclusions about long-term effectiveness or clinical outcomes, they support the perceived feasibility and user acceptance of social robot–mediated serious games from the perspective of dementia caregivers, providing a foundation for future research in this field.

Acknowledgments

We sincerely thank the participating caregivers for their time and valuable contributions to this study. This work was supported by the National Science and Technology Council of Taiwan (grant number NSTC112-2314-B-038 - 024; NSTC114-2314-B-038-102) and Taipei Medical University (grant number TMU111-AE1-B20). No generative AI was used to generate content in the manuscript.

Data Availability

The data will be made available from the corresponding author upon reasonable request.

Authors' Contributions

DB contributed to conceptualization, methodology, formal analysis, resources, funding acquisition, project administration, supervision, and writing – original draft. KTCK contributed to methodology, resources, and writing – review and editing. PYC, IHC, and YLH contributed to writing – review and editing. GHL contributed to methodology, data curation, writing – original draft, and writing – review and editing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey instruments used in the study, including the User Experience Questionnaire, System Usability Scale, and Technology Acceptance Model.
[DOCX File, 43 KB - games_v13i1e76209_app1.docx]

References

1. Dementia care practice recommendations: evidence-based guidelines for quality care. Alzheimer's Association. 2023 Jul 18. URL: https://www.alz.org/professionals/professional-providers/dementia_care_practice_recommendations [accessed 2025-09-13]
2. Hatch S, Seitz DP, Bruneau MA, et al. The Canadian coalition for seniors' mental health Canadian Clinical Practice Guidelines for assessing and managing Behavioural and Psychological Symptoms of Dementia (BPSD). Can Geriatr J 2025 Mar;28(1):91-102. [doi: 10.5770/cgj.28.820] [Medline: 40051590]
3. Meyer C, O'Keefe F. Non-pharmacological interventions for people with dementia: a review of reviews. Dementia (London) 2020 Aug;19(6):1927-1954. [doi: 10.1177/1471301218813234] [Medline: 30526036]
4. Wang G, Albayrak A, van der Cammen TJM. A systematic review of non-pharmacological interventions for BPSD in nursing home residents with dementia: from a perspective of ergonomics. Int Psychogeriatr 2019 Aug;31(8):1137-1149. [doi: 10.1017/S1041610218001679] [Medline: 30334500]
5. Abdulazeem H, Borges do Nascimento IJ, Weerasekara I, et al. Use of digital health technologies for dementia care: bibliometric analysis and report. JMIR Ment Health 2025 Feb 10;12:e64445. [doi: 10.2196/64445] [Medline: 39928936]
6. Chien SY, Zaslavsky O, Berridge C. Technology usability for people living with dementia: concept analysis. JMIR Aging 2024 Jul 3;7:e51987. [doi: 10.2196/51987] [Medline: 38959053]
7. Pappadà A, Chattat R, Chirico I, Valente M, Ottoboni G. Assistive technologies in dementia care: an updated analysis of the literature. Front Psychol 2021;12:644587. [doi: 10.3389/fpsyg.2021.644587] [Medline: 33841281]
8. Mezrar S, Bendella F. A systematic review of serious games relating to cognitive impairment and dementia. J Digit Inf Manag 2022 Dec;20(1):1. [doi: 10.6025/jdim/2022/20/1/01-09]
9. Yu C, Sommerlad A, Sakure L, Livingston G. Socially assistive robots for people with dementia: Systematic review and meta-analysis of feasibility, acceptability and the effect on cognition, neuropsychiatric symptoms and quality of life. Ageing Res Rev 2022 Jun;78:101633. [doi: 10.1016/j.arr.2022.101633] [Medline: 35462001]

XSL•FO
RenderX

10. Chu MT, Khosla R, Khaksar SMS, Nguyen K. Service innovation through social robot engagement to improve dementia care quality. Assist Technol 2017;29(1):8-18. [doi: 10.1080/10400435.2016.1171807] [Medline: 27064692]
11. Moyle W, Jones CJ, Murfield JE, et al. Use of a robotic seal as a therapeutic tool to improve dementia symptoms: a cluster-randomized controlled trial. J Am Med Dir Assoc 2017 Sep 1;18(9):766-773. [doi: 10.1016/j.jamda.2017.03.018] [Medline: 28780395]
12. Fan W, Zhao R, Liu X, Ge L. Intelligent robot interventions for people with dementia: systematic review and meta-analysis of randomized controlled trials. J Med Internet Res 2025 Mar 10;27:e59892. [doi: 10.2196/59892] [Medline: 40063933]
13. Yuan F, Boltz M, Bilal D, et al. Cognitive exercise for persons with Alzheimer's disease and related dementia using a social robot. IEEE Trans Robot 2023 Aug;39(4):3332-3346. [doi: 10.1109/tro.2023.3272846] [Medline: 38495392]
14. Ghafurian M, Hoey J, Dautenhahn K. Social robots for the care of persons with dementia. J Hum-Robot Interact 2021 Dec 31;10(4):1-31. [doi: 10.1145/3469653]
15. Hirt J, Ballhausen N, Hering A, Kliegel M, Beer T, Meyer G. Social robot interventions for people with dementia: a systematic review on effects and quality of reporting. J Alzheimers Dis 2021;79(2):773-792. [doi: 10.3233/JAD-200347] [Medline: 33361589]
16. Shi W, Zhou W, Wang R, Shen H, Xu N, Wang J. Exploring the perceptions and experiences of caregivers with the application of socially assistive robots in dementia care: a systematic review of qualitative studies. Int J Nurs Stud 2025 Jul;167:105084. [doi: 10.1016/j.ijnurstu.2025.105084] [Medline: 40262375]
17. Liao YJ, Jao YL, Boltz M, et al. Use of a humanoid robot in supporting dementia care: a qualitative analysis. SAGE Open Nurs 2023;9:23779608231179528. [doi: 10.1177/23779608231179528] [Medline: 37324571]
18. Harris-Gersten ML, Davagnino JM, Alcorn ER, Hastings SN. Usability and acceptability of social robot pets among community-dwelling veterans living with dementia and their caregivers. Am J Alzheimers Dis Other Demen 2023;38:15333175231200973. [doi: 10.1177/15333175231200973] [Medline: 37688378]
19. Schrepp M, Hinderks A, Thomaschewski J. Construction of a benchmark for the User Experience Questionnaire (UEQ). IJIMAI 2017;4(4):40. [doi: 10.9781/ijimai.2017.445]
20. Klug B. An overview of the system usability scale in library website and system usability testing. Weave 2017;1(6). [doi: 10.3998/weave.12535642.0001.602]
21. Silva P. Davis' technology acceptance model (TAM). In: Khosrow-Pour M, editor. Encyclopedia of Information Science and Technology, 3rd edition: IGI Global; 1989:205-219. [doi: 10.4018/978-1-4666-8156-9.ch013]
22. Thissen D, Steinberg L, Kuang D. Quick and easy implementation of the Benjamini-Hochberg procedure for controlling the false positive rate in multiple comparisons. J Educ Behav Stat 2002 Mar;27(1):77-83. [doi: 10.3102/10769986027001077]
23. StataCorp LP. Stata Statistical Software: Release 11: StataCorp LP; 2009. [doi: 10.1177/1536867X1501500319]
24. Otaka E, Osawa A, Kato K, et al. Positive emotional responses to socially assistive robots in people with dementia: pilot study. JMIR Aging 2024 Apr 11;7:e52443. [doi: 10.2196/52443] [Medline: 38623717]
25. Rettinger L, Fürst A, Kupka-Klepsch E, Mühlhauser K, Haslinger-Baumann E, Werner F. Observing the interaction between a socially-assistive robot and residents in a nursing home. Int J of Soc Robotics 2024 Feb;16(2):403-413. [doi: 10.1007/s12369-023-01088-9]
26. Pu L, Moyle W, Jones C, Todorovic M. The effectiveness of social robots for older adults: a systematic review and meta-analysis of randomized controlled studies. Gerontologist 2019 Jan 9;59(1):e37-e51. [doi: 10.1093/geront/gny046] [Medline: 29897445]
27. Hsieh CJ, Li PS, Wang CH, Lin SL, Hsu TC, Tsai CMT. Socially assistive robots for people living with dementia in long-term facilities: a systematic review and meta-analysis of randomized controlled trials. Gerontology 2023;69(8):1027-1042. [doi: 10.1159/000529849] [Medline: 36871553]
28. Bevilacqua R, Maranesi E, Felici E, et al. Social robotics to support older people with dementia: a study protocol with Paro seal robot in an Italian Alzheimer's day center. Front Public Health 2023;11:1141460. [doi: 10.3389/fpubh.2023.1141460] [Medline: 37415712]
29. Whelan S, Murphy K, Barrett E, Krusche C, Santorelli A, Casey D. Factors affecting the acceptability of social robots by older adults including people with dementia or cognitive impairment: a literature review. Int J of Soc Robotics 2018 Nov;10(5):643-668. [doi: 10.1007/s12369-018-0471-x]
30. Ajzen I. The theory of planned behaviour: reactions and reflections. Psychol Health 2011 Sep;26(9):1113-1127. [doi: 10.1080/08870446.2011.613995] [Medline: 21929476]
31. Chatzoglou PD, Lazaraki V, Apostolidis SD, Gasteratos AC. Factors affecting acceptance of social robots among prospective users. Int J of Soc Robotics 2024 Jun;16(6):1361-1380. [doi: 10.1007/s12369-023-01024-x]

Abbreviations

UEQ: User Experience Questionnaire
SUS: System Usability Scale
TAM: Technology Acceptance Model

XSL•FO
RenderX

Edited by A Coristine; submitted 18.04.25; peer-reviewed by JH Song, T Ling; revised version received 27.08.25; accepted 27.08.25; published 07.10.25.

Please cite as:
Bai D, Tan Cheng Kian K, Chen PY, Hsu YL, Lin GH
Comparing Caregiver Perceptions of a Social Robot and Tablet for Serious Game Delivery in Dementia Care: Cross-Sectional Comparison Study
JMIR Serious Games 2025;13:e76209
URL: https://games.jmir.org/2025/1/e76209
doi:10.2196/76209

© Dorothy Bai, Kelvin Tan Cheng Kian, Po-Yin Chen, Yeh-Liang Hsu, Gong-Hong Lin. Originally published in JMIR Serious Games (https://games.jmir.org), 7.10.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Virtual Reality for Analgesia During Intrauterine Device Insertion: Randomized Controlled Trial

Chloe Higgins[1], DipLang, BMedSci (Hons), MBBS (Hons); Claudia Zecena Morales[1], BBiomedSc, MD, CertWH; Judi Hocking[1], BA; Kate Tyson[1,2], MBBS (Hons), MPH; Cheryl Leung[1], BMedSci, MBBS; Luke Larmour[1], MBBS; Paul Leong[3,4*], MBBS, MPH&TM, CCPU, PhD; Beverley Vollenhoven[1,2*], CREI, PhD

[1]Women's and Newborn Program, Monash Health, 246 Clayton Rd, Melbourne, Australia

[2]Department of Obstetrics and Gynaecology, Monash University, Melbourne, Australia

[3]Department of Lung and Sleep, Monash Health, Melbourne, Australia

[4]Department of Respiratory Research, Monash University, Melbourne, Australia

[*]these authors contributed equally

Corresponding Author:
Chloe Higgins, DipLang, BMedSci (Hons), MBBS (Hons)
Women's and Newborn Program, Monash Health, 246 Clayton Rd, Melbourne, Australia

Abstract

Background: Intrauterine devices (IUDs) are safe and effective long-acting reversible contraceptive therapies that are also used as minimally invasive treatment for heavy menstrual bleeding, endometrial hyperplasia, and early-stage endometrial cancer. Despite many advantages, IUDs are underused predominantly due to patient discomfort. Although many techniques have been explored previously in the literature, there is currently little consensus on effective analgesic strategies. Virtual reality (VR) has demonstrated moderate benefits in acute pain management and has been explored for outpatient hysteroscopy.

Objective: This study aims to explore the effectiveness of VR in improving patient pain and anxiety during outpatient IUD insertion.

Methods: This randomized controlled trial compared the use of a VR headset to standard care during IUD insertion in the outpatient clinic setting. VR content was delivered via smartphone and headset, providing patients with a relaxing 3D video environment. Outcomes measured were patient-reported pain and anxiety, as well as satisfaction reported using a questionnaire. Secondary outcomes included clinician-reported ease of insertion and time required to complete the procedure.

Results: A total of 70 patients were recruited, with 34 randomized to the control group and 36 randomized to VR headset use. Patients with VR headsets reported a mean pain score of 5.5 (SD 3.2) during IUD insertion, which was not significantly different to 4.3 (SD 3.2) for the control group (P=.15). Mean anxiety scores during the procedure were 4 (SD 3) in the VR group, compared to 4.8 (SD 3.5) in the control group, which was also not significantly different (P=.37). Anxiety was the most significant predictor of pain, and this, in turn, significantly increased insertion time (P<.001). Among patients who responded to and benefitted from VR use, baseline anxiety was significantly lower than in those who did not (P<.001). Satisfaction with the use of VR headsets was overall high, and recommendation scores for the use of VR headsets were also high. There were no significant adverse effects experienced with the use of the intervention, with only 1 patient reporting nausea after IUD insertion.

Conclusions: The use of VR headsets did not significantly alter the pain or anxiety experienced by patients during IUD insertion; however, satisfaction and recommendation that others use VR were high, which may suggest other benefits to their use. In addition, preprocedural anxiety appears to have a significant adverse impact on pain scores and the ability of patients to benefit from the VR headsets. This is an important contribution to the previously ambiguous data regarding VR use for gynecological procedures and highlights a new avenue for improving the patient experience.

Trial Registration: Australian New Zealand Clinical Trials Registry (ANZCTR) ACTRN12622000088741; https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=383191

(JMIR Serious Games 2025;13:e72917) doi:10.2196/72917

KEYWORDS

virtual reality; intrauterine device; pain; anxiety; contraception

XSL•FO
RenderX

Introduction

Intrauterine devices (IUDs) are extremely effective long-acting reversible contraceptives that have been shown to be safe and convenient [1-4]. IUDs containing levonorgestrel are used as minimally invasive and fertility-preserving management options for gynecological conditions, including heavy menstrual bleeding, endometrial hyperplasia, and early-stage endometrial cancer [5]. Despite being the most used reversible contraceptive method globally and having substantial potential benefits, IUDs continue to be underused in many settings, including high-income countries [4,6]. Several barriers to IUD use have been hypothesized, and a major factor is the procedural experience, particularly providing accessible and acceptable insertion for patients [2,4].

Gynecological procedures, including IUD insertions, are increasingly being undertaken in the outpatient clinic setting as this is cheaper, more convenient, and avoids the risks of procedural sedation [7]. Despite being more accessible, IUD in this setting is associated with patient discomfort and anxiety, with pain scores usually ranging from 5 to 7 out of 10, which is a significant barrier to patient acceptability [2,8,9]. IUDs are technically feasible to insert for most outpatients, including nulligravida and patients with previous cesarean section [10]; however, these are common predictors of IUD insertion pain, in addition to remoteness from last pregnancy, dysmenorrhea, anticipation, and IUD type used (levonorgestrel being more painful than copper) [3,11].

IUD insertion can be undertaken in the operating theater under general anesthesia or procedural sedation. While this does alleviate patient discomfort, operating theater costs and access make this prohibitive for routine practice [12]. Compounding these reasons, the difficulty of access to operating theaters during the pandemic, and the constant requirement to keep operating theater time as training opportunities for junior doctors, the incentive to optimize the experience of outpatient insertion of IUDs is as strong as ever [12].

There is currently no consensus regarding the analgesic of choice during outpatient IUD insertion. Previously studied methods include oral or intravenous sedation, local anesthetic topically or as a cervical block, oral analgesics, and distraction techniques [7]. Common techniques such as nonsteroidal anti-inflammatory drugs, misoprostol, local anesthetic gel, and intrauterine instillation of local anesthetic have been found to be ineffective [1-3,6,13]. Paracervical local anesthetic injections have been suggested to be of some benefit during insertion; however, the injections themselves can be painful [1]. Tramadol and naproxen have shown possible benefits in small numbers in studies exclusively investigating nulliparous women; however, this was not replicated in follow-up studies [2,11,13]. Further options need to be explored to improve acceptability and patient uptake.

Virtual reality (VR) is gaining momentum as a novel nonpharmacological pain relief option, and its potential role in health care is increasingly relevant, as technological developments have improved its accessibility [14]. VR creates an immersive virtual environment for the user [7]. It can be delivered using a range of hardware, most commonly in the form of a head-mounted display with built-in headphones. It can present as an active VR, where participants interact with the content, such as in games, or passive virtual environments, such as movies. It has been suggested that active VR may have stronger analgesic properties than passive [14].

VR has demonstrated moderate benefits in acute pain management, including burns, labor, dressing changes, and periodontal procedures [15]. It is theorized that VR provides distraction from pain by multisensorial stimulation, perhaps via the gated-pain theory [7,14]. Reported side effects of VR are infrequent, mild, and self-limiting. They include nausea, vomiting, and eyestrain, collectively referred to as “cybersickness” [14]. If any of these events occur, the headset can be removed with rapid resolution of symptoms.

The possible benefits of VR in pain and periprocedural anxiety scores, combined with its minimal risks to patients, highlight the potential of this method and hence the need for further exploration of its use and cost-effectiveness in various settings [7,14,15]. Despite its initial promising results, particularly in pediatric care, it has been hypothesized that the effectiveness of VR is not necessarily generalizable to adult populations and other procedures [15]. The benefits likely vary significantly between patient populations and indications for use, such that studies in specific settings are recommended [14]. It has therefore been suggested for consideration for IUD insertion, with the hypothesis that it may provide an effective distraction technique to help with alleviating pain and anxiety, thereby improving patient access to outpatient IUD insertion [7,16,17].

A search of the literature for VR use in IUD insertion yielded only 2 research trials. One randomized controlled study performed in France with 50 patients per group found no change in perceived pain in women using VR therapy compared to those without, but rather noted the strongest predictor of pain being preprocedure anxiety [16]. The second study performed in Turkey had directly contradicting results, demonstrating an improvement in pain, anxiety, and satisfaction scores in those using a VR headset compared to the control group, with 40 patients in each group [17]. Research performed on other gynecological procedures, such as outpatient hysteroscopy, disagrees on whether pain scores are improved but shows promising improvements in anxiety and patient satisfaction associated with the use of VR [7,18,19].

The objective of this study was therefore to explore the impact of VR on patient-reported pain during IUD insertion in a tertiary gynecology hospital setting in Australia. Secondary aims included the effect of VR on patient anxiety and ease of IUD insertion, while also assessing the feasibility of introducing this as an analgesic method in the outpatient gynecological setting.

Methods

Recruitment

In this single-center randomized controlled trial, patients attending for IUD insertion in the outpatient gynecology clinic were screened for eligibility. Patients were eligible if aged 18 years and older and able to provide written informed consent. Patients were excluded if they had contraindications to IUD

XSL•FO
RenderX

insertion, a history of epilepsy or vestibular disorders contraindicating VR use, an inability to consent, or if they had taken prescription pain relief prior to the procedure (simple analgesia was permitted).

Patients were randomized on the day of IUD insertion in the outpatient clinic to either the intervention group (addition of VR to standard care) or the control group (standard care). A computerized random sequence was generated, with allocations then placed in sequentially numbered opaque envelopes by a blinded third party. These envelopes were opened only after participants were enrolled in the study by the researcher, who was not otherwise involved in the patient's care. Participants were not paid for their participation. Given the obvious nature of VR as an intervention and the subjective nature of pain and anxiety as outcomes, blinding was not possible.

VR digital content was delivered via smartphone and headset during the IUD insertion procedure. The content provided a relaxing, noninteractive 3D video of seals swimming with gentle classical music (Figure 1). The visual content was chosen by the investigators to ensure its suitability for the patient population and minimize any patient reactions to objectionable content. Content containing dialogue was considered inappropriate as it would interfere with clinical communication during this intimate procedure. Standard care included occasional use of a lignocaine spray to the cervix according to the clinician's preference, and this was therefore recorded as part of the data collection.

Several pain scales exist and can be used to quantify the pain intensity from IUD insertion, including the Visual Analog Scale (VAS), the Verbal Rating Scale, and the Numerical Rating Scale [20]. The VAS is reliable, valid, and sensitive to subtle differences and changes in pain intensity [20]. For this reason, it is the most consistently used tool [21] and has been used in prior studies of discomfort associated with IUD insertion [3,5]. VAS differences of 10 mm are considered clinically significant from previous studies [22,23].

Patient demographic information was collected by reviewing the patient's medical records and full consultation notes. In addition, patients were asked to plot their anxiety prior to the procedure, and their pain and anxiety during the procedure on a linear 10-cm VAS. Numerical VAS scores were then obtained by measuring the distance along the line to the nearest millimeter. Patients also completed a questionnaire on their satisfaction, recommendations, and side effects after the procedure. The time required for IUD insertion, clinician-reported ease of insertion on a scale of 1 to 5, and number of attempts required for successful IUD insertion (defined as the number of attempts to pass the IUD through the cervix) were also reported (Multimedia Appendix 1).

Figure 1. Still image captured from virtual reality seals swimming.



Statistical Analysis

Statistical analyses were conducted in SPSS 25 (IBM Corp). Significance was defined as P <.05. Two-tailed t-tests were performed for group-wide comparisons.

Power Calculations

For our power calculation, we assumed an equal SD of 14 mm in VAS pain scores between groups based on previous studies of VAS pain scores in women with pelvic pain during IUD insertion [24]. To achieve 80% of detecting a 10-mm difference between groups with a type 1 error rate of 0.05, we required a sample size of 31 in each group. Accounting for a compliance rate of 90%, we aimed for a sample size of 70 participants.

Ethical Considerations

The study was approved by Monash Health Ethics Committee, Clayton, Victoria, Australia (ethics approval reference number 80700 and Monash Health reference number RES-21-0000-690A). The trial was prospectively registered (Australia and New Zealand Clinical Trial Registry ACTRN12622000088741p). Written informed consent was

XSL•FO
RenderX

obtained from all participants following counseling, with time allowed for patients to review the participant information and consent form. Patients were able to opt out of the study at any stage if they chose to withdraw consent to participate. There were no payments or incentives or compensation provided that could result in pressure to participate. Data collected were deidentified and stored on secure password-protected servers, with all data converted to electronic copies and hard copies destroyed in confidential waste to protect patient privacy.

Results

Participant flow is described in Figure 2.

Of the 70 patients enrolled in the study, there were 34 in the control group and 36 in the VR intervention group. Their baseline characteristics are outlined in Table 1. Of the 34 control group participants, 31 had levonorgestrel-containing devices inserted, and of the 36 VR intervention group participants, 34 had levonorgestrel-containing devices. The remaining participants had copper IUDs.

Of note, despite randomization, there was numerical imbalance with patients in the VR group being more likely to have previously had an IUD, previously used a VR headset, and a history of chronic pain.

There was no significant difference in intraprocedural pain scores with the use of VR headsets (Table 2). Similarly, preprocedural and intraprocedural anxiety scores did not significantly differ (Table 2).

The use of VR headsets also did not appear to have a significant impact on the IUD insertion process as clinician-reported ease of insertion, time to insertion, and the number of insertion attempts did not differ (Table 3).

Satisfaction with the use of VR headsets was overall high (mean 7.3 out of 10, SD 2.99), and recommendation scores for use of VR headsets were also high (mean 8.35 out of 10, SD 2.24). Satisfaction and recommendation were highly correlated, and a high satisfaction score was weakly associated with lower intraprocedural anxiety scores (r=0.33, P=.06).

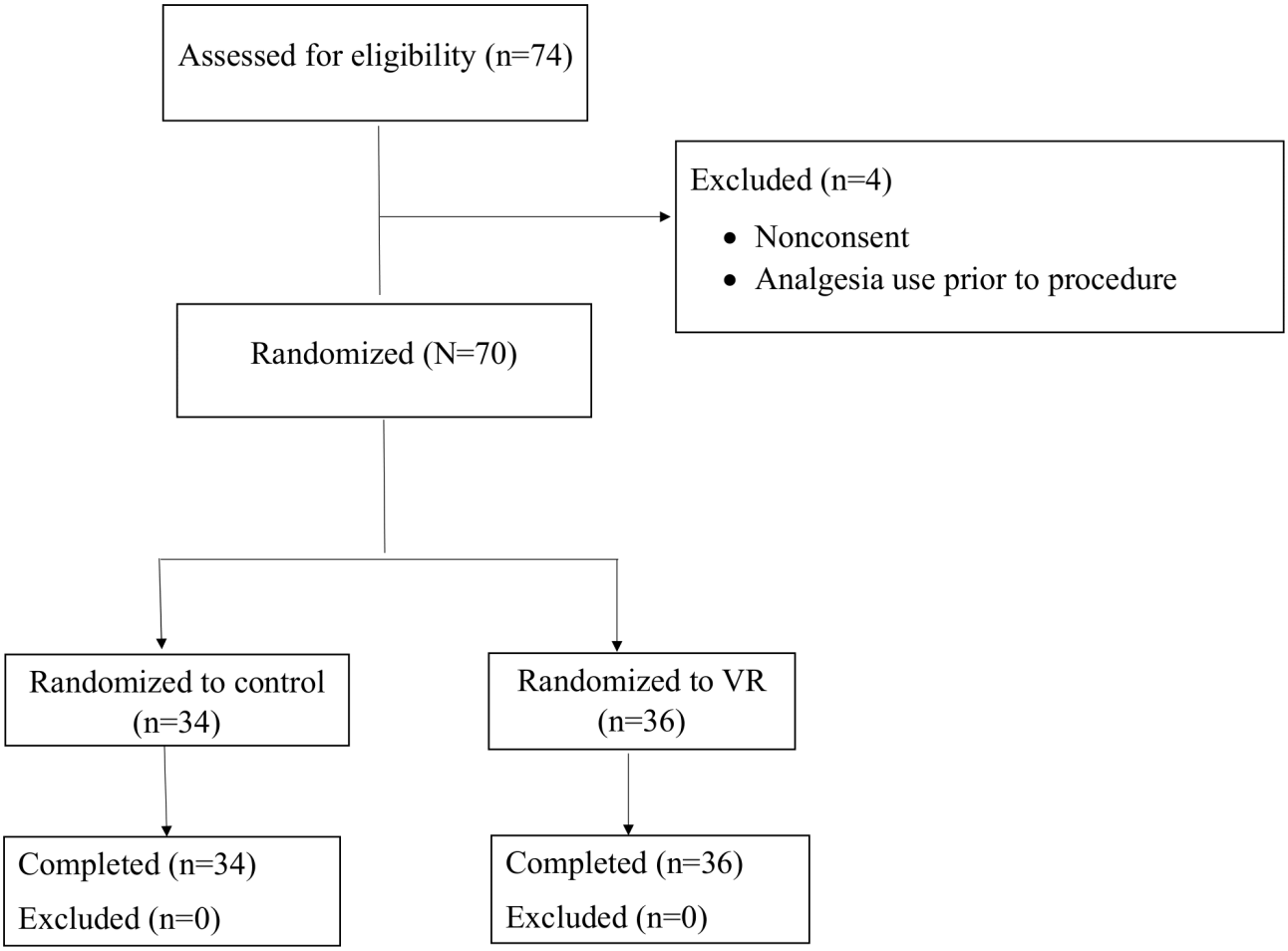
Preprocedural anxiety was significantly correlated with intraprocedural anxiety (P<.001). Overall, anxiety was the most significant predictor of pain during insertion (P=.001). Higher pain was also a significant predictor of increased procedural time (P<.001).

Subgroup analysis was performed to isolate the patients who most benefit from VR. Responders to VR were defined as patients in the VR intervention group who had pain scores of less than or equal to 5. This score was chosen due to literature findings of pain scores of 5 - 7 out of 10 for IUD insertion with placebo use [8]. Patients who were classified as responders were compared with nonresponders (those with pain scores greater than 5 in the VR intervention group). The characteristics of these 2 groups were markedly different, with significantly lower preprocedural and intraprocedural anxiety in the responders than nonresponders (P<.05). Baseline characteristics, including BMI, analgesic spray, prior IUD insertion, and procedure success, were not significantly different between the 2 subgroups.

There were no significant adverse effects experienced with the use of the intervention, with only 1 patient reporting nausea after IUD insertion.

XSL•FO
RenderX

Figure 2. CONSORT (Consolidated Standards of Reporting Trials) diagram of participant inclusion and exclusion. VR: virtual reality.



XSL•FO
RenderX

Table . Baseline population characteristics.

Characteristics	Control (n=34)	VR[b] (n=36)
Age (years), mean (SD)	34 (6)	36 (7)
BMI (kg/m^2), mean (SD)	30 (8)	31 (7)
Gravida, n	2	2
Parity, n	2	2
Previous vaginal deliveries, mean (SD)	1 (1)	1 (1)
Time since last vaginal birth in years, median (IQR)	1.5 (0.21 - 3.25)	0.58 (0.19 - 4)
Previous cervical instrumentation, n (%)	11 (32)	5 (14)
Anatomical distortions, n (%)	6 (18)	6 (17)
Prior IUD[a], n (%)	10 (29)	15 (42)
Previous VR[b] use, n (%)	3 (9)	11 (30)
Use of analgesic spray, n (%)	11 (32)	10 (28)
Indication for IUD[a], n (%)		
Contraception	32 (94)	27 (75)
Abnormal bleeding	0 (0)	5 (14)
Both	2 (6)	2 (6)
Mental health history, n (%)	5 (15)	6 (17)
Chronic pain history, n (%)	0 (0)	5 (14)

[a]IUD: intrauterine device.

[b]VR: virtual reality.

Table . Patient pain and anxiety outcomes.

Characteristics	Control (n=34)	VR[a] (n=36)	*P* value
Pain score, mean (SD)	4.3 (3.2)	5.5 (3.2)	.15
Anxiety prior to procedure, mean (SD)	4.2 (2.8)	4.6 (3.1)	.56
Anxiety during procedure, mean (SD)	4.8 (3.5)	4 (3)	.37

[a]VR: virtual reality.

Table . Effect of virtual reality use on intrauterine device insertion.

Characteristics	Control (n=34)	VR[a] (n=36)	*P* value
Number of attempts to successfully insert IUD[b], mean (SD)	1.09 (0.59)	0.94 (0.65)	.73
Clinician reported ease of insertion (1-5), mean (SD)	2.75 (1.48)	2.75 (1.63)	.41
Time for insertion (minutes:seconds), mean (SD)	12:06 (0:04)	12:06 (0:02)	.61

[a]VR: virtual reality.

[b]IUD: intrauterine device.

XSL•FO
RenderX

Discussion

Overview

Despite the many benefits of IUDs, patient discomfort is a significant barrier to accessing outpatient insertion [2,4]. There is little evidence and no clear consensus on the best methods of addressing this discomfort without this being performed under general anesthesia, which itself presents costs and risks [12].

Principal Findings

This study found no significant difference in anxiety or pain with VR headset use and identified anxiety as the most significant predictor of pain.

Interestingly, despite pain and anxiety scores not significantly differing between the 2 groups, the VR headsets may have benefits that may still be relevant for procedural acceptability. The VR intervention was effective and well-received in a subpopulation of responders, particularly those with reduced levels of pre-existing anxiety. Satisfaction and recommendation for the VR headsets were overall very high, demonstrating patient appreciation of these devices. Patients were provided with the opportunity to comment on their experience, and those who did were overwhelmingly positive. The reasons for high satisfaction and recommendation for VR use appear to be related predominantly to the distraction effect.

Patients' comments particularly highlighted that they appreciated the distraction provided by the headsets, with 1 patient reporting the distraction was helpful for anxiety, while others stated the distraction helped them to forget about the procedure. Other possible reasons for the positive patient feedback for the VR devices may include the novelty of the equipment or other psychological benefits, which, although difficult to quantify, may improve the acceptability of outpatient IUD insertion with VR use despite the lack of improvement in pain scores.

Comparison to Prior Work

As VR grows in its implementation in the health care space, several studies have recently been published on its use in outpatient gynecological procedures. The existing body of evidence relating directly to IUD insertion is scant and contradictory, with the study by Benazzouz et al directly contradicting findings from Oz et al [16,17]. Both studies were randomized trials, recruiting patients in a contraceptive clinic on the day of IUD insertion, applying VR headsets, and collecting similar data measures, including the use of the VAS for pain rating; however, the studies had important methodological differences.

The study by Benazzouz et al [16] was conducted in France in 100 patients with a mean age of roughly 24 years and found no difference in pain or anxiety with VR headset use, but preprocedure anxiety was identified as a significant predictor of pain. Postprocedure mean pain scores in this study were around 5, similar to our study. Most patients underwent copper IUD insertion, which is known to be less painful, and approximately 80% of patients took preprocedural analgesia that included opiates and other prescription medications, factors which may have biased results toward the null.

The study by Oz et al [17] was conducted in Turkey and recruited 80 patients with a mean age of around 24 years. They used copper IUDs and found a reduction in postprocedural pain and anxiety scores with the use of VR. Mean pain scores were around 2 cm. This study's strengths included its uniform use of only 1 VR visual and audio immersion option for all participants, similar to our study. However, the description of inclusion criteria was limited, and confounders such as previous cervical instrumentation, chronic pain, or other concurrent analgesia use were not stated.

Our study is the first to evaluate the use of VR for IUD insertion in an Australian setting. It has multiple similarities to the study by Benazzouz et al [16], including similar mean pain scores and no benefit for pain or anxiety with VR device use. Its prospective design and randomization are strengths and add to the sparse body of literature. We enrolled a representative sample of individuals who would be seen for IUDs in an Australian tertiary hospital context. The IUDs used in our study were predominantly levonorgestrel devices, in contradistinction to the other studies. Other strengths include practical, real-life, broad inclusion criteria, and we excluded people who had taken pain relief other than simple analgesia.

Limitations

This study has limitations. First, although our study was practically oriented and therefore purposefully limited in the data instruments collected, this limited our ability to more deeply understand which subpopulations might benefit from this intervention or not. Deeper characterization of responder and nonresponder populations, particularly with regard to anxiety, would be an important consideration for future studies.

Second, this was a single-center study with multiple operators, so results may not apply to other operators or populations; however, the outcomes observed are plausible and concordant with other data. It is also possible that factors inherent to our study population may have limited the efficacy of VR headset use.

Third, although randomization is supposed to produce confounder balance, there were apparent differences between groups in our study. The VR group was more likely to have IUDs inserted for abnormal uterine bleeding and a history of chronic pain and less likely to have undergone previous cervical instrumentation. These factors could contribute to the patients' interactions with pain experiences and expectations prior to IUD insertion, which in turn can worsen pain. These confounders may have blunted the measured effect in this study. It is also possible that other potential unmeasured confounders may have impacted the outcome (eg, unmeasured history of substance use).

In addition, despite a formal power calculation, we cannot exclude the possibility that this study was underpowered, given that we observed a much higher SD in pain scores than we had anticipated. Pain is an inherently subjective phenomenon, and individual variability in the reporting of these outcomes can be vexing. If future studies are to be performed, these could be on a larger scale in a multicenter context to explore a broader

XSL•FO
RenderX

patient population and improve the generalizability of results while reducing the likely impact of possible confounders.

Another possible limiting factor could be that the type of content provided to patients, though thought to be appropriate, may not have been optimal. By standardizing the content provided, we provided a repeatable intervention, but this may have come at the cost of patient preferences for content and interests, and therefore immersion. There may also have been factors unique to the IUD procedure and our busy clinical environment that may have inhibited the formation of a therapeutic alliance.

Finally, due to the nature of this VR intervention, blinding of both clinicians and patients was not feasible.

Future Directions

Future studies should be directed toward helping to improve the patient experience with IUD insertion. In addition to larger population studies to maximize generalizability of results and ability for subgroup analysis, there are several more specific areas identified by this study that warrant future consideration:

First, further qualitative studies regarding the other possible benefits of VR use for IUD insertion should be considered, given the positive response from patients regarding this despite the lack of improvement in pain and anxiety scores.

Second, the finding that preprocedural anxiety was most predictive of pain raises the question of whether providing anxiety management prior to IUD insertion could be beneficial. It may be that using VR prior to the procedure could be a useful way to reduce anxiety, potentially reduce pain, and ultimately improve procedural tolerability.

Finally, further research is required into optimizing VR content to maximize its interactivity and immersion, particularly in the gynecological context. Co-design with patients should be considered, and a balance will need to be reached between depth of VR immersion while also maintaining adequate situational awareness such that the patient can continue to provide consent to the intimate examinations being undertaken. This particularly requires cautious curation of the audio content to allow clinician communication with the patient. Visual imagery also needs to keep in mind the sensitive nature of this procedure type and therefore should avoid imagery that could be misconstrued or considered triggering for patients with a history of sexual trauma. This is a unique challenge that was felt to limit the appropriate options of VR content in this study and needs to be explored in depth in future VR research. Tailoring content for subgroup-specific features, such as heightened anxiety, or clinical characteristics, such as sexual trauma, may result in greater benefit.

Conclusion

Importantly, this study found that the most significant predictor of pain was patient anxiety, consistent with Bennazouz et al [16] and the broader VR literature. This also predicted the patients who would benefit from VR use with reduced pain scores. This implies that efforts should be focused on managing patient anxiety prior to IUD insertion as a method of alleviating patient pain, and there may be a role for VR in the anxiety-reducing preprocedural setting that should be explored further in future research. The high satisfaction and recommendation for VR during IUD insertion in this trial also highlight that there may be other benefits to VR which have not yet been measured. This may indicate that VR could be considered as a supplemental tool rather than a primary method for pain relief. Ultimately, the goal to facilitate timely, more tolerable outpatient IUD insertions remains a clinical and research priority.

Acknowledgments

Generative AI was not used in any portion of the manuscript generation.

Data Availability

The datasets generated or analyzed during this study are not publicly available because ethics was not specifically granted for this reason during the approval process, but are available from the corresponding author on reasonable request and ethical application.

Authors' Contributions

Conceptualization: PL, KT, CL, LL, BV
Data curation: CH, CZM, JH
Formal analysis: PL, CH, CZM
Investigation: CH, CZM, JH
Methodology: CH, KT, CL, LL, PL, BV
Project administration: PL
Resources: PL, CL, LL
Supervision: PL, BV
Validation: KT, PL, BV
Visualization: CH
Writing – original draft: CH
Writing – review & editing: CH, CZM, JH, KT, CL, LL, PL, BV.

XSL•FO
RenderX

Conflicts of Interest

PL is a co-founder of Smileyscope Pty Ltd. The VR headset was provided by Smileyscope Pty Ltd.

Multimedia Appendix 1
Survey provided to patients and clinicians for data collection.
[PDF File, 46 KB - games_v13i1e72917_app1.pdf]

Checklist 1
CONSORT Checklist.
[PDF File, 1124 KB - games_v13i1e72917_app2.pdf]

References

1. Pergialiotis V, Vlachos DG, Protopappas A, Vlachos GD. Analgesic options for placement of an intrauterine contraceptive: a meta-analysis. Eur J Contracept Reprod Health Care 2014 Jun;19(3):149-160. [doi: 10.3109/13625187.2014.903238] [Medline: 24828514]
2. Miles SM, Shvartsman K, Dunlow S. Intrauterine lidocaine and naproxen for analgesia during intrauterine device insertion: randomized controlled trial. Contracept Reprod Med 2019;4(13):13. [doi: 10.1186/s40834-019-0094-0] [Medline: 31516731]
3. Maguire K, Davis A, Rosario Tejeda L, Westhoff C. Intracervical lidocaine gel for intrauterine device insertion: a randomized controlled trial. Contraception 2012 Sep;86(3):214-219. [doi: 10.1016/j.contraception.2012.01.005] [Medline: 22325115]
4. d'Arcangues C. Worldwide use of intrauterine devices for contraception. Contraception 2007 Jun;75(6 Suppl):S2-S7. [doi: 10.1016/j.contraception.2006.12.024] [Medline: 17531612]
5. Suzuki Y, Ferris JS, Chen L, et al. Fertility-preserving treatment for stage IA endometrial cancer: a systematic review and meta-analysis. Am J Obstet Gynecol 2024 Dec;231(6):599-610. [doi: 10.1016/j.ajog.2024.07.018] [Medline: 39032722]
6. Bednarek PH, Creinin MD, Reeves MF, et al. Prophylactic ibuprofen does not improve pain with IUD insertion: a randomized trial. Contraception 2015 Mar;91(3):193-197. [doi: 10.1016/j.contraception.2014.11.012] [Medline: 25487172]
7. Deo N, Khan KS, Mak J, et al. Virtual reality for acute pain in outpatient hysteroscopy: a randomised controlled trial. BJOG 2021 Jan;128(1):87-95. [doi: 10.1111/1471-0528.16377] [Medline: 32575151]
8. Anthoulakis C, Iordanidou E, Vatopoulou A. Pain perception during levonorgestrel-releasing intrauterine device insertion in nulliparous women: a systematic review. J Pediatr Adolesc Gynecol 2018 Dec;31(6):549-556. [doi: 10.1016/j.jpag.2018.05.008] [Medline: 29890206]
9. Akers AY, Steinway C, Sonalkar S, et al. Reducing pain during intrauterine device insertion: a randomized controlled trial in adolescents and young women. Obstet Gynecol 2017 Oct;130(4):795-802. [doi: 10.1097/AOG.0000000000002242] [Medline: 28885425]
10. The UK medical eligibility for criteria for contraceptive use (UKMEC). The College of Sexual and Reproductive Healthcare. 2016. URL: https://www.fsrh.org/Public/Public/Standards-and-Guidance/uk-medical-eligibility-criteria-for-contraceptive-use-ukmec.aspx [accessed 2025-09-02]
11. Daykan Y, Battino S, Arbib N, et al. Verbal analgesia is as good as oral tramadol prior to intrauterine device (IUD) insertion, among nulliparous women: a randomized controlled trial. Eur J Obstet Gynecol Reprod Biol 2021 Mar;258:443-446. [doi: 10.1016/j.ejogrb.2020.09.019] [Medline: 33187752]
12. Shippert RD. A study of time-dependent operating room fees and how to save $100 000 by using time-saving products. Am J Cosmet Surg 2005 Mar;22(1):25-34. [doi: 10.1177/074880680502200104]
13. Lopez LM, Bernholc A, Zeng Y, et al. Interventions for pain with intrauterine device insertion. Cochrane Database Syst Rev 2015 Jul 29;2015(7):CD007373. [doi: 10.1002/14651858.CD007373.pub3] [Medline: 26222246]
14. Smith V, Warty RR, Sursas JA, et al. The effectiveness of virtual reality in managing acute pain and anxiety for medical inpatients: systematic review. J Med Internet Res 2020 Nov 2;22(11):e17980. [doi: 10.2196/17980] [Medline: 33136055]
15. Mallari B, Spaeth EK, Goh H, Boyd BS. Virtual reality as an analgesic for acute and chronic pain in adults: a systematic review and meta-analysis. J Pain Res 2019;12:2053-2085. [doi: 10.2147/JPR.S200498] [Medline: 31308733]
16. Benazzouz I, Bouhnik C, Chapron A, Esvan M, Lavoué V, Brun T. Effects of virtual reality on pain during intrauterine device insertions: a randomized controlled trial. J Gynecol Obstet Hum Reprod 2024 Jan;53(1):102706. [doi: 10.1016/j.jogoh.2023.102706] [Medline: 38013014]
17. Öz T, Demirci N. The effect of virtual reality glasses applied during intrauterine device insertion on pain, anxiety and satisfaction: randomized controlled study. Scott Med J 2024 May;69(2):37-44. [doi: 10.1177/00369330241234688] [Medline: 38449359]
18. Baradwan S, Alshahrani MS, AlSghan R, et al. The effect of virtual reality on pain and anxiety management during outpatient hysteroscopy: a systematic review and meta-analysis of randomized controlled trials. Arch Gynecol Obstet 2024 Apr;309(4):1267-1280. [doi: 10.1007/s00404-023-07319-8] [Medline: 38165441]

XSL•FO
RenderX

19. Cohen N, Nasra LA, Paz M, Kaufman Y, Lavie O, Zilberlicht A. Pain and anxiety management with virtual reality for office hysteroscopy: systemic review and meta-analysis. Arch Gynecol Obstet 2024 Apr;309(4):1127-1134. [doi: 10.1007/s00404-023-07261-9] [Medline: 37917158]
20. Williamson A, Hoggart B. Pain: a review of three commonly used pain rating scales. J Clin Nurs 2005 Aug;14(7):798-804. [doi: 10.1111/j.1365-2702.2005.01121.x] [Medline: 16000093]
21. Reed MD, Van Nostran W. Assessing pain intensity with the visual analog scale: a plea for uniformity. J Clin Pharmacol 2014 Mar;54(3):241-244. [doi: 10.1002/jcph.250] [Medline: 24374753]
22. Lunde S, Petersen KK, Kugathasan P, Arendt-Nielsen L, Søgaard-Andersen E. Chronic postoperative pain after robot-assisted laparoscopic hysterectomy for endometrial cancer. J Gynecol Surg 2019 Jun 1;35(3):140-146. [doi: 10.1089/gyn.2018.0068]
23. Grundström H, Alehagen S, Berterö C, Kjølhede P. Impact of pelvic pain and endometriosis on patient-reported outcomes and experiences of benign hysterectomy: a study from the Swedish national register for gynecological surgery. J Womens Health (Larchmt) 2018 May;27(5):691-698. [doi: 10.1089/jwh.2017.6546] [Medline: 29148910]
24. Akdemir Y, Karadeniz M. A psychological factor associated with pain during intrauterine device insertion: emotional reactivity. Clin Exp Obstet Gynecol 2020;47(3):335-340. [doi: 10.31083/j.ceog.2020.03.2073]

Abbreviations

IUD: intrauterine device
VAS: visual analog scale
VR: virtual reality

Edited by A Coristine; submitted 23.02.25; peer-reviewed by A Ishola, M Bilal; revised version received 31.05.25; accepted 31.05.25; published 08.09.25.

Please cite as:
Higgins C, Zecena Morales C, Hocking J, Tyson K, Leung C, Larmour L, Leong P, Vollenhoven B
Virtual Reality for Analgesia During Intrauterine Device Insertion: Randomized Controlled Trial
JMIR Serious Games 2025;13:e72917
URL: https://games.jmir.org/2025/1/e72917
doi:10.2196/72917

© Chloe Higgins, Claudia Zecena Morales, Judith Hocking, Kate Tyson, Cheryl Leung, Luke Larmour, Paul Leong, Beverley Vollenhoven. Originally published in JMIR Serious Games (https://games.jmir.org), 8.9.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Exploring the Immediate and Long-Term Effects of Immersive Virtual Reality on Behavioral and Psychological Symptoms of Dementia and Caregiver Burden: Longitudinal Observational Study

Ling-Chun Huang[1,2,3,4], MD, MS; Ching-Fang Chien[1,2,3,5], MD, MS; Yuan-Han Yang[1,2,3,4], MD, MS, PhD

[1]Department of Neurology, Kaohsiung Medical University Gangshan Hospital, Kaohsiung Medical University, No. 8, Jie'an Road, Gangshan District, Kaohsiung, Taiwan

[2]Department of Neurology, Kaohsiung Medical University Hospital, Kaohsiung Medical University, Kaohsiung, Taiwan

[3]Neuroscience Research Center, Kaohsiung Medical University, Kaohsiung, Taiwan

[4]School of Post-Baccalaureate Medicine, College of Medicine, Kaohsiung Medical University, Kaohsiung, Taiwan

[5]School of Medicine, College of Medicine, Kaohsiung Medical University, Kaohsiung, Taiwan

Corresponding Author:
Yuan-Han Yang, MD, MS, PhD
Department of Neurology, Kaohsiung Medical University Gangshan Hospital, Kaohsiung Medical University, No. 8, Jie'an Road, Gangshan District, Kaohsiung, Taiwan

Abstract

Background: Virtual reality (VR) interventions are emerging as promising nonpharmacological strategies for people with dementia, aiming to prevent cognitive decline, reduce behavioral and psychological symptoms of dementia (BPSD), and alleviate caregiver burden. Although some studies have reported beneficial effects, findings remain inconsistent, and little is known about the duration and sustainability of these effects, particularly in real-world care settings.

Objective: This study aimed to examine both the immediate and long-term effects of an immersive VR reminiscence intervention on BPSD and caregiver burden in people with dementia attending day care centers.

Methods: This longitudinal observational study was conducted in 10 dementia day care centers in Kaohsiung, Taiwan. A total of 82 participants with dementia were enrolled. The VR intervention consisted of twice-weekly sessions over one month, featuring culturally familiar live-action 360° scenes filmed in well-known Taiwanese locations. Each session lasted approximately 10 - 12 minutes and included interactive elements. Neuropsychiatric symptoms were assessed using the Neuropsychiatric Inventory Questionnaire, and caregiver burden was assessed using the Zarit Caregiver Burden Interview. Measurements were taken at 3 time points: preintervention, immediately postintervention, and 2 months after the intervention ended. The Wilcoxon signed-rank test was used for statistical comparisons, and rank-biserial correlation was calculated as the effect size.

Results: Significant improvements were observed after 1 month of VR intervention in both caregiver burden (Z=−3.095, *P*=.002, *r*=0.34) and neuropsychiatric symptoms (Z=−2.929, *P*=.003, *r*=0.32). At the two-month follow-up, neuropsychiatric symptoms remained significantly improved (Z=−4.327, *P*<.001, *r*=0.48), although caregiver burden returned to preintervention levels. Regarding specific neuropsychiatric symptoms, significant improvements were observed immediately after the intervention in dysphoria or depression, anxiety, and sleep or nighttime behaviors. These effects were sustained over time, with additional long-term improvements noted in euphoria or elation, apathy or indifference, irritability or lability, aberrant motor behavior, and appetite or eating behaviors.

Conclusions: A 1-month immersive VR reminiscence intervention appears to improve neuropsychiatric symptoms and temporarily reduce caregiver burden in people with dementia, with some symptom improvements lasting up to 2 months. These findings suggest that VR may offer a meaningful therapeutic option in day care settings. Future studies with control groups, including nonimmersive 2D conditions, and comparisons to traditional reminiscence therapy are needed to validate and expand upon these findings.

(JMIR Serious Games 2025;13:e73044) doi:10.2196/73044

KEYWORDS

virtual reality; dementia; behavioral and psychological symptoms of dementia; caregiver burden; long-term effects

XSL•FO
RenderX

Introduction

Over the past few decades, the rapid growth of the older adults population in high-income countries has led to a continuous rise in the prevalence of dementia, posing significant social and economic impacts [1]. Globally, more than 50 million people are estimated to have dementia, and this number continues to rise [2,3]. Behavioral and psychological symptoms of dementia (BPSD), also known as neuropsychiatric symptoms, encompass a wide range of noncognitive symptoms and behaviors, including agitation, aberrant motor behavior, anxiety, elation, irritability, depression, apathy, disinhibition, delusions, hallucinations, and changes in sleep or appetite. BPSD are frequently observed in patients with dementia, leading to distress, reduced quality of life, worsening cognitive and functional impairments, and increased caregiver stress and burden [4-6]. There is evidence that pharmacological treatments for BPSD often come with adverse side effects, including an increased risk of falls, fractures, stroke, and even mortality, while offering limited efficacy in symptom management [7]. Therefore, novel and effective nonpharmacological therapies with fewer side effects are urgently needed.

Virtual reality (VR) is a computer-generated environment that has been developed and refined in recent years. Using headsets, VR provides a fully immersive, highly engaging, and realistic experience. VR technologies are increasingly regarded as valuable tools in dementia research and as safe, well-tolerated nonpharmacological therapies that may enhance the quality of life and well-being of individuals with dementia [8-10]. Most studies investigating the impact of VR therapy on BPSD in people with dementia have been published since 2018 [10]. However, conclusions regarding its effectiveness have been somewhat inconsistent. For example, Rose et al [11] found that BPSD decreased during and after VR intervention compared to preintervention levels. In contrast, Coelho et al [12] reported no significant changes in the psychological and behavioral symptoms of participants following the intervention. Similarly, a scoping review examining empirical evidence on the use of fully immersive VR in reminiscence interventions found that some studies reported positive effects on anxiety, apathy, depression, cognitive function, and caregiver burden. However, these findings were inconsistent across other research [13]. Therefore, further investigation is required to better understand the impact of VR interventions on BPSD and caregiver burden.

Moreover, the long-term effects of VR interventions in people with dementia remain unclear [13]. In our previous study [14], we assessed cognitive function and depressive symptoms in 20 participants before and immediately after a VR intervention, with follow-up assessments conducted in 7 participants 3 to 6 months later. While cognitive abilities showed no significant changes immediately after the intervention, depressive symptoms were significantly reduced. However, cognitive function declined notably over the following months, suggesting that VR interventions may provide temporary cognitive benefits but not long-term preservation. It is important to note that the sample size was small, particularly for follow-up assessments, and only depressive symptoms of BPSD were evaluated in detail.

Therefore, we conducted this study to investigate both the immediate and long-term effects of immersive VR reminiscence therapy on BPSD and caregiver burden in people with dementia, with an increased sample size.

Methods

Study Design

This is a longitudinal observational study conducted in dementia day care centers to evaluate both the immediate and long-term effects of immersive VR reminiscence therapy on BPSD and caregiver burden. Participants were assessed before the intervention, immediately after the 1-month VR intervention, and 2 months postintervention.

Recruitment

Participants were recruited from 10 dementia day care centers in Kaohsiung City, Taiwan. Staff at these centers identified potential participants based on predefined inclusion and exclusion criteria. The inclusion criteria were: (1) a diagnosis of all-cause dementia by experienced physicians, following the National Institute on Aging and Alzheimer's Association diagnostic criteria [15]; and (2) regular attendance at the dementia day care centers between May 2023 and April 2024. Participants were excluded if they: (1) had difficulty recognizing VR images despite adjustments to the headset's positioning; or (2) had cognitive impairment or BPSD severe enough to hinder proper assessment. Baseline demographic and functional information was collected using the Chinese version of the Uniform Data Set (UDS) Form A [16], which includes variables such as age, sex, years of education, marital status, living situation, and functional status (eg, need for assistance with complex or basic activities). The form was completed at enrollment by center staff, based on direct interviews with participants and input from their caregivers or informants.

Ethical Considerations

The participants and their relatives were informed of the details of the study before their inclusion, and appropriate written informed consent was obtained. Given the vulnerability of individuals with dementia, additional precautions were taken to ensure comprehension, including clear verbal explanations and opportunities to ask questions. The Kaohsiung Medical University Hospital Institutional Review Board (KMUHIRB-SV(I)-20210015) approved the study protocol. All data were deidentified, stored securely, and accessible only to authorized researchers. Results are presented in aggregate to ensure confidentiality. Participants did not receive any financial compensation for their participation in this study.

VR Apparatus

The HTC VIVE Pro VR head-mounted display (HMD; HTC Corporation) was used to deliver the immersive VR experience to the participants. It is a stand-alone HMD providing stereoscopic vision at a resolution of 2880×1600 per eye with a 90 Hz refresh rate. Two controllers were paired with the HMD to enable the participants to interact with the VR environment.

XSL•FO
RenderX

Preparation of VR Content Implementation

The VR content consisted of live-action scenes showcasing famous scenic areas and landmarks in Taiwan. These live-action images were captured using a 360° camera at Lianchihtan (Lotus Pond) Scenic Area and the Fo Guang Shan Buddha Museum in Kaohsiung City. Figure 1 displays screenshots from 2 of these scenes.

Figure 1. Scenes from the live-action virtual reality reminiscence intervention. The footage was captured at the (A,B) Lianchihtan (Lotus Pond) Scenic Area and the (C,D) Fo Guang Shan Buddha Museum in Kaohsiung City.



Lianchihtan (Lotus Pond) is one of Kaohsiung's most historic landmarks and was listed as one of the "Eight Sights of Fongshan" during the Qing Dynasty. The Longhu Tower (Dragon Tiger Tower), renowned for its striking architectural design, is the most iconic attraction in Lianchihtan. Built in 1976, the seven-story tower features dragon and tiger sculptures as its symbolic entrance and exit. The walls are adorned with koji pottery created by various artists, depicting stories of religious and cultural significance. Visitors can take in a stunning 360-degree view of Lianchihtan's scenic beauty, immersing themselves in its cultural and historical ambiance. The Buddha Museum fuses traditional and modern elements, offering facilities to assist in spiritual cultivation and the attainment of wisdom, while also serving as a hub for culture and education. The Lianchihtan Scenic Area and Fo Guang Shan Buddha Museum were selected for their cultural familiarity, high recognition, and potential to evoke personal memories in older Taiwanese adults. These well-known sites also offered rich sensory environments, which are beneficial for creating immersive and engaging VR experiences. To capture the essence of this remarkable site, we documented its various scenes through detailed photography. Digital reminiscence therapy recreates past scenes using real-time background removal technology, allowing users' family members to be integrated into the VR environment. This creates a sense of togetherness despite physical distances. Users can be accompanied by family members or guided by a virtual tour guide while experiencing immersive elements like cicada sounds, drawing lots at the Great Buddha City, making wishes at the wishing pond, and tossing coins to ring a bell. In addition, the real-time background removal feature enables quick generation of shared moment photos with family.

Implementation

The VR intervention was administered twice per week over a period of 1 month. The participants viewed and interacted with the VR content for approximately 10 to 12 minutes each time. The HMD images were mirrored to a laptop to enable the researcher to see what the participants were viewing and interacting with. The participants remained seated during the VR intervention to reduce motion sickness when using the HMD and to also minimize the risk of falls in the older adults. During each session, participants were guided through immersive VR scenes while the researcher encouraged reminiscence by prompting discussions about personal experiences related to the viewed locations. Open-ended questions such as "Have you visited this place before?" or "What does this remind you of?" were used to stimulate engagement. While the session structure remained consistent (10 - 12 min, twice weekly), participants' interactions varied based on their individual responses. The participants were provided with the controllers to enable them to interact with the VR space, and the researcher was beside them to assist and guide them when needed.

Assessment of Questionnaire and Scores

Before, immediately after, and 2 months postintervention, the informants completed the Neuropsychiatric Inventory-Questionnaire (NPI-Q) to assess the severity of

XSL•FO
RenderX

neuropsychiatric symptoms and their impact on them. Caregiver burden was evaluated using the Zarit Caregiver Burden Interview (ZBI).

Neuropsychiatric Inventory-Questionnaire

The NPI-Q was developed and cross-validated with the standard NPI to provide a brief assessment of neuropsychiatric symptomatology in patients with dementia for use in both clinical practice and research [17-20]. Adapted from the NPI [21], a validated informant-based interview assessing neuropsychiatric symptoms over the past month, the NPI-Q is designed as a self-administered questionnaire completed by caregivers. Each of the 12 NPI-Q domains contains a survey question that reflects cardinal symptoms of that domain. Initial responses to each domain question are "Yes" (present) or "No" (absent). If the response to the domain question is "No," the informant goes to the next question. If "Yes," the informant then rates both the severity of the symptoms present within the last month on a 3-point scale and the associated impact of the symptom manifestations on them (ie, caregiver distress) using a 5-point scale. The NPI-Q provides symptom severity and distress ratings for each symptom reported, and total severity and distress scores reflecting the sum of individual domain scores. An English version of the NPI-Q is provided in Multimedia Appendix 1.

Zarit Caregiver Burden Interview

The ZBI is a well-known self-reporting measure of perceived burden among caregivers [22], with prior applications in similar research on dementia [14,23]. The instrument measures the caregiver's emotion, psychological health, well-being, social and family life, finances, and degree of control over one's life. The version used contains 22 items, and each item on the questionnaire is a statement that the caregiver is asked to endorse on a 5-point Likert scale (0: never; 1: rarely; 2: sometimes; 3: quite frequently; and 4: nearly always) [22]. Higher scores indicate higher caregiver burden. An English version of the ZBI is provided in Multimedia Appendix 2.

Statistical Analysis

In the results, the data are presented as mean (SD) or as a proportion. Comparisons of scores before and immediately after VR intervention, as well as before intervention and 2 months after stopping intervention, were performed using the Wilcoxon sign-rank test due to the nonnormal distribution of the dependent variables, as determined by the Kolmogorov-Smirnov test. For each Wilcoxon signed-rank test, the test statistic was converted into a standardized Z value, which reflects both the magnitude and direction of change. Effect sizes (r) were calculated using the rank-biserial correlation for each Wilcoxon signed-rank test. Based on Cohen convention [24], r value of approximately 0.1 indicates a small effect, approximately 0.3 a medium effect, and $r \geq 0.5$ a large effect. To control for the false discovery rate resulting from multiple comparisons of the 12 individual NPI-Q domains, the Benjamini-Hochberg procedure was applied, with the false discovery rate set at 0.05. The required sample size was estimated using GPower 3.1 for a Wilcoxon signed-rank test. Assuming a large effect size (r=0.5), a significance level of 0.05, and a power of 0.80, the minimum required sample size was 64 participants. The large effect size was chosen based on Cohen guidelines due to the limited availability of prior data on immersive VR interventions for BPSD and caregiver burden. Analyses were performed using SPSS 26.0 (IBM Corporation). A 2-tailed P value of <.05 was considered to indicate a statistically significant difference.

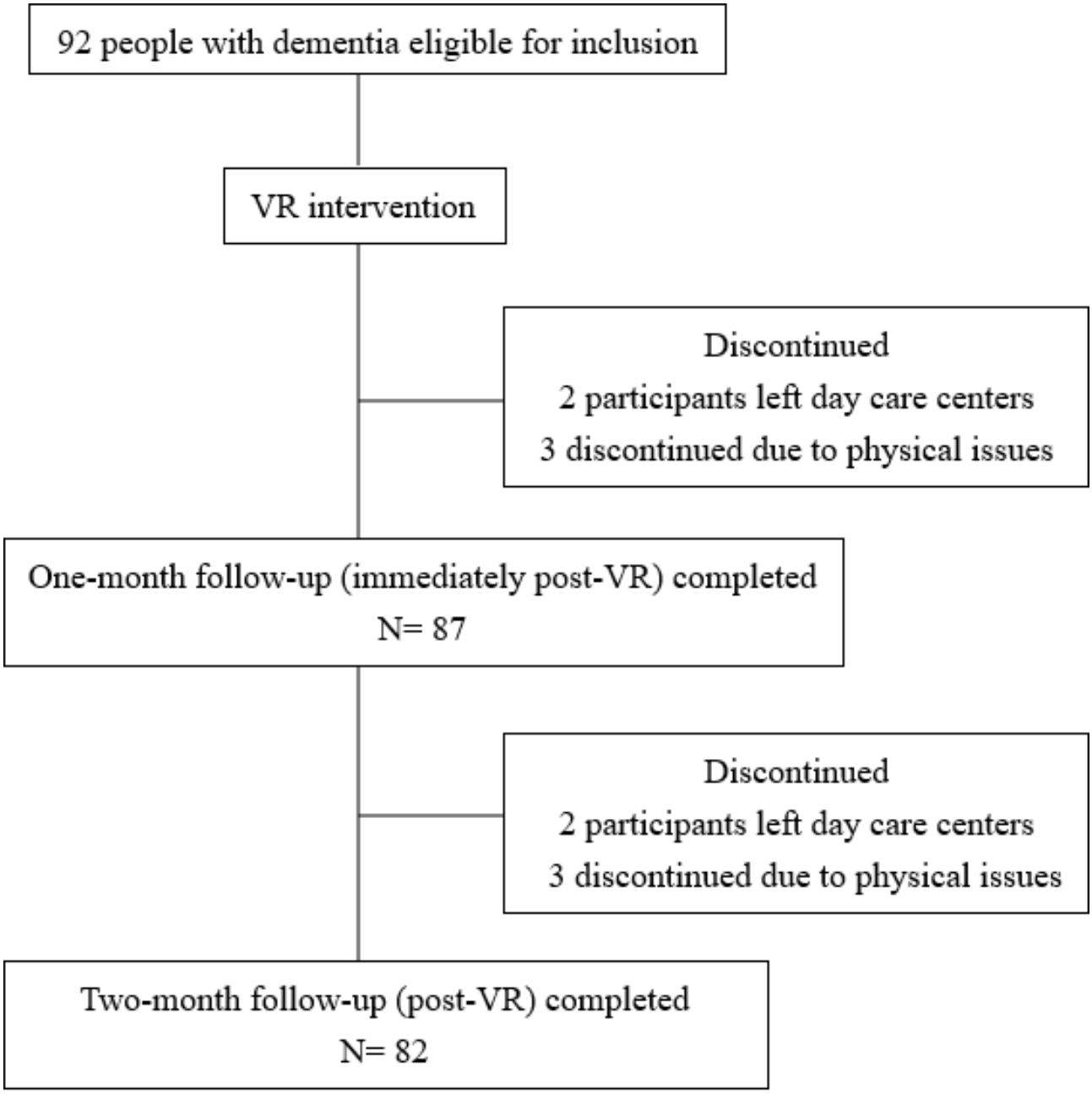
Results

Baseline Demographic Characteristics

A total of 92 individuals were enrolled at the beginning of the study. After the 1-month intervention, 2 participants left the dementia day care centers, and 3 were unable to complete the postintervention evaluation due to physical issues. At the 2-month follow-up, an additional 2 participants left the centers, and 3 were unable to continue due to physical problems (Figure 2).

XSL•FO
RenderX

Figure 2. Diagram of the study design.



VR: Virtual Reality

There were no statistically significant differences between the 2 groups (participants who completed the study versus those who dropped out) in terms of age, sex, education level, or baseline ZBI and NPI-Q scores. A detailed comparison between the 2 groups is presented in Table 1.

XSL•FO
RenderX

Table . Comparison of baseline characteristics between participants who completed the study and those who dropped out.

Variables	Nondrop out group (n=82)	Drop out group (n=10)	*P* value
Age in years, mean (SD)	79.7 (7.8)	81.5 (7.8)	.71
Sex, n (%)			.41
Male	27 (32.9)	2 (20)	
Female	55 (67.1)	8 (80)	
Education in years, mean (SD)	7.6 (3.1)	6.8 (1.5)	.21
Baseline ZBI[a], mean (SD)	44.3 (18.0)	47.4 (17.8)	.46
Baseline NPI-Q[b], mean (SD)	35.0 (23.1	29.7 (20.8)	.73

[a]ZBI: Zarit Caregiver Burden Interview.

[b]NPI-Q: Neuropsychiatric Inventory-Questionnaire.

The baseline demographic characteristics of the 82 participants who completed the study are shown in Table 2. The participants had a mean age of 79.7 years (SD 7.8), with 32.9% (27/82) being male and 67.1% (55/82) being female. The average education level was 7.6 years (SD 3.1). Most participants were categorized as requiring assistance with complex activities (35/82, 42.7%) or with basic activities (38/82, 46.3%). Regarding marital status, 73.2% (60/82) were widowed, and 20.7% (17/82) were married. For living situations, 81.7% (67/82) lived with a spouse, partner, or children; 7.3% (6/82) lived alone; and 7.3% (6/82) lived with a friend or other relative.

Table . Baseline demographic characteristics of the participants.

Variables	Participants (N=82)
Age in years, mean (SD)	79.7 (7.8)
Sex, n (%)	
Male	27 (32.9)
Female	55 (67.1)
Education in years, mean (SD)	7.6 (3.1)
Current marital status, n (%)	
Married	17 (20.7)
Widowed	60 (73.2)
Divorced	1 (1.2)
Separated	3 (3.7)
Never married	1 (1.2)
Level of independence, n (%)	
Independent	4 (4.9)
Assistance with complex activities	35 (42.7)
Assistance with basic activities	38 (46.3)
Completely dependent	5 (6.0)
Living situation, n (%)	
Alone	6 (7.3)
With spouse, partner, or children	67 (81.7)
With friend or other relative	6 (7.3)
Other	3 (3.7)

The Immediate Effect of VR Intervention

Table 3 presents the immediate effect of VR intervention on neuropsychiatric symptoms and caregiver burden. Compared to preintervention levels, caregiver burden (ZBI) significantly decreased after the VR intervention (Z=−3.095, *P*=.002, *r*=0.34), and neuropsychiatric symptoms (NPI-Q) also showed a significant reduction (Z=−2.929, *P*=.003, *r*=0.32). Specifically, significant improvements were observed in dysphoria or

XSL•FO
RenderX

depression (adjusted *P*=.036), anxiety (adjusted *P*=.036), and nighttime behaviors (adjusted *P*=.036).

Table . The immediate effect of virtual reality intervention on neuropsychiatric symptoms and caregiver burden.

Variables	Pre-VR[a], mean (SD)	Immediately post-VR[a], mean (SD)	*P* value[b]
ZBI[c]	44.3 (18.0)	38.0 (19.6)	.002[d]
NPI-Q[e]	35.0 (23.1)	28.6 (21.7)	.003[d]
Delusions	2.0 (4.0)	1.7 (3.4)	.34
Hallucinations	1.8 (3.9)	1.4 (3.3)	.31
Agitation or aggression	1.8 (3.7)	1.4 (3.1)	.48
Dysphoria or depression	3.3 (4.6)	2.0 (3.5)	.04[d]
Anxiety	3.2 (4.4)	2.0 (3.7)	.04[d]
Euphoria or elation	1.3 (3.4)	0.4 (1.3)	.08
Apathy or indifference	1.9 (3.5)	1.5 (2.9)	.61
Disinhibition	1.9 (3.9)	1.0 (3.0)	.10
Irritability or lability	2.3 (3.9)	1.3 (2.9)	.09
Aberrant motor	2.7 (4.5)	2.0 (3.7)	.08
Nighttime behavior	3.1 (5.2)	1.6 (3.3)	.04[d]
Appetite or eating	2.6 (4.9)	1.6 (3.4)	.10

[a]VR: Virtual Reality.

[b]*P* values comparing pre-VR and immediately post-VR using the Wilcoxon signed-rank test.

[c]ZBI: Zarit Caregiver Burden Interview.

[d]*P*<.05.

[e]NPI-Q: Neuropsychiatric Inventory-Questionnaire.

The Long-Term Effect of VR Intervention

To assess the long-term effects of VR intervention, we conducted follow-up evaluations two months after the intervention ended. At the 2-month follow-up, caregiver burden remained unchanged (Z=−1.658, *P*=.097, *r*=0.18), while neuropsychiatric symptoms continued to improve (Z=−4.327, *P*<.001, *r*=0.48), indicating a moderate effect size. As shown in Table 4, significant improvements were observed in dysphoria and depression (adjusted *P*=.012), anxiety (adjusted *P*=.012), euphoria and elation (adjusted *P*=.008), apathy and indifference (adjusted *P*=.042), irritability and lability (adjusted *P*=.008), aberrant motor behavior (adjusted *P*=.015), sleep and nighttime behaviors (adjusted *P*=.008), and appetite and eating behaviors (adjusted *P*=.012).

XSL•FO
RenderX

Table . The long-term effect of virtual reality intervention on neuropsychiatric symptoms and caregiver burden.

Variables	Pre-VR[a], mean (SD)	Two months post-VR[a], mean (SD)	*P* value[b]
ZBI[c]	44.3 (18.0)	41.1 (19.5)	.10
NPI-Q[d]	35.0 (23.1)	25.7 (16.7)	<.001[e]
Delusions	2.0 (4.0)	1.4 (3.3)	.14
Hallucinations	1.8 (3.9)	1.2 (3.1)	.14
Agitation or aggression	1.8 (3.7)	1.0 (2.9)	.14
Dysphoria or depression	3.3 (4.6)	1.9 (3.2)	.01[e]
Anxiety	3.2 (4.4)	1.7 (3.2)	.01[e]
Euphoria or elation	1.3 (3.4)	0.2 (1.0)	.008[e]
Apathy or indifference	1.9 (3.5)	1.1 (3.0)	.04[e]
Disinhibition	1.9 (3.9)	0.9 (2.3)	.12
Irritability or lability	2.3 (3.9)	1.0 (2.3)	.008[e]
Aberrant motor	2.7 (4.5)	1.7 (3.7)	.02[e]
Nighttime behavior	3.1 (5.2)	1.5 (3.2)	.008[e]
Appetite or eating	2.6 (4.9)	1.1 (2.0)	.01[e]

[a]VR: Virtual Reality.

[b]*P*-values comparing pre-VR and two months post-VR using the Wilcoxon signed-rank test.

[c]ZBI: Zarit Caregiver Burden Interview.

[d]NPI-Q: Neuropsychiatric Inventory-Questionnaire.

[e]*P*<.05

Discussion

Principal Results

In this paper, we investigated the potential effects of immersive VR reminiscence therapy on people with dementia and their caregivers, as well as the duration of these effects. After 1 month of VR intervention, both caregiver burden and the severity of patients' neuropsychiatric symptoms significantly decreased, with small to moderate effect sizes. In total, 2 months after the VR intervention ended, caregiver burden remained stable, while patients' neuropsychiatric symptoms continued to improve with a moderate effect size, suggesting the sustained impact of VR therapy. VR was shown to reduce caregiver burden and improve neuropsychiatric symptoms, with symptom stability observed even 2 months postintervention. The improvements in euphoria, apathy, irritability, aberrant motor behavior, and appetite and eating behaviors may suggest delayed therapeutic effects. To the best of our knowledge, this is the first study to investigate not only the potential effects of immersive VR reminiscence on various neuropsychiatric symptoms but also the duration of these effects in individuals with dementia.

Comparison With Previous Work

VR is increasingly recognized as a valuable therapeutic tool for people with dementia. It offers a cost-effective, noninvasive, and ethically acceptable approach to managing BPSD by redirecting patients' attention within a physically safe environment [25]. Immersive VR therapy has shown potential in reducing aggressive behaviors among patients with dementia in acute care hospitals [26]. In addition, BPSD decreased during and after VR interventions in hospitalized individuals with dementia compared to preintervention levels [11]. Our study, conducted in a community care setting, similarly demonstrated a reduction in BPSD following VR intervention. However, a previous study promoting VR reminiscence therapy in individuals with dementia found no significant changes in psychological and behavioral symptoms or quality of life after a short intervention period [12]. Likewise, Mendez et al reported no significant differences in arousal, stress, anxiety, anger, fatigue, or attention between pre- and postintervention sessions [27]. These inconsistencies across studies may stem from small sample sizes, variations in dementia subtypes, differences in study designs, outcome measures, VR content, software delivery methods, and levels of interactivity. Despite these variations, VR-based reminiscence interventions generally foster high levels of interaction and immersion. Older adults, including those with mild cognitive impairment and dementia, often find these interventions enjoyable and engaging, suggesting potential therapeutic benefits.

Most studies focus on the immediate effects of VR interventions, but the duration of these effects remains unclear. We hypothesized that VR would lead to short-term symptom improvement immediately after the intervention and that some effects might persist over time. In our previous study [14], we assessed cognitive function and depressive symptoms in 20 participants before and immediately after VR intervention. In

XSL•FO
RenderX

total, 7 participants were reassessed 3 - 6 months later, revealing continued cognitive decline compared to postintervention levels. In this study, we examined the long-term effects of immersive VR intervention on BPSD and caregiver burden with a larger sample. Given that the NPI-Q assesses neuropsychiatric symptoms over the past month, we selected the time point of 2 months after the intervention ended to minimize measurement interference. Unlike cognitive function, neuropsychiatric symptoms remained improved compared to preintervention levels even 2 months after the intervention ended. This suggests that VR's effects on neuropsychiatric symptoms may persist longer than its impact on cognitive function, lasting approximately twice the duration of the intervention. Further studies with longer follow-ups are needed to confirm this finding. Regarding specific neuropsychiatric symptoms, our results showed significant improvements in dysphoria or depression, anxiety, and sleep or nighttime behaviors immediately after the VR intervention. These improvements persisted over time, with additional long-term benefits observed in euphoria, apathy, irritability, aberrant motor behavior, and appetite and eating behaviors 2 months postintervention. These delayed improvements may reflect gradual engagement of motivational and emotional circuits, psychological adaptation processes, and evolving caregiver-patient dynamics, all of which may require time to unfold. In addition, limited but present neuroplasticity in older adults may contribute to slow-developing affective or behavioral changes [28]. These findings warrant further investigation to clarify the mechanisms driving delayed symptom changes.

Older adults with cognitive impairments may face challenges in accessing experiences beyond their physical surroundings. VR creates a 3D digital space that simulates real-world environments, allowing users to explore new places, revisit past memories, and engage meaningfully in life despite geographical or cognitive limitations [9]. In addition, networked VR allows multiple individuals to engage in shared experiences, such as group travel and coviewing virtual media, regardless of location. Its rich sensory stimuli create an immersive sense of presence, enabling older adults to feel truly connected with their loved ones [29]. In this study, although family members did not participate in the VR experience in real-time, their images were integrated into the virtual environment using real-time background removal technology. This allowed for interaction and created a sense of togetherness despite physical distance. Recent research has found that the use of VR to provide interactions may offer an alternative way of delivering stimulation to people with dementia who do not engage in other lifestyle activities [30]. In our study, participants were provided with controllers, enabling them to interact with the VR space (eg, drawing lots, tossing coins to ring a bell), which further enhanced their engagement in the scene and sparked their interest.

Our findings indicate that VR reminiscence intervention significantly reduced caregiver burden, consistent with the results of Yahara et al [31]. This reduction may be linked to the alleviation of neuropsychiatric symptoms, as these symptoms are strong predictors of increased caregiver burden over time. Longitudinal studies have shown that neuropsychiatric symptoms not only emerge early in dementia [32] but also tend to worsen as the disease progresses [33], significantly impacting caregivers. Among these symptoms, disruptive behaviors such as agitation or aggression, irritability, disinhibition, and aberrant motor behavior contribute the most to caregiver burden, followed by delusions and mood disturbances [34]. Dauphinot et al further found that apathy, agitation or aggression, aberrant motor behavior, appetite changes, and irritability were independently associated with caregiver burden, even after controlling for cognitive decline, functional impairment, and antidepressant use [35]. These findings suggest that VR interventions may help ease caregiver burden by mitigating neuropsychiatric symptoms. However, despite the sustained improvement in neuropsychiatric symptoms, caregiver burden remained unchanged 2 months after the intervention ended. This suggests that other factors may influence caregiver burden beyond symptom severity alone. Further research is needed to identify and address these additional contributing factors.

Limitations

There were some limitations to this study. First, the absence of a control group, including a traditional reminiscence therapy group or a nonimmersive 2D control condition (eg, viewing similar content on a standard screen), requires careful interpretation of the results. Implementing such control conditions posed practical challenges. Assigning participants to wear a VR headset without content may cause discomfort; similarly assigning them to traditional reminiscence therapy or no intervention was deemed impractical due to the group-based nature of dementia day care centers. In addition, caregivers were supportive of the VR intervention, and excluding participants from the activity could have led to dissatisfaction. As a result, it is difficult to determine whether the observed effects were specifically attributable to the immersive nature of VR or to the reminiscence content itself. Moreover, natural symptom fluctuations, regression to the mean, or placebo effects could also have influenced the results. Despite this limitation, our study provides valuable insights into the real-world application of VR for dementia care. Future research should incorporate ethically appropriate control groups, including nonimmersive conditions, to strengthen causal inferences and further validate these findings. Second, potential participants were identified by the staff at dementia day care centers, which may limit the generalizability of our findings to the broader dementia population. In addition, individuals with severe cognitive impairments or BPSD were excluded, which is a common limitation in most VR studies, further restricting the applicability of the results. Third, during the follow-up period, 4 participants left the dementia day care centers, and 6 were unable to complete the follow-up evaluation due to physical problems. This represents a loss to follow-up rate of only 10.9%, which is below the 20% threshold, indicating a minimal risk of bias [36]. Fourth, this study did not include data on dementia severity and dementia subtype classification. However, in Taiwan, most dementia daycare centers recruit individuals with a clinical dementia rating of 1 or higher, with the majority being clinical dementia rating 1 (mild) and the remainder typically clinical dementia rating 2 (moderate). In clinical settings, more than half of dementia cases are attributed to Alzheimer's disease.

XSL•FO
RenderX

Furthermore, according to the study by Lora Appel et al [25], 70% of participants displayed enjoyment or relaxation during the VR session, with dementia severity ranging from mild (20%) to moderate (10%) and advanced (40%). It is worth noting that VR-based reminiscence interventions could be tailored to provide greater assistance for patients in the advanced stage, given their higher prevalence of BPSD. Fifth, caregiver burden in dementia is a complex and multidimensional construct that includes both patient and caregiver-related factors [37]. It has been associated with patients' symptoms (such as BPSD, cognitive performance, and functional impairment), as well as personal characteristics. In this study, the ZBI was primarily completed by regular staff members at the dementia day care centers, as family caregivers were typically not present during assessment periods and often had limited availability. Therefore, the reported burden reflects the perspective of professional caregivers who interacted with the participants on a daily basis, rather than that of family members. Sixth, the observed effect sizes were smaller than the estimated large effect size, which may suggest that our study was underpowered to detect more modest effects. Future studies should consider using more conservative effect size assumptions when calculating sample size. Despite these limitations, to the best of our knowledge, this is the first study to explore not only the potential effects of immersive VR reminiscence but also the duration of these effects in people with dementia, using a relatively large sample size. Future studies could build on these results to develop personalized VR content tailored to different cognitive and emotional needs. Additionally, integrating VR into routine dementia care programs, such as day care centers or home-based interventions, may enhance accessibility and caregiver support. While more large-scale, controlled trials are needed to validate these effects, our study provides foundational evidence that could inform the design of future interventions and practical applications in dementia care.

Conclusions

This study demonstrated that immersive VR reminiscence therapy has the potential to reduce caregiver burden and improve neuropsychiatric symptoms in people with dementia. Notably, the beneficial effects on patients' symptoms persisted even two months postintervention. These findings suggest that VR reminiscence therapy may offer sustained and delayed therapeutic benefits, providing a promising nonpharmacological approach for dementia care.

Acknowledgments

This study was supported by grants from National Health Research Institutes Grant (PH-110-GP-08, PH-111-GP-08, PH-112-GP-08, PH-113-GP-08), the Research Center of Kaohsiung Medical University (KMU-TC113B02), and Kaohsiung Municipal Siaogang Hospital (Grant S-111 - 05). We also thank HamaStar Technology Co, Ltd for their assistance.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request. The VR intervention content (including the 360° video scenes) is also available upon reasonable request for research and non-commercial use, subject to institutional approval and appropriate attribution.

Authors' Contributions

Y-HY contributed to the study conception and design. L-CH, C-FC, and Y-HY contributed to the acquisition of data. L-CH, and Y-HY analyzed and interpreted the data. L-CH conducted the statistical analyses and involved in writing the initial draft of the manuscript. All authors reviewed and revised the manuscript and approved the submitted version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
English version of the Neuropsychiatric Inventory Questionnaire.
[PDF File, 763 KB - games_v13i1e73044_app1.pdf]

Multimedia Appendix 2
English version of the Zarit Caregiver Burden Interview.
[PDF File, 812 KB - games_v13i1e73044_app2.pdf]

References

1. Brookmeyer R, Johnson E, Ziegler-Graham K, Arrighi HM. Forecasting the global burden of Alzheimer's disease. Alzheimers Dement 2007 Jul;3(3):186-191. [doi: 10.1016/j.jalz.2007.04.381] [Medline: 19595937]
2. Dumurgier J, Tzourio C. Epidemiology of neurological diseases in older adults. Rev Neurol (Paris) 2020 Nov;176(9):642-648. [doi: 10.1016/j.neurol.2020.01.356] [Medline: 32145981]

XSL•FO
RenderX

3. GBD 2019 Dementia Forecasting Collaborators. Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: an analysis for the Global Burden of Disease Study 2019. Lancet Public Health 2022 Feb;7(2):e105-e125. [doi: 10.1016/S2468-2667(21)00249-8] [Medline: 34998485]
4. Black W, Almeida OP. A systematic review of the association between the behavioral and psychological symptoms of dementia and burden of care. Int Psychogeriatr 2004 Sep;16(3):295-315. [doi: 10.1017/s1041610204000468] [Medline: 15559754]
5. Devshi R, Shaw S, Elliott-King J, et al. Prevalence of behavioural and psychological symptoms of dementia in individuals with learning disabilities. Diagnostics (Basel) 2015 Dec 2;5(4):564-576. [doi: 10.3390/diagnostics5040564] [Medline: 26854171]
6. Donaldson C, Tarrier N, Burns A. The impact of the symptoms of dementia on caregivers. Br J Psychiatry 1997 Jan;170:62-68. [doi: 10.1192/bjp.170.1.62] [Medline: 9068778]
7. Magierski R, Sobow T, Schwertner E, Religa D. Pharmacotherapy of behavioral and psychological symptoms of dementia: state of the art and future progress. Front Pharmacol 2020;11:1168. [doi: 10.3389/fphar.2020.01168] [Medline: 32848775]
8. Kim O, Pang Y, Kim JH. The effectiveness of virtual reality for people with mild cognitive impairment or dementia: a meta-analysis. BMC Psychiatry 2019 Jul 12;19(1):219. [doi: 10.1186/s12888-019-2180-x] [Medline: 31299921]
9. Manera V, Chapoulie E, Bourgeois J, et al. A feasibility study with image-based rendered virtual reality in patients with mild cognitive impairment and dementia. PLoS One 2016;11(3):e0151487. [doi: 10.1371/journal.pone.0151487] [Medline: 26990298]
10. Appel L, Ali S, Narag T, et al. Virtual reality to promote wellbeing in persons with dementia: a scoping review. J Rehabil Assist Technol Eng 2021;8:20556683211053952. [doi: 10.1177/20556683211053952] [Medline: 35024166]
11. Rose V, Stewart I, Jenkins KG, Tabbaa L, Ang CS, Matsangidou M. Bringing the outside in: the feasibility of virtual reality with people with dementia in an inpatient psychiatric care setting. Dementia (London) 2021 Jan;20(1):106-129. [doi: 10.1177/1471301219868036] [Medline: 31510801]
12. Coelho T, Marques C, Moreira D, et al. Promoting reminiscences with virtual reality headsets: a pilot study with people with dementia. Int J Environ Res Public Health 2020 Dec 12;17(24):9301. [doi: 10.3390/ijerph17249301] [Medline: 33322679]
13. Lu Z, Wang W, Yan W, Kew CL, Seo JH, Ory M. The application of fully immersive virtual reality on reminiscence interventions for older adults: scoping review. JMIR Serious Games 2023 Oct 6;11:e45539. [doi: 10.2196/45539] [Medline: 37801360]
14. Huang LC, Yang YH. The long-term effects of immersive virtual reality reminiscence in people with dementia: longitudinal observational study. JMIR Serious Games 2022 Jul 25;10(3):e36720. [doi: 10.2196/36720] [Medline: 35877169]
15. McKhann GM, Knopman DS, Chertkow H, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011 May;7(3):263-269. [doi: 10.1016/j.jalz.2011.03.005] [Medline: 21514250]
16. Wang WF, Chiu PY, Lin YT, Hu CJ, Fuh JL, Yang YH. Registration of Alzheimer's disease in Taiwan: patient and informant. Am J Alzheimers Dis Other Demen 2014 Feb;29(1):18-22. [doi: 10.1177/1533317513504818] [Medline: 24085249]
17. Kaufer DI, Cummings JL, Ketchel P, et al. Validation of the NPI-Q, a brief clinical form of the Neuropsychiatric Inventory. J Neuropsychiatry Clin Neurosci 2000;12(2):233-239. [doi: 10.1176/jnp.12.2.233] [Medline: 11001602]
18. Schaffer Aguzzoli C, Ferreira PCL, Povala G, et al. Neuropsychiatric symptoms and microglial activation in patients with Alzheimer disease. JAMA Netw Open 2023 Nov 1;6(11):e2345175. [doi: 10.1001/jamanetworkopen.2023.45175] [Medline: 38010651]
19. Chi HC, Ma LZ, Wang ZB, et al. Associations of frailty with neuropsychiatric symptoms of Alzheimer's disease: a longitudinal study. J Alzheimers Dis 2024;98(2):629-642. [doi: 10.3233/JAD-231111] [Medline: 38427482]
20. Sánchez-Nieto D, Castaño-Castaño S, Navarro-Martos R, Obrero-Gaitán E, Cortés-Pérez I, Nieto-Escamez F. An intervention on anxiety symptoms in moderate Alzheimer's disease through virtual reality: a feasibility study and lessons learned. Int J Environ Res Public Health 2023 Feb 3;20(3):2727. [doi: 10.3390/ijerph20032727] [Medline: 36768091]
21. Cummings JL, Mega M, Gray K, Rosenberg-Thompson S, Carusi DA, Gornbein J. The Neuropsychiatric Inventory: comprehensive assessment of psychopathology in dementia. Neurology (ECronicon) 1994 Dec;44(12):2308-2314. [doi: 10.1212/wnl.44.12.2308] [Medline: 7991117]
22. Hébert R, Bravo G, Préville M. Reliability, validity and reference values of the Zarit Burden Interview for assessing informal caregivers of community-dwelling older persons with dementia. Can J Aging 2000;19(4):494-507. [doi: 10.1017/S0714980800012484]
23. Goodall G, Ciobanu I, Taraldsen K, et al. The use of virtual and immersive technology in creating personalized multisensory spaces for people living with dementia (SENSE-GARDEN): protocol for a multisite before-after trial. JMIR Res Protoc 2019 Aug 19;8(9):e14096. [doi: 10.2196/14096] [Medline: 31538942]
24. Cohen J. Statistical Power Analysis for the Behavioral Sciences: Routledge; 2013. [doi: 10.4324/9780203771587]

XSL•FO
RenderX

25. Appel L, Kisonas E, Appel E, et al. Administering virtual reality therapy to manage behavioral and psychological symptoms in patients with dementia admitted to an acute care hospital: results of a pilot study. JMIR Form Res 2021 Feb 3;5(2):e22406. [doi: 10.2196/22406] [Medline: 33533720]
26. Appel L, Appel E, Kisonas E, et al. evaluating the impact of virtual reality on the behavioral and psychological symptoms of dementia and quality of life of inpatients with dementia in acute care: randomized controlled trial (VRCT). J Med Internet Res 2024 Jan 30;26:e51758. [doi: 10.2196/51758] [Medline: 38289666]
27. Mendez MF, Joshi A, Jimenez E. Virtual reality for the assessment of frontotemporal dementia, a feasibility study. Disabil Rehabil Assist Technol 2015 Mar;10(2):160-164. [doi: 10.3109/17483107.2014.889230] [Medline: 24524440]
28. Park DC, Bischof GN. The aging mind: neuroplasticity in response to cognitive training. Dialogues Clin Neurosci 2013 Mar;15(1):109-119. [doi: 10.31887/DCNS.2013.15.1/dpark] [Medline: 23576894]
29. Cummings JJ, Bailenson JN. How immersive is enough? A meta-analysis of the effect of immersive technology on user presence. Media Psychol 2016 Apr 2;19(2):272-309. [doi: 10.1080/15213269.2015.1015740]
30. Baker S, Waycott J, Robertson E, et al. Evaluating the use of interactive virtual reality technology with older adults living in residential aged care. Information Processing & Management 2020 May;57(3):102105. [doi: 10.1016/j.ipm.2019.102105]
31. Yahara M, Niki K, Ueno K, et al. Remote reminiscence using immersive virtual reality may be efficacious for reducing anxiety in patients with mild cognitive impairment even in COVID-19 Pandemic: a case report. Biol Pharm Bull 2021;44(7):1019-1023. [doi: 10.1248/bpb.b21-00052] [Medline: 34193684]
32. Gaugler JE, Kane RL, Kane RA, Newcomer R. The longitudinal effects of early behavior problems in the dementia caregiving career. Psychol Aging 2005 Mar;20(1):100-116. [doi: 10.1037/0882-7974.20.1.100] [Medline: 15769217]
33. Gaugler JE, Davey A, Pearlin LI, Zarit SH. Modeling caregiver adaptation over time: the longitudinal impact of behavior problems. Psychol Aging 2000 Sep;15(3):437-450. [doi: 10.1037//0882-7974.15.3.437] [Medline: 11014707]
34. Cheng ST. Dementia caregiver burden: a research update and critical analysis. Curr Psychiatry Rep 2017 Aug 10;19(9):64. [doi: 10.1007/s11920-017-0818-2] [Medline: 28795386]
35. Dauphinot V, Delphin-Combe F, Mouchoux C, et al. Risk factors of caregiver burden among patients with Alzheimer's disease or related disorders: a cross-sectional study. J Alzheimers Dis 2015;44(3):907-916. [doi: 10.3233/JAD-142337] [Medline: 25374109]
36. Dettori JR. Loss to follow-up. Evid Based Spine Care J 2011 Feb;2(1):7-10. [doi: 10.1055/s-0030-1267080] [Medline: 22956930]
37. van der Lee J, Bakker TJEM, Duivenvoorden HJ, Dröes RM. Multivariate models of subjective caregiver burden in dementia: a systematic review. Ageing Res Rev 2014 May;15:76-93. [doi: 10.1016/j.arr.2014.03.003] [Medline: 24675045]

Abbreviations

BPSD: Behavioral and psychological symptoms of dementia
HMD: Head-mounted display
NPI-Q: Neuropsychiatric Inventory-Questionnaire
VR: Virtual reality
UDS: Uniform Data Set
ZBI: Zarit Caregiver Burden Interview

Edited by A Coristine; submitted 24.02.25; peer-reviewed by K Elechi, KW Tay, R Tripathi; revised version received 23.05.25; accepted 24.05.25; published 16.07.25.

Please cite as:
Huang LC, Chien CF, Yang YH
Exploring the Immediate and Long-Term Effects of Immersive Virtual Reality on Behavioral and Psychological Symptoms of Dementia and Caregiver Burden: Longitudinal Observational Study
JMIR Serious Games 2025;13:e73044
URL: https://games.jmir.org/2025/1/e73044
doi:10.2196/73044

© Ling-Chun Huang, Ching-Fang Chien, Yuan-Han Yang. Originally published in JMIR Serious Games (https://games.jmir.org), 16.7.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

A Virtual Reality Force Control Training System on Brain Activation: Functional Near-Infrared Spectroscopy (fNIRS) Study

Luigi Gan[1], MSc; Chien-Ju Lin[2], PhD; Hsiao-Feng Chieh[2], PhD; Kai-Nan An[2,3], PhD; Fong-Chin Su[1,2], PhD

[1]Department of Biomedical Engineering, National Cheng Kung University, No. 1, Dasyue Rd, East District, Tainan, Taiwan

[2]Medical Device Innovation Center, National Cheng Kung University, Tainan, Taiwan

[3]Department of Orthopedic Surgery, Mayo Clinic, Rochester, MN, United States

Corresponding Author:
Fong-Chin Su, PhD
Department of Biomedical Engineering, National Cheng Kung University, No. 1, Dasyue Rd, East District, Tainan, Taiwan

Abstract

Background: Aging can bring upon several effects that can hinder one's quality of life. One of the effects is the decline in one's ability to perform activities of daily living, which is caused by the loss of hand function due to aging. To mitigate this, several virtual reality (VR)-based training or rehabilitation systems that use hand tracking were developed. Although these systems are effective, immersive, and can promote motivation, they are mostly limited to providing range of motion exercises. The addition of a force control component to the hand tracking of these systems could make them even more effective at improving or restoring hand function, as the majority of activities of daily living require a degree of force control.

Objective: This study aimed to compare the effects of 2 VR input systems: regular hand tracking and the novel VR input system in this study, which incorporate force control to regular hand tracking on the brain activity of younger and older adults. The degree of cortical activity during a training or rehabilitation task is linked to better functional outcomes and improvements of neuroplasticity.

Methods: Twelve younger adults (mean age 25.00, SD 4.50 years) and 12 older adults (mean age 73.00, SD 3.6 years) were recruited to play a game specifically developed for this study using 2 VR input systems. Brain activity during gameplay was recorded using functional near-infrared spectroscopy over the following cortical regions: prefrontal cortex (PFC), premotor cortex (PMC), supplementary motor area (SMA), and primary motor cortex (M1).

Results: Compared with the regular hand-tracking system, adding a force control component increased average oxygenated hemoglobin (HbO) concentrations and decreased deoxygenated hemoglobin (HbR) concentrations in key brain regions. In young adults, these changes were observed in the right PMC and right M1. In older adults, higher HbO and lower HbR concentrations appeared in the right PFC, bilateral M1, and right SMA (HbR only). The force control component also led to more widespread activity across all ROIs.

Conclusions: The novel input system in this study can be used for improving or restoring hand function. The results of this study can be used as a reference for the development of better VR-based training or rehabilitation systems.

Trial Registration: ClinicalTrials.gov NCT06412887; https://clinicaltrials.gov/study/NCT06412887

(JMIR Serious Games 2025;13:e63874) doi:10.2196/63874

KEYWORDS

pinch force control; virtual reality; VR; functional near-infrared spectroscopy; aging; rehabilitation; exercises; exercise; elder; older adults; geriatric; neural plasticity; prefrontal cortex; PFC; premotor cortex; PMC; motor

Introduction

Aging is a significant global issue, with the World Health Organization predicting that by the year 2030, 1 in 6 people will be aged 60 years or older [1]. These lead to impaired hand function, which can decrease one's level of independence and ability to perform activities of daily living (ADLs).

Nowadays, the use of virtual reality (VR) as a platform for rehabilitation and training is becoming increasingly prevalent, as these systems can be effective [2-4] and are better at increasing one's motivation [5] to undergo training or rehabilitation than conventional methods. This may be due in part to the platform's ability to realistically simulate ADLs, reduce boredom and monotony compared with conventional methods, and provide user-adaptable difficulty levels [6,7]. Some notable examples of VR-based rehabilitation system that incorporate the VR features are the Amadeo hand rehabilitation device (Tyromotion GmbH) and the lab custom Tipr device. The Amadeo device is an interactive robotic device for hand

XSL•FO
RenderX

function assessment, range of motion exercises, and force control rehabilitation. The Tipr, on the other hand, is a tabletop device specifically designed for older adult individuals and patients with mild cognitive impairment to engage in game-based interventions, using the forces from each finger as input. Trials for both devices show effectiveness in improving cognition, fine motor skills, and manual dexterity in users [8-11]. However, due to their tabletop design, such devices are limited to exercises where the hands remain fixed in a certain position during rehabilitation.

One of the advantages of using VR as a platform for providing rehabilitation and training is its flexibility in interactions methods, input systems, and feedback mechanisms, all of which can affect rehabilitation outcomes [12]. Among the various input systems for VR rehabilitation, hand tracking—where user's hands are captured and displayed in the VR environment via cameras, allowing interaction through grasps, pokes, and so forth—is preferred by users with motor difficulties [13]. However, hand tracking alone limits the system to range of motion exercises. Adding a force control component could enhance these systems by making VR interactions more lifelike and similar to ADLs.

The opposable thumb, along with its ability to touch the other 4 fingers, makes human hands unique. Tasks involving the thumb and forefinger opposition, which accounts for a large portion of ADLs (dexterous manipulation of small objects, feeding, grooming, etc), are often accomplished by the precise control of pinch force of human fingers.

One method for evaluating the effects of different input systems, interaction methods, and game modes on upper extremity function [12,14,15], involves measuring brain activity using functional near-infrared spectroscopy (fNIRS). fNIRS is a noninvasive brain-imaging tool that assesses cortical activity by detecting changes in blood oxygen concentration in response to a stimulus. Higher brain activity during training or rehabilitation as indicated by changes in oxygenated (HbO) and deoxygenated hemoglobin (HbR) is associated with better functional outcomes and the promotion of neural plasticity [16-18].

This study aims to verify the effects of adding a force control component to a standard hand-tracking system on hand function recovery and improvement in younger adult and older adult populations. This is achieved by comparing the brain activity between the standard hand-tracking system (without force control [woFC]) and the novel hand-tracking system with force control (wFC) while playing a game developed for this study, with the hypothesis that playing the game wFC would elicit greater brain activity. The game simulates a feeding task, where the user must use a VR fork to pick up and deliver food to a baby. A more detailed explanation of the game is provided in the following sections.

Methods

Study Design and Setting

This study is a prospective, quantitative, and nonrandomized observational study examining the effects of 2 VR input systems (wFC and woFC) on brain activation in both younger adults and older adults. The research was conducted at the Motion Analysis Laboratory and the Medical Device Innovation Center of National Cheng Kung University (NCKU) in Tainan, Taiwan, from July 10, 2023, to August 11, 2023. The reporting adheres to the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines for observational research [19].

Subject Recruitment

Participants were recruited through posters displayed around the NCKU campuses and local Line groups. The eligibility criteria included being at least 21 years of age, not having conditions that would prevent normal use of the VR system (such as finger or hand amputations), and not experiencing motion sickness when using the VR system.

Ethical Considerations

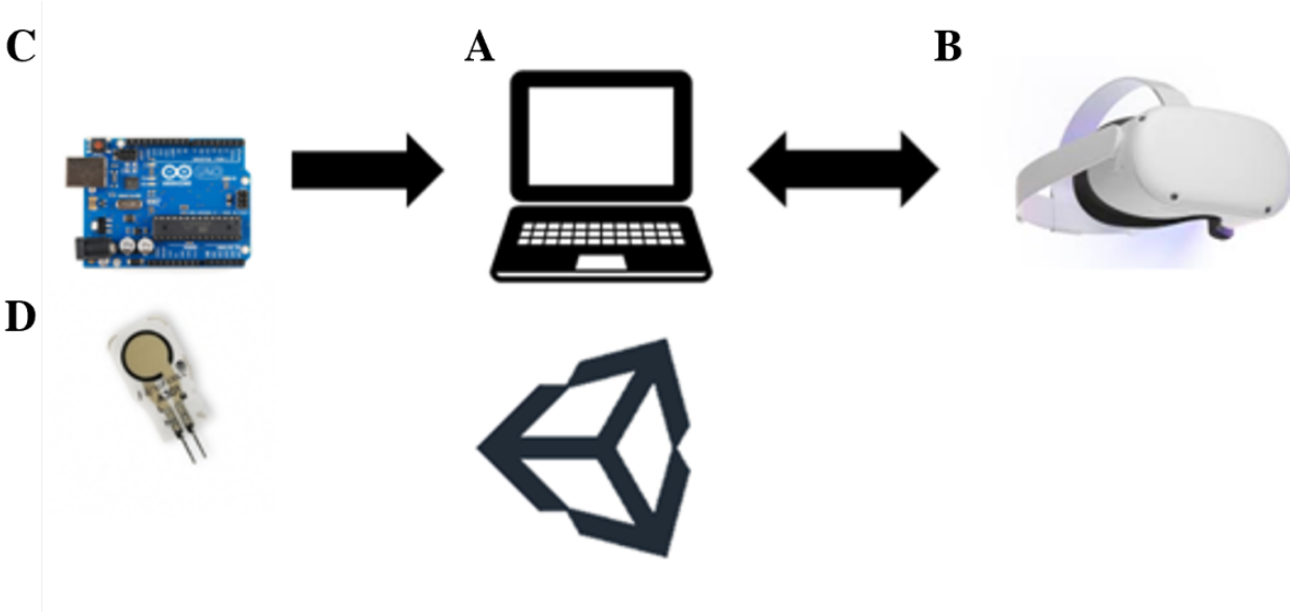
The study was approved by the Human Research Ethics Review Committee of NCKU (approval NCKU HREC-E-112-181-2) and was also registered in ClinicalTrials.gov (NCT06412887). Informed consent was obtained from all participants before data collection began, including consent for the primary or secondary analysis and publication of research data. All data were kept anonymous. Participants received a compensation of US $18 upon completing the experiment.

VR Force Control Training System—Hardware

The hardware part of the VR Force Control Training System consists of 4 main components. A laptop computer running Unity (Unity Technologies) is labeled A in Figure 1. A head-mounted display, specifically the Meta Quest 2 (Meta Platforms Inc), is labeled B in Figure 1 and is used to display the game. A force-sensing system, built around a Flexiforce A301 force-sensitive resistor (FSR) from Tekscan, Inc, is labeled D in Figure 1. This FSR was calibrated using test weights ranging from 300 g to 7 kg with 5 trials per weight, each lasting 10 seconds. An Arduino Uno, labeled C in Figure 1, converts and sends the FSR readings to the laptop computer running Unity. Figure 1 illustrates the VR force control training system used in this study. The system's stability was previously evaluated by measuring 3 different test weights over 3 days at 3-hour time intervals. The resulting intraclass correlation coefficient over 3 days was 0.75.

XSL•FO
RenderX

Figure 1. Hardware components of the virtual reality (VR) force control training system. (A) A laptop computer for running the VR game. (B) Meta Quest 2 head-mounted display for displaying the game and capturing the position and poses of the user's hands. (C) Arduino Uno for detecting, converting, and sending the changes in the resistance values of the force-sensitive resistor (FSR) to the laptop computer running Unity. (D) FSR for capturing and measuring changes in the pressure and force exerted between the thumb and the index finger.



VR Force Control Training System—Software

As mentioned in the previous sections, the game consists of 2 tasks that can be played using 2 input systems, both based on the hand-tracking features provided by the Oculus Interaction SDK. The first task requires users to memorize a 5-element list containing the order of the food to be fed to the baby. The second task involves picking up an interactive fork, selecting the appropriate food, and feeding it to the baby. In the woFC system, the user simply performs a thumb-index finger opposition to pick up the fork. In contrast, the wFC system is designed with the same thumb-index finger opposition in woFC system but with an additional isometric pinch contraction at 70 (SD 10%) maximum voluntary contraction (MVC), a value derived from a prior pilot study.

Once the fork is picked up in the wFC system, the user's pinch force is continuously monitored and displayed on a force bar and provides real-time feedback as depicted in Figure 2. Exceeding or falling below the allowable pinch force range causes the fork to reset, requiring the users to initiate the grasp again. These modifications in the wFC system introduce an added layer of difficulty immersion compared with the woFC system. To achieve this functionality, custom code was written to modify the hand-tracking features of the Oculus Interaction SDK. Differences between wFC and woFC systems are summarized by Figure 3.

XSL•FO
RenderX

Figure 2. A user maintaining the pinch force at the allowable range, represented by the green area of the force bar, going above or below the allowable range displays the red and white areas, respectively.



XSL•FO
RenderX

Figure 3. Flowchart illustrating the differences between without force control (left) and with force control (right).

XSL•FO
RenderX

Experiment Setup and Procedures

Brain activity was measured using NIRScout system (NIRx Medical Technologies). The setup consisted of 8 sources and 16 detector optodes, spaced 3 cm apart, forming 30 channels. These were arranged according to the international 10 - 20 system Figure 4 to cover the regions of interest (ROIs) involved with hand movements, specifically the prefrontal cortex (PFC), premotor cortex (PMC), supplementary motor area (SMA), and primary motor cortex (M1). Data were collected over 2 wavelengths (760 nm and 850 nm) with a sampling rate of 7.81 Hz. The experiment followed a block-related design as illustrated in Figure 5. A trial in this study consisted of a single gameplay loop, followed by a rest event, each lasting 30 seconds. Measurements were taken during 5 consecutive trials using the gameplay loop wFC, followed by a rest event and then 5 consecutive trials using the gameplay loop woFC. During each gameplay loop the participants played the game using the assigned input system. In contrast, during the rest events, they were asked to close their eyes and enter a resting state.

Figure 4. Arrangement of the source and detectors optodes for this study. Source and detector optodes are represented by red and blue dots, respectively. Regions of interest are formed from multiple color-coded channels with the PFC, PMC, SMA, and M1 formed by white, yellow, green, and violet channels respectively. M1: primary motor cortex; PFC: prefrontal cortex; PMC: premotor cortex; SMA: supplementary motor area.

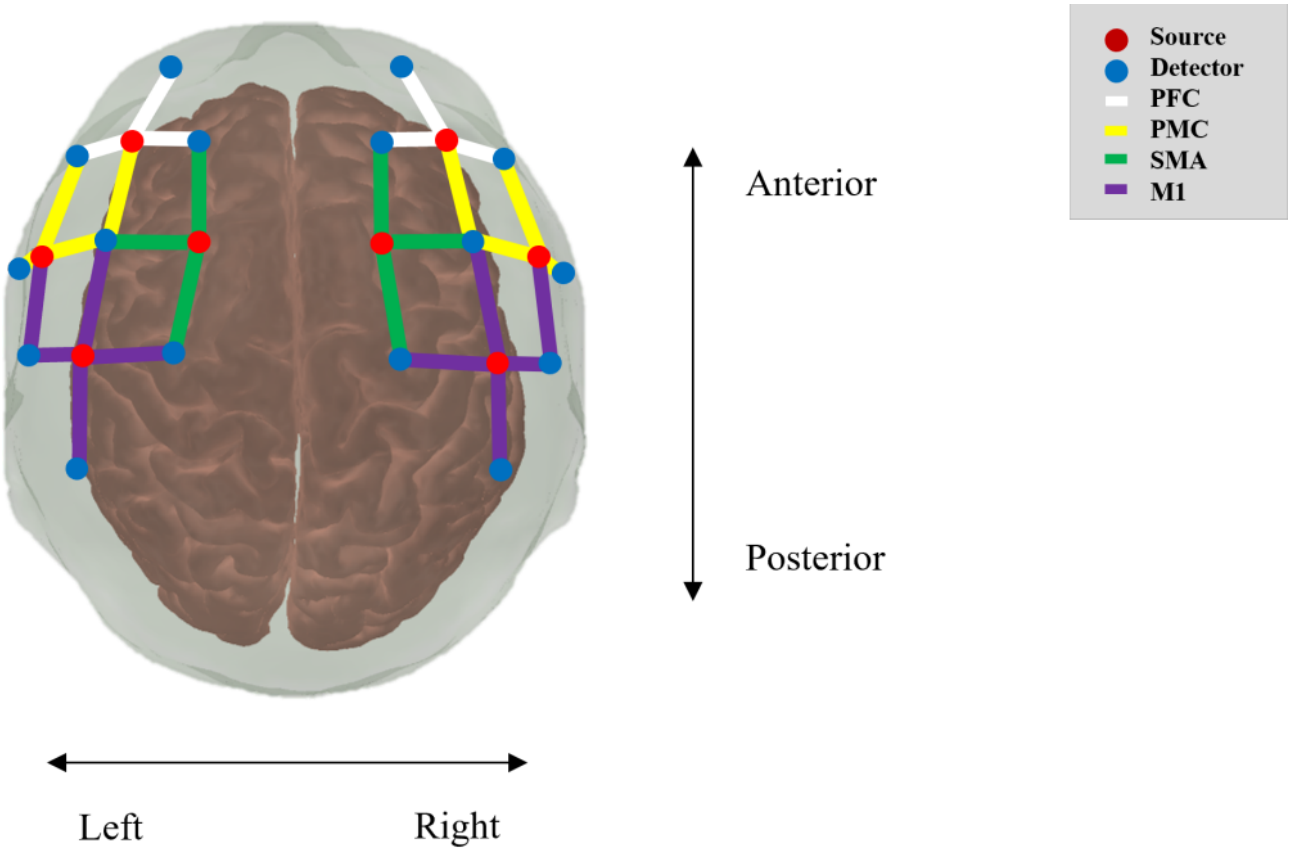
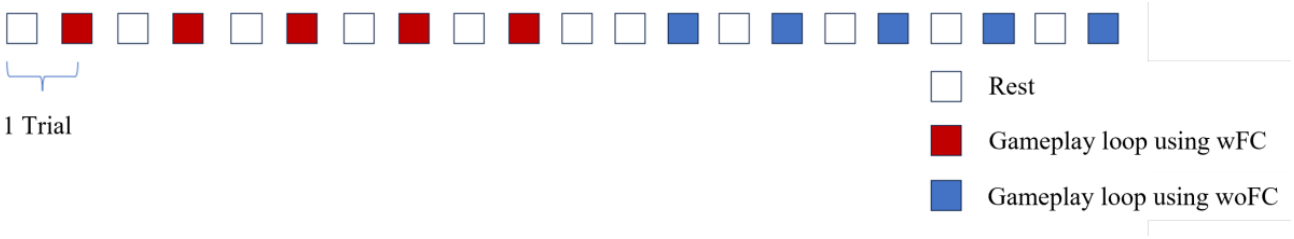


Figure 5. The event-related paradigm used in this study. Rest, gameplay loop using wFC, and gameplay loop using woFC are represented by white, red, and blue squares, respectively. wFC: with force control; woFC: without force control.



The experiment procedures were as follows: First, participants entered an isolated room where they were briefed about the experiment and provided informed consent. Second, participants familiarized themselves with both input systems of the game, and the in-game maximum pinch voluntary contraction was recorded to calculate each user's force threshold. Third, the fNIRS system was calibrated using the NIRStar software (NIRx Medical Technologies) to ensure acceptable signal quality across all channels. Finally, data collection proceeded according to the block-related design outlined in Figure 5, while Figure 6 shows a participant performing the experiment.

XSL•FO
RenderX

Figure 6. A participant playing the game using the wFC (with force control).



Data Processing and Analysis

Game performance was quantified using 3 parameters: the amount of grasp initiated, pinch force coefficient of variation (COV), and memory task accuracy. As the amount of grasp initiated, a unitless metric was derived by counting the number of times the participants needed to pick up the fork in each gameplay loop, reflecting the number of times the participants'

XSL•FO
RenderX

pinch force exceeded or fell below the threshold. The pinch force COV was calculated by dividing the SD of the pinch force by the average pinch force during each gameplay loop, resulting in a unit of Newton/Newton. Memory task accuracy was calculated by dividing the number of correct foods fed by the total number of foods fed in each gameplay loop.

Raw fNIRS signals were processed using MATLAB R2023b (Mathworks) running the open-source app Homer3 (version 1.80.2) [20]. The signals were preprocessed manually and by using functions inside Homer3. These include visually inspecting the raw signals for noise and discontinuities, channel pruning, conversion from intensity to optical density, motion corrections, bandpass filtering, and conversion from optical density to concentration. Channel pruning was performed to eliminate signals that were too weak, too strong, or had a high SD. Motion correction was done using the Wavelet-based algorithms [21]. Bandpass filter was set at 0.01 Hz and 0.1 Hz was applied to eliminate the effects of heartbeat, breathing, and other Mayer waves [22]. The optical densities were then converted into concentration using the modified Beer-Lambert Law with the default partial path length factor of 1.

The hemodynamic response function (HRF) of HbO and HbR was estimated with general linear modeling using the ordinary least squares method [23]. The HRF was modeled by a consecutive sequence of Gaussian functions, with SD of 1 and a temporal spacing of 1 second between the mean over a period of −5 to 35 seconds [24]. The baseline period was set to 5 seconds before stimulus onset, and its mean was subtracted from the whole HRF (−5 seconds to 35 seconds) to perform baseline normalization. The mean HbO and HbR concentration for each ROI was calculated by averaging across the channels in each ROI from the period of 10 seconds to 35 seconds owing to the delay in movement-induced hemodynamic response [22]. The mean HbO and HbR concentration was then projected to the cortical surface of the Colin Atlas by registering the optode and channel locations in AtlasViewer (version 2.44) [25], a MATLAB toolbox for the anatomical interpretation of fNIRS data processed in Homer3.

The Pearson correlation between each ROI's processed HbO and HbR data was calculated for each group, resulting in four 8×8 matrices representing 4 full network connections. For a more comprehensive evaluation of the network, 2 features were extracted for statistical analysis: connection ratio (CR) and connection strength (CS) [26]. To obtain CR and CS, the Pearson correlation coefficient of each matrix was first converted to normal distribution through the Fisher *z* transformation [27]. CR was defined as the ratio between the significant connections and the total number of connections in each network, while CS was defined as the average of absolute *z* values of a network. A connection was considered significant when its absolute *z* value was greater than 0.7.

The Mann-Whitney *U* test was used for comparing the game performance between younger adult and older adult populations. The Wilcoxon signed rank test was used to assess the statistical significance of differences in HbO and HbR responses, CS, and CR between the 2 input systems. The Spearman correlation was used to identify correlations between game scores and HbO and HbR responses. The level of significance was set at .05. For data involving multiple comparisons, the Benjamini-Hochberg procedure [28] was used to control the false discovery rate with a Benjamini-Hochberg critical value of 0.05.

Results

This section aims to present the main results of this study, namely, game performance, mean HbO concentration, temporal characteristics of HRF, and the cortical activation maps.

Participants

The study included 12 eligible younger adults (mean age 25.00, SD 4.5 years) and 12 eligible older adults (mean age 73.00, SD 3.6 years). Of the participants, 23 were right-handed; the dominant hand of 1 participant was not specified.

Game Performance

Table 1 shows the average game scores per run and its comparison between both groups. Only the pinch force COV showed a statistically significant difference.

Table . Game performance of both groups.

	Younger adults	Older adults	*P* value
Amount of grasp initiated[a], mean (SD)	5.02 (3.13)	5.26 (2.79)	.67
Pinch force coefficient of variation[b], mean (SD)	1.23 (0.40)	1.6 (0.98)	.03
Memory task accuracy[c], mean (SD)	0.83 (0.12)	0.72 (0.23)	.11

[a]Amount of grasp initiated is a unitless metric.

[b]Pinch force coefficient of variation is in the unit of Newton/Newton.

[c]Memory task accuracy is a unitless metric.

Mean HbO Concentration

The mean HbO concentration for the younger adults and the older adults during gameplay is shown in Tables 2 and 3, respectively. Some ROIs show significantly higher HbO concentration when playing the game using the wFC, namely, the right premotor cortex (rPMC); right primary motor cortex (rM1) for the younger adults; and left primary motor cortex (lM1), right prefrontal cortex (rPFC), and rM1 for the older adults.

XSL•FO
RenderX

Table . Mean values of oxygenated hemoglobin concentration for the younger adult group under while playing the game both with force control and without force control input systems[a].

Younger adult group	lPFC[b]	lPMC[c]	lSMA[d]	lM1[e]	rPFC[f]	rPMC[g]	rSMA[h]	rM1[i]
wFC[j] HbO[k] mean	32.88 (9.49)	70.83 (20.45)	39.51 (15.49)	53.66 (15.49)	41.73 (12.05)	81.14 (23.42)	16.62 (4.80)	35.15 (10.15)
woFC[l] HbO mean	23.18 (6.69)	33.66 (9.72)	26.82 (7.74)	26.97 (7.78)	25.67 (7.41)	16.60 (4.79)	1.48 (0.43)	−0.44 (0.13)
P value	.583	.034	.272	.050	.182	.003	.084	.003
Adjusted *P* value	.583	.907	.312	.100	.243	.012	.134	.012

[a]Data are presented in mean (SE) of the mean. The prefixes l and r stand for left and right, respectively. Units are in μM-mm

[b]lPFC: left prefrontal cortex.

[c]lPMC: left premotor cortex.

[d]lSMA: left supplementary motor area.

[e]lM1: left primary motor cortex.

[f]rPFC: right prefrontal cortex.

[g]rPMC: right premotor cortex.

[h]rSMA: right supplementary motor area.

[i]rM1: right primary motor cortex.

[j]wFC: with force control.

[k]HbO: oxygenated hemoglobin.

[l]woFC: without force control.

Table . Mean values of oxygenated hemoglobin concentration for the older adult group while playing the game under both with force control and without force control input systems[a].

Older adult group	lPFC[b]	lPMC[c]	lSMA[d]	lM1[e]	rPFC[f]	rPMC[g]	rSMA[h]	rM1[i]
wFC[j] HbO[k] mean	38.58 (11.14)	41.67 (12.03)	51.99 (15.01)	62.63 (18.08)	37.89 (10.94)	45.19 (13.05)	58.81 (16.98)	59.56 (17.19)
woFC[l] HbO mean	25.17 (7.27)	33.48 (9.66)	48.35 (13.96)	42.37 (12.23)	16.51 (4.77)	30.13 (8.70)	35.92 (10.37)	37.84 (10.92)
P value	.158	.875	.583	.015	.015	.272	.041	.012
Adjusted *P* value	.253	.875	.666	.040	.040	.363	.082	.040

[a]Data are presented in mean (SE) of the mean. The prefixes l and r stand for left and right, respectively. Units are in μM-mm.

[b]lPFC: left prefrontal cortex.

[c]lPMC: left premotor cortex.

[d]lSMA: left supplementary motor area.

[e]lM1: left primary motor cortex.

[f]rPFC: right prefrontal cortex.

[g]rPMC: right premotor cortex.

[h]rSMA: right supplementary motor area.

[i]rM1: right primary motor cortex.

[j]wFC: with force control.

[k]HbO: oxygenated hemoglobin.

[l]woFC: without force control.

Mean HbR Concentration

The mean HbR concentration for the younger adults and the older adults during gameplay is shown in Tables 4 and 5, respectively. Some ROIs show significantly lower HbR concentration when playing the game using the wFC, namely, the rPMC; rM1 for the younger adults; and lM1, rPFC, right supplementary motor area (rSMA), and rM1 for the older adults.

XSL•FO
RenderX

Table . Mean values of deoxygenated hemoglobin concentration for the younger adult group while playing the game under both with force control and without force control input systems[a].

	lPFC[b]	lPMC[c]	lSMA[d]	lM1[e]	rPFC[f]	rPMC[g]	rSMA[h]	rM1[i]
wFC[j] HbR[k] mean	−19.85 (5.73)	−42.72 (12.33)	−29.06 (8.39)	−29.77 (8.59)	−18.05 (5.21)	−51.11 (14.75)	−4.77 (1.38)	−18.36 (5.30)
woFC[l] HbR mean	−9.73 (2.81)	−20.06 (5.79)	−20.23 (5.84)	−14.45 (4.17)	−10.42 (3.01)	−12.83 (3.70)	4.19 (1.21)	−0.96 (0.28)
P value	.239	.023	.308	.060	.182	.003	.117	.004
Adjusted *P* value	.273	.062	.308	.120	.243	.016	.187	.016

[a]Data are presented in mean (SE) of the mean. The prefixes l and r stand for left and right, respectively. Units are in μM-mm.
[b]lPFC: left prefrontal cortex.
[c]lPMC: left premotor cortex.
[d]lSMA: left supplementary motor area.
[e]lM1: left primary motor cortex.
[f]rPFC: right prefrontal cortex.
[g]rPMC:right premotor cortex.
[h]rSMA: right supplementary motor area.
[i]rM1: right primary motor cortex.
[j]wFC: with force control.
[k]HbR: deoxygenated hemoglobin.
[l]woFC: without force control.

Table . Mean values of deoxygenated hemoglobin concentration for the older adult group while playing the game under both with force control and without force control input systems[a].

	lPFC[b]	lPMC[c]	lSMA[d]	lM1[e]	rPFC[f]	rPMC[g]	rSMA[h]	rM1[i]
wFC[j] HbR[k] mean	−14.53 (4.19)	−21.96 (6.34)	−19.83 (5.72)	−21.83 (6.30)	−12.60 (3.64)	−23.07 (6.66)	−17.68 (5.10)	−19.10 (5.51)
woFC[l] HbR mean	−8.84 (2.55)	−16.17 (4.67)	−19.78 (5.71)	−14.52 (4.19)	−4.03 (1.16)	−13.56 (3.92)	−9.16 (2.65)	−12.12 (3.50)
P value	.158	.638	.875	.023	.015	.308	.010	.008
Adjusted *P* value	.253	.729	.875	.046	.040	.411	.040	.040

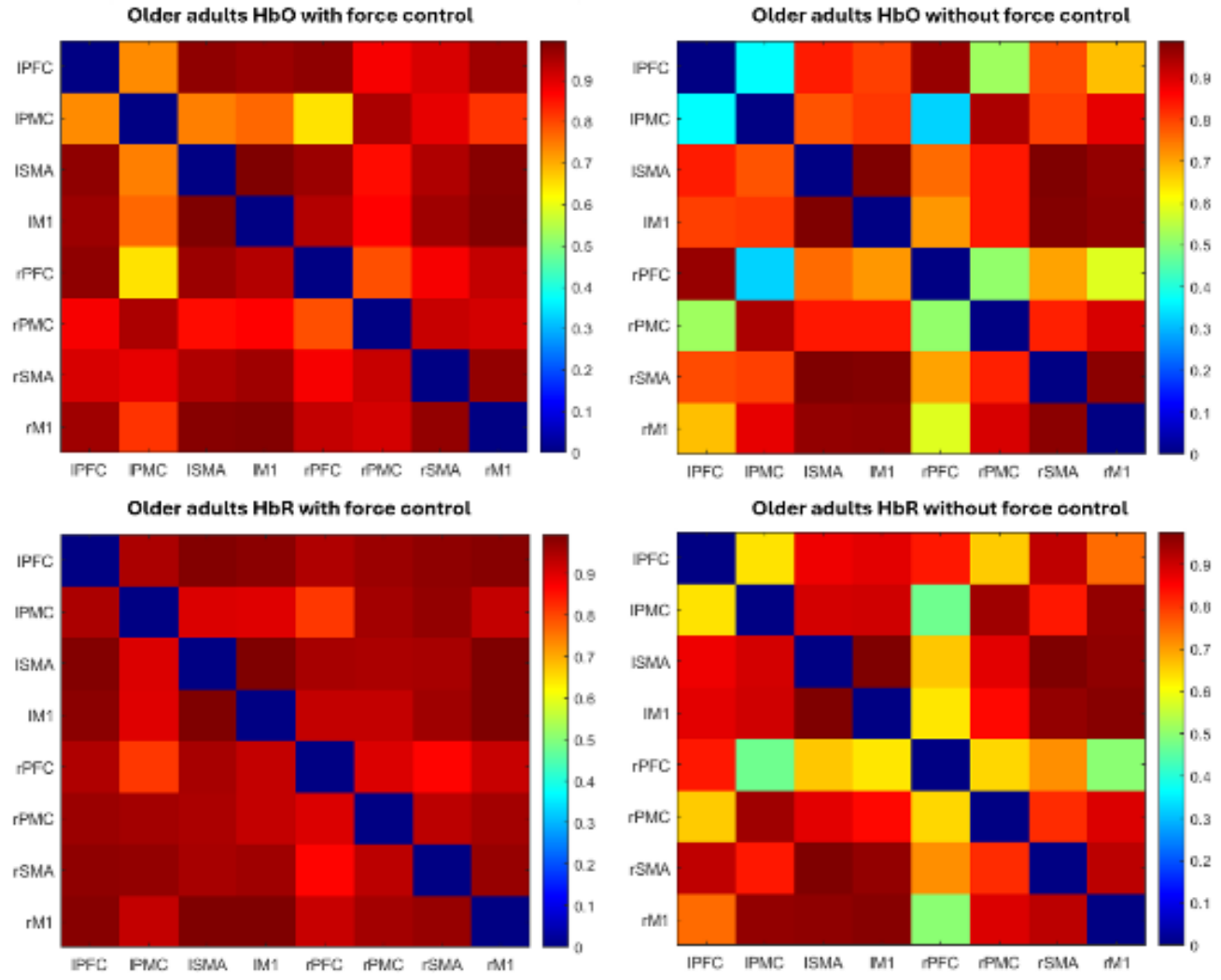
[a]Data are presented in mean (SE) of the mean. The prefixes l and r stand for left and right, respectively. Units are in μM-mm.
[b]lPFC: left prefrontal cortex.
[c]lPMC: left premotor cortex.
[d]lSMA: left supplementary motor area.
[e]lM1: left primary motor cortex.
[f]rPFC: right prefrontal cortex.
[g]rPMC: right premotor cortex.
[h]rSMA: right supplementary motor area.
[i]rM1: right primary motor cortex.
[j]wFC: with force control.
[k]HbR: deoxygenated hemoglobin.
[l]woFC: without force control.

Correlation Matrices Between ROIs

The correlation matrices between the HbO and HbR data of each ROI for both groups, while playing the game with and without force control (wFC and woFC), are presented in Figures 7 and 8, respectively. ROIs with higher correlations appear in reddish colors, while ROIs with less correlations appear in bluish colors.

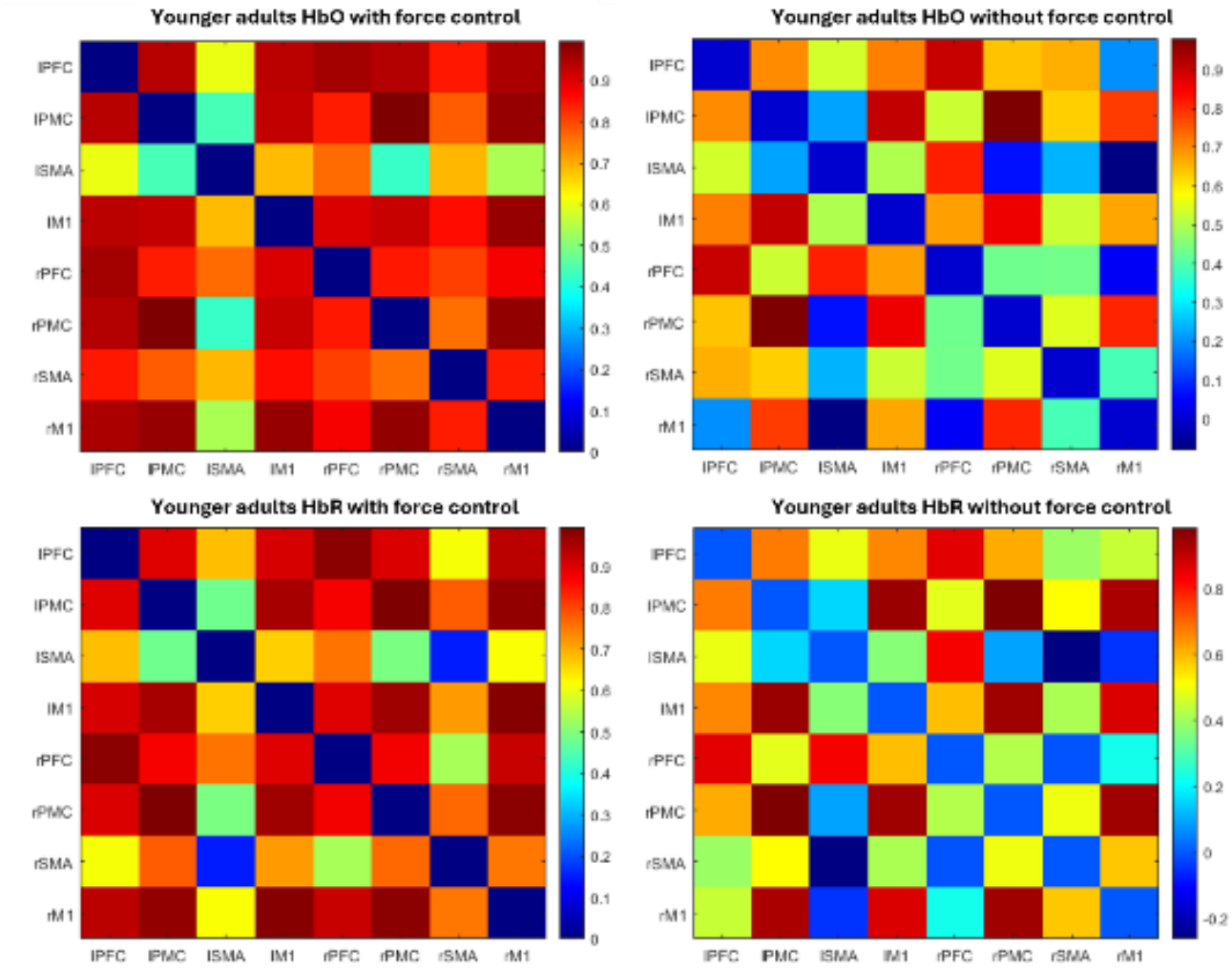
XSL•FO
RenderX

Figure 7. HbO (top) and HbR (bottom) correlation matrices of the older adults during with force control (left) and without force control (right) gameplay. HbO: oxygenated hemoglobin; HbR: deoxygenated hemoglobin; IM1: left primary motor cortex; IPFC: left prefrontal cortex; IPMC: left premotor cortex; ISMA: left supplementary motor area; rM1: right primary motor cortex; rPFC: right prefrontal cortex.rPMC: right premotor cortex.rSMA: right supplementary motor area.



XSL•FO
RenderX

Figure 8. HbO (top) and HbR (bottom) correlation matrices of the younger adults during with force control (left) and without force control (right) gameplay. HbO: oxygenated hemoglobin; HbR: deoxygenated hemoglobin; IM1: left primary motor cortex; IPFC: left prefrontal cortex; IPMC: left premotor cortex; ISMA: left supplementary motor area; rM1: right primary motor cortex; rPFC: right prefrontal cortex.rPMC: right premotor cortex.rSMA: right supplementary motor area.



Connection Strength and Connection Ratio

The CS and CR for the younger adults and the older adults are shown in Table 6. The older adult group exhibited higher CS and CR when playing the game using wFC as compared with woFC.

Table . Network features of both groups while playing the game under with force control and without force control input systems[a].

	Older adult		Younger adult	
	Connection strength	Connection ratio	Connection strength	Connection ratio
wFC[b]	1.15 (0.33)	0.76 (0.22)	0.79 (0.23)	0.48 (0.14)
woFC[c]	0.94 (0.27)	0.62 (0.18)	0.69 (0.20)	0.43 (0.13)
P value	.010	.018	.117	.373

[a]Data are presented in mean (SE) of the mean.

[b]wFC: with force control.

[c]woFC: without force control.

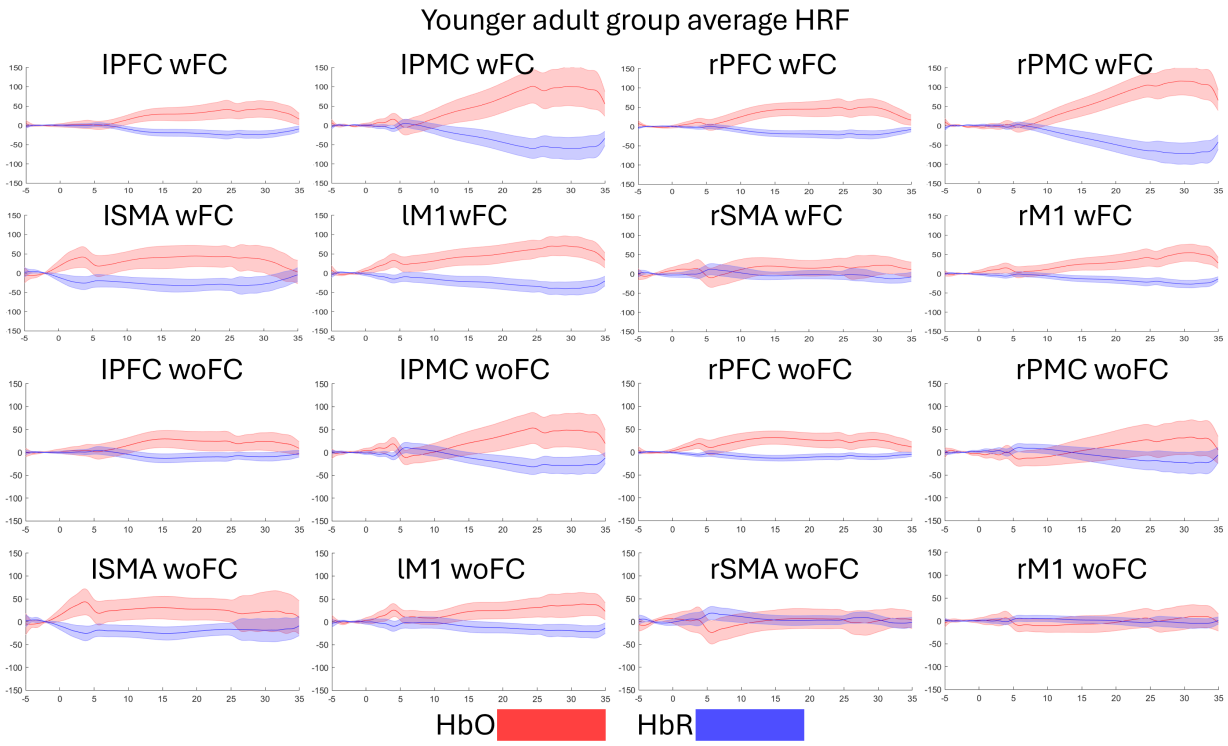
Temporal Characteristics of HRF

Figures 9 and 10 present the temporal characteristics of the HRF (HbO and HbR) of both groups during gameplay. In Figure 9, the HbO temporal characteristics of the PFC, PMC, and M1 during wFC follow a rise after task onset that slowly decreases as the task progresses. The HbR follows the opposite pattern. As for the SMA, the left side, regardless of input system, shows a spike in the HbO and the HbR, followed by a constantly elevated response, while the HbO and the HbR of the right side stay at around baseline levels. Regarding the M1 during woFC, the HbO and the HbR on both the left and right sides show little

XSL•FO
RenderX

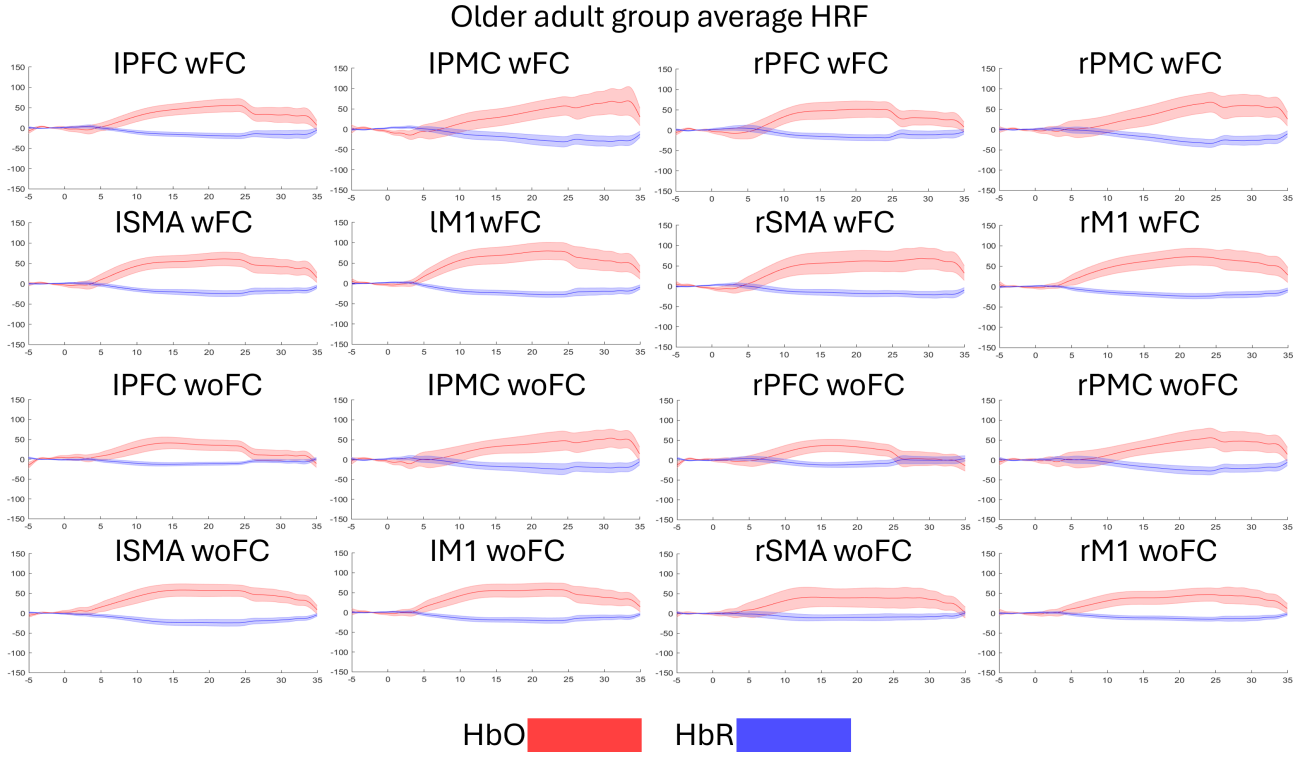
to no response. For the older adult group, Figure 10 shows that all ROIs follow an increase in HbO after task onset, which slowly decreases as the task progresses, while the HbR follows an opposite pattern.

Figure 9. Temporal characteristics of the HRF of the younger adult group. Red represents the HbO, while blue represents the HbR. The shaded regions are the SE of mean. The figure is divided into the HRF between the left (first to second columns) and right (third to fourth columns) hemispheres and between wFC (first to second rows) and woFC (third to fourth rows). The *x*-axis represents time in seconds, while the *y*-axis represents HRF concentration in μM-mm. HbO: oxygenated hemoglobin; HbR: deoxygenated hemoglobin; HRF: hemodynamic response function; lM1: left primary motor cortex.; lPFC: left prefrontal cortex; lPMC: left premotor cortex; lSMA: left supplementary motor area; PFC: prefrontal cortex; rm1: right primary motor cortex; rPMC: right premotor cortex; rSMA: right supplementary motor area; wFC: with force control; woFC: without force control.



XSL•FO
RenderX

Figure 10. Temporal characteristics of the HRF of the older adult group. Red represents the HbO, while blue represents the HbR. The shaded regions are the SE of mean. The figure is divided into the HRF between the left (first to second columns) and right (third to fourth columns) hemispheres and between wFC (first to second rows) and woFC (third to fourth rows). The *x*-axis represents time in seconds, while the *y*-axis represents HRF concentration in µM-mm.HbO: oxygenated hemoglobin; HbR: deoxygenated hemoglobin; HRF: hemodynamic response function; lM1: left primary motor cortex.; lPFC: left prefrontal cortex; lPMC: left premotor cortex; lSMA: left supplementary motor area; PFC: prefrontal cortex; rm1: right primary motor cortex; rPMC: right premotor cortex; rSMA: right supplementary motor area; wFC: with force control; woFC: without force control.



Cortical Activation Maps

Figures 11 and 12 show the cortical activation maps during gameplay under both input systems for the younger adult and older adult groups, respectively. The red color signifies increases in HbO while blue signifies decreases in HbO. Based on both figures, the wFC elicits a larger area of activation than the woFC, irrespective of age groups. Age-related differences occur in the laterality of activation, with the older adults experiencing a more bilateral activation across 2 input systems.

XSL•FO
RenderX

Figure 11. Cortical activation maps of the younger adult group during gameplay using both input systems. Top 3 images are the cortical activation maps during gameplay using with force control, while the bottom 3 images are the cortical activation maps during gameplay using without force control.

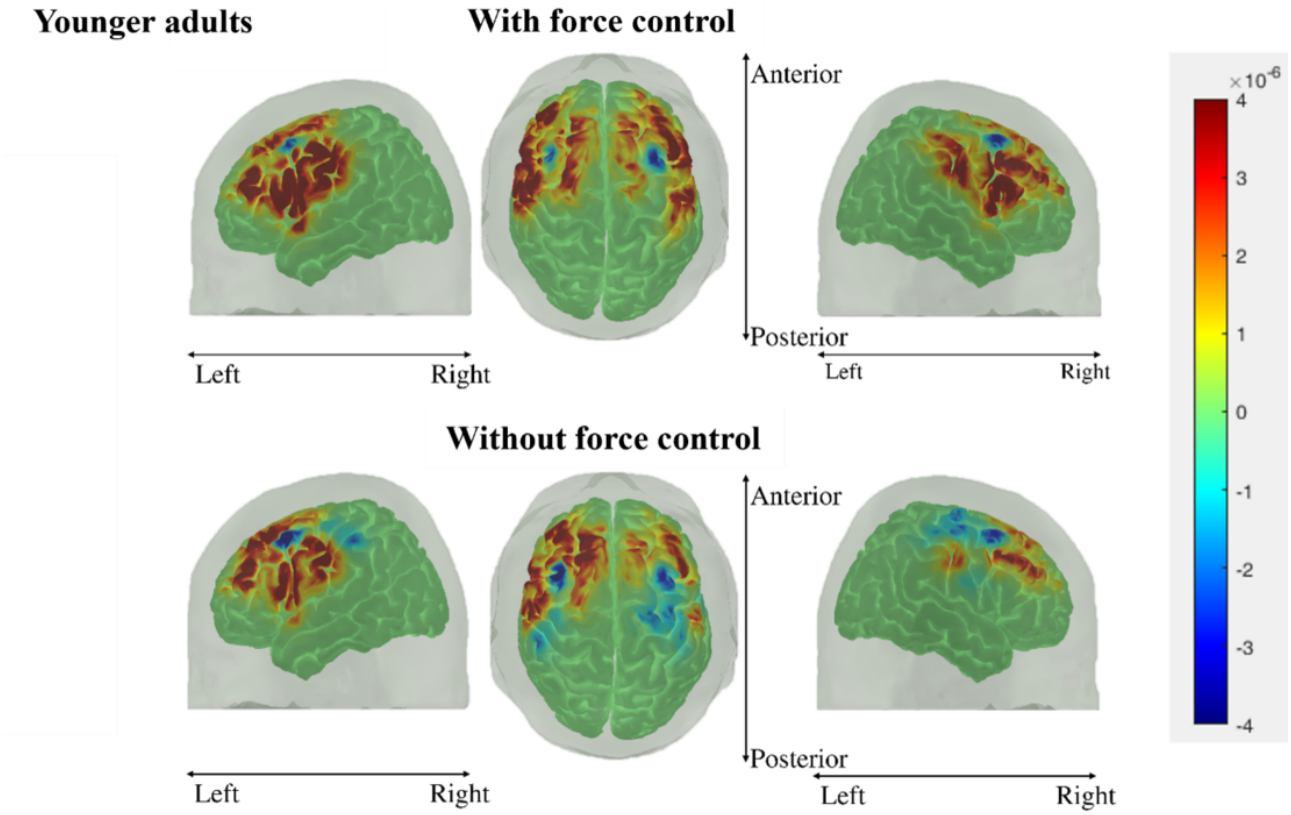
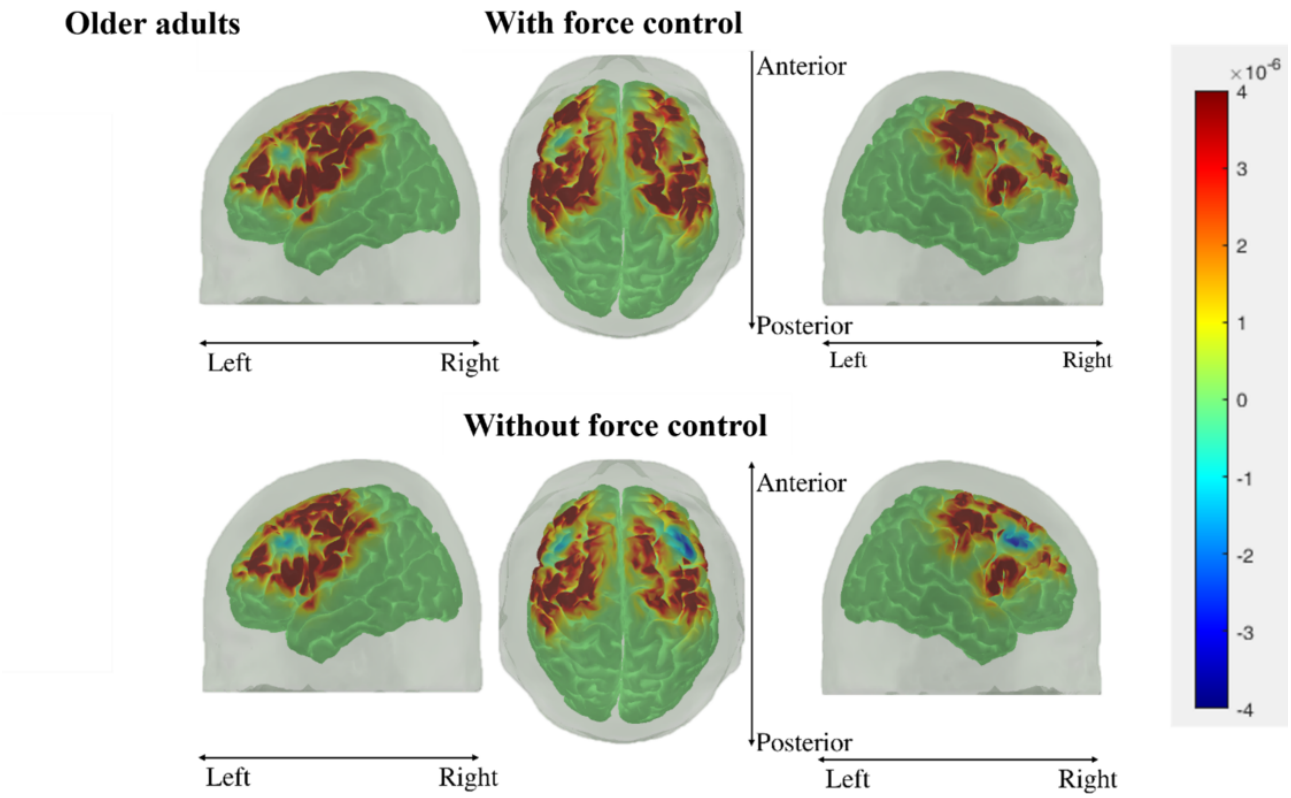


Figure 12. Cortical activation maps of the older adult group during gameplay using both input systems. Top 3 images are the cortical activation maps during gameplay using with force control, while the bottom 3 images are the cortical activation maps during gameplay using without force control.



XSL•FO
RenderX

Discussion

Principal Findings

This study investigated the effects of integrating force control into a VR hand-tracking system on brain activity in the younger adults and the older adults. Participants played the same game using 2 different control systems (wFC and woFC). Despite studies showing that there is a clear difference in game performance between younger adults and older adults, whether it be commercial games [29], exergames [30,31], cognitive games [32], and so forth, the results of this study did not (amount of grasp initiated and memory task accuracy). In the aforementioned studies, aging-related changes and the level of difficulty were identified as the main factors for the differences in game performance. As for this study, the difficulty influenced by the force threshold and number of elements in the food list could not have been challenging enough, which resulted in similar performance across age groups.

Comparison to Prior Work

The findings of this study align with previous fNIRS research [12,33] on finger opposition movement as a primary input and the effects of different interaction modes in both immersive and nonimmersive VR games. These studies suggest that changes in interaction mode or input system can influence movement patterns, posture, perceived difficulty, and enjoyment, all of which may contribute to variations in brain activity intensity and the areas of activated brain regions. Looking further into how the 2 input systems (wFC and woFC) affected each ROIs reveals the following:

First, the addition of force control can introduce elements that can alter the movement strategies of the participants as reflected by the higher right PMC activity in the younger adults and higher rSMA activity in the older adults during wFC gameplay. Aside from motor planning, the PMC is also responsible for action observation and transforming visual information into motor commands. This includes the anticipation of the force output required to lift an object based on visual cues [34]. Furthermore, the precision of one's pinch force control can also greatly affect the PMC activation, with tasks requiring a higher precision of pinch force control, such as holding a small object with a sustained force level of just above the point before the object slips, producing a much higher right PMC activity as compared with holding the small object with one's normal and automatically scaled pinch force [35]. The right PMC activity of the younger adult group in this study also followed the behavior found in the previous studies. The lack of difference in the right PMC activation of the older adults when playing the game in both control systems can be due to their right PMC already being recruited to play the game using the woFC as a compensatory mechanism by the aging brain to recruit additional cortical areas when performing simple motor tasks [36,37]. Furthermore, this mechanism, along with the wFC's requirement to follow a precise movement sequence (thumb and index finger opposition followed by sustained isometric contraction at a strict time window ~0.5 second), likely contributed to the higher rSMA activity exhibited by the older adults, as well as the weak to moderate correlation between the pinch force COV-HbO (r=0.393) and pinch force COV-HbR (r=−0.494). The SMA is directly involved in planning complex movements that require temporal accuracies, as revealed by studies requiring users to press a button 2.5 seconds after a stimulus [38-40] and in studies disrupting the function of the SMA via repetitive transcranial magnetic stimulation, revealing worse performance in a sequential pressing task.

Second gameplay using the wFC induces greater ipsilateral M1 activity in both groups than gameplay using the woFC. In the young and healthy population, simple movements involving the hands and fingers exclusively activate the contralateral M1, while older adults feature more bilateral activity in the motor regions in response to motor tasks [36,41-43]. Aside from being age dependent, the ipsilateral activity of the M1 could also be influenced by task requirements and complexity [44-47]. Tasks that require sustained isometric contraction can also induce bilateral activity in the M1, detectable by both fNIRS and functional magnetic resonance imaging (fMRI) [48-50]. The magnitude of the force during these tasks ranges from 20% to 80% MVC and lasts from 10 up to 180 seconds per trial. Results from these studies point out that the bilateral activity of the M1 can be achieved by finger tasks accompanied by 20%, 40%, and 60% MVC. The contralateral activity of the M1 does not scale directly with force output, and the ipsilateral M1 is directly affected by the magnitude of the force, as well as the duration of the muscle contraction. This study also obtained consistent results with the previously mentioned studies. The results of this study aligning with previous findings show that wFC gameplay induces higher rM1 activity than woFC gameplay in both groups. This can be attributed to the wFC requiring users to exert an isometric contraction at approximately 70% MVC and maintain it for a few seconds to play the game effectively. Additionally, a positive correlation was observed between pinch force COV and HbO, with higher pinch force COV scores being associated with greater rM1 activity (r=0.494). One thing to take note of is that only the older adults exhibited higher activity in the lM1 in addition to the rM1 during gameplay using the wFC as compared with the woFC. This finding is also consistent with the previous studies mentioned, where older adults generally exhibit more bilateral activation in the motor regions than younger adults performing the same task.

Finally, rPFC in the older adults is recruited to keep up with the higher demands of playing the VR game when using the wFC. Aside from being involved with cognitive functions [51,52] and motor tasks [53], numerous studies also link increasing attentional demands, dual tasking costs, task difficulty, and complexity to the degree of PFC activity [54-57]. The functions of the PFC can also be divided further between the left and right hemispheres, with the rPFC being implicated for both externally and internally cued [58] real-time monitoring of movement and for the retrieval of memory [59,60]. As for this study, in addition to the memory and motor tasks when using the woFC, the introduction of a force control task adds a layer of complexity to the game when using the wFC. This could have caused the older adult group to allocate more attentional and memory resources for adjusting their force output through monitoring of the force bar, as reflected by the similar number of grasps initiated and the higher pinch force COV as compared

XSL•FO
RenderX

with the younger adults, while at the same time recalling the order of the food list. fMRI studies [61,62] also found rPFC activity during a precision force control task and cited the visual feedback as a contributing factor to the rPFC activity.

Stronger CS and CR in previous studies reflect an increase in coordinating the senses and motor control function [63] and an additional cognitive load [64] in response to task complexity. In this study's wFC, the additional requirement of maintaining one's pinch force, alongside the compensatory mechanism of the aging brain mentioned in the previous sections, may explain higher CS and CR in the older adults during wFC.

The temporal characteristics of the HRF for both groups tend to follow the typical hemodynamic response as those found in the literature [14,17] in which activated brain regions experienced a rise in the concentration of HbO, accompanied by a drop in the concentration of HbR.

When considering the cortical activation maps, differences can be observed between the 2 input systems and between the 2 study groups. Comparing the cortical activation maps of the 2 input systems shows that the wFC activates a larger spatial activity in the brain. fMRI and positron emission tomography studies [50,65-69] have also established a link between the extent of the spatial activity of the brain to the force level and the presence of an isometric contraction during a handgrip task. These studies suggest that recruiting larger areas of the cortex is an adaptation mechanism by the brain for sustaining isometric force tasks. The same process could have also occurred in this study, as gameplay using the wFC requires an isometric pinch as opposed to the woFC system where closing the gap between the index finger and the thumb is enough to play the game properly. On the other hand, a comparison of the cortical activation maps between the 2 study groups reveals that the older adult group experienced more bilateral brain activity irrespective of control systems. fMRI studies [43,70], where subjects were asked to perform finger opposition and hand movements, also obtained comparable results with their older adult groups exhibiting a more bilateral activity in several cortical regions (PMC, SMA, M1, etc) than their younger adult counterparts. The effects of aging on the corpus callosum are cited to be one of the culprits to the reduction of hemispheric lateralization in the older adults.

Limitations and Future Directions

This study is limited by the following factors: First, the study population consisted of only healthy individuals, so the effects of the developed system on the Hb response of patients with neurological symptoms cannot be determined. Second, the experiment used a fixed block design—5 consecutive wFC gameplay sessions followed by 5 consecutive woFC sessions—to minimize the time interval between conditions. This design was chosen to reduce delays, as removing the sensor was quicker than reapplying it midexperiment, which could have introduced systematic bias. Third, the single game difficulty level was used in the study, as different game difficulty levels [71] can directly affect the Hb responses of the brain. Fourth, the long-term training or rehabilitation effects of this system cannot be determined in this study as this study was a one-time intervention. Fifth, the usage of a single sensor in the system, although transposable, limits it to only accepting inputs between a single thumb and opposing finger pair, and modification to the system needs to be done to let it accept inputs from different thumb and opposing finger pairs. And finally, the limited number of source detector pairs used in fNIRS data acquisition may not be enough to cover the whole motor area region of the brain.

To address these limitations, a long-term follow-up study containing training sessions, adjustable difficulty levels, modified experiment block design, and so forth, can be performed to give more insight regarding the effectiveness of adding a force control component to rehabilitation outcomes of VR-based programs.

Conclusions

In this study, the effects between the regular hand-tracking system of VR head-mounted displays and the novel VR input system on brain activity were compared. To the author's best knowledge, this is the first system specifically designed for training or rehabilitation in the literature to incorporate a force control component to a VR hand-tracking system. The fNIRS results from this study suggest that the addition of a force control component to VR hand-tracking systems could make them more effective for training or rehabilitation by inducing greater and more widespread activity in the cortex, thus making it better for promoting neural plasticity. Additionally, the results from this study can be used as a reference for the development of new VR-based training or rehabilitation systems, which incorporate force control as a gameplay component.

Acknowledgments

This research was funded by the Taiwan Ministry of Science and Technology (grant no NSTC 112 - 2321-B-006 - 017).

Data Availability

The datasets analyzed during this study are available in Google Drive [72].

Conflicts of Interest

None declared.

XSL•FO
RenderX

References

1. Steverson M. Ageing and Health. : WHO; 2022.
2. Lin CW, Kuo LC, Lin YC, Su FC, Lin YA, Hsu HY. Development and testing of a virtual reality mirror therapy system for the sensorimotor performance of upper extremity: a pilot randomized controlled trial. IEEE Access 2021;9:14725-14734. [doi: 10.1109/ACCESS.2021.3050656]
3. Hsu HY, Kuo LC, Lin YC, Su FC, Yang TH, Lin CW. Effects of a virtual reality-based mirror therapy program on improving sensorimotor function of hands in chronic stroke patients: a randomized controlled trial. Neurorehabil Neural Repair 2022 Jun;36(6):335-345. [doi: 10.1177/15459683221081430] [Medline: 35341360]
4. Hsu HY, Lin CW, Lin YC, et al. Effects of vibrotactile-enhanced music-based intervention on sensorimotor control capacity in the hand of an aging brain: a pilot feasibility randomized crossover trial. BMC Geriatr 2021 Nov 23;21(1):660. [doi: 10.1186/s12877-021-02604-0] [Medline: 34814839]
5. Fan SC, Su FC, Chen SS, et al. Improved intrinsic motivation and muscle activation patterns in reaching task using virtual reality training for stroke rehabilitation: a pilot randomized control trial. J Med Biol Eng 2014;34(4):399. [doi: 10.5405/jmbe.1502]
6. Afridi A, Malik AN, Tariq H, Rathore FA. The emerging role of virtual reality training in rehabilitation. J Pak Med Assoc 2022 Jan;72(1):188-191. [doi: 10.47391/JPMA.22-006] [Medline: 35099468]
7. Burdea GC. Virtual rehabilitation—benefits and challenges. Methods Inf Med 2003;42(5):519-523. [doi: 10.1055/s-0038-1634378]
8. Calabrò RS, Accorinti M, Porcari B, et al. Does hand robotic rehabilitation improve motor function by rebalancing interhemispheric connectivity after chronic stroke? Encouraging data from a randomised-clinical-trial. Clin Neurophysiol 2019 May;130(5):767-780. [doi: 10.1016/j.clinph.2019.02.013] [Medline: 30904771]
9. Takahashi CD, Der-Yeghiaian L, Le V, Motiwala RR, Cramer SC. Robot-based hand motor therapy after stroke. Brain (Bacau) 2008 Feb;131(Pt 2):425-437. [doi: 10.1093/brain/awm311] [Medline: 18156154]
10. Ip LS. The effect of finger fine motor training device with music on older adult. National Cheng Kung University. 2022. URL: https://thesis.lib.ncku.edu.tw/thesis/detail/17874bbe70eb47c15065ffbce15223f5/ [accessed 2025-04-09]
11. Wang YT. Music assisted hand training system for brain activation. National Cheng Kung University. 2022. URL: https://thesis.lib.ncku.edu.tw/thesis/detail/0a95a0351148b1a1f565b090bb37932d/ [accessed 2025-04-09]
12. Ge R, Wang Z, Yuan X, et al. The effects of two game interaction modes on cortical activation in subjects of different ages: a functional near-infrared spectroscopy study. IEEE Access 2021;9:11405-11415. [doi: 10.1109/ACCESS.2021.3050210]
13. Juan MC, Elexpuru J, Dias P, Santos BS, Amorim P. Immersive virtual reality for upper limb rehabilitation: comparing hand and controller interaction. Virtual Real 2023;27(2):1157-1171. [doi: 10.1007/s10055-022-00722-7] [Medline: 36475065]
14. Bae S, Park HS. Development of immersive virtual reality-based hand rehabilitation system using a gesture-controlled rhythm game with vibrotactile feedback: an fNIRS pilot study. IEEE Trans Neural Syst Rehabil Eng 2023;31:3732-3743. [doi: 10.1109/TNSRE.2023.3312336] [Medline: 37669214]
15. Zheng J, Ma Q, He W, et al. Cognitive and motor cortex activation during robot-assisted multi-sensory interactive motor rehabilitation training: an fNIRS based pilot study. Front Hum Neurosci 2023;17:1089276. [doi: 10.3389/fnhum.2023.1089276] [Medline: 36845877]
16. Hara Y. Brain plasticity and rehabilitation in stroke patients. J Nippon Med Sch 2015;82(1):4-13. [doi: 10.1272/jnms.82.4] [Medline: 25797869]
17. Cheng PY, Chieh HF, Lin CJ, et al. Digit force controls and corresponding brain activities in finger pressing performance: a comparison between older adults and young individuals. J Aging Phys Act 2020 Jan 1;28(1):94-103. [doi: 10.1123/japa.2018-0252] [Medline: 31629354]
18. Lubrini G, Martín-Montes A, Díez-Ascaso O, Díez-Tejedor E. Brain disease, connectivity, plasticity and cognitive therapy: A neurological view of mental disorders. Neurología (Engl Ed) 2018 Apr;33(3):187-191. [doi: 10.1016/j.nrleng.2017.02.001]
19. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Ann Intern Med 2007 Oct 16;147(8):573-577. [doi: 10.7326/0003-4819-147-8-200710160-00010]
20. Huppert TJ, Diamond SG, Franceschini MA, Boas DA. HomER: a review of time-series analysis methods for near-infrared spectroscopy of the brain. Appl Opt 2009 Apr 1;48(10):D280-D298. [doi: 10.1364/ao.48.00d280] [Medline: 19340120]
21. Molavi B, Dumont GA. Wavelet-based motion artifact removal for functional near-infrared spectroscopy. Physiol Meas 2012 Feb;33(2):259-270. [doi: 10.1088/0967-3334/33/2/259] [Medline: 22273765]
22. Lin TY, Wu JS, Lin LL, Ho TC, Lin PY, Chen JJJ. Assessments of muscle oxygenation and cortical activity using functional near-infrared spectroscopy in healthy adults during hybrid activation. IEEE Trans Neural Syst Rehabil Eng 2016 Jan;24(1):1-9. [doi: 10.1109/TNSRE.2015.2429655] [Medline: 25974942]
23. Ye JC, Tak S, Jang KE, Jung J, Jang J. NIRS-SPM: statistical parametric mapping for near-infrared spectroscopy. Neuroimage 2009 Jan 15;44(2):428-447. [doi: 10.1016/j.neuroimage.2008.08.036] [Medline: 18848897]
24. Gagnon L, Perdue K, Greve DN, Goldenholz D, Kaskhedikar G, Boas DA. Improved recovery of the hemodynamic response in diffuse optical imaging using short optode separations and state-space modeling. Neuroimage 2011 Jun 1;56(3):1362-1371. [doi: 10.1016/j.neuroimage.2011.03.001] [Medline: 21385616]

XSL•FO
RenderX

25. Aasted CM, Yücel MA, Cooper RJ, et al. Anatomical guidance for functional near-infrared spectroscopy: AtlasViewer tutorial. Neurophotonics 2015 Apr;2(2):020801. [doi: 10.1117/1.NPh.2.2.020801] [Medline: 26157991]
26. Nguyen T, Babawale O, Kim T, Jo HJ, Liu H, Kim JG. Exploring brain functional connectivity in rest and sleep states: a fNIRS study. Sci Rep 2018 Nov 1;8(1):16144. [doi: 10.1038/s41598-018-33439-2]
27. Corey DM, Dunlap WP, Burke MJ. Averaging correlations: expected values and bias in combined Pearson r s and Fisher's z transformations. J Gen Psychol 1998 Jul;125(3):245-261. [doi: 10.1080/00221309809595548]
28. Singh AK, Dan I. Exploring the false discovery rate in multichannel NIRS. Neuroimage 2006 Nov 1;33(2):542-549. [doi: 10.1016/j.neuroimage.2006.06.047] [Medline: 16959498]
29. Lin P, Chen H, Yeh P. The effects of age group, interface type, and key area of mobile phones on game playing performance, usability, and visual fatigue: a case study of a rhythm game. Hum Factors Ergon Manuf Service Industry 2019 May;29(3):275-282. [doi: 10.1002/hfm.20783]
30. Lamoth CJC, Alingh R, Caljouw SR. Exergaming for elderly: effects of different types of game feedback on performance of a balance task. Stud Health Technol Inform 2012;181:103-107. [Medline: 22954837]
31. Brauner P, Ziefle M. Serious motion-based exercise games for older adults: evaluation of usability, performance, and pain mitigation. JMIR Serious Games 2020 Apr 1;8(2):e14182. [doi: 10.2196/14182] [Medline: 32234703]
32. Wang JY. Designing brain training games and evaluating the usability between young and elderly. Presented at: 2016 3rd International Conference on Information Science and Control Engineering (ICISCE); 2016; Beijing, China p. 308-312. [doi: 10.1109/ICISCE.2016.75]
33. Wang Z, Liao M, Li Q, et al. Effects of three different rehabilitation games' interaction on brain activation using functional near-infrared spectroscopy. Physiol Meas 2020 Dec 31;41(12):125005. [doi: 10.1088/1361-6579/abcd1f] [Medline: 33227728]
34. Chouinard PA, Leonard G, Paus T. Role of the primary motor and dorsal premotor cortices in the anticipation of forces during object lifting. J Neurosci 2005 Mar 2;25(9):2277-2284. [doi: 10.1523/JNEUROSCI.4649-04.2005] [Medline: 15745953]
35. Ehrsson HH, Fagergren A, Jonsson T, Westling G, Johansson RS, Forssberg H. Cortical activity in precision- versus power-grip tasks: an fMRI study. J Neurophysiol 2000 Jan;83(1):528-536. [doi: 10.1152/jn.2000.83.1.528] [Medline: 10634893]
36. Mattay VS, Fera F, Tessitore A, et al. Neurophysiological correlates of age-related changes in human motor function. Neurology (ECronicon) 2002 Feb 26;58(4):630-635. [doi: 10.1212/wnl.58.4.630] [Medline: 11865144]
37. Riecker A, Gröschel K, Ackermann H, Steinbrink C, Witte O, Kastrup A. Functional significance of age-related differences in motor activation patterns. Neuroimage 2006 Sep;32(3):1345-1354. [doi: 10.1016/j.neuroimage.2006.05.021]
38. Coull JT, Vidal F, Nazarian B, Macar F. Functional anatomy of the attentional modulation of time estimation. Science 2004 Mar 5;303(5663):1506-1508. [doi: 10.1126/science.1091573]
39. Macar F, Anton JL, Bonnet M, Vidal F. Timing functions of the supplementary motor area: an event-related fMRI study. Brain Res Cogn Brain Res 2004 Oct;21(2):206-215. [doi: 10.1016/j.cogbrainres.2004.01.005] [Medline: 15464352]
40. Gerloff C, Corwell B, Chen R, Hallett M, Cohen LG. Stimulation over the human supplementary motor area interferes with the organization of future elements in complex motor sequences. Brain (Bacau) 1997 Sep;120 (Pt 9)(9):1587-1602. [doi: 10.1093/brain/120.9.1587] [Medline: 9313642]
41. Ward NS. Compensatory mechanisms in the aging motor system. Ageing Res Rev 2006 Aug;5(3):239-254. [doi: 10.1016/j.arr.2006.04.003] [Medline: 16905372]
42. Sailer A, Dichgans J, Gerloff C. The influence of normal aging on the cortical processing of a simple motor task. Neurology (ECronicon) 2000 Oct 10;55(7):979-985. [doi: 10.1212/WNL.55.7.979]
43. Naccarato M, Calautti C, Jones PS, Day DJ, Carpenter TA, Baron JC. Does healthy aging affect the hemispheric activation balance during paced index-to-thumb opposition task? An fMRI study. Neuroimage 2006 Sep;32(3):1250-1256. [doi: 10.1016/j.neuroimage.2006.05.003]
44. van den Berg FE, Swinnen SP, Wenderoth N. Excitability of the motor cortex ipsilateral to the moving body side depends on spatio-temporal task complexity and hemispheric specialization. PLoS One 2011 Mar 9;6(3):e17742. [doi: 10.1371/journal.pone.0017742] [Medline: 21408031]
45. Chen R, Cohen LG, Hallett M. Role of the ipsilateral motor cortex in voluntary movement. Can J Neurol Sci 1997 Nov;24(04):284-291. [doi: 10.1017/S0317167100032947]
46. Chen R, Gerloff C, Hallett M, Cohen LG. Involvement of the ipsilateral motor cortex in finger movements of different complexities. Ann Neurol 1997 Feb;41(2):247-254. [doi: 10.1002/ana.410410216] [Medline: 9029074]
47. Barany DA, Revill KP, Caliban A, et al. Primary motor cortical activity during unimanual movements with increasing demand on precision. J Neurophysiol 2020 Sep 1;124(3):728-739. [doi: 10.1152/jn.00546.2019] [Medline: 32727264]
48. Shibuya K, Kuboyama N, Tanaka J. Changes in ipsilateral motor cortex activity during a unilateral isometric finger task are dependent on the muscle contraction force. Physiol Meas 2014 Mar;35(3):417-428. [doi: 10.1088/0967-3334/35/3/417] [Medline: 24521545]
49. Shibuya K, Kuboyama N, Yamada S. Complementary activation of the ipsilateral primary motor cortex during a sustained handgrip task. Eur J Appl Physiol 2016 Jan;116(1):171-178. [doi: 10.1007/s00421-015-3262-1] [Medline: 26377003]

XSL•FO
RenderX

50. Dai TH, Liu JZ, Sahgal V, Brown RW, Yue GH. Relationship between muscle output and functional MRI-measured brain activation. Exp Brain Res 2001 Oct;140(3):290-300. [doi: 10.1007/s002210100815] [Medline: 11681304]
51. Curtis CE, D'Esposito M. Persistent activity in the prefrontal cortex during working memory. Trends Cogn Sci 2003 Sep;7(9):415-423. [doi: 10.1016/s1364-6613(03)00197-9] [Medline: 12963473]
52. Frith C, Dolan R. The role of the prefrontal cortex in higher cognitive functions. Brain Res Cogn Brain Res 1996 Dec;5(1-2):175-181. [doi: 10.1016/s0926-6410(96)00054-7] [Medline: 9049084]
53. D'Esposito M, Ballard D, Zarahn E, Aguirre GK. The role of prefrontal cortex in sensory memory and motor preparation: an event-related fMRI study. Neuroimage 2000 May;11(5 Pt 1):400-408. [doi: 10.1006/nimg.2000.0571] [Medline: 10806027]
54. Tung C, Lord SR, Pelicioni PHS, Sturnieks DL, Menant JCC. Prefrontal and motor planning cortical activity during stepping tasks is related to task complexity but not concern about falling in older people: a fNIRS study. Brain Sci 2023 Dec 5;13(12):1675. [doi: 10.3390/brainsci13121675]
55. Kotegawa K, Yasumura A, Teramoto W. Changes in prefrontal cortical activation during motor imagery of precision gait with age and task difficulty. Behav Brain Res 2021 Feb 5;399:113046. [doi: 10.1016/j.bbr.2020.113046] [Medline: 33279636]
56. Saraiva M, Paszkiel S, Vilas-Boas JP, Castro MA. Influence of cognitive task difficulty in postural control and hemodynamic response in the prefrontal cortex during static postural standing. Appl Sci (Basel) 2022;12(13):6363. [doi: 10.3390/app12136363]
57. Mandrick K, Derosiere G, Dray G, Coulon D, Micallef JP, Perrey S. Prefrontal cortex activity during motor tasks with additional mental load requiring attentional demand: a near-infrared spectroscopy study. Neurosci Res 2013 Jul;76(3):156-162. [doi: 10.1016/j.neures.2013.04.006] [Medline: 23665138]
58. Picton TW, Stuss DT, Shallice T, Alexander MP, Gillingham S. Keeping time: effects of focal frontal lesions. Neuropsychologia 2006;44(7):1195-1209. [doi: 10.1016/j.neuropsychologia.2005.10.002] [Medline: 16271270]
59. Henson RNA, Shallice T, Dolan RJ. Right prefrontal cortex and episodic memory retrieval: a functional MRI test of the monitoring hypothesis. Brain (Bacau) 1999 Jul;122(7):1367-1381. [doi: 10.1093/brain/122.7.1367]
60. Stuss DT, Alexander MP. Is there a dysexecutive syndrome? Phil Trans R Soc B 2007 May 29;362(1481):901-915. [doi: 10.1098/rstb.2007.2096]
61. Beck MM, Spedden ME, Dietz MJ, Karabanov AN, Christensen MS, Lundbye-Jensen J. Cortical signatures of precision grip force control in children, adolescents, and adults. Elife 2021 Jun 14;10:e61018. [doi: 10.7554/eLife.61018] [Medline: 34121656]
62. Vaillancourt DE, Thulborn KR, Corcos DM. Neural basis for the processes that underlie visually guided and internally guided force control in humans. J Neurophysiol 2003 Nov;90(5):3330-3340. [doi: 10.1152/jn.00394.2003] [Medline: 12840082]
63. Koch G, Franca M, Del Olmo MF, et al. Time course of functional connectivity between dorsal premotor and contralateral motor cortex during movement selection. J Neurosci 2006 Jul 12;26(28):7452-7459. [doi: 10.1523/JNEUROSCI.1158-06.2006] [Medline: 16837593]
64. Ye Y, Zhou T, Zhu Q, Vann W, Du J. Brain functional connectivity under teleoperation latency: a fNIRS study. Front Neurosci 2024;18:1416719. [doi: 10.3389/fnins.2024.1416719] [Medline: 39605793]
65. Keisker B, Hepp-Reymond MC, Blickenstorfer A, Meyer M, Kollias SS. Differential force scaling of fine-graded power grip force in the sensorimotor network. Hum Brain Mapp 2009 Aug;30(8):2453-2465. [doi: 10.1002/hbm.20676] [Medline: 19172654]
66. Braaß H, Feldheim J, Chu Y, et al. Association between activity in the ventral premotor cortex and spinal cord activation during force generation—a combined cortico-spinal fMRI study. Hum Brain Mapp 2023 Dec 15;44(18):6471-6483. [doi: 10.1002/hbm.26523] [Medline: 37873743]
67. Thickbroom GW, Phillips BA, Morris I, Byrnes ML, Mastaglia FL. Isometric force-related activity in sensorimotor cortex measured with functional MRI. Exp Brain Res 1998 Jul;121(1):59-64. [doi: 10.1007/s002210050437] [Medline: 9698191]
68. Korotkov A, Radovanovic S, Ljubisavljevic M, et al. Comparison of brain activation after sustained non-fatiguing and fatiguing muscle contraction: a positron emission tomography study. Exp Brain Res 2005 May;163(1):65-74. [doi: 10.1007/s00221-004-2141-5] [Medline: 15645226]
69. Cramer SC, Weisskoff RM, Schaechter JD, et al. Motor cortex activation is related to force of squeezing. Hum Brain Mapp 2002 Aug;16(4):197-205. [doi: 10.1002/hbm.10040] [Medline: 12112762]
70. Taniwaki T, Okayama A, Yoshiura T, et al. Age-related alterations of the functional interactions within the basal ganglia and cerebellar motor loops in vivo. Neuroimage 2007 Jul 15;36(4):1263-1276. [doi: 10.1016/j.neuroimage.2007.04.027] [Medline: 17524667]
71. Girouard A, Solovey ET, Hirshfield LM, et al. Distinguishing Difficulty Levels with Non-Invasive Brain Activity Measurements: Springer; 2009:440-452. [doi: 10.1007/978-3-642-03655-2_50]
72. Gan L, Lin CJ, Chieh HF, An KN, Su FC. Experiment data. Google Drive. URL: https://drive.google.com/drive/folders/1ImFUHK6DNTjRwgxOgG61JuTJ-ACU555w?usp=drive_link [accessed 2025-06-26]

XSL•FO
RenderX

Abbreviations

ADL: activity of daily living
COV: coefficient of variation
CR: connection ratio
CS: connection strength
fMRI: functional magnetic resonance imaging
fNIRS: functional near-infrared spectroscopy
FSR: force-sensitive resistor
HbO: oxygenated hemoglobin
HbR: deoxygenated hemoglobin
HRF: hemodynamic response function
lM1: left primary motor cortex
M1: primary motor cortex
MVC: maximum voluntary contraction
NCKU: National Cheng Kung University
PFC: prefrontal cortex
PMC: premotor cortex
rM1: right primary motor cortex
ROI: region of interest
rPFC: right prefrontal cortex
rPMC: right premotor cortex
rSMA: right supplementary motor area
SMA: supplementary motor area
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology
VR: virtual reality
wFC: with force control input system
woFC: without force control input system

Edited by A Coristine; submitted 02.07.24; peer-reviewed by C Shi, T Zhang; revised version received 08.04.25; accepted 08.04.25; published 14.07.25.

Please cite as:
Gan L, Lin CJ, Chieh HF, An KN, Su FC
A Virtual Reality Force Control Training System on Brain Activation: Functional Near-Infrared Spectroscopy (fNIRS) Study
JMIR Serious Games 2025;13:e63874
URL: https://games.jmir.org/2025/1/e63874
doi:10.2196/63874

© Luigi Gan, Chien-Ju Lin, Hsiao-Feng Chieh, Kai-Nan An, Fong-Chin Su. Originally published in JMIR Serious Games (https://games.jmir.org), 14.7.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Adolescents' and Parents' Perspectives on Using the MedSMARxT Families Intervention in Emergency Departments for Opioid Medication Safety Education: Mixed Methods Study

Olufunmilola Abraham[1], BPharm, MSc, PhD; Sara Nadi[2], MPH, PharmD; Irene Hurst[3], MSci, MD

[1]Pharmacy Practice Science, College of Pharmacy, University of Kentucky, 789 S. Limestone Street, TODD 292K, Lexington, KY, United States

[2]Social and Administrative Sciences Division, School of Pharmacy, University of Wisconsin–Madison, Madison, WI, United States

[3]Division of Pediatric Emergency Medicine, Department of Emergency Medicine, University of Wisconsin–Madison, Madison, WI, United States

Corresponding Author:
Olufunmilola Abraham, BPharm, MSc, PhD
Pharmacy Practice Science, College of Pharmacy, University of Kentucky, 789 S. Limestone Street, TODD 292K, Lexington, KY, United States

Abstract

Background: The opioid crisis has significantly impacted adolescents and their families. This is attributed in part to increased opioid prescriptions in pediatric emergency departments (EDs) due to acute pain conditions and injuries. Although EDs frequently prescribe opioids, no effective preventative interventions have been implemented to educate adolescents and their families on safe opioid use. This study evaluates the MedSMA T Families intervention, which consists of an engaging serious game, Adventures in PharmaCity, and a personalized Family Medication Safety Plan (FMSP) with the aim of reducing opioid misuse and promoting opioid medication safety. The MedSMA T Families intervention was developed to educate adolescents and adults prescribed opioids on safe practices such as opioid storage and disposal.

Objective: This study aimed to explore and characterize adolescents' and parents' experiences and perspectives on implementing the MedSMA T Families intervention in the ED to improve opioid education and safety among adolescents.

Methods: A total of 93 participants, including 16 children and 77 parents, were recruited from the pediatric ED at a tertiary academic hospital to play the MedSMA T game in the ED. A total of 16 participants, including 8 children and 8 parents, were followed up with interviews to gather qualitative feedback. Participants engaged with the MedSMA T game—Adventures in PharmaCity—and the FMSP. Data were collected through gameplay observation and 75-minute semistructured interviews via Zoom. Quantitative in-game data were analyzed using descriptive analysis and qualitative data were analyzed using thematic analysis with NVivo (version 14; Lumivero).

Results: Parents spent an average of 22.16 (SD 4.97) minutes playing the game, while children spent an average of 21.99 (SD 8.06) minutes. Families appreciated the game's design and noted usability challenges and suggested enhancements for clearer gameplay instructions. Participants reported increased knowledge of opioid safety, highlighted the importance of communication with health care providers, and a desire for a mobile app to assist with medication management. The FMSP was perceived as valuable for promoting awareness of safe practices and connected well to the knowledge gained from the game.

Conclusions: The MedSMA T Families intervention was well received as a beneficial educational tool to educate adolescents and their families on safe opioid use. Additionally, it highlights a clear need for more accessible digital tools to increase opioid education. This feedback indicates a strong interest in improving educational resources to ensure safe opioid management within families.

(JMIR Serious Games 2025;13:e68814) doi:10.2196/68814

KEYWORDS

opioid medication safety; family medication safety; educational intervention; serious games; emergency department; MedSMA T; MedSMARxT

Introduction

Pediatric emergency departments (EDs) are considered the best practice models for the emergency care of children [1-5]. Patients frequently visit the ED for acute pain [6], including headaches, back pain, stomach discomfort, musculoskeletal injuries, postoperative pain, and pain associated with cancer. The use of opioids, conventional nonnarcotic medications (such as acetaminophen and nonsteroidal anti-inflammatory medicines), and nonpharmacological treatments including distraction techniques are among the most common treatments

XSL•FO
RenderX

used by emergency physicians for pain management. Despite alternative pain control methods, opioids are commonly used in and prescribed from the ED for moderate to severe pain [6]. Concordantly, adolescent prescription opioid misuse is an increasing concern in the United States [7]. Nearly 60,000 opioid-related pediatric visits occurred throughout the US EDs from 2014 to 2017 [8]. In 2017, a public health emergency was declared due to the increased use and overdose incidences of opioids [9]. It is posited that an increased regulatory focus on pain management, aggressive pharmaceutical marketing, and the initial low concern for opioid addiction from prescription opioids have contributed to this crisis [10-15].

The use of prescription opioids increases the risk of opioid abuse in the future [16-21]. Persons with opioid use disorder commonly report that they were first exposed to opioids through a legitimate prescription, frequently from an ED [16-21]. As a result, the future risks and exposure to prescription opioids have significantly increased for individuals in their youth and adolescence. From 2019 to 2020, prescription opioid-involved death rates increased by 17% [22,23]. Adolescents who report having a valid prescription for a narcotic painkiller by the time they are in the 12th grade are more likely to abuse these prescription drugs later on [24,25]. Youth who report using opioids for nonmedical purposes before reaching adulthood are similarly at risk for abuse [26]. The rate of opioid misuse and opioid-related deaths among pediatric patients has increased along with the rise in opioid prescribing [17], in addition to increased hospitalizations attributable to opioid use [27].

Given the opioid crisis among adolescents in the United States and the increased risk of opioid misuse in this age group, it is crucial to address these issues effectively. Past interventions aimed at minimizing the risks of opioid misuse among adolescents have included universal prevention programs such as the Strengthening Families Program and the Life Skills Training Program, which focus on enhancing protective factors and reducing risk [19]. However, selective prevention strategies specifically tailored for at-risk adolescents and emerging adults remain limited, as recent studies have shown that most interventions fail to address the unique needs and contexts of this population [19]. Our two-part intervention, MedSMA T Families, aims to reduce the epidemic of opioid misuse by educating adolescents and their families who have been prescribed opioids from the ED. This intervention consists of an adolescent-tailored serious game entitled MedSMA T: Adventures in PharmaCity, as well as a family-focused tool called the Family Medication Safety Plan (FMSP), which is a personalized tool for family education [28]. Together, these resources help increase awareness about safe opioid use and provide educational information on safe opioid storage and disposal. There has been previous work regarding the MedSMA T Families intervention, such as quantitative and qualitative work investigating parental, adolescent, and staff perceptions regarding the game. However, the focus of this paper is the implementation of the MedSMA T game in the ED setting, specifically exploring the perspectives of the end users of this intervention [29]. This study explores parents' and adolescents' perspectives on the implementation of the MedSMA T Families intervention in the ED by examining the perceptions of adolescents and their families regarding the MedSMA T game and the FMSP.

Methods

Recruitment

The University of Wisconsin-Madison Institutional Review Board approved this study, and written consent was obtained from all participants. Families were recruited from a pediatric ED at an academic hospital in Wisconsin. Eligibility criteria included the ability to speak, read, and understand English, access to a computer, tablet, or mobile phone with videoconferencing ability, and having been prescribed an opioid medication from the ED. Data were collected through mobile gameplay, surveys for demographic information, interviews, and the creation of an FMSP with child-parent dyads. Recruitment flyers were displayed in high-volume areas throughout the department. Emergency Department Research Coordinators presented the study and distributed fliers at pre-shift huddles. If a child between the ages of 12 and 18 years in the ED was prescribed an opioid medication, the Emergency Department Research Coordinator team screened and approached families for participation. If families consented, the research team contacted participants by email at least 3 times to arrange a Zoom gameplay and interview session.

The MedSMA T: Adventures in PharmaCity serious game (Figure 1) follows the story of an anthropomorphized sheep as they navigate making safe decisions regarding prescription opioids [28]. Correct opioid-related choices progress the story forward, while unsafe opioid-related choices rewind the story to allow the player to re-attempt the level and learn from their mistakes. In level 1, the player learns about safe opioid storage and the consequences of sharing opioids with friends. Level 2 provides background information and enhances the game storyline by adding a real-life scene about being in pain and forgetting there was an important assignment due that day. In level 3, the player learns not to take others' prescription opioid medications. Level 4 teaches the player about the purpose of Narcan (a medicine that rapidly reverses an opioid overdose) and demonstrates the consequences of sharing opioids with others. Level 5 outlines safe opioid disposal. The game experience is inherently designed to improve adolescents' decision-making about safe opioid use.

XSL•FO
RenderX

Figure 1. MedSMA T: Adventures in PharmaCity serious game.



The personalized FMSP is a tool for families to record important information about their medications and create a plan for safe use, storage, and disposal [28]. The FMSP is designed to promote communication among families and providers or parent-child dyads as they complete and review the plan. It also serves as a tool for families to reference important medication information in one location and note any questions they have for their health care team.

Five MedSMA T Families features facilitate family communication and adolescent opioid safety. These include (1) engaging and interactive delivery through game-based learning; (2) realistic gameplay scenarios where adolescents and parents practice making safe medication choices in various community settings, including at school, in their neighborhoods, and at home; (3) a space for important medication information such as side effects, drug-drug interactions, dosage instructions, and reasons for use; (4) a section for proper storage and disposal information and planning; and (5) an area for positive communication, allowing families to ask questions about medications for their providers. The FMSP is displayed in Multimedia Appendix 1.

Study Protocol

Data were collected between February 2023 and June 2023 until data saturation was reached. Similar to a previous study on pharmacist's perspectives of MedSMA T, data saturation was achieved at 12 participants [30]. However, data collection continued to 16 participants to ensure no new data or themes emerged from the data set. Participants completed a survey about their demographic information before the gameplay. A member of the research team observed child-parent dyads as they played the game and became familiar with the FMSP for 30-minutes during the web-based meeting sessions. Following that, there was a 75-minute semistructured follow-up interview with the participants. The goal of observation during games was to record any problems that arose during gameplay and to step in if there was a technological error. Each session was documented by noting the occurrence of specific events, such as gameplay interruptions, participant questions, and instances of technological errors. Zoom was used to record both the interview and the gameplay. The interview audio recordings were professionally transcribed verbatim.

Statistical Analysis

A descriptive analysis was conducted to analyze the quantitative in-game data. This included reporting the mean and the SD of the total time spent playing the game, the number of distinct levels played, and the number of distinct levels completed. It also included the number of opioid-related decisions or actions that resulted in game failure. Additionally, it reported the number of opioid-related decisions or actions that contributed to game progress and the total time spent playing levels 1 - 5.

Measures

A standard semantic inductive technique was used to examine qualitative data to investigate aspects associated with child-parent dyads' viewpoints regarding the design and implementation of the intervention. With this method, implementation and iterative design patterns could be explored and defined based on participant answers.

One of the authors began by immersing in the interview transcripts and in-game data to gain a comprehensive understanding of the content. A coding framework was developed based on the initial read-throughs of the data. This included generating initial codes that encapsulated meaningful segments of the data using inductive thematic analyses via NVivo (version 14; Lumivero), a software program used for qualitative research. The initial codes were collated and organized into potential themes. This process involved examining the relationships between codes and identifying overarching patterns that captured the experiences of participants regarding the MedSMA T intervention. The themes were refined to ensure they accurately represented the data. This involved checking if the themes worked in relation to the coded extracts and the entire dataset. Each theme was clearly defined and named to reflect its content, ensuring clarity in communicating

XSL•FO
RenderX

key findings. The final themes were reported along with illustrative quotes from participants to provide context and enhance the richness of the findings.

Reflexivity was considered throughout the study process. The authors engaged in regular discussions to reflect on their biases and preconceptions that could influence data interpretation. To enhance confirmability, the study team maintained detailed records of coding decisions and the evolution of themes. Dependability was addressed by ensuring that consistent procedures were followed throughout the study process through documentation of decisions made during data collection, analysis, and theme development, allowing for replication of the study if desired. While this study was conducted within a pediatric ED in Wisconsin, efforts were made to provide rich descriptions of the participants and the intervention. The diverse demographics of participants and the inclusion of both adolescents and parents enhance the transferability of our results.

Ethical Considerations

This study has been reviewed and approved by the Institutional Review Board of the University of Wisconsin-Madison (2022 - 0895). Informed consent was obtained from participants during the primary data collection process. Study data are anonymized and deidentified to protect the privacy and confidentiality of all human participants involved. Data handling processes are in place to ensure no identifiable information is disclosed. Participants in this study received compensation of US $40 gift card for their participation in the research.

Results

User Statistics

A total number of 93 participants, including 16 children and 77 parents, played the MedSMA T game in the ED and provided the in-game data. The mean age among children was 15 years, 68% (n=11) of children were females, 27% (n=4) of children were males, and 5% (n=1) of children were nonbinary. A total of 73% (n=12) of the children were White or Caucasian, 11% (n=2) of children were Black or African American, 9% (n=1) of children were Hispanic or Latinx, and 9% (n=1) of children were Asian. The mean age among parents was 46 years, 65% (n=50) of parents were females, 33% (n=25) of parents were males, and 2% (n=2) of parents were nonbinary. A total of 79% (n=61) of parents were White or Caucasian, 14% (n=11) of parents were Black or African American, and 7% (n=5) of parents were Hispanic or Latinx. Sixteen participants, including 8 children and 8 parents, were followed up with interviews. The average time spent on introducing the game and explaining the instructions on how to play was 15 minutes. The average time spent playing the game from in-game data for children was 21.99 (SD 8.06) minutes with an average of completing 3 out of the 5 levels in the game. On average, child participants made 3 opioid-related decisions/actions during the game process that led to game progress from correct decisions. Alternately, this group made 2 opioid-related decisions/actions that resulted in game failure indicating incorrect decisions, based on different scenarios in the game.

Regarding the in-game data for parents, the average time spent playing the game was 22.16 minutes (SD 4.97) with an average of completing 4 out of the total 5 levels in the game. On average, participants made 5 opioid-related decisions/actions during the game process that led to game progress from correct decisions. This same group made, on average, 3 opioid-related decisions/actions that resulted in game failure indicating incorrect decisions based on different scenarios in the game. Table 1 represents the in-game data analysis results for child participants and Table 2 represents the in-game data analysis results for parents.

Table . In-game data: children's data (16 participants).

In-game data	Value, mean (SD)
Total time playing the game (minutes)	21.99 (8.06)
Number of distinct levels played	3.49 (1.54)
Number of distinct levels completed	2.97 (1.87)
Number of decisions/actions that lead to game failure	4.68 (2.82)
Number of opioid-related decisions/actions that lead to game failure	2.15 (1.42)
Number of decisions/actions that lead to game progress	3.57 (2.25)
Number of opioid-related decisions/actions that lead to game progress	3.48 (2.08)
Total time playing level 1 (minutes)	10.84 (6.58)
Total time playing level 2 (minutes)	1.26 (0.74)
Total time playing level 3 (minutes)	7.35 (5.65)
Total time playing level 4 (minutes)	1.28 (1.43)
Total time playing level 5 (minutes)	1.24 (1.92)

XSL•FO
RenderX

Table . In-game data: parents' data (77 participants).

In-game data	Value, mean (SD)
Total time playing the game (minutes)	22.16 (4.97)
Number of distinct levels played	4.53 (0.85)
Number of distinct levels completed	4.16 (1.17)
Number of decisions/actions that lead to game failure	5.87 (2.52)
Number of opioid-related decisions/actions that lead to game failure	2.87 (1.45)
Number of decisions/actions that lead to game progress	5.00 (1.41)
Number of opioid-related decisions/actions that lead to game progress	5.18 (1.47)
Total time playing level 1 (minutes)	7.46 (2.90)
Total time playing level 2 (minutes)	1.40 (0.51)
Total time playing level 3 (minutes)	8.82 (2.51)
Total time playing level 4 (minutes)	1.77 (0.81)
Total time playing level 5 (minutes)	2.69 (2.77)

Evaluation Outcomes

Four main themes were identified in the MedSMA T game. The first theme was titled "User Experience." Three subthemes were then extracted from the data which were: "Usability," "Suggestions for Improvement," and "Target Audience."

Regarding "Usability," parents mostly expressed having a hard time figuring out how to play the game and what they were supposed to do as the next steps. Children tended to have a more positive experience once they became familiar with the game instructions.

> *I really liked the art style. It was really cute. I think the controls were easy to manage and to get the hang of.* [Child 6]

The second subtheme was "Suggestions for Improvement" which included providing more interaction with scenarios throughout the game, more complexity in gameplay, more educational content on Narcan, and providing more guidance on emergency situations related to that.

The last subtheme was "Target Audience." Feedback emphasized that the game's target demographic seemed more suitable for children under 12 y old.

> *…I think if you're looking for giving it to demographics that are above the age of 12, or adults, it needs to be a game that would capture that attention.* [Parent 5]

The second main theme identified was "Educational Learning and Value." Two subthemes extracted from this theme were "Learnings" and "Perceived Goals." In the "learnings" category, participants highlighted significant takeaways from their experience playing the game and increasing their knowledge on important educational content. This included proper and safe disposal, storage methods, and side effects of opioids. One parent admitted to the following:

> *I liked just learning more about the opioids. I did not know that you could take unused or expired prescriptions back to a pharmacy.* [Parent 7]

Regarding the "Perceived Goals," the educational aims were positively received such as teaching the significance of not letting peer pressure decide especially when children are at school, and educating on safe medication management.

> *I thought it was a good way to teach people how to maintain and manage their medications and keep it out of the hands of other kids and the scenarios of not using other kids' medications in school or in the home or on the bus.* [Parent 4]

The third theme included "In-Game Elements" with subthemes of "Character Perception" and "Scenario Evaluation." For "Character Perception," while the child participants appreciated the realistic and fun character design, several parents found them too juvenile and more appropriate and appealing for 12 years or younger.

In the "Scenario Evaluation," participants perceived the game scenarios to be realistic and relatable, and simulated real-world situations specifically different scenarios that could happen at schools. This reportedly made it a more engaging learning experience.

> *I feel like they were pretty good and just showing different situations that you could easily have any common ones at school.* [Child 8]

The last theme recognized was "Educational Gaming Experiences" with 2 subthemes as "Interest in Educational Games" and "Perceived Value." Regarding the first subtheme, "Interest in Educational Games," general enthusiasm was found for educational gaming among both parents and children. Participants mostly indicated that playing educational games is enjoyable if they find an element of fun and engagement in the gameplay.

> *I don't mind playing games that are educational as long as they are fun.* [Child 1]

According to the last subtheme of "Perceived Value," participants valued the game's educational potential and the effective incorporation of educational content in an engaging

XSL•FO
RenderX

format that captures attention and interest while increasing knowledge and awareness.

> *I've played some other educational games, and I think there are just many different pathways you can go down with different consequences.* [Child 8]

Table 3 presents further feedback on themes, subthemes, and additional illustrative quotes for the MedSMA T game.

XSL•FO
RenderX

Table . Themes and subthemes for the MedSMA T game: parents and children data.

Theme and subtheme	Illustrative quotes
User experience	
Usability	• "I thought it was hard to maneuver. It was also hard to understand what I was supposed to do. Once I got going, I kind of understood, but some of the verbiage in it was kind of, left me wondering what I should do." [Parent 1] • “Well, I thought it was really cute and simple once I figured it out.” [Child 2]
Suggestions for improvement	• "I think the Narcan wasn’t touched on, but I don’t think, I mean, people don’t, most people aren’t carrying Narcan around in their pockets. So, I don’t know maybe talking a bit more about what to do if somebody does take medication they’re not supposed to, like calling 911 if they’re unconscious, or you know, calling urgent care or something if they took something but they’re still talking to you.” [Parent 4] • “Maybe a little bit more detailed instructions. Maybe like less like dialogue and more like direction kind of.” [Child 1] • "Maybe if there were more like, paths to go through? Like, I feel like I said there were, I think I just kind of breezed past most of them pretty quickly. Maybe I would like if there were more like, I don’t know, difficult options that were like, it was harder to find the correct answer. So, then there was like, more interaction with the cutscenes and stuff.” [Child 5]
Target audience	• "I mean, probably younger kids that are more versed in games and just learning about, you know, the, the basics like maybe, you know, more like the preschool ages, you know, elementary, like the really young that need to learn more about the risks and, and could find those characters more appealing.” [Parent 8] • "The characters I feel are kind of childish, if you’re looking for like teens or above age groups. If you’re looking to get, like maybe 12 and under, I think they’re fitting for that.” [Parent 7]
Educational learning and value	
Learnings	• “I thought you could flush down the toilet, but, I guess I learned that is probably a no-no.” [Parent 3] • "It’s just a useful tool like, kind of approaches different scenarios and allows kids to think through and adults to think through, what should I do in this scenario?” [Parent 6] • "I feel like it would be useful for people to learn about like drug use. So, they didn’t know much about that. Or if they’re going through something like this. I think it’s a fun way to get people to learn about it. And yeah, I think I did learn quite a bit, because I felt like I didn’t really know anything to begin with so yeah.” [Child 4] • "Well, before I wasn’t really sure about, like, any of the side effects like of opioids, or like any of the treatments or cures of that. I kind of knew that prescriptions weren’t supposed to be shared, but like I just guess I didn’t really know, exactly, like, why, but I guess, like, I guess the game taught me that, like, you know, it’s more because of like, different people need different amounts of stuff. And for one person, it can be an overdose can be like, way lower than with another person. And then also, I didn’t really know like, how strict you should be about disposing of the medications.” [Child 3]

XSL•FO
RenderX

Theme and subtheme	Illustrative quotes
Perceived goals	• "Um, just how to not let peer pressure decide, you know, what's right and wrong, and not to just assume that you should, just because something's okay for you, that's okay for everyone else, and to be responsible with things that are prescribed directly to you." [Parent 5] • "I feel like it could be a good educational experience for people who don't have this awareness. I mean, whether it's at the clinic or the pharmacy, when they go to pick up a prescription, it could be fun, especially if there's kids in the house, or if it's prescribed to a child or, or an adolescent." [Parent 7] • "I thought it was a good way to teach people how to maintain and manage their medications and keep it out of the hands of other kids and the scenarios of not using other kids' medications in school or in the home or on the bus." [Parent 4] • "To probably to make like right choices and see what the choices are that you can make in situations." [Child 7]
In-game elements	
Character perception	• "I think the characters are kind of childish if you're looking for a teen or above age groups." [Parent 3] • "I feel like they're pretty realistic to an actual, like, high school student." [Child 1] • "I liked kind of how they weren't like human. I think that made it a little bit more fun. I don't know, I guess like, I did like the storyline where like, he had to do the speech and like, he was a bit nervous and stuff. And then, like, how he was in pain there. And then like, the dilemma of him, like, should I take this? Or should I not?" [Child 3]
Scenario evaluation	• "I thought they were really good. Um, I guess the friend offering medicine and then the person asking for medicine were very salient points." [Parent 8] • "I liked the scenarios. I thought they were all really good. Um, yeah, I guess the characters were cute." [Parent 6] • "I felt like they were pretty like realistic to what could possibly happen in real life revolving around like a medication or drug." [Child 3]
Educational gaming experiences	
Interest in educational games	• "I don't mind them if they are something that you know, is kind of that fast moving or keep your interest kind of thing. I don't mind educational games. But, like I said it would have to kind of keep my interest." [Parent 4] • "I think they're great." [Parent 2] • "I think they're okay. They're like fun and not boring." [Child 5] • "I think that it's great if the game can be like, fun and educational at the same time. But a lot of the times when I'm playing games, I'm more in it for the fun rather than the education." [Child 7]
Perceived value	• "I mean, I would think word games to be educational. But I like it. I'd like I mean, I know that like playing word games keeps your mind sharp, especially as you get older." [Parent 1] • "I like how um it teaches you like stuff in a different way. And um I sometimes don't like the topic that we learn about but." [Child 6]

Three main themes were identified as a result of thematic analysis for the FMSP. The first theme was "Routine Medication Management" with 2 subthemes, "Medication Safety Practices at Home" and "Communications with Healthcare Providers." "Medication Safety Practices at Home" reported various practices in terms of the storage and disposal of opioids to maintain safety and secure medication management strategies.

> *I have a lockbox that has some actually, old medicines that need to be taken to a drop box in it. But otherwise, everything is pretty accessible.* [Parent 5]
>
> *We keep them in a cabinet or a special door. We don't really take them unless we need to.* [Child 7]

For "Communications with Healthcare Providers," parents highlighted the significance of effective communication with health care providers regarding medications and different

XSL•FO
RenderX

questions they may have. They mostly reported using digital platforms such as MyChart to communicate with their doctors and reaching out to their pharmacists asking prescription-related questions.

> *…it's mostly by MyChart, you can just, send the doctor a message and they're pretty fast at responding back. Unless it's something you need an immediate [response], then you call.* [Parent 4]

> *I used to talk to the pharmacist about it. I usually also ask about if I am on any other medications or vitamins if it is all okay, whether or not I have to take it with food, trying to figure out a schedule of when I need to take it, writing it down as to when I took it last, if I forget, I can look back at that log to see when the last time is that I took it.* [Parent 2]

The second theme was "FMSP Usability," including a subtheme identified as "FMSP Implementation Setting." Participants had a perception that health care providers would play an important role in implementing the FMSP. Many participants shared different ways for providing information about the FMSP to patients, such as adding prescriptions from doctors' offices, implementing through MyChart in after-visit summaries, as a printout in your medication bags at the pharmacy, or being advertised at schools.

> *… it could be something in your after-visit summary that they print out or put on MyChart for you. Your pharmacist could print one out and put it in your medication bags.* [Parent 6]

> *I feel like pharmacies or doctor's offices [would] be good? Even if there was an ad or commercial, could bring it up at school too, or health class or advisory?* [Child 6]

The second subtheme was "Recommendations for Platform Enhancement." Participants provided insightful comments on improving the usability of the FMSP. Many stated that an app on mobile devices would facilitate medication tracking, making it more convenient to have medication reminders and logs accessible.

> *I think an app would be most useful because you have your phone with you all the time, and if you are away from home, and you have medication with you, that you take, you would be able to track it, when you took it. If you had any side effects, you could maybe even set an alarm on your phone, through the app that would tell you or remind you now is time to take your medication. and then you can maybe check it off so that you know that you took it at that time, and you can look back and see the times that you have taken it.* [Parent 2]

A child also suggested improvements as follows:

> *Well if it's in an app, on a phone or computer, or if you have an Apple Watch.* [Child 4]

The third and last theme identified was "FMSP Usage," with 2 subthemes of "Awareness and Education" and "The Correlation Between the MedSMA T Game and the FMSP." For "Awareness and Education," participants perceived the FMSP as a useful and beneficial tool for increasing their awareness and knowledge about medication safety practices such as safe disposal, storage, and use of opioids. Children also echoed the desire for similar tools for safety.

> *Knowing where else to dispose of them. It'd be easier for me to go to the pharmacy because I go there to pick up medications and I can just dispose of them through the pharmacy. And for some people working through it with their pharmacist or even a pharmacy tech for there's a lot of low health literacy out there.* [Parent 4]

'The Correlation Between the MedSMA T Game and the FMSP" subtheme indicated a perceived connection between the knowledge gained form the MedSMA T game and the principles of the FMSP.

> *…it made it easier because already have to think about medication storage and disposal because you'd already gain some knowledge.* [Parent 8]

Participants emphasized the educational value of the game and that learning from the game reinforced key safety messages, such as not sharing opioids and adhering to prescribed dosages. Table 4 presents themes, subthemes, and additional illustrative quotes for the FMSP.

XSL•FO
RenderX

Table . Themes and subthemes for the FMSP: parents and children data.

Theme and subtheme	Illustrative quotes
Routine medication management	
Medication safety practices at home	• "So, the kind of non-prescription ones are in the cabinet. Um, the prescription ones are in a, in a cabinet they're tucked away for just us and then when we are done using them or they need to be disposed of, I just kind of collect them in the safe until I know I'm gonna go somewhere to get rid of them so." [Parent 1] • "Normally know where it's kept and when to take it." [Child 3]
Communications with health care providers	• "Usually through MyChart is kind of, you know, the way I do it, if it's like more of an urgent matter then I would try calling them first if it's like Um, I don't know, I guess I haven't really had a circumstance or then real urgent. I would call like a nurse on call or something. But I, otherwise it's just been, like waiting for someone to revive you through like later." [Parent 6] • "Sometimes if it's like, just a medication question like, Um, whether or not they have to have it with food or, or things that sometimes I'll just call the pharmacy, and I'll talk with them about that versus the doctor. You know, if it's not clear on the paperwork that came home, the prescription or things, there's a discrepancy of what we thought the doctor said, versus the paperwork says, I'll usually reach out to the pharmacist." [Parent 1] • "I would ask about side effects, safety, dosage, time to take, storage." [Parent 3] • "Um probably just about like, how often I should be taking it and the side effects." [Child 7]
FMSP[a] usability	
FMSP implementation setting	• "I would say both the doctor and the pharmacy. Of course, you would then have it actually, have it on your phone or whatever. But like, information about it would be useful both at the doctor and the pharmacy." [Parent 1] • "I mean, I think especially if it's a high-risk medication, giving it to somebody when they pick up the medication from the pharmacy would be a good idea. And, and then I think maybe it's more for children and their medications and management." [Parent 8] • "Um, I would say, you know, probably more like the, like pediatric, like the checkups, like when doctors' offices, like when you would go in, when your kids are young like maybe first starting out like when, you know, medications like you might start getting some of those drugs prescribed." [Parent 3] • "At the pharmacy and doctor's office." [Child 2]
Recommendations for platform enhancement	• "I think it depends on the person. I can see some people needing to have paper pencil, but then other people would just lose their papers. So, having it electronically, and everyone has a computer in their pocket all the time now. So, something that is mobile friendly or maybe even you could, you could develop an app." [Parent 5] • "So, I keep coming at this as helping other people, but you know, maybe having something on my phone, that would just make it right there. It could be one of those, you know, how families have their family sharing. It can be like a shared calendar, but a shared medication management app or something. And then the app could also have alarms to remind people to take their medications." [Parent 2] • "On a website? Maybe like an app or something?" [Child 3] • "I think an app would be good." [Child 1]
FMSP usage	
Awareness and education	• "Yeah, I, I mean, I guess I, I was one of those that that you could flush it down the toilet. So, I mean, I guess, you know, just kind of knowing a little bit more like, what is safer for getting rid of." [Parent 2] • "Just to like keep it, like to use it. Make sure you're using it safely and how to keep yourself and others safe." [Child 4]

XSL•FO
RenderX

Theme and subtheme	Illustrative quotes
The correlation between the MedSMA T game and the FMSP	• "Um, I guess I would just maybe be a little more, not that I blocked it away, Um, but just, you know, maybe, I guess maybe a little more verbalizing, you know, like talking to the kids more about usage, and, and just safety." [Parent 5] • "Um, I mean, just the kind of some of the, I mean, like I already knew, but it's still reminder the common sense, like don't share it. Um, you know, just don't take, you know, more than prescribed because, you know, sometimes yeah, when you're in a lot of pain, you sometimes you don't, you don't care because it hurt, you know, so that you, you might, I guess just do it without thinking, so I think just, you know, staffing and realizing to think a little bit more about things." [Parent 3] • "I think it looks good. I think it considers like all the different things you would need to know, to just kind of have it laid out." [Child 2]

[a]FMSP: Family Medication Safety Plan.

Discussion

Principal Results

As the opioid crisis continues, opioid education and prescribing practices continue to be a highlight of investigation and improvement in EDs nationwide. This study sought to understand adolescents' and parents' perceptions of the MedSMA T Families intervention for both gameplay and the FMSP. This study addresses this gap by exploring their valuable and novel insights regarding this intervention. It also demonstrates the general acceptability and usefulness of the MedSMA T Families intervention in educating parents and children presenting to an ED about opioid safety. Considering the observations during gameplay, it is noteworthy that both parents and adolescents made errors, which provided significant opportunities for teaching and reinforcing important concepts related to opioid safety. These errors, rather than being viewed negatively, revealed crucial insights into the participants' knowledge gaps and misconceptions regarding safe opioid use. The findings from our study suggest that users found the game straightforward and smooth to play once they became familiar with the instructions. Participants were willing to have direct communication with health care providers and use digital tools to ensure appropriate medication management and opioid safety. Respondents' feedback suggested enhancements for the platform to be more user-friendly, flexible, and tech-savvy to support and facilitate medication management. Our results also indicate a strong desire for ongoing education regarding medications and especially opioid management.

Comparison With Prior Work

A previous study was conducted to explore ED staff perceptions on implementing the MedSMA T Families intervention in the ED setting and assessing its feasibility within the context of the ED environment. It was reported that the game is more fun and age-appropriate for adolescents compared with the current educational materials used in the ED. ED staff believed that an interactive design and real-life scenarios helped with creating a unique approach to involve adolescents, as they are interested in technology-based games. Game availability as a phone app and translation to other commonly used languages among non-English speaking patients such as Spanish and Hmong were suggested. Additionally, it was reported that the FMSP is a good resource for families at home as it provides detailed information regarding safe opioid use, disposal, and storage that is often missed in patient education [31].

Results are also consistent with a previous study on assessing adolescents' experiences and their suggestions for the use of the MedSMA T game. This identified that adolescents were successful in identifying the game's aim which is promoting opioid medication safety. They had positive impressions of the game's level, graphics, and characters. Participants recommended more instructions for how to play the game and create more game levels, which were implemented at that time [32].

In another study that explored parents' perceptions of the MedSMA T game, participants expressed positive reactions to the game characters and scenes depicted in the game. They thought scenarios were appropriate and realistic for adolescents and raised their awareness about safe strategies to store and dispose of opioids. Similar to the current results in our study, parents of adolescents who were prescribed opioids in ED reported some challenges in navigating through the game. The slow pace of the game was viewed as a significant difficulty for gameplay. However, parents still recommended implementing the game in both health care and nonhealth care settings [33].

Pharmacists' perceptions about the MedSMA T game have also been previously examined. Results highlighted the age-appropriate language of the game for adolescents and the use of relatable characters in the game design. Pharmacists valued the interactive nature of the gameplay, which led to active learning and recalling educational content [30].

Adolescents have also found the FMSP to be an acceptable and useful tool for opening conversations about opioids and other medications with their parents [34]. Parents' also viewed using the FMSP to promote proper opioid prescription practices with adolescents as demonstrated in this study, and reinforced the significance of a personalized family plan for safe medication management and practices at home [35]. Similarly, pharmacists' perceived the FSMP as a tool to encourage interactive opioid conversations between adolescents, families, and pharmacists as well. They believed that patients might use the FMSP as a

XSL•FO
RenderX

visual cue to help think of what questions families should ask pharmacists [36].

Our study reinforces previous results that demonstrate significant gaps in knowledge regarding opioid safety among adolescents and their families. However, novel themes were also identified that emerged specifically from the ED setting, which differs from other settings in prior studies. While there are similarities between other previous studies and our study, findings suggest significant positive impressions and key recommendations for future improvement of the intervention. Our study also suggests that parents and adolescents found the interactive nature of the game more engaging and suitable for their learning needs compared with traditional educational materials. This aligns with findings from ED staff that indicated a preference for interactive, technology-based educational tools [31]. Participants also highlighted important considerations for leveraging health care professionals to implement the intervention in their health care settings. This would facilitate effective communication with health care providers and the use of this intervention for families.

Serious games are used as the main educational tool in the MedSMA T Families intervention as well as the research study by Aneni et al [37] and Pendergrass et al [38]. Adolescents, a demographic population, who do not typically use traditional instructional materials for medicines, were the target audience for all 3 interventions discussed in these studies. The gameplay's interactive features encourage engagement and improve learning outcomes. In line with the goals of the MedSMA T game, the 2 other serious games also aimed to raise awareness about the risks of opioid misuse and safe storage and disposal. The 2 other studies used additional resources to strengthen the gaming experience, as the MedSMA T Families intervention incorporates the FMSP. These additional materials frequently support continuing education on safe medication practices and enhance family conversations.

Aneni et al [37] focus on older adolescents, possibly excluding the parental perspective [37], whereas the MedSMA T study specifically includes both adolescents and their parents, establishing a family-centered approach. The school-based health settings described by Pendergrass et al [38], on the other hand, place more emphasis on educational outcomes in school settings rather than on direct family interactions [38]. The MedSMA T game scenes and features encourage adolescents to learn by making mistakes and were specifically made to simulate real-life situations with the use of opioids. Aneni et al [37], on the other hand, investigates alternative design approaches or gaming mechanisms that prioritize role-playing or narrative components, altering how players engage with the content. Opioid medications are administered in a variety of contexts in these studies. Pediatric EDs, an acute care setting that might affect how information is received and the urgency attached to it, were used to test the MedSMA T Families intervention. Conversely, Aneni et al [37] and Pendergrass et al [38] explore settings such as schools, which can change the approach and interaction style influenced by the institutional environment and common teenage experiences.

Limitations

As child-parent dyad interviews were recorded via Zoom, participants might have been more likely to respond in a socially desirable manner, as opposed to nonvisual phone interviewing. Another limitation is that all the participants interviewed were from one academic ED, which may not represent the sentiments of all EDs across the United States. Given the unique characteristics of ED settings and their vital role in intervention implementation, further research using a more diverse sample of ED settings is warranted to gain organizational-related information and adaptations. Additionally, one of the limitations is to address the feedback that the game was appropriate for kids younger than 12 years which could be improved in future versions of the game.

Conclusions

In conclusion, participants found the MedSMA T intervention to be a useful educational tool for adolescents and families who were prescribed opioids from an ED about safe opioid use, storage, and disposal. By focusing on ED participants, we emphasize the unique context of the ED setting, which affects the attitudes and experiences of adolescents and their families. This study uniquely contributes to the literature by demonstrating how engaging and interactive educational tools, such as the MedSMA T game and the FMSP, can effectively support opioid safety education in a high-stress environment.

Participants offered insightful feedback on user experience and suggested future improvements regarding content and gameplay instructions. Participants appreciated the FMSP as an instructional tool. Future improvements were recommended, including making it available on mobile devices. There is a clear need for improved digital tools that promote opioid safety and highlight the significance of knowledge, instruction, and useful tactics for efficiently managing opioids and other medications in the home.

Acknowledgments

The authors thank Tyler J McCarthy for helping with the data collection. Funding for this project was provided by the University of Wisconsin-Madison (UW) School of Medicine and Public Health from the Wisconsin Partnership Program (WPP 5129) through a grant to the UW Institute for Clinical and Translational Research (UW ICTR). ICTR also received funding from National Institutes of Health National Center for Advancing Translational Sciences (NIH-NCATS) Clinical and Translational Science Award (CTSA; 1UL1TR002373).

XSL•FO
RenderX

Data Availability

All data generated or analyzed during this study are included in this published article and its supplementary information files in Multimedia Appendices 1-3. The MedSMA T game is downloadable and can be made available on request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Family Medication Safety Plan.
[PDF File, 216 KB - games_v13i1e68814_app1.pdf]

Multimedia Appendix 2
Parent interview questions.
[PDF File, 168 KB - games_v13i1e68814_app2.pdf]

Multimedia Appendix 3
Child interview questions.
[PDF File, 194 KB - games_v13i1e68814_app3.pdf]

References

1. Michelson KA, Hudgins JD, Monuteaux MC, Bachur RG, Finkelstein JA. Cardiac arrest survival in pediatric and general emergency departments. Pediatrics 2018 Feb;141(2):e20172741. [doi: 10.1542/peds.2017-2741] [Medline: 29367204]
2. Remick K, Gausche-Hill M, Joseph MM, et al. Pediatric readiness in the emergency department. Pediatrics 2018 Nov;142(5):e20182459. [doi: 10.1542/peds.2018-2459] [Medline: 30389843]
3. Remick KE, Bartley KA, Gonzales L, MacRae KS, Edgerton EA. Consensus-driven model to establish paediatric emergency care measures for low-volume emergency departments. BMJ Open Qual 2022 Jul;11(3):e001803. [doi: 10.1136/bmjoq-2021-001803] [Medline: 35803615]
4. Auerbach M, Brown L, Whitfill T, et al. Adherence to pediatric cardiac arrest guidelines across a spectrum of fifty emergency departments: a prospective, in situ, simulation-based study. Acad Emerg Med 2018 Dec;25(12):1396-1408. [doi: 10.1111/acem.13564] [Medline: 30194902]
5. Auerbach M, Whitfill T, Gawel M, et al. Differences in the quality of pediatric resuscitative care across a spectrum of emergency departments. JAMA Pediatr 2016 Oct 1;170(10):987-994. [doi: 10.1001/jamapediatrics.2016.1550] [Medline: 27570926]
6. Motov SM, Vlasica K, Middlebrook I, LaPietra A. Pain management in the emergency department: a clinical review. Clin Exp Emerg Med 2021 Dec;8(4):268-278. [doi: 10.15441/ceem.21.161] [Medline: 35000354]
7. Selby S, Anderson J, Silva LO, et al. 234 Epidemiology of emergency department pediatric and adolescent opioid use across the United States. Ann Emerg Med 2021 Oct;78(4):S95. [doi: 10.1016/j.annemergmed.2021.09.246]
8. Champagne-Langabeer T, Cardenas-Turanzas M, Ugalde IT, et al. The impact of pediatric opioid-related visits on U.S. emergency departments. Children (Basel) 2022 Apr 7;9(4):524. [doi: 10.3390/children9040524] [Medline: 35455568]
9. Determination that a public health emergency exists. US Department of Health and Human Services. URL: https://www.hhs.gov/sites/default/files/opioid%20PHE%20Declaration-no-sig.pdf [accessed 2024-08-01]
10. National Academies of Sciences, Engineering, and Medicine, Health and Medicine Division, Board on Health Sciences Policy, Committee on Pain Management and Regulatory Strategies to Address Prescription Opioid Abuse. Evidence on strategies for addressing the opioid epidemic. In: Phillips JK, Ford MA, Bonnie RJ, editors. Pain Management and the Opioid Epidemic: Balancing Societal and Individual Benefits and Risks of Prescription Opioid Use: National Academies Press (US); 2017. URL: https://www.ncbi.nlm.nih.gov/books/NBK458653/ [accessed 2024-10-20]
11. Consumption of prescription opioids for pain: a comparison of opioid use in the united states and other countries. : CRS Products (Library of Congress); 2021 URL: https://crsreports.congress.gov/product/pdf/R/R46805 [accessed 2024-10-21]
12. Lee B, Zhao W, Yang KC, Ahn YY, Perry BL. Systematic evaluation of state policy interventions targeting the US opioid epidemic, 2007-2018. JAMA Netw Open 2021 Feb 1;4(2):e2036687. [doi: 10.1001/jamanetworkopen.2020.36687] [Medline: 33576816]
13. National Academies of Sciences, Engineering, and Medicine, Health and Medicine Division, Board on Health Sciences Policy, Committee on Pain Management and Regulatory Strategies to Address Prescription Opioid Abuse. Pain management and the intersection of pain and opioid use disorder. In: Phillips JK, Ford MA, Bonnie RJ, editors. Pain Management and the Opioid Epidemic: Balancing Societal and Individual Benefits and Risks of Prescription Opioid Use: National Academies Press (US); 2017. URL: https://www.ncbi.nlm.nih.gov/books/NBK458655/ [accessed 2024-10-20]

XSL•FO
RenderX

14. Fischer B, Rehm J. Revisiting the “paradigm shift” in opioid use: developments and implications 10 years later. Drug Alcohol Rev 2018 Apr;37 Suppl 1(suppl 1):S199-S202. [doi: 10.1111/dar.12539] [Medline: 28334467]
15. Warren LK, Adams J, Bobashev G. Trends in opioid misuse among individuals aged 12 to 21 years in the US. JAMA Netw Open 2023 Jun 1;6(6):e2316276. [doi: 10.1001/jamanetworkopen.2023.16276] [Medline: 37261827]
16. Dowell D, Ragan KR, Jones CM, Baldwin GT, Chou R. CDC clinical practice guideline for prescribing opioids for pain - United States, 2022. MMWR Recomm Rep 2022 Nov 4;71(3):1-95. [doi: 10.15585/mmwr.rr7103a1] [Medline: 36327391]
17. McCabe SE, West BT, Veliz P, McCabe VV, Stoddard SA, Boyd CJ. Trends in medical and nonmedical use of prescription opioids among US adolescents: 1976-2015. Pediatrics 2017 Apr;139(4):e20162387. [doi: 10.1542/peds.2016-2387] [Medline: 28320868]
18. Wilson JD, Abebe KZ, Kraemer K, et al. Trajectories of opioid use following first opioid prescription in opioid-naive youths and young adults. JAMA Netw Open 2021 Apr 1;4(4):e214552. [doi: 10.1001/jamanetworkopen.2021.4552] [Medline: 33885777]
19. Bonar EE, Coughlin L, Roche JS, et al. Prescription opioid misuse among adolescents and emerging adults in the United States: a scoping review. Prev Med 2020 Mar;132:105972. [doi: 10.1016/j.ypmed.2019.105972] [Medline: 31904397]
20. Meisel ZF, Lupulescu-Mann N, Charlesworth CJ, Kim H, Sun BC. Conversion to persistent or high-risk opioid use after a new prescription from the emergency department: evidence from Washington medicaid beneficiaries. Ann Emerg Med 2019 Nov;74(5):611-621. [doi: 10.1016/j.annemergmed.2019.04.007] [Medline: 31229392]
21. Barnett ML, Olenski AR, Jena AB. Opioid-prescribing patterns of emergency physicians and risk of long-term use. N Engl J Med 2017 Feb 16;376(7):663-673. [doi: 10.1056/NEJMsa1610524] [Medline: 28199807]
22. Teen newsletter: opioids. Centers for Disease Control and Prevention. 2022. URL: https://www.cdc.gov/museum/education/newsletter/2022/july/index.html [accessed 2024-10-21]
23. Groenewald CB. Opioid-prescribing patterns for pediatric patients in the United States. Clin J Pain 2019 Jun;35(6):515-520. [doi: 10.1097/AJP.0000000000000707] [Medline: 30985396]
24. McCabe SE, Schulenberg J, McCabe VV, Veliz PT. Medical use and misuse of prescription opioids in US 12th-grade youth: school-level correlates. Pediatrics 2020 Oct;146(4):e20200387. [doi: 10.1542/peds.2020-0387] [Medline: 32913134]
25. Jones CM, Clayton HB, Deputy NP, et al. Prescription opioid misuse and use of alcohol and other substances among high school students - youth risk behavior survey, United States, 2019. MMWR Suppl 2020 Aug 21;69(1):38-46. [doi: 10.15585/mmwr.su6901a5] [Medline: 32817608]
26. Shanahan L, Hill SN, Bechtiger L, et al. Prevalence and childhood precursors of opioid use in the early decades of life. JAMA Pediatr 2021 Mar 1;175(3):276-285. [doi: 10.1001/jamapediatrics.2020.5205] [Medline: 33369615]
27. Kane JM, Colvin JD, Bartlett AH, Hall M. Opioid-related critical care resource use in US children’s hospitals. Pediatrics 2018 Apr;141(4):e20173335. [doi: 10.1542/peds.2017-3335] [Medline: 29507166]
28. Abraham O, McCarthy TJ. An overview of the CRoME Lab’s MedSMA T families intervention for promoting opioid medication safety among adolescents and parents. Innov Pharm 2023;14(3):11. [doi: 10.24926/iip.v14i3.5678] [Medline: 38487388]
29. Abraham O, Rosenberger C, Tierney K, Birstler J. Investigating the use of a serious game to improve opioid safety awareness among adolescents: quantitative study. JMIR Serious Games 2021 Dec 23;9(4):e33975. [doi: 10.2196/33975] [Medline: 34747712]
30. Abraham O, Slonac E, Paulsen Z. Pharmacists’ perspectives on MedSMA T: a serious game to educate youth about opioid safety. J Am Pharm Assoc (2003) 2023;63(4):1087-1094. [doi: 10.1016/j.japh.2023.04.020] [Medline: 37116795]
31. Nadi S, Hurst IA, Abraham O. Using MedSMA T families for opioid safety education: emergency department staff perspective. Pediatr Open Sci 2025 Mar 1;1(1):1-9. [doi: 10.1542/pedsos.2024-000172]
32. Abraham O, Koeberl CR, McCarthy TJ. MedSMA T adventures in PharmaCity game: youth experiences and recommendation for use in opioid safety education. Pharmacy (Basel) 2023 Sep 10;11(5):143. [doi: 10.3390/pharmacy11050143] [Medline: 37736915]
33. Abraham O, Nixon GA, Seitz LL. Parents’ perceptions of a serious game for educating families on prescription opioid safety: qualitative pilot study of MedSMARxT: adventures in PharmaCity. JMIR Serious Games 2023 Sep 12;11:e49382. [doi: 10.2196/49382] [Medline: 37698899]
34. Abraham O, Peterson KA, Thao SS, McCarthy TJ. Adolescents’ assessment of a family medication safety plan for opioid prescriptions: a qualitative usability study. J Am Pharm Assoc (2003) 2023;63(6):1768-1775. [doi: 10.1016/j.japh.2023.08.021] [Medline: 37673284]
35. Abraham O, Klubertanz GC, Maday EN. Parents’ perspectives on using a family medication safety plan designed to promote appropriate opioid prescription practices with adolescents. J Am Pharm Assoc (2003) 2024;64(3):102024. [doi: 10.1016/j.japh.2024.01.020] [Medline: 38309416]
36. Abraham O, Peters J, Peterson KA. Characterizing pharmacist perspectives on utilizing a personalized family medication safety plan for opioid education with adolescents and parents. Pharmacy (Basel) 2023 Jan 24;11(1):22. [doi: 10.3390/pharmacy11010022] [Medline: 36827660]

XSL•FO
RenderX

37. Aneni K, Fernandes CSF, Hoerner LA, Szapary C, Pendergrass Boomer TM, Fiellin LE. A video game intervention to prevent opioid misuse among older adolescents: development and preimplementation study. JMIR Serious Games 2023 Nov 3;11:e46912. [doi: 10.2196/46912] [Medline: 37921851]
38. Pendergrass Boomer TM, Hoerner LA, Fernandes CSF, et al. A digital health game to prevent opioid misuse and promote mental health in adolescents in school-based health settings: protocol for the PlaySmart game randomized controlled trial. PLoS ONE 2023;18(9):e0291298. [doi: 10.1371/journal.pone.0291298] [Medline: 37683047]

Abbreviations

ED: emergency department
FMSP: Family Medication Safety Plan

Edited by A Coristine; submitted 14.11.24; peer-reviewed by AL Drendel, LI Kelley-Quon, M Agbede; revised version received 25.04.25; accepted 25.04.25; published 26.06.25.

Please cite as:
Abraham O, Nadi S, Hurst I
Adolescents' and Parents' Perspectives on Using the MedSMARxT Families Intervention in Emergency Departments for Opioid Medication Safety Education: Mixed Methods Study
JMIR Serious Games 2025;13:e68814
URL: https://games.jmir.org/2025/1/e68814
doi:10.2196/68814

© Olufunmilola Abraham, Sara Nadi, Irene Hurst. Originally published in JMIR Serious Games (https://games.jmir.org), 26.6.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Factors Influencing Virtual Reality as a Distraction Tool for Venipuncture in Children: Observational Pilot Feasibility Study

Chris Worth[1]; Leyi Yang[1]; Catherine Fullwood[2,3]; Indraneel Banerjee[1,4]

[1]Department of Paediatric Endocrinology, Royal Manchester Children's Hospital, Oxford Road, Manchester, United Kingdom

[2]Research and Innovation, Manchester University NHS Foundation Trust, Manchester, United Kingdom

[3]Centre for Biostatistics, Division of Population Health, Health Services Research and Primary Care, School of Health Sciences, University of Manchester, Manchester, United Kingdom

[4]Faculty of Biology, Medicine and Health, University of Manchester, Manchester, United Kingdom

Corresponding Author:
Chris Worth
Department of Paediatric Endocrinology, Royal Manchester Children's Hospital, Oxford Road, Manchester, United Kingdom

Abstract

Background: Virtual reality (VR) is increasingly used as a distraction tool for painful procedures in children. Studies have shown variable benefit but have not identified factors to maximize utility.

Objective: This study aimed to undertake a feasibility study to investigate factors influencing virtual reality headset (VRH) utility for venipuncture and cannulation.

Methods: Children admitted as inpatients were recruited and given a VRH during anticipated venipuncture or cannulation. Feedback from participants, parents, and operators was obtained through questionnaires and the Wong-Baker Faces Scale (WBFS).

Results: Thirteen children of a target 32 (41%), aged a median of 7 (range 5-12) years, were recruited to the study; 9 (69%) parents and 7 (54%) participants reported a positive VR experience, found VRH comfortable, and wanted repeat application for future venisection or cannulation. However, patient recruitment was suboptimal at 41% (binomial 95% CI 24% - 59%) of the target, as busy operators undertaking venisection or cannulation on eligible patients were unable to spare time for VRH use during the COVID-19 pandemic. The preprocedure time spent with VRH was associated with participants' positive experience of VR distraction (median 15, IQR 2.5 - 50 vs 180, IQR 120-450) seconds (P=.02). Five (38.4%) participants removed the VRH prior to procedure; these were relatively young compared to those who continued VRH (median 6, IQR 5.00 - 7.00 vs 10, IQR 6.75 - 12.00 years), suggesting better acceptance of VR in older children. There was no significant difference in WBFS pain ratings before (median 0, IQR 0 - 10) or after the procedure (median 0, IQR 0 - 6), with many children choosing 0 ("no hurt") at initial assessment. By contrast, parent and doctor or phlebotomist responses indicated that VR reduced pain and anxiety (n=9, 69%), in agreement with participant perception (n=7, 54%; Cohen κ=0.68).

Conclusions: VR as a distraction tool in children is influenced by age and preprocedure familiarity, suggesting that the optimal use is in older children with greater cognitive and emotional maturity. Multidimensional feedback from participants, parents, and investigators should be obtained to test the true efficacy of VR in future studies.

Trial Registration: ClinicalTrials.gov NCT04472507; https://clinicaltrials.gov/study/NCT04472507

(JMIR Serious Games 2025;13:e66656) doi:10.2196/66656

KEYWORDS

virtual reality; pain; VR; computer-generated simulations; simulations; digital worlds; virtual environments; digital environments; virtual tour; venipuncture; children; VR headset; head-mounted display; cannulation; inpatients; questionnaires; pediatric care; phlebotomy

Introduction

Overview

Virtual reality (VR) is a relatively new technology that is increasingly used as a distraction tool due to its immersive nature and capacity to engage multiple senses simultaneously [1]. VR utilizes the notion that attention is finite and can be diverted from pain and discomfort and therefore serves as effective cognitive distraction [2]. VR provides immersive engagement, isolating users from the real-world environment and associated pain and anxiety [1].

VR has been used as a nonpharmacological adjunct in pediatric care in a variety of clinical scenarios [3-9]. Children are often fearful of invasive procedures such as venisections and intravenous cannulation, which may need to be undertaken at frequent intervals depending on clinical needs; these procedures

XSL•FO
RenderX

are necessary but reinforce and amplify pain and anxiety, adversely impacting on their hospital experience. Reduction of pain and anxiety is essential to continue effective clinical management for patient benefit and is currently provided by a range of pharmacological and nonpharmacological methods, including VR.

Studies have shown variable benefit of VR technology as a standalone tool, in comparison with alternative distraction techniques. However, most studies seeking to establish efficacy evaluated virtual reality headset (VRH) use in grouped study participants without qualification based on patient choice or other factors. A recent systematic review and meta-analysis [10] collated 11 randomized controlled trials in children under 12 years of age to conclude that VR was more useful than other nonpharmacological methods to relieve pain and anxiety for various procedures in children. Another recent meta-analysis [11] of 10 randomized controlled trials evaluated the efficacy of VR use in children aged 3 to 12 years to conclude favorably on the efficacy and benefit in pediatric venipuncture. However, results from such studies provide aggregated outcomes and are unable to investigate ancillary factors that encourage VRH adherence or dissuade participation. Therefore, such unqualified and indiscriminate evaluation may mask true benefits by failing to explore factors influencing the use of VR to maximize its distraction potential. There is a need to explore individual patient factors such as age, ethnicity, and prior experience of VR in additional studies that may qualify and add further strength to the adoption of this new technology in children's health care.

Furthermore, the majority of previous VR studies have offered VRH to children during outpatient and planned procedures. The feasibility of VRH in the inpatient setting of a busy hospital, with less advance notice of procedures, has not been evaluated. It is important to investigate VRH use in high-pressure clinical areas to provide a more realistic and nuanced estimate of clinical benefit. We have undertaken a feasibility study to assess the practicalities of recruiting to such a study and to provide exploratory information around factors influencing VRH use in busy inpatient clinical areas in a children's hospital.

Aims

We aimed to explore the feasibility of VRH use and factors influencing VRH utility for venipuncture and cannulation within the inpatient pediatric setting.

Methods

Ethical Considerations

We undertook an observational, feasibility study (UK ethics approval 20/NE/0072 by the Health Research Authority; ClinicalTrials.gov NCT04472507) during the 2020 - 2021 COVID-19 pandemic to review participant, parent and operator (doctor and phlebotomist) perspectives and investigate the optimal utility of VR distraction for inpatient venipuncture and cannulation. Study information was provided to ward-based operators undertaking venipuncture and cannulation. Potential patients and their parents also received study information prior to the procedure but the timing of the procedure was not confirmed, as it was dependent on clinical circumstances and operator availability. As patients were under 12 years old, parents (1 parent per child) consented for their participation in the study. Patients older than 6 years provided additional assent. Data were anonymized prior to write up. No compensation was provided to participants.

Study Design and Analysis

Participants and parents (1 parent per child) were selected from all pediatric inpatients aged 5 - 12 years, in a large tertiary hospital in the United Kingdom. All ethnic groups were eligible for participation. However, considering the unavailability of interpreters for various languages at the time of the COVID-19 pandemic, only English-speaking children and their families were selected. Operators performing inpatient venipuncture and cannulation informed the investigator team of eligible participants less than 6 hours before the procedure.

Prior experience of VRH was considered if a child had used a VRH in a nonclinical setting outside the hospital over the previous year. Information on prior experience was recorded by the investigator in a questionnaire administered to the parent. Experience of VRH use by a parent did not count as prior experience.

Participants were shown demonstrations of the OculusGo (Meta Reality Labs) VRH (and a choice of age-appropriate, single hand–operated software) and spent a median of 2 (range 0.1 - 25.0) minutes using the VRH prior to their procedure, which lasted for 3 (range 0.1 - 16.0) minutes. Each VRH was thoroughly cleaned after each procedure, which increased total procedural time. For the children's feedback, the Wong-Baker Faces Scale (WBFS; scored from 0 - 10) was chosen to record perceived pain before and after the procedure. Parent and operator feedback perspectives were captured by separate questionnaires and free text. The study design was principally quantitative and descriptive to provide the basis for the design of future larger trials using VR technologies. Therefore, no qualitative data were gathered in this feasibility study.

A target of 32 children to be recruited over a 4-week period was calculated to give 95% CI bounds of 56.6% - 88.5% (Clopper-Pearson exact) around the acceptability target of 75%. Statistical analysis was undertaken using IBM SPSS statistics, version 29.0 (descriptive data), and the statistical computing package R, version 4.2.1 (analytical data), without imputation for missing data. All data were presented with appropriate descriptive statistics: absolute values and percentages, means and SDs, or medians and IQRs or ranges. The Wilcox signed-rank test was used post hoc to assess differences in the nonparametric continuous variables—time and age—with subgroups defined by experience, while the demographic variables—gender and ethnicity—were assessed against experience using a Fisher exact test.

Results

Thirteen children were recruited to the study during 3 week-long recruitment periods between January and May 2021. The median age at recruitment was 7 (range 5-12) years with a greater proportion of girls (n=8, 62%). Most participants (n=9, 69%) were White and of European ethnic origin, 2 were Black and

XSL•FO
RenderX

of African origin, and 2 were ethnically South Asian (Table 1); all children and their families spoke English.

Table . Demographic information of all participants in the study.

		Value (n=13)
Age years, median (range)		7 (5-12)
Sex (female), n (%)		8 (62)
Ethnicity and race, n (%)		
	Asian Muslim	1 (8)
	Bangladeshi	1 (8)
	Black African	2 (15)
	White European	9 (69)
School stage, n (%)		
	Infant	4 (31)
	Junior	5 (38)
	Senior	4 (31)
Used virtual reality before, n (%)		6 (46)

Participant recruitment was challenging and considerably less than expected, with reluctance from busy operators who were concerned at introducing additional delay for eligible patients. Their feedback indicated the requirement for greater availability of VRH and advance planning. Recruitment peaked at 41% of the target (binomial 95% CI 24% - 59%), despite the addition of an extra week of recruitment beyond those initially planned.

Child-reported WBFS pain ratings did not demonstrate a change between preprocedure (median 0, IQR 0 - 10) and postprocedure scores (median 0, IQR 0 - 6), as many children chose 0 ("no hurt") at the initial assessment. Nonetheless, more than half of participants (n=7, 54%) reported a positive experience, felt VRH was comfortable, and wanted to repeat the experience. Comments included "I loved it," "it was really great" and "it was sick." No child felt unwell during the procedure. There were no differences in participants' VR distraction experience by either sex (P=.82) or ethnicity (P=.05). However, preprocedure time spent with VRH was found to be associated with participants' positive experience of VR distraction (median 15, IQR 2.5 - 50 vs 180, IQR 120-450 seconds; Wilcox signed-rank P=.02).

Although the VRH was provided for all participants, 5 (38%) children removed the VRH during the procedure citing that they "could not see mother" and "wanted to see arm" and that it "made them more anxious" and was "associated with blood sample." Those who removed the VRH were relatively young compared to those who continued (median 6, IQR 5.00 - 7.00 vs 10, IQR 6.75 - 12.00 years). These children were all of primary school age, whereas those who engaged with the VRH were spread across both primary and secondary school age groups (UK school year for those who removed the VRH: median 2, IQR 1.00 - 3.00 vs UK school year for those who remained engaged: median 6, IQR 3.75 - 7.25; P=.03), suggesting better acceptability of VR in older children.

Over half of parents (n=8, 60%) perceived a reduction in both pain and anxiety in their children using VR distraction. This showed agreement with overall participant-reported positive experience and comfort and operator perception, with only 1 child reporting enjoyment contradicting operators' perceptions. Parents reported a low perceived median pain score of 2 (range 0 - 6), showing a moderate correlation (Cohen κ=0.57) with pre-VR participant-reported scores, although the same correlation was not seen postprocedure. Overall, parents felt their children had a positive VR experience during the procedure, in agreement with the participants' perceptions (n=9, 69% and n=7, 54%, respectively; Cohen κ=0.68). Operators perceived VR to have reduced procedure time in only 3 (23%) children; however, they also noted that a majority of participants and parents had a positive experience (n=9, 69%).

Discussion

Principal Findings

Our exploratory results for the use of VR as a distraction tool for venipuncture and cannulation in children reported a positive experience in more than half of participants. While VRH use did not reduce procedure time, the overall experience of both children and their parents suggests benefits from use without adverse effects. While previous studies have reported aggregated benefits, our study has uncovered participant factors that have not been previously described, which are important for clinical application and for future study design. Importantly, our findings demonstrate the need for familiarization with the VRH and software prior to an invasive procedure. Our study also notes that more than a third of children removed their VRH; these children were relatively young and had greater parental attachment, suggesting that younger children and those with greater emotional needs may require additional strategies to remain connected in the VR environment to optimize VRH use. While children had individual reasons for removing their VRH, specific quotes from parent suggest reduced suitability in those with greater cognitive and emotional immaturity. By contrast,

XSL•FO
RenderX

the acceptance and success of VRH in older children with greater psychological maturity is reassuring.

In previously published studies, VRH has been evaluated in all grouped participants [2,3,5,6] with suboptimal distraction utility, leading to the inevitable conclusion that further studies are required to achieve the full benefit [5]. Systematic reviews and meta-analyses have demonstrated aggregated benefit but have not identified individual factors that increase VR success [1,4]. Our study identified specific participant factors that suggest targeting VR for maximal gain: older age and prior experience and greater familiarity with a VRH. A multifaceted feedback approach, studying perspectives of parents, operators, and children, may assist the appreciation of different dimensions of pain, anxiety, and experience in VR distraction.

The study was designed as a feasibility study to assess the acceptability of the concept alongside a limited number of key variables in preparation for a future study in larger groups of children. Therefore, it has not been possible to assess the whole range of factors impacting VRH uptake and use. We expect to increase both the number of participants and the number of variables in a comprehensive study to critically test nuances around benefit and risk. Such a study will require an adequate sample size powered to detect differences among several variables, be standardized for VR software and content, and provide a preprocedure time for familiarization with VRH as well as consideration of additional operator time and how to minimize this in the inpatient setting.

We experienced reluctance from operators to engage with this study, citing the additional time required for VRH use. This extra time pressure must be considered when evaluating the feasibility of VRH in future inpatient pediatric settings. We also noted unusual responses from participants on the WBFS, with some children recording 0 (“no hurt”) at the initial assessment. While the use of the scale is justified in this study, we recognize its limitations, particularly in young children with learning disabilities. As the WBFS has a choice of only a few options, there may be limited sensitivity compared to other scales with a larger number of options. Additionally, children may have misinterpreted the crying faces on the WBFS to indicate emotional affect, instead of pain intensity [12,13].

Our study is limited by size owing to the reduced uptake of participants during the COVID-19 pandemic as well as the reluctance on behalf of operators undertaking venipuncture and cannulation services in a busy hospital ward setting. During the pandemic, patient-facing time was generally expedited to minimize potential spread of infection; this may have contributed to hesitance and reluctance of recruitment in both patients and operators. Therefore, conclusions drawn from this relatively small exploratory study will need to be carefully considered and tested for replication in larger cohorts alongside some preliminary qualitative data. Nonetheless, our data are important to assist the design of larger VR distraction trials to allow the redundancy of eligible participants being enrolled in the study and to optimize VRH use in high volume clinical settings.

The study is also constrained by the lack of a qualitative component that may have identified common themes from this relatively small sample. However, as the study was exploratory, we chose to focus on quantitative variables before designing a larger study with both quantitative and qualitative arms.

Despite study constraints, our results have uncovered real-world problems that will need to be addressed in future studies. Our study design did not provide a mandatory familiarization period, resulting in a rejection by some children, the majority of whom were under 7 years old. Therefore, for future studies, it would be important to stipulate a mandatory period to allow children and their parents to become familiar with the VRH and software, prior to study participation. Additionally, study design should consider strategies to ensure young children and those with greater parental attachment and emotional needs remain connected in the VR environment to optimize VRH use.

We acknowledge several sources of bias that may have influenced study findings. These include recruitment of relatively young patients, children with greater parental attachment, and those with or without previous experience of VRH use. It is possible that children and families with greater interest and experience of modern digital technology may have chosen to participate, thereby reducing the participation of those with a more cautious attitude to digital adoption. Although parental reports were mostly consistent with their children’s report, we recognize the possibility of reporting bias that may have influenced study findings. Another source of bias lies within the study population, the diversity of which was limited by COVID-19 restrictions on translators. We expect these biases to be addressed in the future design of more comprehensive studies.

We accept that VR technologies are expensive and not accessible to several families or even available in all hospitals. Some families may be inherently cautious of digital technologies and may resent the adoption of immersive distraction tools. VRH rejections in our study suggest that cultural and socioeconomic nuances may need to be addressed in future study designs and clinical application.

Conclusion

Specific participant factors favor VR distraction in children. Future VR studies should target older children and those with greater cognitive and emotional maturity as the technology is more readily adopted in such groups. Furthermore, studies should introduce mandatory preprocedure VR familiarization time periods to ensure the standardization and optimal uptake of VRH. They should also include investigator and parent feedback in addition to child-provided pain ratings and questionnaires and ensure sufficient operational times are factored in the inpatient setting.

XSL•FO
RenderX

Acknowledgments

This study was funded by a grant from Research & Innovation, Manchester University Hospitals Foundation Trust.

Authors' Contributions

CW and IB designed, organized and led the study. LY conducted the study. CF provided statistical analysis for the study. All authors reviewed and approved the final version of the manuscript.

Conflicts of Interest

None declared.

References

1. Buche H, Michel A, Blanc N. Use of virtual reality in oncology: from the state of the art to an integrative model. Front Virtual Real 2022 Aug 4;3:894162. [doi: 10.3389/frvir.2022.894162]
2. Gupta A, Scott K, Dukewich M. Innovative technology using virtual reality in the treatment of pain: does it reduce pain via distraction, or is there more to it? Pain Med 2018 Jan 1;19(1):151-159. [doi: 10.1093/pm/pnx109] [Medline: 29025113]
3. Olbrecht VA, O'Conor KT, Williams SE, et al. Transient reductions in postoperative pain and anxiety with the use of virtual reality in children. Pain Med 2021 Nov 26;22(11):2426-2435. [doi: 10.1093/pm/pnab209] [Medline: 34175959]
4. Czech O, Rutkowski S, Kowaluk A, Kiper P, Malicka I. Virtual reality in chemotherapy support for the treatment of physical functions, fear, and quality of life in pediatric cancer patients: a systematic review and meta-analysis. Front Public Health 2023;11:1039720. [doi: 10.3389/fpubh.2023.1039720] [Medline: 37124795]
5. Fabricant PD, Gross PW, Mackie AT, et al. Virtual reality distraction is no better than simple distraction techniques for reducing pain and anxiety during pediatric orthopaedic outpatient procedures: a randomized controlled trial. Clin Orthop Relat Res 2024 May 1;482(5):854-863. [doi: 10.1097/CORR.0000000000002889] [Medline: 37939199]
6. Gruber N, Shemesh-Iron M, Kraft E, et al. Virtual reality's impact on children with type 1 diabetes: a proof-of-concept randomized cross-over trial on anxiety, pain, adherence, and glycemic control. Acta Diabetol 2024 Feb;61(2):215-224. [doi: 10.1007/s00592-023-02195-9] [Medline: 37845502]
7. Ozturk CS, Toruner EK. Effectiveness of virtual reality in anxiety and pain management in children and adolescents receiving cancer treatment: a systematic review and meta-analysis of randomized controlled trials. J Med Syst 2023 Oct 10;47(1):37815610. [doi: 10.1007/s10916-023-01995-4] [Medline: 37815610]
8. Rygh L, Johal S, Johnson H, Karlson CW. Virtual reality for pediatric oncology port-a-cath access: a pilot effectiveness study. J Pediatr Hematol Oncol Nurs 2023;40(6):379-385. [doi: 10.1177/27527530221147875] [Medline: 37386775]
9. van den Berg S, Hoogeveen MO, van Winden TMS, Chegary M, Genco MS, Jonkman NH. Virtual reality hypnosis for needle-related procedural pain and fear management in children: a non-inferiority randomized trial. Eur J Pediatr 2023 Oct;182(10):4421-4430. [doi: 10.1007/s00431-023-05116-8] [Medline: 37486409]
10. Niaz F, Tariq S, Ahmed AM, et al. Virtual reality for pain and anxiety management in children undergoing venipuncture procedures: a systematic review and meta-analysis. Global Pediatrics 2023 Jun;4:100060. [doi: 10.1016/j.gpeds.2023.100060]
11. Wei Q, Sun R, Liang Y, Chen D. Virtual reality technology reduces the pain and anxiety of children undergoing vein puncture: a meta-analysis. BMC Nurs 2024 Aug 7;23(1):541. [doi: 10.1186/s12912-024-02184-5] [Medline: 39112987]
12. Miró J, Castarlenas E, de la Vega R, et al. Validity of three rating scales for measuring pain intensity in youths with physical disabilities. Eur J Pain 2016 Jan;20(1):130-137. [doi: 10.1002/ejp.704] [Medline: 25833415]
13. Tomlinson D, von Baeyer CL, Stinson JN, Sung L. A systematic review of faces scales for the self-report of pain intensity in children. Pediatrics 2010 Nov;126(5):e1168-e1198. [doi: 10.1542/peds.2010-1609] [Medline: 20921070]

Abbreviations

VR: virtual reality
VRH: virtual reality headset
WBFS: Wong-Baker Faces Scale

XSL•FO
RenderX

Edited by M Schmid; submitted 19.09.24; peer-reviewed by A Khalili, D Poddighe, E Chan, J Camacho-Cruz; revised version received 04.03.25; accepted 07.03.25; published 22.05.25.

Please cite as:
Worth C, Yang L, Fullwood C, Banerjee I
Factors Influencing Virtual Reality as a Distraction Tool for Venipuncture in Children: Observational Pilot Feasibility Study
JMIR Serious Games 2025;13:e66656
URL: https://games.jmir.org/2025/1/e66656
doi:10.2196/66656

© Chris Worth, Leyi Yang, Catherine Fullwood, Indraneel Banerjee. Originally published in JMIR Serious Games (https://games.jmir.org), 22.5.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Active Video Games Training for Older Adults: Comparative Study of User Experience, Workload, Pleasure, and Intensity

Néva Béraud-Peigné[1,2], BSc, MSc; Alexandra Perrot[1], BSc, MSc, PhD; Pauline Maillot[2], BSc, MSc, PhD

[1]Université Paris Saclay, Complexité, Innovation et Activités Motrices et Sportives, 335, rue Pierre de Coubertin, Bures-sur-Yvette, France

[2]Université Paris Cité, Institut des Sciences du Sport-Santé de Paris, Paris, France

Corresponding Author:
Néva Béraud-Peigné, BSc, MSc
Université Paris Saclay, Complexité, Innovation et Activités Motrices et Sportives, 335, rue Pierre de Coubertin, Bures-sur-Yvette, France

Abstract

Background: Given the appeal of active video games (AVG), many tools are now being used for combined training in older adults. However, there is a lack of comparative data to determine which type of AVG is better suited to older adults.

Objective: The purpose of this study was to compare user experience (UX), workload, pleasure, and intensity of three different experiences: (1) an Immersive and Interactive Wall Exergame (I2WE), (2) a consumer device (SWITCH), and (3) a combination of video games and physical stimulation (biking and videogaming, BIKE-VG) for older adults. I2WE and SWITCH are categorized as Moving While Thinking training, meaning that the cognitive task is integrated into the motor or physical task. In contrast, BIKE-VG is categorized as Thinking While Moving training, where the cognitive and motor or physical tasks are not interconnected. The nature of the cognitive, physical, and motor combinations also differentiates them. I2WE is multi-domain training, while BIKE-VG is physical-cognitive training, and SWITCH is motor-cognitive training.

Methods: A total of 90 older adults (mean [SD] 69.49 [5.78]) were divided into 3 groups (I2WE, SWITCH, and BIKE-VG). Each participant completed a 45-minute group session and then filled out questionnaires to evaluate UX, workload, pleasure, and intensity.

Results: The UX was positive for I2WE and SWITCH, and neutral for BIKE-VG. It was higher for I2WE than for BIKE-VG (t_{87}=2.83; P=.02; d=0.70; 95% CI 0.15 - 1.69). The workload was moderate across all 3 groups. The intensity was moderate for all groups, ranging between 50% and 70% of the maximum heart rate, and approached high intensity for the I2WE and SWITCH groups. It was significantly higher for I2WE than for BIKE-VG (t_{66}= 2.86; P=.01; d=0.70; 95% CI 1.04 - 11.43). The perceived pleasure was significantly higher for I2WE (t_{87}=3.63; P=.001; d=0.9;95% CI 2.74 - 13.23) and SWITCH (t_{87}=3.11; P=.01; d=0.87; 95% CI 1.82 - 13.69) compared with BIKE-VG.

Conclusions: The UX and perceived enjoyment are higher for the Moving While Thinking training compared with the Thinking While Moving training. This indicates that the I2WE and SWITCH training approaches are promising and motivating options for combined training for older adults.

(JMIR Serious Games 2025;13:e67314) doi:10.2196/67314

KEYWORDS

combined training; experience; video game; older adult; elderly; user experience; AVG; UX; physical stimulation; exergame; exergaming; game; gamification; motor; physical; physical activity

Introduction

The interest in training with active video games (AVG) for older adults has been growing over the past few decades [1 - 2]. AVG are good opportunities to combine cognitive and physical (or motor) stimulation. Moreover, they could increase healthy lifestyle adoption [3], counteract sedentary lifestyles, and enhance individual health behavior, as technology programs provide more enjoyment than traditional interventions [4]. Given the appeal of AVG, many tools are beginning to be used for training, including consumer devices (eg, Xbox, Wii, Switch, and Kinect). The Nintendo Wii, released in 2006, was one of the first consumer devices and was later replaced by the Nintendo Switch (Nintendo Co) in 2017. The Nintendo Wii or Switch console has a wireless handheld pointing device with embedded sensors that detect direction, speed, and acceleration changes in three dimensions. The games offer immediate visual and auditory performance feedback and progress information. More specifically, the cognitive task is incorporated into the motor task by the Moving While Thinking approach [5]. Thus, the 2 tasks are linked and interactive, which could benefit older adults' cognitive functions since it can increase the synergistic

XSL•FO
RenderX

effect of the combined training tenfold [6]. This consumer device allows collective interactions with cooperative and opposition games. In the 2020s, a new type console was developed. These Exergames are assisted by an Immersive and Interactive Wall (I2WE), such as LUMOplay (Lumo Interactive), Play Lü (Lü Interactive Playground) [7], iWall (CSE Entertainment) [8], ExerCube (Sphery AG) [9,10] and Neo One (NeoXperiences) [11,12]. In comparison with consumer devices, I2WE is more extensive and more immersive. I2WE creates simultaneous multi-domain training, with complete physical stimulation (ie, aerobic, muscle strengthening), motor stimulation (complex motor skills), and freedom of movement, in addition to a virtual continuous cognitive stimulation. These characteristics could produce more well-rounded and intensive training than consumer devices.

Other experiences were created with a combination of physical exercise (eg, cycling, kayaking) and virtual stimulation (eg, video games, interactive maps) [13-15]. Cognitive stimulation is sometimes linked (Moving While Thinking, eg, cycling with an interactive map), but other times, they are not (Thinking While Moving, eg, cycling with unrelated video games).

Alongside developing these various tools, many studies have been undertaken to measure their health effects on older adults, including physical [1,16], dual-task [12], cognitive [2,17], and social benefits [18]. However, the positive impacts of AVG on individual cognitive domains are inconsistent across studies [19]. This may result from devices' varying forms and definitions of AVG's physical and cognitive demands. Indeed, AVG could generate different solicitations like physical-cognitive training, motor-cognitive training, and multi-domain training [20]. Torre and Temprado [20] described physical-cognitive training as the sequential or simultaneous combination of endurance (aerobic) and muscular resistance exercises with cognitive tasks. Motor-cognitive training, on the other hand, involves pairing complex motor skills training with cognitive tasks that are typically added separately from the motor task. However, we propose that motor-cognitive training can also involve simultaneous motor and cognitive stimulation (eg, consumer device like Nintendo Switch). Multi-domain training integrates aerobic exercises and muscular resistance, complex motor skills, and cognitive tasks. Moreover, the interaction, immersion, controller (ie, body controller, object controller, button controller, mixed button, and gesture controller), and games may differ. As Li et al [18] recommended in their systematic review, more studies are needed to compare the effects of different exergaming platforms. Herold et al [21] indicated that future studies should directly compare the effectiveness of various digital and home-based physical training interventions.

It is highly recommended that AVG for public health and disease prevention be designed to meet the expectations and needs of its target users [22]. Before conducting full-scale studies, Campbell [23] indicated that the acceptability and feasibility should be tested with older adults. The workload must be appropriate to provide sufficient stimulation without causing excessive stress. The same goes for the intensity, which should be moderate according to recommendations [19]. More specifically, to make AVG appealing to older adults, it needs to be designed to encourage initial acceptance and sustained use over time. This requires a user-centered design approach, including thoroughly evaluating the exergames' user experience (UX) [24]. UX, which results from the interaction between the user, the system, and the context, is crucial in ensuring high-quality user-system interaction. Consequently, UX is increasingly used to assess AVG for the elderly [24-27]. The concept of UX builds upon the foundations of usability, which alone are now seen as inadequate for predicting and explaining technology acceptance. UX expands usability by incorporating emotional, subjective, and temporal dimensions, defining an individual's technology interaction. These aspects are crucial for designing suitable interactive systems. Hassenzahl [28] uses a pragmatic–hedonic model of UX, in which individuals perceive the quality of their interactions with products according to two types of attributes: pragmatic qualities, directly related to task accomplishment and product use, and hedonic qualities, which emphasize the psychological well-being of individuals. These 2 attributes combine to generate a global evaluation of product attractiveness. In actual studies that evaluated active video game UX among older adults, UX is positive [29,30]. Perceived enjoyment is also often evaluated to complete the UX assessment [31-33]. More enjoyable AVG can lead to more intense activity [34,35].

Numerous studies have been conducted on AVG involving older adults, but few focus on comparing different experiences with each other. Studies have mainly focused on single gaming platforms or compared 2 modes with short game exposures [26] or have lacked participants [30].

Studies agree that AVG creates positive training experiences. However, there is a lack of comparative information to determine whether one type of AVG would be more beneficial than others. The context of this study was primarily focused on a preventive approach, aimed at reducing the risk of physical and cognitive decline associated with aging. The goal was to explore how innovative strategies, such as the use of active video games, could be used as a primary prevention strategy to preserve cognitive function before the onset of impairments. The purpose of this study was to compare three different experiences: (1) I2WE, (2) consumer device (SWITCH), and (3) a combination of video games and physical stimulation (biking and gaming, BIKE-VG). We expected the 3 AVG training sessions create a physical intensity in line with the recommendations for older adults (moderate to high), with a correct session workload and a positive UX. We hypothesized that I2WE and SWITCH would generate a greater UX and intensity than BIKE-BG. Moreover, a higher workload was expected for the Thinking While Moving training (ie, BIKE-VG) compared to the 2 Moving While Thinking trainings (ie, I2WE, SWITCH).

Methods

Ethical Considerations

This study was approved by the Ethics and Research Committee of Université Paris-Saclay (CER-Paris-Saclay-2023 - 084). Informed consent was obtained from all participants prior to their inclusion in the study. Participants were clearly informed

XSL•FO
RenderX

of their right to withdraw from the study at any time without consequence. All data collected during the study were fully anonymized. No financial compensation was provided to participants for their involvement in the study.

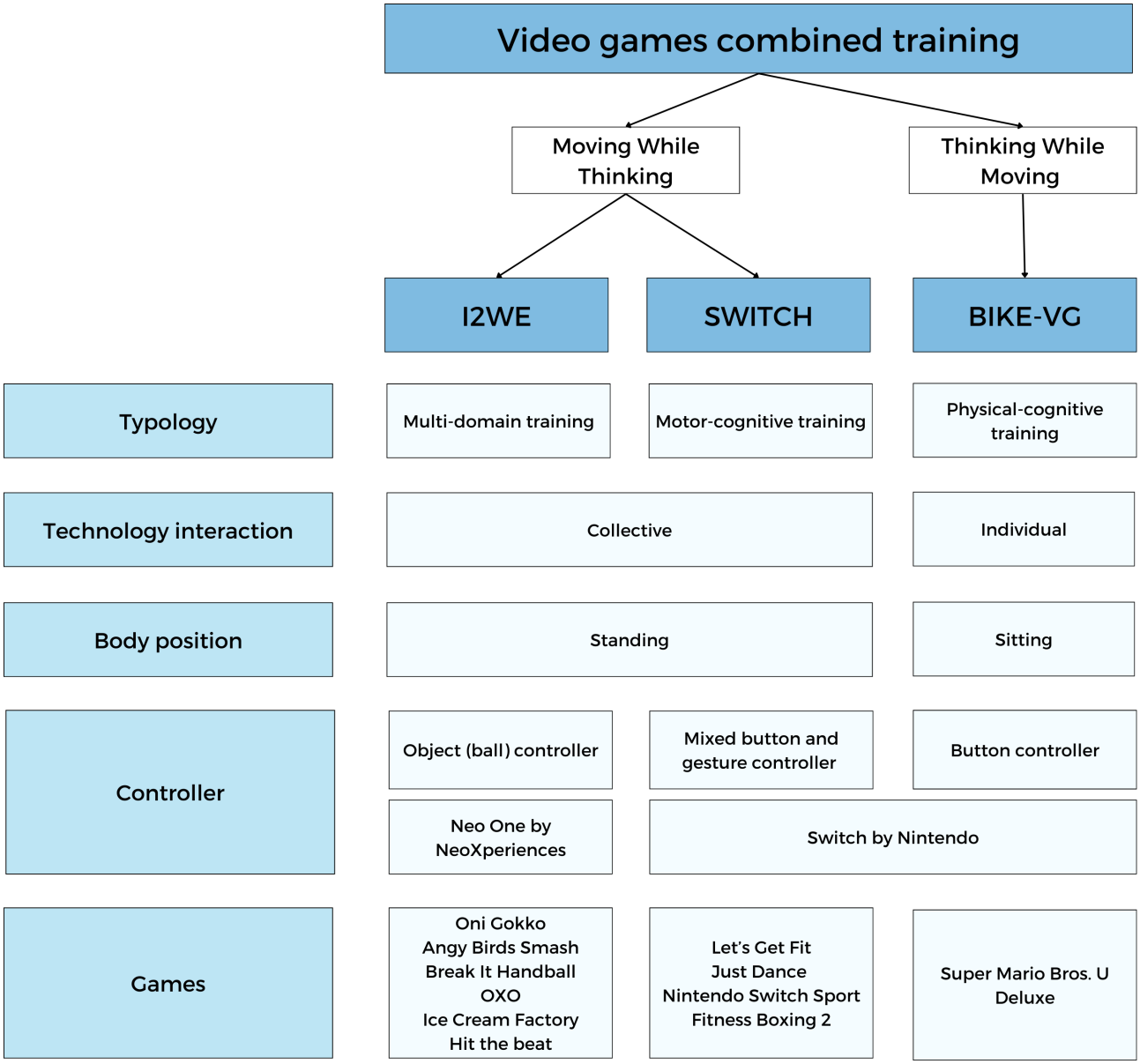
Recruitment

Potential participants responded positively to a call for participation by email or flyer distributed at associations and clubs in Ile-de-France (France). A power analysis using the G*Power version 3.1.9.7 (Heinrich-Heine-University), based on the primary statistical test planned for this study (ANOVA: Fixed effects, omnibus, one-way), revealed that a sample size of 66 participants was needed (α=.05; power=0.80; effect size: f=0.40). All participants met the following criteria: (1) be over 60 years of age, (2) score three or higher on a subjective health scale (1=verypoor-5=excellent), (3) have no major pathology, and (4) have normal or corrected vision and hearing.

Protocol

Participants were split into 3 groups (ie, I2WE, SWITCH, and BIKE-VG). The characteristics (ie, physical/motor stimulation, interaction, body position, controller, and games) are described in Figure 1.

Figure 1. Description of the 3 video games combined training Notes. Three types of video game combined training were compared: (A) Immersive and Interactive Wall Exergame, (B) SWITCH games, and (C) Bike-based Video Games group. I2WE and SWITCH were classified as Moving While Thinking training, meaning that the cognitive task is integrated into the motor/physical task. BIKE-VG was considered Thinking While Moving training, where the cognitive and motor/physical tasks were not interconnected [5]. To define the different types of combined training, we followed the modified definition by Torre and Temprado [20]. I2WE was categorized as multi-domain training (integrates aerobic and muscular resistance exercises, complex motor skills, and cognitive tasks). SWITCH was a motor-cognitive training that paired complex motor skills training with cognitive tasks. BIKE-VG was defined as physical-cognitive training, which involved the combination of endurance and muscular resistance training with cognitive tasks.



For each group, the protocol was the same: the participants were divided into groups of four participants per session (45 min) and then filled in a battery of questionnaires. The session was always preceded by ten minutes of exercises, during which the

XSL•FO
RenderX

participants gradually warmed up (joint mobilization – eg, turning the head, shoulders, and ankles – walking around the room at various paces, and squat movements). A sports therapist supervised the session. The supervisor ensured participants' safety by monitoring correct posture. For I2WE and SWITCH, the breaks between the games lasted a few seconds, just enough time to switch to the next game.

I2WE Group

A Neo One device was used (Figure 2A), an I2WE equipped with a video projector and infrared sensors that allow playing without headphones or equipment, so the body controls games. Neo One was conducted standing, offering multi-domain training with collective interactions (ie, cooperation, opposition) between players (multiplayer). Neo One has about 20 games, including sports and puzzle games, that can be played in cooperation or opposition. For this study, the session was composed of 6 games of 5 minutes. The instructions and rules of the games were displayed on the screen and re-explained by the supervisor if necessary. The first 3 games were cooperative, while the next 2 were competitive (2 vs 2). The final game was also competitive but played 1 versus 1.

Figure 2. Photography of each group (A) Immersive and Interactive Wall Exergames, (B) SWITCH games, (C) Bike-based Video Games group .



Oni Gokko

Participants had to complete several mini games, including Hangman, Hidden Object, Memory, and Word Scramble. The goal was to score as many points as possible.

Angry Birds Smash

In this game, players had to destroy as many fortresses and pigs on the screen as possible to progress through levels. The objective was to complete the levels and achieve the highest score within the allotted time.

Break It Handball

In this game, players aimed at various bricks displayed on the screen to advance through levels. The objective was to complete the levels and achieve the highest score within the time limit.

OXO

This puzzle game was based on the mechanics of tic-tac-toe. The goal was to align three symbols on the grid to win.

Ice Cream Factory

In this game, 2 teams competed to fulfill ice cream orders as quickly as possible and earned coins. The team with the most coins at the end of the time won.

Hit the Beat

In this game, participants had to hit targets at the right moment as they appeared in rhythm with the music.

SWITCH Group

The Nintendo Switch device was a motion-sensitive controller to detect three dimensions (Figure 2B). It was an exergame connected with a 35-inch screen. The controller has mixed buttons and gestures. Exergames were conducted standing, offering motor-cognitive solicitation, with collective interactions (ie, cooperation and opposition) – multiplayer. A total of 5 games of 5 - 7 minutes were used. For the first two games, participants played all 4 together. For the last three games, participants were separated into 2 groups of two but played simultaneously.

Let's Get Fit

This game consisted of following the movements of an avatar and performing sports movements (eg, high knees, heel kicks, and squats).

Just Dance

The players followed dance choreographies on screen by mimicking the movements of virtual dancers as best as possible. Around 2 pieces of music were played: *"Rain Over Me" and "Le Bal Masqué."*

Football on Nintendo Switch Sport

In this game, 2 participants competed against each other. The aim was to score more goals than their opponent. Each participant moved their avatar with the Joy-Stick and shot the ball with a Joy-Con placed on their leg.

XSL•FO
RenderX

Badminton on Nintendo Switch Sport

In this game, 2 players competed against each other. The game used motion controls and button inputs to simulate swings and shots. The aim was to score more points than their opponent.

Fitness Boxing 2

It combined boxing-inspired workouts with music and rhythm gameplay elements. Players used the Joy-Con controllers to simulate a series of boxing movements (ie, forehand, hook, uppercut) in time with the music and on-screen prompts.

BIKE-VG Group

Participants had to play the *Super Mario Bros.U Deluxe* game on the Nintendo Switch while pedaling on a stationary bike (Ultra Green R 2.0 by MOOVYOO), offering physical-cognitive training (Figure 2C). Each participant played individually (single-player). Joy-Con controllers were attached to the console in the participant's hands, so the controller was only by button. The game involved navigating through side-scrolling levels filled with enemies, obstacles, and secrets. After explaining how to use the game console and the game instructions, a 5-minute warm-up at minimum resistance was performed on the bike. Next, participants had to pedal and play for 40 minutes, with resistance incrementing every 3 minutes to a minimum level of 4. If participants wanted to put on more resistance, they were free to do so.

Outcomes

Heart Rate

The participant's heart rate was continuously monitored throughout the session with a Polar Verity Sense heart rate monitor. Theoretical Maximum Heart Rate (TMHR) was determined using the formula developed by Gellish and colleagues [36]: TMHR =192 - 0.007x age². The subsequent calculation defines the average intensity experienced by the individual during the session: average intensity = (average heart rate x 100) / TMHR.

Questionnaires

QAPPA (in French : Questionnaire d'Activité Physique pour les Personnes Agées – in English: Physical Activity Questionnaire for Older Adults) [37] was a questionnaire developed to assess weekly physical activity levels, tailored specifically for older adults.

AttrakDiff 2 [38] which assessed user experience (UX), was translated and validated in French [39]. The questionnaire items were categorized into four scales: (1) Pragmatic Qualities, which evaluated usability and perceived ease of use; (2) Hedonic Quality – Stimulation, which assessed the support for stimulation with new, attractive, and stimulating content, features, and interaction styles; (3) Hedonic Quality – Identity, which evaluated the support for social function and communication of user identity; and (4) Overall Attractiveness, which gauged the overall perceived value of the product based on its pragmatic and hedonic qualities. For each scale, the score ranges from −3 (low score) and 3 (high score).

NASA-Task Load index (NASA-TLX) [40] assessed the workload associated with tasks. Participants rated six dimensions of workload (ie, mental demand, physical demand, time pressure, performance, effort, and frustration) on a scale from 0 to 100. Due to the multidimensional nature of NASA-TLX, some dimensions may contribute more importantly to the total score, prompting weighting procedures to assign coefficients to each dimension. However, these procedures were criticized for being time-consuming [41]. Alternatively, Bittner et al [42] proposed a more straightforward approach by averaging the scores across all six dimensions, termed the "Raw" Task Load Index (RTLX). They demonstrated a high correlation between the RTLX and the traditional calculated score. Cegarra and Morgado [43] confirmed this finding with the French version of the scale.

The physical activity enjoyment scale [44], consisting of ten items translated and validated in French [45] on a scale ranging from 7 to 70 (best score) assessed the perceived enjoyment experienced during physical exercise.

The modified Borg Scale [46] gauged perceived exertion during exercise. It used a numerical scale ranging from 0 to 10, with corresponding verbal anchors (eg, very light, difficult, painful effort) to help individuals rated their exertion level.

Want to Continue to Play Active Video Games

Participants were asked if they would like to continue playing AVG in the future (strongly agreed, agreed, don't know, disagreed, strongly disagreed).

Statistical Analysis

The analyses were performed using JASP software (0.18.3 version). Descriptive data (ie, mean, SD) were reported for all variables. Personal variables group differences were compared with a one-way analysis of variance (ANOVA) for continuous variables (ie, age, BMI, QAPPA) and χ^2 test for categorical variables (ie, sex, educational level, active video games practice). To determine if the UX for each AVG was positive, a one-sample Student *t* test was used to compare the mean of each sample to a reference value. ANOVAs were employed to compare perceived enjoyment, session workload, physical intensity, and UX between groups. 95% CI completed them for the mean difference. All statistical analyses with a *P* value <.05 were considered significant. A Tukey Post Hoc test was realized when a significant interaction was found. Effect Sizes were calculated to study the power of the results: Eta-Squared (η^2) for ANOVA and Cohen *d* for *t* test. As indicated in the effect size guidelines from Cohen [47], η^2 is considered low if the effect size is between 0.01 and 0.06, medium if it is between 0.06 and 0.14, and large if it is superior to 0.14. Cohen *d* is considered low if the effect size is between 0.20 and 0.50, medium between 0.50 and 0.80, and large if superior to 0.80.

Results

Overview

Personal information for each group is given in Table 1. There were no significant differences for all variables.

XSL•FO
RenderX

Table . Personal variables.

Personal variables	Switch (n=23)	I2WE[a] (n=38)	Bike-vg[b] (n=29)	*P* value
Age (years), mean (SD)	69.39 (6.42)	68.26 (5.53)	71.17 (5.34)	.12
BMI (score)	25.09 (4.94)	24.08 (3.33)	24.45 (2.59)	.57
QAPPA[c] (n)	877.39 (712.01)	1008.95 (683.62)	857.93 (705.43)	.63
Sex, n (%)				.58
Male	4 (17.39)	9 (23.68)	4 (13.79)	
Female	19 (82.61)	29 (76.32)	25 (86.21)	
Level of education, n (%)				.10
No diploma	0 (0)	4 (10.53)	0 (0)	
Certificate	2 (8.7)	5 (13.16)	1 (3.45)	
Vocational certificate	7 (30.43)	12 (31.58)	3 (10.34)	
High school diploma	2 (8.7)	1 (2.63)	3 (10.34)	
Associate’s degree	4 (17.39)	8 (21.05)	5 (17.24)	
Bachelor’s degree	4 (17.39)	7 (18.42)	8 (27.59)	
Master’s degree	4 (17.39)	1 (2.63)	7 (24.14)	
Doctorate degree	0 (0)	0 (0)	2 (6.9)	
Active video games practice (%)				.74
More than one a week	0 (0)	0 (0)	0 (0)	
Between once a week and once a month	0 (0)	1 (2.63)	0 (0)	
Between one a month and once a year	3 (13.04)	4 (10.53)	2 (6.90)	
Never	20 (86.96)	33 (86,84)	27 (93.10)	

[a]I2WE: Immersive and Interactive Wall Exergame

[b]BIKE-VG: Bike-based Video Game group

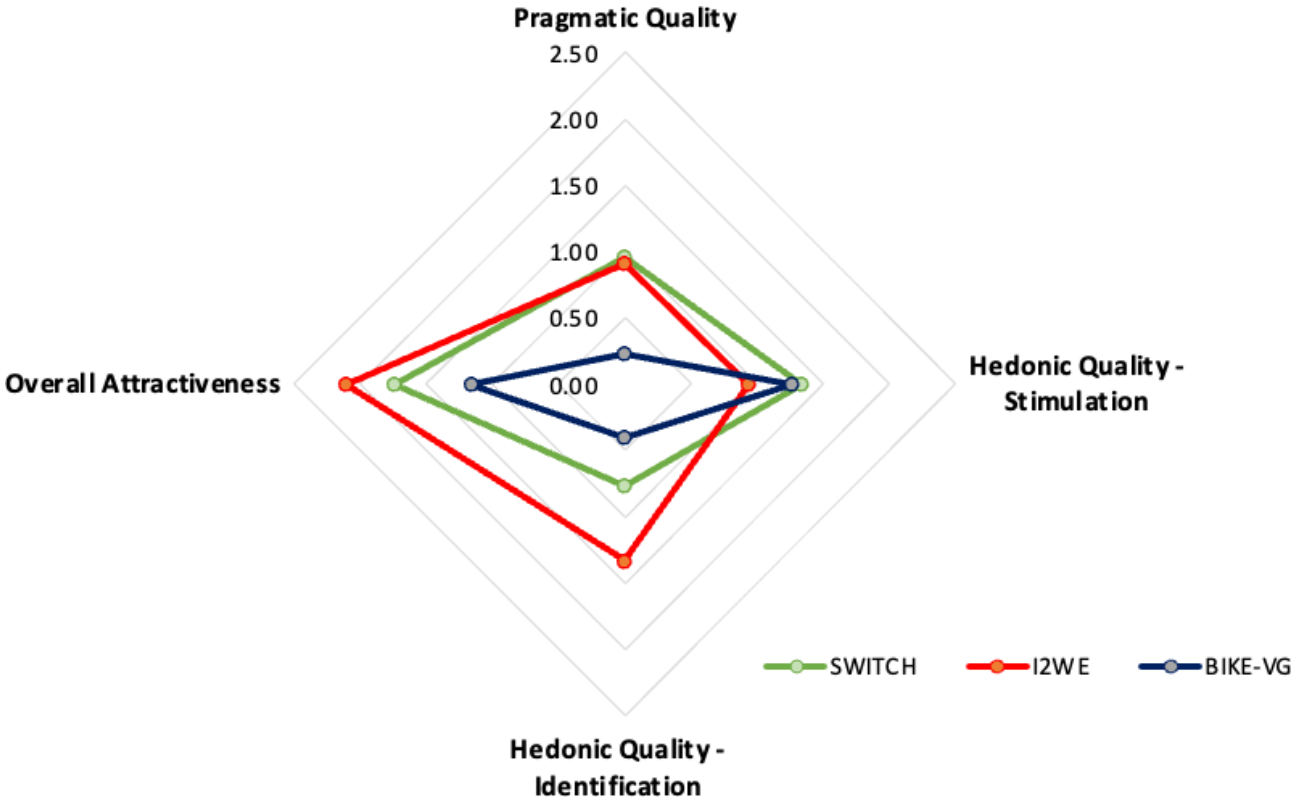
[c]QAPPA: Questionnaire d'Activité Physique Pour les Personnes Âgées (Physical Activity Questionnaire for Older Adults)

User Experience

The UX, depending on the group, are given in Figure 3.

XSL•FO
RenderX

Figure 3. User experience depending on groups.



Comparison at the Threshold

The area of positivity of the AttrakDiff2 items is between 1 and 3, neutrality between −1 and 1, and negativity between −1 and −3. To know if the scores were significantly positive, a *t* test was performed to compare the averages for the four AttrakDiff2 items with a value of 1 for each group (Table 2).

XSL•FO
RenderX

Table . Score difference from standard 1.

Score	Mean (SD)	*t* test	*P* value	*d*
I2WE[a]				
Pragmatic quality (score)	0.91 (0.89)	0.65	.52	0.10
Hedonic quality stimulation (score)	1.68 (0.81)	5.18	<.001	0.84
Hedonic quality identity(score)	1.32 (0.69)	2.90	.01	0.47
Total attractiveness (score)	2.08 (0.79)	8.44	<.001	1.37
SWITCH				
Pragmatic quality (score)	0.96 (0.79)	0.22	.82	0.05
Hedonic quality stimulation (score)	1.53 (1.01)	2.49	.02	0.52
Hedonic quality identity (score)	0.91 (0.89)	0.47	.65	0.10
Total attractiveness (score)	1.93 (1.17)	3.78	.001	0.79
BIKE-VG[b]				
Pragmatic quality (score)	0.23 (1.09)	3.79	<.001	0.70
Hedonic quality stimulation (score)	1.26 (0.92)	1.52	.14	0.28
Hedonic quality identity (score)	0.40 (1.22)	2.66	.01	0.49
Total attractiveness (score)	1.16 (1.85)	0.47	.64	0.09

[a]I2WE: Immersive and Interactive Wall Exergame

[b]BIKE-VG: Bike-based Video Game group

The Pragmatic Quality is significantly smaller than 1 for BIKE-VG (t_{28}=3.79; P<.001; d=0.70). There was no significant difference to 1 for I2WE and SWITCH.

The Hedonic Quality - Stimulation is significantly greater than 1 for I2WE (t_{37}=5.18; P<.001; d=0.84) and SWITCH (t_{22}=2.49; P=.02; d=0.52).

The Hedonic Quality - Identity is significantly greater than 1 for I2WE (t_{37}=2.90; P=.01; d=0.47) and significantly lesser than 1 for BIKE-VG (t_{28}=2.66; P=.01; d=0.49).

The Total Attractiveness is significantly greater than 1 for I2WE (t_{37}=8.44; P<.001; d=1.37) and SWITCH (t_{22}=3.78; P=.001; d=0.79).

Comparison Between Groups

ANOVA analysis revealed a significant group effect for three items: Pragmatic Quality ($F_{2,87}$=5.45; P=.01; η^2=.11), Hedonic Quality Identity ($F_{2,87}$=780; P<.001; η^2=0.16), and Total Attractiveness ($F_{2,87}$=4.31; P=.02; η^2=.09). Post Hoc analysis indicated that for Pragmatic Quality, the score is significantly better for SWITCH than BIKE-VG (t_{87}=2.79; P=.02; d=0.78; 95% CI 0.11 - 1.36), and for I2WE than BIKE-VG (t_{87}=2.92; P=.01; d=0.72; 95% CI 0.12 - 1.23); for Hedonic Quality Identity, the score is significantly better for I2WE than BIKE-VG (t_{87}=4; P<.001; d=0.99; 95% CI 0.37 - 1.48) and for Total Attractiveness, the score is significantly better for I2WE than BIKE-VG (t_{87}=2.83; P=.02; d=0.70; 95% CI 0.15 - 1.69).

Scores on the physical activity enjoyment scale, the RTLX, the modified Borg scale, and the intensity of the session are provided in Table 3.

XSL•FO
RenderX

Table . Perceived enjoyment scores, workload, physical intensities.

Score	SWITCH	I2WE[a]	BIKE-VG[b]	*P* value	η^2
Perceived Enjoyment (score)	60.48 (8.93)	60.71 (6.04)	52.72 (11.67)	**<.001**	**.15** [c]
RTLX[d] (score)	54.46 (18.54)	53.46 (14.28)	59.14 (16.36)	.34	.02
Mental demand (score)	62.39 (23.25)	63.16 (25.8)	75.00 (18.03)	.07	.06
Physical demand (score)	60.65 (23.47)	62.37 (26.65)	51.72 (29.74)	.25	.03
Temporal demand (score)	63.26 (24.24)	64.47 (23.79)	61.72 (23.19)	.90	.00
Effort (score)	56.96 (22.50)	57.63 (22.5)	57.07 (23.59)	.99	.00
Performance (score)	44.78 (24.47)	39.74 (16.52)	54.66 (25.46)	**.03**	**.08**
Frustration (score)	38.70 (31.7)	33.42 (23.11)	54.66 (27.48)	**.01**	**.11**
Modified Borg Scale (score)	3.83 (1.11)	4.37 (1.55)	3.45 (1.11)	**.03**	**.08**
Mean Intensity (percent of TMHR)[e]	68.01 (8.2)	69.64 (9.11)	63.41 (8.95)	**.02**	**.09**

[a]I2WE: Immersive and Interactive Wall Exergame

[b]BIKE-VG: Bike-based Video Game group

[c]Bold values are significant.

[d]RTLX: ''Raw'' Task Load Index

[e]TMHR: theoretical maximal heart rate

Perceived Enjoyment

There was a significant group effect ($F_{2,87}$=7.71, *P*<.001, η^2=.15) for *Perceived Enjoyment*. Post Hoc analysis revealed a significant difference between I2WE and BIKE-VG (t_{87}=3.63, *P*=.001, *d*=0.90, 95% CI 2.74 - 13.23) and between SWITCH and BIKE-VG (t_{87}=3.11, *P*=.01, *d*=0.87, 95% CI 1.82 - 13.69).

Session's Workload

The session's workload, evaluated by the RTLX, was correct for the three groups. The three dimensions related to the task (ie, mental demand, physical demand, and temporal demand) were balanced, while frustration was the lowest dimension, except for BIKE-VG, where the frustration was higher. There was a significant group effect ($F_{2,87}$=5.30; *P*=.01; η^2=.11), and the post hoc test revealed a significant difference between I2WE and BIKE-VG (t_{87}=3.20; *P*=.01; *d*=0.79; 95% CI 5.41 - 37.06).

There was also a significant group effect ($F_{2,87}$=3.88; *P*=.03; η^2=.08) for performance. The post hoc test indicated a significant difference between I2WE and BIKE-VG (t_{87}=2.77; *P*=.02; *d*=0.68; 95% CI 2.09 - 27.75).

Physical Intensity

The objective physical intensity was moderated for the three groups, but there was a significant group effect ($F_{2,87}$=4.22; *P*=.02; η^2=.09) for mean intensity. The post hoc test showed a significant difference between I2WE and BIKE-VG (t_{87} = 2.86; *P*=.01; *d*=0.70; 95% CI 1.04 - 11.43)

The intensity was perceived between very easy and easy for SWITCH (*M*=3.93) and BIKE-VG (*M*=3.45) and between easy and moderate for I2WE (*M*=4.37). There was a significant group effect for subjective intensity ($F_{2,87}$=3.58; *P*=.03; η^2=.08). I2WE was significantly greater than BIKE-VG (t_{87}=2.64; *P*=.03; *d*=0.65;95% CI 0.09 - 1.75).

Want to Continue to Play Active Video Games

In all 3 groups, most participants wanted to continue: 84% (32 out of 38 for I2WE; 58% completely agreed, 26% agreed), and 65% for SWITCH (15 out of 23; 43% completely agreed, 22% agreed). However, these figures were lower for BIKE-VG with 59% (17 out of 29; 21% completely agreed, 38% agreed). A nonnegligible proportion disagreed (17%) or completely disagreed (14%).

Discussion

Principal Findings and Comparison With Previous Works

This study aimed to compare three AVG experiences (I2WE, SWITCH, BIKE-VG) on older adults. Concerning UX, the overall attractiveness, which is the total perceived value of the product based on the perception of its pragmatic and hedonic qualities, was positive for I2WE and SWITCH. Games-based programs could offer an enjoyable and stimulating exercise experience [4]. AVG provide instant visual and auditive feedback, motivational messages such as bonuses, or encouraging comments that make the activity entertaining and interactive [48]. However, the overall attractiveness was not positive for the three experiences, given that it was neutral for BIKE-VG, with a significant difference between I2WE and BIKE-VG. According to Lallemand and Gronier [49], the neutral

XSL•FO
RenderX

zone indicates that the product fulfills its purpose without causing any negative impact. The difference between the two other groups may be explained by the combination of physical and cognitive stimulation, which is less natural for BIKE-VG than for I2WE and SWITCH: the physical task (ie, cycling) is not directly linked to the cognitive task (ie, video games). Exercises involving Thinking While Moving may be less instinctive and ecological. In contrast, I2WE and SWITCH involved Moving While Thinking tasks: the cognitive task was incorporated into the physical tasks. Moving While Thinking training may be easier, especially since this was the first experience with AVG for most participants in this study. Although there was no significant difference between the groups, this could explain why the BIKE-VG program, a Thinking While Moving program, had a high cognitive scale score on the workload scale (75/100) compared with the other two Moving While Thinking training programs, with 62.39/100 for the SWITCH group and 63.16/100 for the I2WE group. The difference in overall attractiveness may also be partially attributed to the variation in interaction on with other participants: BIKE-VG used single-player game, while SWITCH and I2WE involved multiplayer games.

Moreover, the controller may be the origin of this dissimilarity, reflected by differences in the Pragmatic Quality scores. Its evaluation represents users' perceived ability to use products that allow them to manipulate the environment to accomplish their objectives. The Pragmatic Quality score was neutral for the three groups; however, the score was significantly lower for BIKE-VG than for I2WE and SWITCH groups. A button controller may be more complex to use than another controller (ballon controller or mixed button and gesture controller) and need more learning effort [50]. This may be creating an alteration of the experience. In their study, Pham and Theng [50] found that the older participants preferred to play games with fewer or no controlling devices. The Gesture Controller (Kinect), which was reported to require much physical movement, scored the highest in overall attractiveness, while participants gave negative feedback to the button controller, which requires fewer body movements. A difficulty in using the buttons can explain lower score for the BIKE-VG group and can, therefore, generate more frustration and worse performance than the body or mixed button and gesture controller. Indeed, the BIKE-VG group had a higher frustration score at NASA-TLX than the I2WE group. Moreover, it was the same for performance. However, the global workload score was correct for the three groups, with no significant difference between the three groups.

AVG promise to reduce loneliness, increase social connection, and foster positive attitudes toward others [18]. However, AVG may also provide opportunities for social engagement with peers and family members [48]. In this study, this is reflected by the positive score of the I2WE groups for the Hedonic Quality Identity score, which is the capacity to support a social function and communicate a specific user identity. This AVG provides a lot of communication and interaction between players. The BIKE-VG group, where the participants played alone (single-player), obtained a neutral score significantly lower from I2WE (multiplayer). Moreover, SWITCH and I2WE supported the need for stimulation with new, attractive, and stimulating content, features, and interaction styles, given that the two groups had positive scores at Hedonic Quality Stimulation. The score was neutral for BIKE-VG, but no significant difference existed between the groups. It is possible that the BIKE-VG session was too monotonous because it contained only one game, compared to SWITCH and I2WE, which included several games.

The intensity was moderated for the three groups, between 50% and 70% of the maximum heart rate, and almost high intensity for the I2WE and SWITCH groups. This is in line with the ACSM recommendations. Therefore, these three AVG could be used in a training program to improve physical and cognitive abilities. However, the physical intensity was significantly greater for I2WE than for BIKE-VG. The physical solicitation of I2WE may explain this difference: the whole-body multi-domain training and the freedom of movement offered by the system may increase intensity, as Martin-Niedecken showed with the ExerCube [9]. Moreover, Moving While Thinking training may increase intensity when two tasks are linked (there is no task prioritization). Subjectively, the intensity was considered very easy for SWITCH and BIKE-VG and easy to moderate for I2WE, with a significant difference between I2WE and BIKE-VG. The physical stimulation, which was intermittent for I2WE and SWITCH and continuous for BIKE-VG, may explain these differences.

As indicated in many studies [27,33,51,52] and the present work, the pleasure generated by AVG is high. However, it was significantly higher for I2WE and SWITCH than BIKE-VG. Some AVG create more enjoyment than others, depending on the features of the interfaces and the interactions between players. Indeed, as noted by Caserman et al [53], social interaction, one of the many quality criteria required for a game, plays an important role in enhancing enjoyment. Concerning the interface, Kim et al [35] showed that the level of interface embodiment influences the level of enjoyment on young adults (mean age =24.39). They compared three levels of interface embodiment (low, medium, and high), finding that the high level provided significantly more enjoyment than the other two.

The individual characteristics of participants may influence each user's experience. In this study, no differences were found between age, BMI, physical activity level, gender, education level, and previous experience with active video games. Although most participants (82%-90% depending on the group) had never played active video games, players' preferences and profiles may still affect UX, enjoyment, or physical intensity.

According to our knowledge, this study was the first to compare three AVG experiences in older adults. Thus, the comparison with the literature is thin. For example, studies often compare only 2 devices [54], or sometimes 3, but with sessions lasting a few minutes [26] that do not reflect a complete session.

Conclusion

In light of the present study, Moving While Thinking training (eg, I2WE, SWITCH) appears promising for stimulating physical, cognitive, and dual-task capacities in older adults. An interventional approach is needed to compare the effects of

XSL•FO
RenderX

these programs to Thinking While Moving programs (eg, BIKE-VG).

Data Availability

Deidentified data that support the findings of this study are available on request from NBP.

Authors' Contributions

NBP performed data curation, formal analysis, investigation, and methodology. NBP, AP, and PM performed project administration. NBP, AP, and PM wrote the original draft and performed review and editing. AP and PM performed supervision.

Conflicts of Interest

None declared.

References

1. Deng N, Soh KG, Abdullah BB, et al. Active video games for improving health-related physical fitness in older adults: a systematic review and meta-analysis. Front Public Health 2024;12:1345244. [doi: 10.3389/fpubh.2024.1345244] [Medline: 38694976]
2. Soares VN, Yoshida HM, Magna TS, et al. Comparison of exergames versus conventional exercises on the cognitive skills of older adults: a systematic review with meta-analysis. Arch Gerontol Geriatr 2021;97:104485. [doi: 10.1016/j.archger.2021.104485] [Medline: 34293715]
3. DeSmet A, Van Ryckeghem D, Compernolle S, et al. A meta-analysis of serious digital games for healthy lifestyle promotion. Prev Med 2014 Dec;69:95-107. [doi: 10.1016/j.ypmed.2014.08.026] [Medline: 25172024]
4. Valenzuela T, Okubo Y, Woodbury A, et al. Adherence to technology-based exercise programs in older adults: a systematic review. J Geriatr Phys Ther 2018;41(1):49-61. [doi: 10.1519/JPT.0000000000000095] [Medline: 27362526]
5. Herold F, Hamacher D, Schega L, et al. Thinking while moving or moving while thinking - concepts of motor-cognitive training for cognitive performance enhancement. Front Aging Neurosci 2018;10:228. [doi: 10.3389/fnagi.2018.00228] [Medline: 30127732]
6. Fissler P, Küster O, Schlee W, et al. Novelty interventions to enhance broad cognitive abilities and prevent dementia: synergistic approaches for the facilitation of positive plastic change. Prog Brain Res 2013;207:403-434. [doi: 10.1016/B978-0-444-63327-9.00017-5] [Medline: 24309264]
7. Roure C, Lentillon-Kaestner V, Fargier P. The effects of a play lü inter-team competition exergame on adolescents' moderate-to-vigorous physical activity and situational interest. Presented at: ISBNPA XChange Initiative - "Thinking independently together – Let's dance around the world!"; Vancouver, Canada URL: http://hdl.handle.net/20.500.12162/4973
8. Röglin L, Martin-Niedecken AL, Ketelhut S. Exercising digitally: a multi-perspective analysis of exergames for physical activity and health promotion. In: Brooks AL, editor. Creating Digitally: Springer International Publishing; 2023, Vol. 241:79-116. [doi: 10.1007/978-3-031-31360-8_4]
9. Martin-Niedecken AL, Rogers K, Turmo Vidal L, et al. ExerCube vs. personal trainer: evaluating a holistic, immersive, and adaptive fitness game setup. Presented at: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems; Glasgow Scotland Uk. [doi: 10.1145/3290605.3300318]
10. Martin-Niedecken AL, Mekler ED. The exercube: participatory design of an immersive fitness game environment. In: Göbel S, Garcia-Agundez A, Tregel T, Ma M, Baalsrud Hauge J, Oliveira M, editors. Presented at: International Joint Conference on Serious Games; Darmstadt, Germany. [doi: 10.1007/978-3-030-02762-9_28]
11. Béraud-Peigné N, Maillot P, Perrot A. The user experience of an immersive and interactive wall exergame in older adults. Games Health J 2023 Jun;12(3):220-227. [doi: 10.1089/g4h.2022.0075] [Medline: 36251866]
12. Béraud-Peigné N, Maillot P, Perrot A. The effects of a new immersive multidomain training on cognitive, dual-task and physical functions in older adults. Geroscience 2024 Apr;46(2):1825-1841. [doi: 10.1007/s11357-023-00952-w] [Medline: 37743415]
13. Anderson-Hanley C, Arciero PJ, Brickman AM, et al. Exergaming and older adult cognition: a cluster randomized clinical trial. Am J Prev Med 2012 Feb;42(2):109-119. [doi: 10.1016/j.amepre.2011.10.016] [Medline: 22261206]
14. Anderson-Hanley C, Barcelos NM, Zimmerman EA, et al. The aerobic and cognitive exercise study (ACES) for community-dwelling older adults with or at-risk for mild cognitive impairment (MCI): neuropsychological, neurobiological and neuroimaging outcomes of a randomized clinical trial. Front Aging Neurosci 2018;10:76. [doi: 10.3389/fnagi.2018.00076] [Medline: 29780318]
15. Barcelos N, Shah N, Cohen K, et al. Aerobic and cognitive exercise (ACE) pilot study for older adults: executive function improves with cognitive challenge while exergaming. J Int Neuropsychol Soc 2015 Nov;21(10):768-779. [doi: 10.1017/S1355617715001083] [Medline: 26581789]

XSL•FO
RenderX

16. Maillot P, Perrot A, Hartley A. Effects of interactive physical-activity video-game training on physical and cognitive function in older adults. Psychol Aging 2012 Sep;27(3):589-600. [doi: 10.1037/a0026268] [Medline: 22122605]
17. Gallou-Guyot M, Mandigout S, Bherer L, et al. Effects of exergames and cognitive-motor dual-task training on cognitive, physical and dual-task functions in cognitively healthy older adults: an overview. Ageing Res Rev 2020 Nov;63:101135. [doi: 10.1016/j.arr.2020.101135] [Medline: 32768594]
18. Li J, Erdt M, Chen L, et al. The Social effects of exergames on older adults: systematic review and metric analysis. J Med Internet Res 2018 Jun 28;20(6):e10486. [doi: 10.2196/10486] [Medline: 29954727]
19. Stojan R, Voelcker-Rehage C. A systematic review on the cognitive benefits and neurophysiological correlates of exergaming in healthy older adults. J Clin Med 2019 May 23;8(5):734. [doi: 10.3390/jcm8050734] [Medline: 31126052]
20. Torre MM, Temprado JJ. A review of combined training studies in older adults according to A new categorization of conventional interventions. Front Aging Neurosci 2021;13:808539. [doi: 10.3389/fnagi.2021.808539] [Medline: 35177975]
21. Herold F, Theobald P, Gronwald T, et al. Alexa, let's train now! - a systematic review and classification approach to digital and home-based physical training interventions aiming to support healthy cognitive aging. J Sport Health Sci 2024 Jan;13(1):30-46. [doi: 10.1016/j.jshs.2023.01.004] [Medline: 36736727]
22. Gerling K, Mandryk R. Custom-designed motion-based games for older adults: A review of literature in human-computer interaction. Gerontechnology 2014;12(2):68-80. [doi: 10.4017/gt.2013.12.2.001.00]
23. Campbell M, Fitzpatrick R, Haines A, et al. Framework for design and evaluation of complex interventions to improve health. BMJ 2000 Sep 16;321(7262):694-696. [doi: 10.1136/bmj.321.7262.694] [Medline: 10987780]
24. Nawaz A, Skjæret N, Ystmark K, Helbostad JL, Vereijken B, Svanæs D. Assessing seniors' user experience (UX) of exergames for balance training. 2014 Oct 26 Presented at: NordiCHI '14: Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational; Oct 26-30, 2014; Helsinki Finland p. 578-587. [doi: 10.1145/2639189.2639235]
25. Pyae A, Liukkonen TN, Mika L, et al. Investigating the Finnish elderly people's user experiences in playing digital game-based skiing exercise: a usability study. Gerontechnology 2017 Aug 9;16(2):65-80. [doi: 10.4017/gt.2017.16.2.002.00]
26. Thin AG, Brown C, Meenan P. User experiences while playing dance-based exergames and the influence of different body motion sensing technologies. International Journal of Computer Games Technology 2013;2013:1-7. [doi: 10.1155/2013/603604]
27. Vaziri DD, Aal K, Ogonowski C, et al. Exploring user experience and technology acceptance for a fall prevention system: results from a randomized clinical trial and a living lab. Eur Rev Aging Phys Act 2016;13:6. [doi: 10.1186/s11556-016-0165-z] [Medline: 27293489]
28. Hassenzahl M. The thing and I: understanding the relationship between user and product. In: Blythe MA, Overbeeke K, Monk AF, Wright PC, editors. Funology: Dordrecht: Springer Netherlands; 2003, Vol. 3:31-42. [doi: 10.1007/1-4020-2967-5_4]
29. Adcock M, Sonder F, Schättin A, Gennaro F, de Bruin ED. A usability study of a multicomponent video game-based training for older adults. Eur Rev Aging Phys Act 2020;17:3. [doi: 10.1186/s11556-019-0233-2] [Medline: 31938075]
30. Gerling KM, Schild J, Masuch M. Exergaming for elderly: analyzing player experience and performance. In: Eibl M, editor. Mensch & Computer 2011: Oldenbourg Wissenschaftsverlag GmbH; 2011:401-411. [doi: 10.1524/9783486712742.401]
31. Adcock M, Thalmann M, Schättin A, et al. A pilot study of an in-home multicomponent exergame training for older adults: feasibility, usability and pre-post evaluation. Front Aging Neurosci 2019;11:304. [doi: 10.3389/fnagi.2019.00304] [Medline: 31824295]
32. Hoda M, Alattas R, El Saddik A. Evaluating player experience in cycling exergames. Presented at: 2013 IEEE International Symposium on Multimedia (ISM); Dec 9-11, 2013; Anaheim, CA, USA p. 415-420. [doi: 10.1109/ISM.2013.81]
33. Soltani P, Figueiredo P, Vilas-Boas JP. Does exergaming drive future physical activity and sport intentions? J Health Psychol 2021 Oct;26(12):2173-2185. [doi: 10.1177/1359105320909866] [Medline: 32114831]
34. Lyons EJ, Tate DF, Ward DS, et al. Engagement, enjoyment, and energy expenditure during active video game play. Health Psychol 2014 Feb;33(2):174-181. [doi: 10.1037/a0031947] [Medline: 23527520]
35. Kim SY(, Prestopnik N, Biocca FA. Body in the interactive game: How interface embodiment affects physical activity and health behavior change. Comput Human Behav 2014 Jul;36:376-384. [doi: 10.1016/j.chb.2014.03.067]
36. Gellish RL, Goslin BR, Olson RE, et al. Longitudinal modeling of the relationship between age and maximal heart rate. Med Sci Sports Exerc 2007 May;39(5):822-829. [doi: 10.1097/mss.0b013e31803349c6] [Medline: 17468581]
37. de Souto Barreto P, Ferrandez AM, Saliba-Serre. B. Questionnaire d'activité physique pour les personnes âgées (QAPPA): validation d'un nouvel instrument de mesure en langue française [Article in French]. Sci Sports 2011 Feb;26(1):11-18. [doi: 10.1016/j.scispo.2010.09.006]
38. Hassenzahl M, Burmester M, Koller F. AttrakDiff: ein fragebogen zur messung wahrgenommener hedonischer und pragmatischer qualität [Book in German]. In: Szwillus G, Ziegler J, editors. Mensch &Amp; Computer 2003: Wiesbaden: Vieweg+Teubner Verlag; 2003:187-196. [doi: 10.1007/978-3-322-80058-9_19]
39. Lallemand C, Koenig V, Gronier G, Martin R. Création et validation d'une version française du questionnaire AttrakDiff pour l'évaluation de l'expérience utilisateur des systèmes interactifs [Article in French]. European Review of Applied Psychology 2015 Sep;65(5):239-252. [doi: 10.1016/j.erap.2015.08.002]

XSL•FO
RenderX

40. Hart SG, Staveland LE. Development of NASA-TLX (task load index): results of empirical and theoretical research. In: Advances in Psychology: Elsevier; 1988, Vol. 52:139-183. [doi: 10.1016/S0166-4115(08)62386-9]
41. Luximon A, Goonetilleke RS. Simplified subjective workload assessment technique. Ergonomics 2001 Feb 20;44(3):229-243. [doi: 10.1080/00140130010000901] [Medline: 11219757]
42. Bittner AC, Byers JC, Hill SG, et al. Generic workload ratings of a mobile air defense system (LOS-F-H). Proceedings of the Human Factors Society Annual Meeting 1989 Oct;33(20):1476-1480. [doi: 10.1177/154193128903302026]
43. Cegarra J, Morgado N. Actes du 5ème colloque EPIQUE (psychologie ergonomique) [article in french]. 2009 Presented at: 5ème Colloque EPIQUE (Ergonomie Psychologique et Ingénierie des Connaissances Utiles en Conception); Nice, France URL: https://arpege-recherche.org/user/pages/06.activites/03.colloques-epique/10.5e-colloque-epique/ActesEPIQUE2009.pdf [accessed 2025-05-22]
44. Kendzierski D, DeCarlo KJ. Physical activity enjoyment scale: two validation studies. J Sport Exerc Psychol 1991;13(1):50-64. [doi: 10.1123/jsep.13.1.50]
45. Delignières D, Pérez S. Le plaisir perçu dans la pratique des activités physiques et sportives: élaboration d'un outil d'évaluation [Article in French]. STAPS 1998;19(45):7-18. [doi: 10.3406/staps.1998.1106]
46. Borg GAV. Psychophysical bases of perceived exertion. Medicine & Science in Sports & Exercise 1982 May;14(5):377. [doi: 10.1249/00005768-198205000-00012]
47. Cohen J. Statistical Power Analysis for the Behavioural Sciences: L Erlbaum Associates; 1988:102.
48. Agmon M, Perry CK, Phelan E, et al. A pilot study of Wii Fit exergames to improve balance in older adults. J Geriatr Phys Ther 2011;34(4):161-167. [doi: 10.1519/JPT.0b013e3182191d98] [Medline: 22124415]
49. Lallemand C, Gronier G. Méthodes de Design UX: 30 Méthodes Fondamentales Pour Concevoir et Évaluer Les Systèmes Interactifs [Book in French]: Paris: Eyrolles; 2015. URL: https://www.utstat.toronto.edu/brunner/oldclass/378f16/readings/CohenPower.pdf
50. Pham TP, Theng YL. Game controllers for older adults: experimental study on gameplay experiences and preferences. Presented at: FDG '12: Proceedings of the International Conference on the Foundations of Digital Games; Raleigh North Carolina p. 284-285. [doi: 10.1145/2282338.2282401]
51. Rebsamen S, Knols RH, Pfister PB, et al. Exergame-driven high-intensity interval training in untrained community dwelling older adults: a formative one group quasi- experimental feasibility trial. Front Physiol 2019;10:1019. [doi: 10.3389/fphys.2019.01019] [Medline: 31440168]
52. Schoene D, Lord SR, Delbaere K, et al. A randomized controlled pilot study of home-based step training in older people using videogame technology. PLoS ONE 2013;8(3):e57734. [doi: 10.1371/journal.pone.0057734] [Medline: 23472104]
53. Caserman P, Hoffmann K, Müller P, et al. Quality criteria for serious games: serious part, game part, and balance. JMIR Serious Games 2020 Jul 24;8(3):e19037. [doi: 10.2196/19037] [Medline: 32706669]
54. Li J, Li L, Huo P, et al. Wii or Kinect? A pilot study of the exergame effects on older adults' physical fitness and psychological perception. Int J Environ Res Public Health 2021 Dec 8;18(24):12939. [doi: 10.3390/ijerph182412939] [Medline: 34948547]

Abbreviations

AVG: active video game
BIKE-VG: bike and video games
I2WE: immersive and interactive wall exergames
IRRID: International Registered Report Identifier
NASA-TLX: NASA-Task Load Index
PACES: Physical Activity Enjoyment Scale
QAPPA: Questionnaire d'Activité Physique pour les Personnes Âgées
RTLX: Raw Task Load Index
TMHR: Theoretical Maximum Heart Rate
UX: User Experience

Edited by A Coristine, E Oliveira; submitted 08.10.24; peer-reviewed by KM Naugle, S Göbel; revised version received 23.01.25; accepted 11.02.25; published 26.05.25.

Please cite as:
Béraud-Peigné N, Perrot A, Maillot P
Active Video Games Training for Older Adults: Comparative Study of User Experience, Workload, Pleasure, and Intensity
JMIR Serious Games 2025;13:e67314
URL: https://games.jmir.org/2025/1/e67314
doi:10.2196/67314

XSL•FO
RenderX

© Néva Béraud-Peigné, Alexandra Perrot, Pauline Maillot. Originally published in JMIR Serious Games (https://games.jmir.org), 26.5.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Physical Activity and Enjoyment in Active Virtual Reality Games in Youth: Comparative Analysis of Gorilla Tag and Beat Saber

Brenden Boots[1], BS; Daniel Berg[1], BS; Easton Hewitt[2]; Keith Naugle[1], MS, PhD; Kelly Naugle[1], MS, PhD

[1]School of Health and Human Sciences, Indiana University Indianapolis, 901 East New York St, Indianapolis, IN, United States

[2]School of Public Health, Indiana University, Bloomington, IN, United States

Corresponding Author:
Brenden Boots, BS
School of Health and Human Sciences, Indiana University Indianapolis, 901 East New York St, Indianapolis, IN, United States

Abstract

Background: Virtual reality (VR) active gaming is growing in popularity, but little is known about physical activity during gameplay. Two popular VR games are Gorilla Tag (Another Axiom Inc) and Beat Saber (Beat Games). Little is known about physical activity during these games in youth.

Objective: The purpose of this study was to investigate the enjoyment, intensity, and amount of physical activity while playing Gorilla Tag and Beat Saber in early adolescent youth.

Methods: Sixteen participants, 13 males and 3 females with an average age of 10.7 (SD 0.34) years, played 2 VR games (Gorilla Tag and Beat Saber) in a single session. Both games followed the same procedure: a maximum of 10-minute familiarization period, 5 minutes of rest, 15 minutes of gameplay, and 10 minutes of rest. Participants wore a heart rate monitor to track heart rate reserve (%HRR) and accelerometers on the wrist and waist to monitor time in sedentary activity, light physical activity, and moderate to vigorous physical activity of the arm and whole body. The Physical Activity Enjoyment Scale–Child Version (PACES) and ratings of perceived exertion (RPE) were completed after each game. Dependent *t* tests compared measures between games.

Results: The results revealed that average and maximum %HRR were significantly higher during Gorilla Tag than during Beat Saber, with heart rate–based physical activity intensity reaching light for Beat Saber and moderate for Gorilla Tag. Arm moderate to vigorous physical activity and whole-body moderate to vigorous physical activity and light physical activity were greater during Gorilla Tag than during Beat Saber. Arm and whole-body sedentary time were significantly lower during Gorilla Tag than during Beat Saber. Gorilla Tag and Beat Saber were rated as highly enjoyable. There were no differences between games for maximum (P=.352) or average (P=.362) RPE. Both games were rated as light intensity for average RPE (Gorilla Tag: mean 4.3, SD 1.9; Beat Saber: mean 4.7, SD 2.3) and moderate intensity for maximum RPE (Gorilla Tag: mean 5.4, SD 1.9; Beat Saber: mean 5.8, SD 2.4).

Conclusions: These results suggest that Beat Saber produced light-intensity physical activity and Gorilla Tag produced light- to moderate-intensity physical activity in early adolescent youth, with both games rated as highly enjoyable.

(JMIR Serious Games 2025;13:e66593) doi:10.2196/66593

KEYWORDS

active gaming; movement; Gorilla Tag; Beat Saber; virtual reality; VR; physical activity; youth; early adolescents; young adults; gaming; heart rate; exergame; enjoyment

Introduction

Sedentary lifestyles have been on the rise in the United States over the last few decades, with some sources estimating only 20% of children or adolescents meet the minimum amount of recommended physical activity (60 minutes of moderate to vigorous physical activity per day) [1,2]. Alarmingly, risk factors for chronic diseases (heart disease, hypertension, and type 2 diabetes) are increasing in children [3]. Regular physical activity decreases the likelihood that these risk factors will develop [2]. One potential reason for the shift into a sedentary lifestyle in adolescent youth could be the dramatic increase in screen time [4]. For example, approximately 70% of individuals below the age of 18 years play video games regularly [5]. However, video games could also offer a solution. Active video gaming, in which the user moves in coordination with the system, has increased in popularity in the last 20 years. Up to this point, research has shown mixed results on whether active video gaming can produce moderate-intensity physical activity, which is beneficial for cardiovascular health [6-10]. For example, one study conducted by Graf et al [7] compared treadmill walking to games on the Nintendo such as Dance Dance Revolution. The results showed that active gaming elicited similar energy expenditure to moderate walking on a treadmill in children. In contrast, Ufholz et al [11] evaluated active gaming as an

XSL•FO
RenderX

intervention in which children were required to play active video games for 3 - 4 weeks. Video game play reached light physical activity levels but never reached moderate physical activity. Overall, based on the current literature, the intensity of gameplay during active gaming is somewhat dependent on the game being played.

In the last 10 years, a new form of active video gaming has emerged involving the use of virtual reality (VR). This new form of gaming allows users to wear a head-mounted display and use handheld controllers to interact with a virtual environment using movement in all 3 planes of motion. VR is relatively new compared to other forms of active gaming, so little is known about the physical activity levels reached during gameplay of commercial VR games. Further, most active VR studies have focused on adults [6,12-14]. Currently, two of the most popular commercially available games are Beat Saber (Beat Games) and Gorilla Tag (Another Axiom Inc), with millions of active users for these games [15,16]. Both games require significant movement to play the game. Gorilla Tag elicits upper and lower body movements as the player, a gorilla, plays tag with other gorillas using their arms to run, jump, or climb in the VR environment. Beat Saber is primarily an upper body game that uses the hands and arms to slice oncoming blocks to music. Currently, limited studies have evaluated both physical activity and enjoyment of either game in youth. Most recently, Godfrey and colleagues [17] evaluated the physical activity intensity of youth while playing Beat Saber and Thrill of Flight VR games. Similar to a study in adults [6], Beat Saber was played at a light intensity. However, data were analyzed for only 4 minutes of playtime, and participants were connected to tubing to measure VO2, which may have restricted movement. To our knowledge, no published studies have evaluated Gorilla Tag in youth or adults. Therefore, the purposes of this study were (1) to determine the intensity of physical activity via heart rate (HR), ratings of perceived exertion (RPE), and total body movements during gameplay and (2) to evaluate enjoyment levels of Beat Saber and Gorilla Tag in early adolescent youth. Based on prior research [6,17] and the significant upper body movement required by these games, the hypotheses were that Gorilla Tag and Beat Saber would produce light to moderate physical activity and high levels of enjoyment during gameplay.

Methods

Participants

Sixteen youths (average age of 10.7 [SD 0.34] years) participated in this study and completed the 2023 Physical Activity Readiness Questionnaire (PAR-Q) to determine if they were able to perform physical activity without medical supervision [18]. A power analysis using G Power 3.1 was used to estimate the sample size needed for detecting a within-subject difference in raw HR between games and across minutes of gameplay using repeated measures ANOVA. With α set at .05, power at 0.8, a 0.5 correlation among repeated measures, and an estimated effect size of f=0.25, the power analysis determined that 16 participants were needed. Other exclusion criteria included self-reported motion sickness, seizures, and epilepsy. Participants were recruited through email and newspaper advertisements around Indianapolis.

Ethical Considerations

This study was approved by the Indiana University Institutional Review Board (IRB protocol #18787). All participants signed an IRB-approved assent form. Parents or guardians of participants signed an IRB-approved informed consent form. All paper (except informed consent or assent forms) and computer records containing information about participants were identified only by a participant number rather than by participant name. Additionally, all paper records were maintained in a locked filing cabinet in a locked room and were only accessible to the study investigators. Computer data files (without name identifiers) were stored on computer servers with secure passwords or encrypted electronic storage devices. All participants who completed the study were compensated with US $25 in gift cards at the end of the study session.

Procedure

Participants attended one study session that lasted approximately 2 hours. See Figure 1 for an overview of study session events. The beginning of the session included completing the informed consent and assent, inclusion or exclusion criteria assessment, and demographic questionnaire that included height, weight, ethnicity, age, sex, and VR experience level. Participants rated their experience playing each game using the following options: never, ≤1 time per week, 2 - 4 times per week, and ≥5 times per week. Participants were fitted with an HR monitor wrapped around the chest with the sensor at the xiphoid process and accelerometers placed on the dominant wrist and wrapped around the waist with the sensor placed on the right side at hip level. Participants were asked to sit quietly for 10 minutes to collect resting HR. Participants were fitted with the VR headset and shown how to use the VR controllers, with the remainder of the session devoted to playing both Gorilla Tag and Beat Saber. The order of the games was counterbalanced among participants. First, one game was played for a 5 - 10-minute familiarization period. Most participants received 5 minutes, but a few participants who did not understand the movement system received more time at the discretion of the researchers. Participants rested for 5 minutes and then completed a 15-minute data collection period of VR gameplay. Participants played at a self-selected intensity. Next, participants rested for 10 minutes and then completed the same procedure for the second game. RPEs were collected at the end of each 15-minute gameplay, and an enjoyment questionnaire was completed following each game. Physical activity intensity during gameplay was measured with an HR watch, chest strap, and 2 accelerometers during each session.

XSL•FO
RenderX

Figure 1. Order of events during the study session.

		Game 1				Game 2		
Informed consent and screening	Equipment set-up	Game familiarization (5-10 min)	5-min rest period	15-min bout of game play	10-min rest	Game familiarization (5-10 min)	5-min rest period	15-min bout of game play

VR Instrumentation

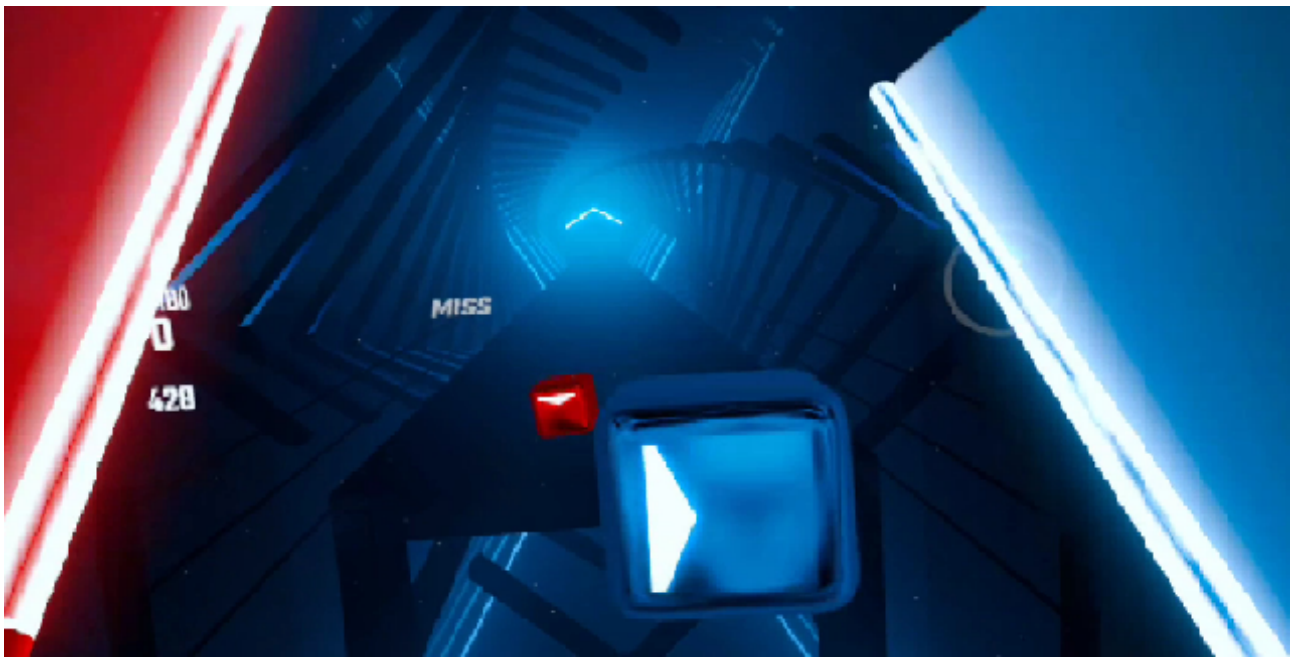
VR Equipment

The Meta Quest 2 VR system includes a head-mounted display and 2 handheld controllers. The Meta Quest 2 uses motion tracking technology in which the movement of the head and controllers is translated into the 3D environment displayed on the headset. Each participant had the system verbally explained to them and was familiarized with the equipment. Manufacturer guidelines suggest that the VR system is for users aged ≥10 years. Participants played 2 separate games (described below) in a 6.5 ft × 8.5 ft space.

Beat Saber

Beat Saber is a music-based game that places players in the path of oncoming colored blocks. Players must cut or slice the blocks in the correct direction, indicated by a white arrow on the blocks, using controllers, which are colored light sabers in the game (see Figure 2). The goal of the game is to successfully cut a majority of the blocks, or the song will fail. However, for this study, we used the no-fail setting to minimize interruption of gameplay. Additionally, participants played the same 5 songs in the same order and could select the difficulty level they played following the first song, which was played on “normal.”

Figure 2. Picture of gameplay during Beat Saber. Participants control the red and blue light sabers with the VR controllers and attempt to cut the blocks in the direction indicated by the arrow.



Gorilla Tag

Gorilla Tag is an online multiplayer game that is based on a unique movement system that has players, as virtual gorillas, use their arms to propel themselves around the map by virtually running, climbing, or jumping. The participants played a tag-like game mode (ie, infection mode) in the forest map where they either had to attempt to tag other players or avoid being tagged (see Figure 3). Other live players were present during the game; therefore, we muted all other online players so that participants would not hear any inappropriate language.

XSL•FO
RenderX

Figure 3. Gorilla tag. Top left: picture of a virtual gorilla using the arms to move. Top right: picture of a gorilla who has been tagged and thus appears with a lava skin to indicate they are a tagger. Bottom: picture of forest map.



Outcome Measures

HR Measure

A Polar HR monitor measured HR right before the start of gameplay for a baseline measurement and during every second of gameplay. Polar equipment consisted of a Polar Unite series watch paired with a Polar H10 HR Sensor. Raw HR values were averaged for minutes 1 - 4, 5 - 9, and 10 - 14. The maximum and average HR values for minutes 1 - 14 were also recorded. We calculated HR reserve (HRR) percentages to assess intensity levels during gameplay. The following formula was used: [(average HR during activity – resting HR)/HRR]×100 [19]. As recommended by the American College of Sports Medicine, the HRR% ranges used to determine exercise intensity were light (30% - 39%), moderate (40% - 59%), and vigorous (≥60%) [20,21]. We also calculated the percentage of time that participants were in moderate to vigorous intensity (ie, the percentage of HRR values that were ≥40% for each second of gameplay) during the 15-minute bout.

RPE Measure

RPE was measured using the 0 - 10 children's OMNI RPE scale [22]. Intensity was described to participants as "not tired at all" for 0 and "very, very tired" for 10. Participants were also given the following instructions on how to use the scale, as recommended by Robertson et al [22]:

> *Perceived exertion is how tired your body feels during exercise. Please use the numbers on the picture to tell us how your body feels when you are doing the activity. Look at the person at the bottom of the hill. If you feel like this person you will be "not tired at all," so you should point to the 0 (zero). Now look at the person who is at the top of the hill. If you feel like this person you will be "very, very tired," so you should point to number 10. If you fall somewhere in between, point to a number between 0 and 10. We want you to tell us how your whole body feels, and*

XSL•FO
RenderX

remember there are no right or wrong answers. Use both the pictures and the words to help you choose.

Participants gave a maximum and average RPE following gameplay while looking at the scale. This scale has been used in a prior youth active VR gaming study [17] and has been validated for preadolescent youth on a range of exercise-related activities [22-24].

Accelerometry

The ActiGraph GT3X+ accelerometer was worn on the hip and dominant wrist during gameplay. Activity data were captured in 1-second epochs. Accelerometer data used for analysis were calculated from minutes 1 - 14 of each gameplay session. ActiLife software was used to process the data and calculate the following variables: percentage of time in sedentary time (%SED), percentage of time in light physical activity (%LPA), and percentage of time in moderate to vigorous physical activity (%MVPA). The data for the waist were processed using the Romanzini et al [25] cut points for vector magnitude counts, while the wrist data were processed with the Crouter et al [26] cut points for vector magnitude. The ActiGraph GT3X+ is an accurate and reliable tool that has been previously used in active gaming studies to measure physical activity [27-29].

Enjoyment

Participants completed the child-modified Physical Activity Enjoyment Scale (PACES), which uses a 16-question bipolar scaling method [30]. Each statement includes a 5-point Likert scale ranging from “disagree a lot”=1 to “agree a lot”=5. Questions assessed enjoyment during physical activity (score is average of 16 items) with each statement beginning with the stem “When I am physically active...” For this study, we modified the stem statement to “When I am playing Gorilla Tag/Beat Saber...” The PACES is a validated tool for children to measure enjoyment of activity [31].

Data Analysis

Descriptive statistics were calculated for each outcome measure for each game. The outcome variables for HR, RPE, the accelerometer, and PACES were analyzed with dependent *t* tests to determine whether differences existed between games. We conducted a 2 (game) × 4 (time: baseline, 1 - 4 min, 5 - 9 min, and 10 - 14 min) repeated measures ANOVA on the raw HR data to examine whether HR increased from baseline (HR just prior to gameplay) to during gameplay for each game. Post hoc analyses were conducted with simple effects tests for significant interactions and *t* tests with Bonferroni corrections for significant main and simple effects. For the raw HR analysis, we also reported partial eta-square (η_p^2), which is an effect size in ANOVA models. Partial eta-square values range from 0 to 1 and can be interpreted as 0.01 being a small effect, 0.06 being a medium effect, and ≥0.14 being a large effect. The frequency of responses for each item of the PACES was calculated to obtain a detailed description of whether participants enjoyed each game. Alpha was set at $P<.05$.

Results

Descriptive Statistics

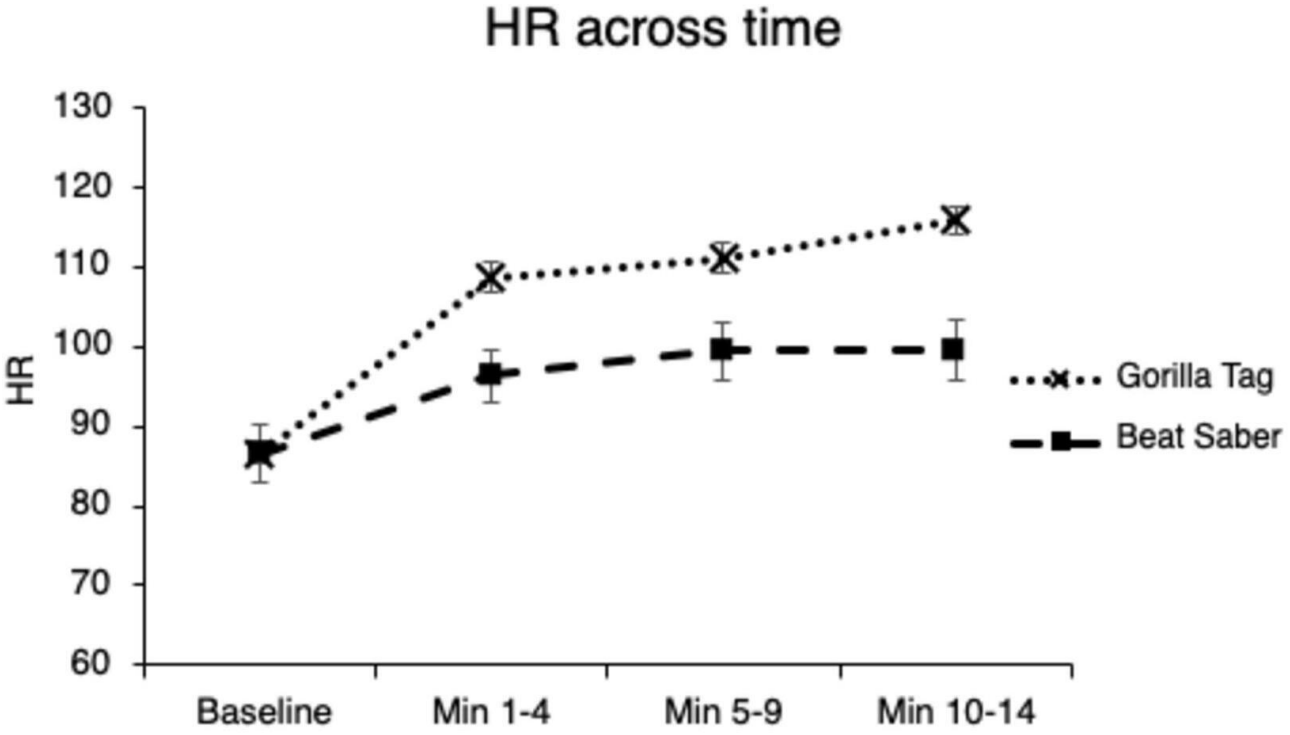
Sixteen participants (13 male) enrolled in this study. The average age of participants was 10.7 (SD 0.34) years. Fourteen participants identified as Caucasian and 2 identified as African American. Eight participants had prior experience playing Gorilla Tag, while 11 participants had experience playing Beat Saber. The heart rate was not collected properly for 1 participant (male); thus, this participant was not included in the HR data analysis.

HR Across Time

The repeated measures ANOVA revealed a significant main effect of the game ($P=.004$, $\eta_p^2=0.46$) and time ($P<.001$, $\eta_p^2=0.64$). The main effects were superseded by a significant game × time interaction, $P=.009$ ($\eta_p^2=0.34$). The simple effects tests of the game within each timepoint were significant (P's<.01), except at baseline ($P=.952$). HR was significantly greater during Gorilla Tag at each timepoint during gameplay compared to Beat Saber. The simple effects tests of time were significant within Gorilla Tag ($P=.003$) and Beat Saber ($P=.012$). For Gorilla Tag, HR during all timepoints of gameplay (1 - 4 min, 5 - 9 min, and 10 - 14 min) was significantly greater than baseline. For Beat Saber, HR at 5 - 9 minutes and 10 - 14 minutes was greater than baseline. In addition, HR at 5 - 9 minutes was greater than HR at 1 - 4 minutes. See Figure 4 for HR across time for each game.

XSL•FO
RenderX

Figure 4. Heart rate (HR) across time for each game. Error bars represent standard error. Min=minutes.



Percentage of HRR

See Table 1 for the means and SDs for the %HRR variables, RPE measures, and PACES scores for each game. The dependent *t* tests revealed significantly higher average %HRR (*P*=.004) and maximum %HRR (*P*=.011) for Gorilla Tag compared to Beat Saber. For Gorilla Tag, the average %HRR was equivalent to light physical activity, while the maximum %HRR was equivalent to moderate-intensity physical activity. Based on the %HRR values, Beat Saber reached a maximum of light-intensity physical activity. The analysis conducted on the percentage of time in MVPA (based on %HRR) was significant (*P*=.036), with participants obtaining a greater amount of MVPA during Gorilla Tag compared to Beat Saber.

Table . Mean, SD, and 95% CI for % heart rate reserve, ratings of perceived exertion, and PACES[a] score.

Variable	Gorilla Tag			Beat Saber			*t* test[b] (*df*)	*P* value
	Mean	SD	95% CI	Mean	SD	95% CI		
Average %HRR[c]	31.5	13.4	24.1 to 38.9	21.3	8.6	16.6 to 26.1	3.48 (14)	.004
Maximum %HRR	49.8	15.9	41.0 to 58.7	37.1	9.1	32.0 to 42.1	−2.93 (14)	.011
Time in MVPA[d], %	27.1	34.8	3.1 to 4.9	5.4	34.9	3.3 to 5.9	2.32 (15)	.036
Average RPE[e]	4.3	1.9	4.3 to 6.3	4.7	2.3	4.3 to 6.8	−0.94 (15)	.362
Maximum RPE	5.5	1.9	7.8 to 46.4	5.8	2.4	−2.5 to 13.2	−0.96 (15)	.352
PACES score	4.1	0.6	3.8 to 4.4	4.1	0.7	3.8 to 4.5	0.09 (15)	.931

[a]PACES: Physical Activity Enjoyment Scale–Child Version.

[b]All *t* tests are 2-tailed.

[c]HRR: heart rate reserve.

[d]MVPA: moderate to vigorous physical activity; time in MVPA is based on %HRR.

[e]RPE: ratings of perceived exertion.

RPE and Enjoyment

The dependent *t* tests did not show significant differences between games for maximum RPE (*P*=.352) and average RPE (*P*=.362). No significant differences existed between games for PACES score, *P*=.931. The frequency of item responses for each item of the PACES is presented in Tables 2 and 3.

XSL•FO
RenderX

Table . Frequency of responses for each item of the PACES[a] for Gorilla Tag.

	Disagree a lot	Disagree	Not sure	Agree	Agree a lot
I enjoy it	0	0	1	6	9
I feel bored	7	7	1	1	0
I dislike it	10	4	2	0	0
I find it pleasurable	0	1	2	5	8
It's no fun at all	11	4	1	0	0
It gives me energy	0	1	4	5	6
It makes me sad	12	3	1	0	0
It's very pleasant	0	0	5	6	5
My body feels good	0	1	3	8	4
I get something out of it	1	1	3	8	4
It's very exciting	0	0	3	5	8
It frustrates me	5	3	3	3	2
It's not at all interesting	10	5	1	0	0
It gives me a strong feeling of success	1	2	5	5	3
It feels good	0	1	3	6	6
I feel as though I would rather be doing something else	6	6	3	1	0

[a]PACES: Physical Activity Enjoyment Scale–Child Version.

XSL•FO
RenderX

Table . Frequency of responses for each item of the PACES[a] for Beat Saber.

	Disagree a lot	Disagree	Not sure	Agree	Agree a lot
I enjoy it	0	0	1	5	11
I feel bored	8	5	2	1	0
I dislike it	6	8	2	0	0
I find it pleasurable	0	0	3	7	6
It's no fun at all	11	3	2	0	0
It gives me energy	1	0	3	8	4
It makes me sad	13	2	1	0	0
It's very pleasant	0	2	3	6	5
My body feels good	1	2	4	5	4
I get something out of it	0	1	5	5	5
It's very exciting	0	0	3	7	6
It frustrates me	8	2	0	4	2
It's not at all interesting	10	4	1	1	0
It gives me a strong feeling of success	0	2	4	4	6
It feels good	1	0	6	2	7
I feel as though I would rather be doing something else	6	5	1	4	0

[a]PACES: Physical Activity Enjoyment Scale–Child Version.

Accelerometer Variables

Wrist

See Table 4 for the means and SDs for the accelerometer data for each game. The dependent *t* tests showed significant differences between games for %SED (P=.004), %LPA (P=.03), and %MVPA (P=.003). Participants spent more time in sedentary time and LPA for Beat Saber and more time in MVPA for Gorilla Tag.

Table . Mean, SD, and 95% CI for accelerometer variables.

Variable	Gorilla Tag			Beat Saber			*t* test[a] (*df*)	*P* value
	Mean	SD	95% CI	Mean	SD	95% CI		
Wrist								
% Sedentary	2.5	2.4	1.2-4.0	9.7	8.3	5.2-14.6	−3.42 (15)	.004
% Light PA[b]	0.5	0.4	0.3-0.7	0.9	0.5	0.6-1.2	−0.4 (15)	.03
% MVPA[c]	96.9	2.7	95.3-98.4	89.4	8.4	84.4-94.0	3.5 (15)	.003
Waist								
% Sedentary	59.7	16.7	49.3-67.6	87.0	11.7	80.4-93.8	−6.09 (15)	<.001
% Light PA	20.7	6.0	18.2-24.4	6.8	6.5	3.0-10.5	7.19 (15)	<.001
% MVPA	20.0	13.1	12.9-27.6	6.2	5.7	2.9-9.4	4.4 (15)	<.001

[a]All *t* tests are 2-tailed.

[b]PA: physical activity.

[c]MVPA: moderate to vigorous physical activity.

XSL•FO
RenderX

Waist

The dependent *t* tests showed significant differences between games for %SED ($P<.001$), %LPA ($P<.001$), and %MVPA ($P<.001$). Participants spent more time in %SED for Beat Saber and more time in %LPA and %MVPA for Gorilla Tag.

Discussion

Rationale

With our society becoming increasingly sedentary, innovative strategies are needed to increase physical activity. Given the increasing popularity of video games among children and adolescents, VR active games could be used to increase physical activity in this age group. However, since this technology is new, a lack of research exists on whether these newer games engage participants in MVPA. To our knowledge, our study is the first to examine physical activity, RPE, and enjoyment in early adolescent youths playing the widely popular games Gorilla Tag and Beat Saber. Our study found that both VR games produced light- to moderate-intensity physical activity while being rated as highly enjoyable.

Principal Findings

The first hypothesis stated Gorilla Tag and Beat Saber would produce light to moderate physical activity. The HR data partially supported this hypothesis, with both games increasing participants' HR significantly from baseline to during gameplay. Gorilla Tag elicited significantly higher %HRR and higher %MVPA than Beat Saber. Based on %HRR data, physical activity during Gorilla Tag was on-average light and a maximum of moderate, with over 25% of playtime spent in MVPA (based on HR data). Beat Saber elicited very light physical activity to light physical activity, with only an average of 5% of playtime spent in MVPA. These results are similar to prior studies in adults, which indicate that adults play Beat Saber at a very light to light intensity [6,13]. For example, one study examining actual versus perceived exertion found that while playing Beat Saber actual and perceived exertion were in the light to very light range [13]. Similar to our study, Godfrey et al [17] found that youths aged 8 - 12 years had significantly higher HR while playing Beat Saber compared to rest. In addition, based on a 4-minute period of gameplay, participants had %HRR and metabolic equivalents of task values consistent with light intensity. In the Godfrey study, participants were connected to tubing to measure VO2, which may have restricted movement. However, our study's results suggest that even when youths are less restricted in movement, Beat Saber is still only played at a light intensity.

Our accelerometer data also provided insight into the type and intensity of movement during both games. The accelerometer data showed similar trends to HR with significantly more moderate to vigorous movement of the arm and whole body (ie, waist accelerometer) during Gorilla Tag compared to Beat Saber. For Gorilla Tag, participants spent significant time in upper-body movement (~97%) and just under half the time in whole-body movement. During Beat Saber, participants spent most of their time (≥89%) in MVPA for the arm but little time in whole-body movement (~13%). This is likely because Beat Saber is an upper-body-based game requiring very little lower-body activity to be successful. The Beat Saber accelerometer data were similar to Evans et al [6], who found adults spent a high percentage of time in arm MVPA but little time in whole-body MVPA while playing Beat Saber. Likely, the significant upper-body movement required to play Beat Saber does not, by itself, induce physiological changes consistent with higher cardiovascular physical activity intensity levels (ie, intensity based on HR). Indeed, several studies found that whole-body movement versus arm movement measured via accelerometers is more important for energy expenditure during active gaming [29,32]. For example, Naugle et al [29] revealed that whole-body MVPA (waist accelerometer), but not arm MVPA (wrist accelerometer), predicted energy expenditure during active gaming as measured by a portable metabolic measurement system during gameplay.

The perceived intensity of physical activity was measured during gameplay via RPE with the children's OMNI scale. The RPE data indicated that participants rated their exertion toward the middle of the scale as "getting more tired" to "tired" for both Gorilla Tag and Beat Saber. In contrast to the HR data, no differences were found in perceived intensity between games. Several explanations could explain these contrasting results. First, even though the OMNI RPE scale has been validated in children [22], it is possible participants did not fully understand the scale. Participants could have been rated on the difficulty of the game to be successful rather than on the physical exertion of playing the game. Additionally, prior studies show that VR games cause inaccurate perceptions of exertion during physical activity via distraction from bodily sensations [12,13,33]. Further, because RPE ratings were taken at the end of gameplay, it is possible that the ratings could have been influenced by fatigue, depending on the fitness of the participants. Finally, several other potential factors that were unexplored in this study may have contributed to this discrepancy between the HR and RPE data, including age, gender, and experience. Future research should explore the potential factors underlying the discrepancy between HR and RPE data during VR games.

The final hypothesis was that playing both Gorilla Tag and Beat Saber would elicit high levels of enjoyment. The PACES data supported this hypothesis, with both games having an average score above 4 (5 representing the maximum enjoyment) with no difference between games. In a prior study of over 600 children using the PACES, participants rated enjoyment related to physical activity just under a score of 4 [31]. Perhaps active VR games could enhance physical activity enjoyment in youth. Enjoying physical activity is extremely important for physical activity participation in youth. Indeed, DiLorenzo and colleagues [34] revealed physical activity enjoyment was the only consistent predictor of physical activity levels among 5th- and 6th-grade children. Future research should examine whether youth enjoy the active VR games more than traditional exercises completed at a similar intensity.

Limitations

A few limitations of this study should be noted. First, the game modes, maps, and settings were all constrained to attempt to provide some standardization across the games for each

XSL•FO
RenderX

participant. Using different game settings could have potentially provided different results, such as greater physical exertion or enjoyment levels. In addition, given that Gorilla Tag is an online multiplayer game, enjoyment may also significantly increase in Gorilla Tag if participants were able to play online with their friends. Second, the sample was small and comprised primarily of male and Caucasian children. Thus, whether these results generalize to females and children of other races is unknown. Third, prior active gaming studies have found prior game experience and skill level can contribute to physical activity levels during gameplay [35,36]. Participants in this study had a mix of VR experiences. Fourth, this study did not include measurement for the frequency or habitual activity level of the participants. Fifth, there was only one session of gameplay, and how youth play the games may change over time with repeated play. Prior studies suggest motivation to play active video games may decline over time after the novelty of the games wears off [37]. However, this has not been tested with active VR games that allow for a range of experiences. Lastly, Beat Saber provides a proficiency score following each song indicating the number of blocks successfully cut as a percentage. These data were not collected and could have been used to explore correlations with enjoyment or activity level.

Future Directions

To broaden the applicability of the results, future studies evaluating Gorilla Tag or Beat Saber should compare gameplay at different game settings (Beat Saber: normal vs hard difficulty levels; Gorilla Tag: playing with a friend vs strangers) among participants. Future research should also explore how intensity and movement patterns change with more experience in Gorilla Tag and Beat Saber, as well as compare the gameplay of individuals who are active compared to nonactive. Furthermore, longitudinal studies should assess physical activity and enjoyment during gameplay across multiple sessions of gameplay.

Conclusions

In conclusion, Beat Saber produced light-intensity physical activity, and Gorilla Tag produced light- to moderate-intensity physical activity in early adolescent youth, with both games rated as highly enjoyable. Beat Saber primarily involved upper-extremity movement, while Gorilla Tag elicited upper-extremity and whole-body movement. Future research should evaluate more VR active games as a potential mode of physical activity in youth, particularly games that incorporate whole-body movement. Active VR games could potentially be a part of a larger program for exercise-adverse youth in many settings. Enjoyment of physical activity through VR games may foster greater intrinsic motivation in youth to play these active games and hence enhance physical activity participation. However, more research is needed to confirm these hypotheses.

Acknowledgments

We would like to acknowledge Carolyn Boone, who assisted with data collection for this study. This study was funded by the IUPUI Undergraduate Research Opportunity Program (KEN), which had no involvement in the review or approval of the manuscript.

Data Availability

The raw data supporting the conclusions of this study will be made available by the authors upon request, without undue reservation.

Authors' Contributions

Conceptualization: Keith Naugle, Kelly Naugle
Formal analysis: Keith Naugle, BB
Funding acquisition: Kelly Naugle
Investigation: BB, DB, EH
Methodology: Keith Naugle, Kelly Naugle
Project administration: Keith Naugle
Resources: Keith Naugle, Kelly Naugle
Supervision: Keith Naugle, Kelly Naugle
Writing – original draft: BB
Writing –review & editing: Keith Naugle, Kelly Naugle, DB, EH

Conflicts of Interest

None declared.

References

1. Pleis JR, Schiller JS, Benson V. Summary health statistics for U.S. adults: National Health Interview Survey, 2000. Vital Health Stat 2003 Dec;10(215):1-132. [Medline: 15792070]
2. Physical Activity Guidelines for Americans, 2nd edition: U.S. Department of Health and Human Services; 2018.

XSL•FO
RenderX

3. Freedman DS, Dietz WH, Srinivasan SR, Berenson GS. The relation of overweight to cardiovascular risk factors among children and adolescents: the Bogalusa heart study. Pediatrics 1999 Jun;103(6 Pt 1):1175-1182. [doi: 10.1542/peds.103.6.1175] [Medline: 10353925]
4. Barnett TA, Contreras G, Ghenadenik AE, et al. Identifying risk profiles for excess sedentary behaviour in youth using individual, family and neighbourhood characteristics. Prev Med Rep 2021 Dec;24:101535. [doi: 10.1016/j.pmedr.2021.101535] [Medline: 34987952]
5. 2020 Essential Facts About the Video Game Industry. Entertainment Sofware Association. 2020 Jul 15 URL: https://www.theesa.com/2020-essential-facts/ [accessed 2025-03-20]
6. Evans E, Naugle KE, Kaleth AS, Arnold B, Naugle KM. Physical activity intensity, perceived exertion, and enjoyment during head-mounted display virtual reality games. Games Health J 2021 Oct;10(5):314-320. [doi: 10.1089/g4h.2021.0036] [Medline: 34449262]
7. Graf DL, Pratt LV, Hester CN, Short KR. Playing active video games increases energy expenditure in children. Pediatrics 2009 Aug;124(2):534-540. [doi: 10.1542/peds.2008-2851] [Medline: 19596737]
8. Graves L, Stratton G, Ridgers ND, Cable NT. Comparison of energy expenditure in adolescents when playing new generation and sedentary computer games: cross sectional study. BMJ 2007 Dec 22;335(7633):1282-1284. [doi: 10.1136/bmj.39415.632951.80] [Medline: 18156227]
9. Mellecker RR, McManus AM. Energy expenditure and cardiovascular responses to seated and active gaming in children. Arch Pediatr Adolesc Med 2008 Sep;162(9):886-891. [doi: 10.1001/archpedi.162.9.886] [Medline: 18762609]
10. Pollock ML, Gaesser GA, Butcher JD, et al. ACSM Position Stand: the recommended quantity and quality of exercise for developing and maintaining cardiorespiratory and muscular fitness, and flexibility in healthy adults. Med Sci Sports Exerc 1998 Jun;30(6):975-991. [doi: 10.1097/00005768-199806000-00032] [Medline: 9624661]
11. Ufholz KE, Flack KD, Roemmich JN. The influence of active video game play upon physical activity and screen-based activities in sedentary children. PLoS ONE 2022;17(6):e0269057. [doi: 10.1371/journal.pone.0269057] [Medline: 35700210]
12. Gomez DH, Bagley JR, Bolter N, Kern M, Lee CM. Metabolic cost and exercise intensity during active virtual reality gaming. Games Health J 2018 Oct;7(5):310-316. [doi: 10.1089/g4h.2018.0012] [Medline: 30325233]
13. Stewart TH, Villaneuva K, Hahn A, et al. Actual vs. perceived exertion during active virtual reality game exercise. Front Rehabil Sci 2022;3:887740. [doi: 10.3389/fresc.2022.887740] [Medline: 36189005]
14. Yoo S, Ackad C, Heywood T, Kay J. Evaluating the actual and perceived exertion provided by virtual reality games. 2017 May 6 Presented at: Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems—CHI EA '17; Denver, Colorado, USA p. 3050-3057. [doi: 10.1145/3027063.3053203]
15. Kerr C. VR title gorilla tag has earned over $100 million since 2021. Game Developer. 2024 Jun 19. URL: https://www.gamedeveloper.com/business/vr-title-gorilla-tag-has-earned-over-100-million-since-2021 [accessed 2025-03-20]
16. Hayden S. 'Beat Saber' reportedly generated over a quarter billion dollars in lifetime sales. Road to VR. 2023 Apr 17. URL: https://www.roadtovr.com/report-beat-saber-unit-sales-meta-quest-2 [accessed 2025-03-20]
17. Godfrey CA, Oody JF, Conger SA, Steeves JA. Active virtual reality games: comparing energy expenditure, game experience, and cybersickness to traditional gaming and exercise in youth aged 8-12. Games Health J 2025 Feb;14(1):42-48. [doi: 10.1089/g4h.2024.0034] [Medline: 39093835]
18. Warburton DER, Jamnik VK, Bredin SSD, Gledhill N. The Physical Activity Readiness Questionnaire for Everyone (PAR-Q+) and Electronic Physical Activity Readiness Medical Examination (ePARmed-X+). Health Fit J Can 2011;4(2):3-17. [doi: 10.14288/hfjc.v4i2.103]
19. Liguori G, American College of Sports Medicine. ACSM's Guidelines for Exercise Testing and Prescription, 11th edition: Wolters Kluwer; 2021.
20. Strath SJ, Swartz AM, Bassett DR Jr, O'Brien WL, King GA, Ainsworth BE. Evaluation of heart rate as a method for assessing moderate intensity physical activity. Med Sci Sports Exerc 2000 Sep;32(9 Suppl):S465-S470. [doi: 10.1097/00005768-200009001-00005] [Medline: 10993416]
21. Swain DP, Leutholtz BC. Heart rate reserve is equivalent to %VO2Reserve, not to %VO2max. Med Sci Sports Exerc 1997 Mar;29(3):410-414. [doi: 10.1097/00005768-199703000-00018] [Medline: 9139182]
22. Robertson RJ, Goss FL, Boer NF, et al. Children's OMNI scale of perceived exertion: mixed gender and race validation. Med Sci Sports Exerc 2000 Feb;32(2):452-458. [doi: 10.1097/00005768-200002000-00029] [Medline: 10694131]
23. Gammon C, Pfeiffer KA, Pivarnik JM, Moore RW, Rice KR, Trost SG. Age-related differences in OMNI-RPE scale validity in youth: a longitudinal analysis. Med Sci Sports Exerc 2016 Aug;48(8):1590-1594. [doi: 10.1249/MSS.0000000000000918] [Medline: 27433962]
24. Utter AC, Robertson RJ, Nieman DC, Kang J. Children's OMNI Scale of Perceived Exertion: walking/running evaluation. Med Sci Sports Exerc 2002 Jan;34(1):139-144. [doi: 10.1097/00005768-200201000-00021] [Medline: 11782659]
25. Romanzini M, Petroski EL, Ohara D, Dourado AC, Reichert FF. Calibration of ActiGraph GT3X, Actical and RT3 accelerometers in adolescents. Eur J Sport Sci 2014;14(1):91-99. [doi: 10.1080/17461391.2012.732614] [Medline: 24533499]
26. Crouter SE, Flynn JI, Bassett DR Jr. Estimating physical activity in youth using a wrist accelerometer. Med Sci Sports Exerc 2015 May;47(5):944-951. [doi: 10.1249/MSS.0000000000000502] [Medline: 25207928]

XSL•FO
RenderX

27. Aadland E, Ylvisåker E. Reliability of the Actigraph GT3X+ accelerometer in adults under free-living conditions. PLoS ONE 2015;10(8):e0134606. [doi: 10.1371/journal.pone.0134606] [Medline: 26274586]
28. Kelly LA, McMillan DG, Anderson A, Fippinger M, Fillerup G, Rider J. Validity of actigraphs uniaxial and triaxial accelerometers for assessment of physical activity in adults in laboratory conditions. BMC Med Phys 2013 Nov 26;13(1):5. [doi: 10.1186/1756-6649-13-5] [Medline: 24279826]
29. Naugle KE, Carey C, Ohlman T, Godza M, Mikesky A, Naugle KM. Improving active gaming's energy expenditure in healthy adults using structured playing instructions for the Nintendo Wii and Xbox Kinect. J Strength Cond Res 2019 Feb;33(2):549-558. [doi: 10.1519/JSC.0000000000002997] [Medline: 30531415]
30. Motl RW, Dishman RK, Saunders R, Dowda M, Felton G, Pate RR. Measuring enjoyment of physical activity in adolescent girls. Am J Prev Med 2001 Aug;21(2):110-117. [doi: 10.1016/s0749-3797(01)00326-9] [Medline: 11457630]
31. Moore JB, Yin Z, Hanes J, Duda J, Gutin B, Barbeau P. Measuring enjoyment of physical activity in children: validation of the physical activity enjoyment scale. J Appl Sport Psychol 2009 Jan 1;21(S1):S116-S129. [doi: 10.1080/10413200802593612] [Medline: 20209028]
32. Jordan M, Donne B, Fletcher D. Only lower limb controlled interactive computer gaming enables an effective increase in energy expenditure. Eur J Appl Physiol 2011 Jul;111(7):1465-1472. [doi: 10.1007/s00421-010-1773-3] [Medline: 21165640]
33. Naugle KE, Cervantes XA, Boone CL, Wind B, Naugle KM. Exploring actual and perceived levels of physical activity intensity during virtual reality active games. Front Sports Act Living 2024;6:1349521. [doi: 10.3389/fspor.2024.1349521] [Medline: 38406766]
34. DiLorenzo TM, Stucky-Ropp RC, Vander Wal JS, Gotham HJ. Determinants of exercise among children. II. A longitudinal analysis. Prev Med 1998;27(3):470-477. [doi: 10.1006/pmed.1998.0307] [Medline: 9612838]
35. Bosch PR, Poloni J, Thornton A, Lynskey JV. The heart rate response to Nintendo Wii boxing in young adults. Cardiopulm Phys Ther J 2012 Jun;23(2):13-29. [Medline: 22833705]
36. Sell K, Lillie T, Taylor J. Energy expenditure during physically interactive video game playing in male college students with different playing experience. J Am Coll Health 2008;56(5):505-511. [doi: 10.3200/JACH.56.5.505-512] [Medline: 18400662]
37. Barnett A, Cerin E, Baranowski T. Active video games for youth: a systematic review. J Phys Act Health 2011 Jul;8(5):724-737. [doi: 10.1123/jpah.8.5.724] [Medline: 21734319]

Abbreviations

%LPA: percentage of time in light physical activity
%MVPA: percentage of time in moderate to vigorous physical activity
%SED: percentage of time in sedentary
HR: heart rate
HRR: heart rate reserve
IRB: institutional review board
LPA: light physical activity
MVPA: moderate to vigorous physical activity
PACES: Physical Activity Enjoyment Scale
PAR-Q: Physical Activity Readiness Questionnaire
RPE: ratings of perceived exertion
VR: virtual reality

Edited by A Coristine; submitted 17.09.24; peer-reviewed by JA Steeves, W Yi-Syuan; revised version received 19.02.25; accepted 20.02.25; published 01.04.25.

Please cite as:
Boots B, Berg D, Hewitt E, Naugle K, Naugle K
Physical Activity and Enjoyment in Active Virtual Reality Games in Youth: Comparative Analysis of Gorilla Tag and Beat Saber
JMIR Serious Games 2025;13:e66593
URL: https://games.jmir.org/2025/1/e66593
doi:10.2196/66593

© Brenden Boots, Daniel Berg, Easton Hewitt, Keith Naugle, Kelly Naugle. Originally published in JMIR Serious Games (https://games.jmir.org), 1.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic

XSL•FO
RenderX

information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Comparative Efficacy of Video Games Versus Midazolam in Reducing Perioperative Anxiety in Pediatric Patients: Systematic Review and Meta-Analysis

Ziyue Luo[1*], BSc; Sisi Deng[1*], BSc; Ruihao Zhou[1], BSc; Ling Ye[2], MD; Tao Zhu[1], MD; Guo Chen[1], MD

[1]Department of Anesthesiology, National Clinical Research Center for Geriatrics, West China Hospital, Sichuan University, No. 37 Guoxue Lane, Wuhou District, Sichuan Province, Chengdu, China

[2]Department of Pain Management, National Clinical Research Center for Geriatrics, West China Hospital, Sichuan University, Chengdu, China

[*]these authors contributed equally

Corresponding Author:
Guo Chen, MD
Department of Anesthesiology, National Clinical Research Center for Geriatrics, West China Hospital, Sichuan University, No. 37 Guoxue Lane, Wuhou District, Sichuan Province, Chengdu, China

Abstract

Background: Pediatric patients undergoing surgery frequently experience significant anxiety, which can result in adverse effects such as prolonged sedation and behavioral changes associated with pharmacological interventions such as oral midazolam. Video games offer a nonpharmacological distraction method that shows promise in alleviating procedural anxiety without significant adverse effects. However, the effectiveness of video games compared to midazolam in managing perioperative anxiety remains uncertain.

Objective: This study aimed to evaluate the effectiveness of video game interventions in reducing perioperative anxiety in pediatric patients undergoing general anesthesia.

Methods: We conducted a comprehensive search across PubMed, Embase, Web of Science, and the Cochrane Library, supplemented by reference screening. Primary outcomes included anxiety levels assessed during parent separation and mask induction procedures, while secondary outcomes encompassed emergence delirium, postoperative behavior, and length of stay in the postanesthesia care unit (PACU). The risk of bias was assessed using the Risk of Bias 2 scale. Data were synthesized descriptively and through meta-analysis, with the certainty of the evidence evaluated using the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) criteria.

Results: Six randomized controlled trials involving 612 participants were included in the analysis. Children who participated in video game interventions reported significantly lower anxiety levels during parent separation (standardized mean difference, SMD −0.31, 95% CI −0.50 to −0.12; *P*=.001), with high certainty, and during mask induction (SMD −0.29, 95% CI −0.52 to −0.05; *P*=.02), with moderate certainty, compared to those receiving oral midazolam. Additionally, significant differences in postoperative behavior changes in children were observed compared to oral midazolam (SMD −0.35, 95% CI −0.62 to −0.09; *P*=.008). Children in the video game intervention groups also had a shorter length of stay in the PACU (mean difference, MD −19.43 min, 95% CI −31.71 to −7.16; *P*=.002). However, no significant differences were found in emergence delirium (MD −2.01, 95% CI −4.62 to 0.59; *P*=.13).

Conclusions: Video game interventions were more effective than midazolam in reducing perioperative anxiety among pediatric patients, improving postoperative behavior, and shortening the length of stay in the PACU. However, video games alone did not outperform midazolam in managing emergence delirium. Further high-quality research is needed for more conclusive results.

Trial Registration: PROSPERO CRD42023486085; https://tinyurl.com/yc3suavb

(JMIR Serious Games 2025;13:e67007) doi:10.2196/67007

KEYWORDS

video games; midazolam; perioperative period; anxiety; meta-analysis; pediatric patients

XSL•FO
RenderX

Introduction

Background

Perioperative anxiety is a significant concern for pediatric patients undergoing surgery [1]. This anxiety, characterized by worry, nervousness, or unease about uncertain outcomes, can lead to various negative effects, including increased preoperative distress, postoperative pain, and longer recovery times [2]. Research shows that high levels of preoperative anxiety are linked to greater postoperative pain and higher analgesic consumption [3], delayed hospital discharge [4], and the emergence of negative behavioral changes [5], such as nightmares, separation anxiety, and increased fear of medical procedures. Therefore, effectively managing perioperative anxiety is crucial for improving surgical outcomes and enhancing overall patient well-being.

Traditional methods for managing perioperative anxiety often rely on pharmacological interventions, with oral midazolam being a commonly used anxiolytic agent. Administered at a typical dosage of 0.25 to 0.5 mg/kg, midazolam is effective in reducing anxiety in children, inducing sedation and anxiolysis within 20 to 30 minutes [6,7]. Studies have shown that midazolam premedication can significantly improve cooperation during anesthesia induction and decrease postoperative behavioral disturbances [8,9]. However, the use of pharmacological agents is not without drawbacks. Potential adverse effects of midazolam include respiratory depression [10], paradoxical reactions [11], prolonged recovery periods [12], and postoperative cognitive dysfunction [13]. These concerns highlight the need to explore alternative, nonpharmacological interventions that can effectively manage perioperative anxiety without adverse effects.

In recent years, there has been growing interest in nonpharmacological interventions for managing anxiety in children undergoing surgery, such as music [14], clown doctors [15], preoperative preparation videos [16], virtual reality tools [17], augmented reality tools [18], and video games [19]. Video games, in particular, have been shown to be engaging and effective in reducing anxiety levels in various medical contexts [20]. A study indicated that video games can decrease both pain and anxiety in pediatric surgery patients [18]. Their interactive and immersive nature captivates children's attention, providing a sense of control and normalcy in a potentially intimidating hospital environment. Patel et al [21] reported that distraction with hand-held video games significantly reduced preoperative anxiety levels compared to standard care. Additionally, video games offer the benefits of no adverse effects and enhancing patient cooperation during medical procedures.

The potential benefits of video games in clinical settings extend beyond distraction. They can also enhance patient engagement and compliance [22], reduce the need for sedative medications [15], and improve overall patient satisfaction with the surgical experience [23]. The versatility and appeal of video games make them a promising tool for anxiety management, warranting a systematic comparison with traditional pharmacological treatments such as oral midazolam. Given their increasing popularity and potential benefits, it is essential to compare their efficacy against traditional pharmacological treatments such as oral midazolam.

Objectives

The primary objective of this systematic review and meta-analysis is to compare the efficacy of video games and midazolam in reducing perioperative anxiety in pediatric patients during critical moments, such as parent separation and mask induction. Additionally, we conducted a subgroup analysis of midazolam dosages related to anxiety. Furthermore, the review aims to assess the impact of these interventions on secondary outcomes, including emergence delirium, postoperative behavior, and length of stay in the postanesthesia care unit (PACU). This is the first systematic review and meta-analysis to evaluate the effects of video games on outcomes related to managing perioperative anxiety in children.

Methods

This systematic review follows the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 guidelines (Checklist 1) [24]. Additionally, we formally registered this systematic review and meta-analysis with the International Prospective Register of Systematic Reviews (PROSPERO, CRD42023486085).

Search Strategy

We conducted a comprehensive literature search in PubMed, Embase, Web of Science, and the Cochrane Library up to June 30, 2024. The search strategy included Medical Subject Headings and text words related to "video game," "midazolam," and "anxiety." Search terms were tailored for each database to ensure comprehensiveness. The detailed search strategies for each database are provided in Multimedia Appendix 1. Furthermore, the reference lists of the included studies were scrutinized for articles not initially identified in the primary search. There were no restrictions on the publication date, but articles were limited to those published in English.

Study Eligibility Criteria

Inclusion criteria for this study were as follows: (1) Population: Children (≤18 years old) undergoing surgical procedures under general anesthesia. (2) Intervention: Video games utilized as a perioperative intervention to alleviate anxiety. (3) Comparators: Midazolam as a perioperative intervention for anxiety relief. (4) Outcomes: Primary outcomes included anxiety levels during parent separation and mask induction, while secondary outcomes comprised emergence delirium, postoperative behavior, and length of stay in the PACU. Various measurement methods were accepted, including self-report, proxy report, and observation. (5) Study design: Only randomized controlled trials were considered eligible for inclusion.

The exclusion criteria were as follows: (1) studies not involving pediatric surgical patients under general anesthesia; (2) patients who received anxiolytic premedication or had cognitive impairments such as psychiatric disorders or autism; and (3) nonrandomized controlled trials, reviews, meta-analyses, and single case reports.

XSL•FO
RenderX

Study Selection and Data Extraction

We combined the search results from the four databases and removed duplicate articles. Two investigators (ZYL and SSD) independently screened titles/abstracts and full texts using a shared spreadsheet (Microsoft Excel). To implement the double-blind process, each reviewer maintained a separate decision sheet. Discrepancies in evaluations triggered a re-examination of the conflicting articles. The investigators then discussed the articles' eligibility for inclusion or exclusion. If disagreements persisted, a third senior author was consulted for resolution during the systematic review.

For the review, two investigators (ZYL and SSD) extracted the data from the eligible articles. The data included the name of the first author, publication year, participant information (sample size and age), study location (country), intervention and control group details, and instruments and time points of outcome assessment. Disagreements between reviewers were resolved through discussion. In cases of missing or unclear data, the study authors were contacted twice via email.

Study Quality Assessment

The revised Cochrane risk of bias tool for randomized trials was employed to analyze the risk of bias in randomized studies included in this review, as recommended elsewhere [25]. The tool comprises five domains with different questions: (1) bias arising from the randomization process; (2) bias due to deviations from intended interventions; (3) bias due to missing outcome data; (4) bias in measurement of the outcome; and (5) bias in selection of the reported result. The risk of bias in each domain was categorized into three levels: "low risk of bias," "some concerns," and "high risk of bias."

Data Synthesis and Analysis

For the meta-analysis, eligible studies were analyzed to combine results, ensuring clinical and methodological homogeneity of the intervention and follow-up period. We utilized Review Manager 5.4 (The Cochrane Collaboration) for conducting the meta-analysis and heterogeneity testing. Where necessary, data transformations were performed prior to analysis, including converting standard errors of the mean (SEMs) to standard deviations (SDs) using the formula SD=SEM×n and transforming medians (IQRs) to means (SDs) with the method given by Wan et al [26], assuming normal or log-normal distributions, to ensure compatibility for meta-analysis. For outcome-specific effect measures, standardized mean differences (SMDs) were calculated for perioperative anxiety and postoperative behavioral outcomes due to heterogeneity in measurement scales across studies, while mean differences (MDs) were used for emergence delirium and length of stay in the PACU, as these outcomes shared consistent measurement units; all effect estimates are reported with 95% CIs. To enhance clinical interpretability, SMDs were converted back to the original measurement scale by multiplying them with the pooled standard deviation (SD_{pooled}) of the included studies. Forest plots were generated to visualize the results. Given the anticipated clinical and methodological diversity across studies, including variations in participant characteristics, surgical procedures, cultural contexts, and intervention protocols, we selected a random-effects model a priori for all meta-analyses. Heterogeneity among the studies was assessed using the χ^2 test, with a significance threshold set at $P<.10$ [27]. Additionally, the I^2 test was employed to quantify the extent of variability, categorizing it as 0% - 40% as potentially unimportant, 30% - 60% as moderate, 50% - 90% as substantial, and 75% - 100% as considerable, taking into account the magnitude and direction of effects as well as the strength of the evidence [28] . Statistical significance was determined through two-tailed tests, with a threshold of $P<.05$. Additionally, we conducted a subgroup analysis of midazolam dosages related to perioperative anxiety.

We utilized the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) framework to assess the evidence level of the included outcomes [29]. This evaluation was performed with GRADEpro software (McMaster University and Evidence Prime Inc). The following five domains of the GRADE criteria were analyzed: methodological limitations (risk of bias), inconsistency, indirectness, imprecision, and publication bias. Each domain involves a qualitative assessment of the evidence for each outcome, allowing the classification of confidence in the estimated effects as high, moderate, low, or very low.

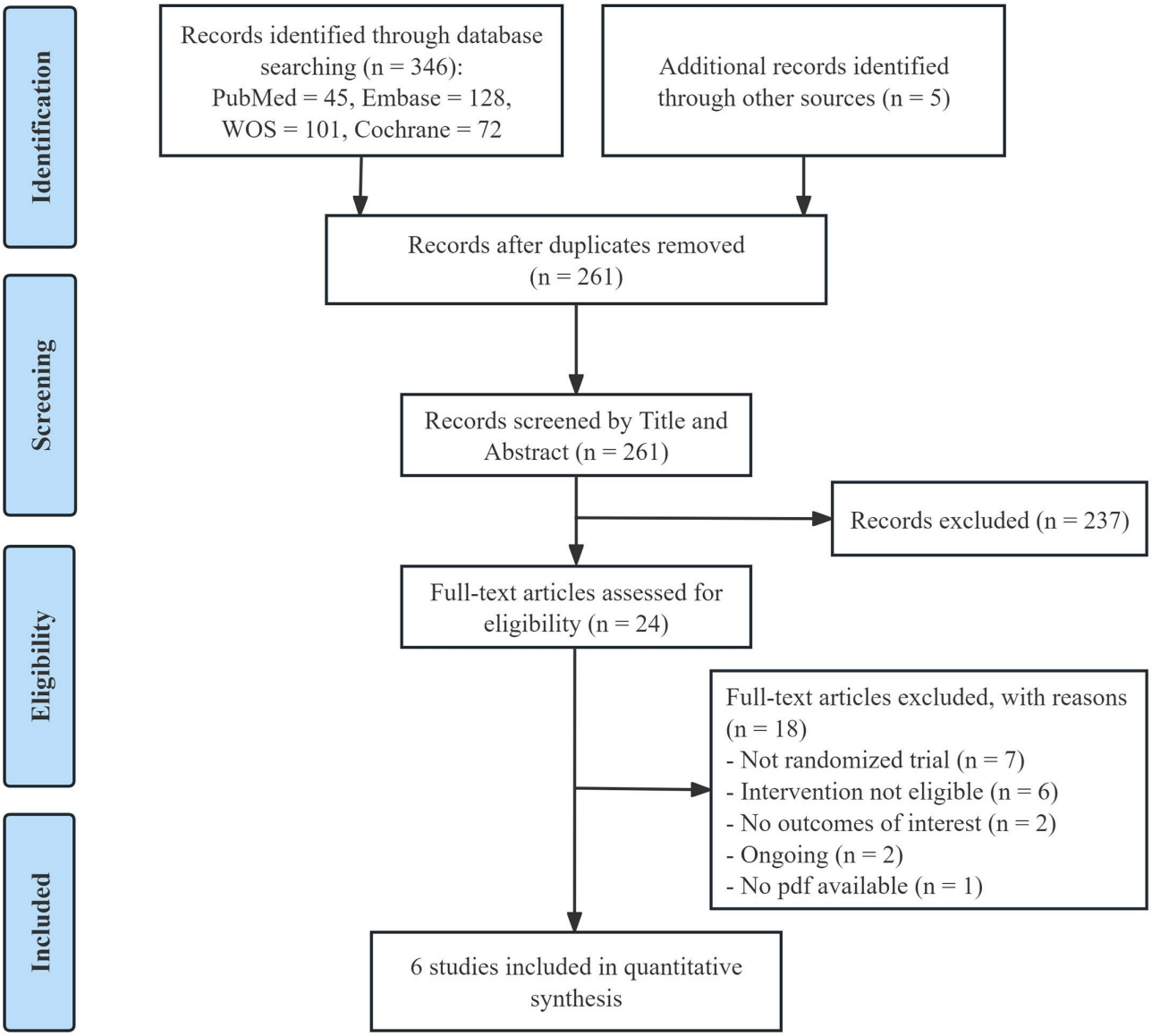
Results

Search Results and Selection

We conducted a comprehensive search across four databases, identifying a total of 346 articles. After removing duplicates, 256 studies remained for screening. Following a review of titles and abstracts, irrelevant studies were excluded, leaving 24 full-text articles for further assessment. Ultimately, 6 studies met the inclusion criteria and were included in our meta-analysis [21,23,30-33]. The detailed screening process is presented in Figure 1.

XSL•FO
RenderX

Figure 1. Study flow diagram of the search strategy and included studies.



Characteristics and Results of Individual Studies

The meta-analysis encompasses 6 studies from France (2 studies) and the United States (4 studies) that assessed interventions aimed at reducing preoperative anxiety and managing postoperative behavior among pediatric patients. The patient cohort, totaling 612 subjects aged 3 to 12 years, was predominantly classified as having an American Society of Anesthesiologists score of I-II. These studies compared the effectiveness of age-appropriate video games (316 participants) against standard midazolam dosages of 0.3 mg/kg or 0.5 mg/kg (296 participants) administered before anesthesia. Anxiety levels were systematically assessed at multiple time points utilizing either the Modified Yale Preoperative Anxiety Scale (mYPAS) or its abbreviated version, the mYPAS-Short Form. The mYPAS includes 27 items related to activity, arousal, vocalization, dependence on parents, and emotional status. The total score ranges from 23 to 100, with scores above 30 indicating anxiety and scores above 40 indicating high anxiety [34]. Postoperative behavioral outcomes were evaluated through the implementation of either the posthospital behavior questionnaire (PHBQ) or the posthospitalization behavior change questionnaire for ambulatory surgery. The PHBQ comprises 27 items among 6 subscales (general anxiety and regression, separation anxiety, eating disturbance, aggression toward authority, apathy/withdrawal, and anxiety about sleep) [35]. Caregivers were also given the response option of not applicable. Furthermore, the incidence of emergence delirium was quantitatively measured using the validated pediatric anesthesia emergence delirium (PAED) scale. The PAED Scale [36] consists of 5 items scored from 0 to 4 (with 3 reverse-scored items). The 5 items relate to the observation of eye contact with caregivers, purposeful actions, awareness of surroundings, restlessness, and inconsolability. Detailed characteristics and results of these studies are summarized in Multimedia Appendix 2.

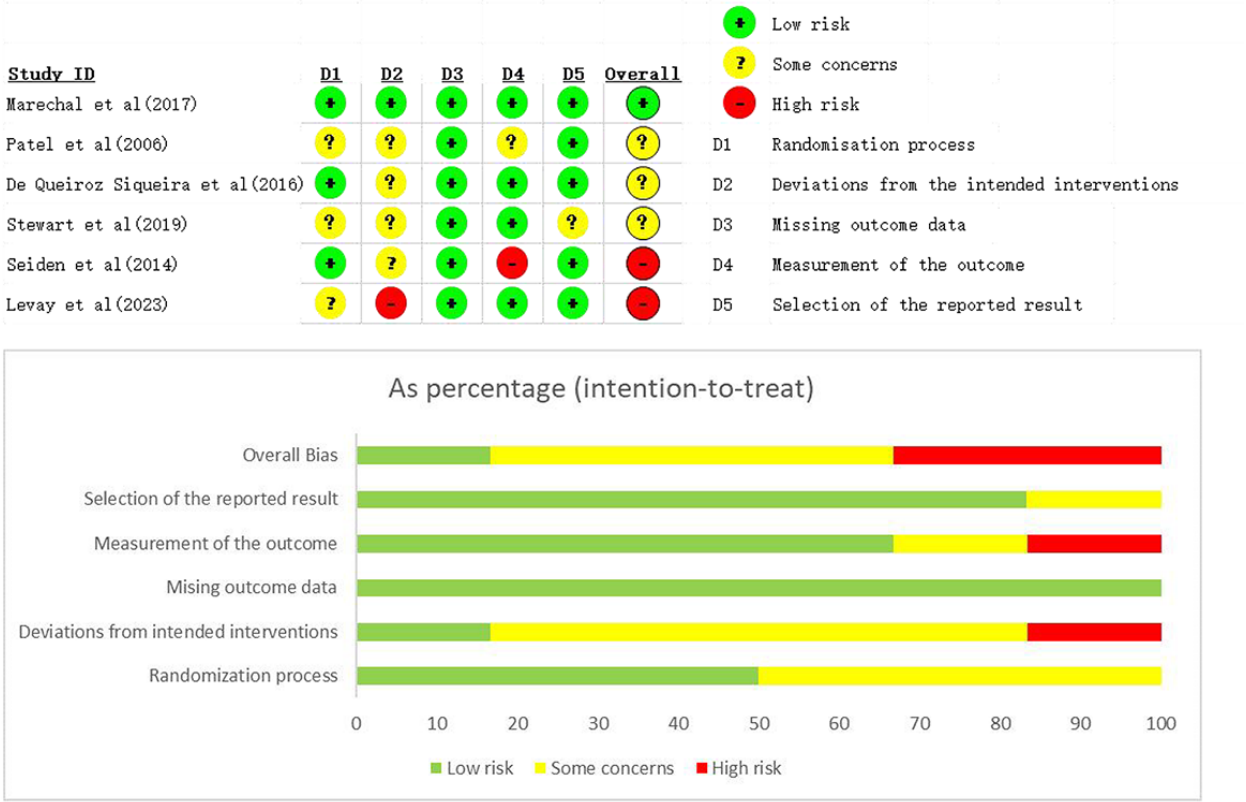
Risk of Bias Assessment

We assessed the risk of bias for the 6 included randomized controlled trials using the RoB 2 tool (version 2 of the Cochrane risk-of-bias tool for randomized trials). The analysis revealed that 1 article was rated as low risk [31], 3 articles exhibited

XSL•FO
RenderX

some concerns and were thus categorized as medium risk [21,30,32], and 2 articles were classified as high risk [23,33]. The concerns predominantly stemmed from issues such as inadequate blinding of participants and deviations from intended interventions. Specifically, the 2 studies identified as having a high risk of bias faced significant issues related to deviations from intended interventions and inaccuracies in the measurement of outcomes [23,33]. These findings are detailed in Figure 2.

Figure 2. Risk of bias assessment of the included studies [21,23,30-33] using version 2 of the Cochrane risk-of-bias tool for randomized trials tool.



Meta-Analysis

Pediatric Anxiety

Data on pediatric anxiety at the time of parent separation were obtained from 4 studies, totaling 226 participants [23,30-32]. Overall, children who engaged in video games exhibited significantly lower anxiety levels compared to those in the control group, showing a statistically significant effect (SMD −0.31, 95% CI −0.50 to −0.12; *P*=.001) with high certainty (Table 1). According to the Cohen criteria [37], this corresponds to a small-to-moderate effect size. After conversion to the original anxiety scale, the estimated MD was −5.23 points. Additionally, no significant heterogeneity was found among the included studies (I^2=0%; *P*=.39). Subgroup analysis further confirmed that children playing video games reported lower anxiety levels in both the 0.3 mg/kg midazolam subgroup (SMD −0.22, 95% CI −0.44 to −0.01; *P*=.04), corresponding to a small effect size and translating to a MD of −3.81 points on the original anxiety scale, and the 0.5 mg/kg midazolam subgroup (SMD −0.58, 95% CI −0.97 to −0.20; *P*=.003), corresponding to a moderate-to-large effect size and equating to a MD of −8.98 points. The heterogeneity test for subgroup differences indicated some variability (I^2=60.7%; *P*=.11) (Figure 3A).

Data on pediatric anxiety during mask induction were collected from 6 studies involving 316 participants [21,23,30-33]. The results indicated a significant difference in anxiety reduction between children playing video games and those receiving midazolam (SMD −0.29, 95% CI −0.52 to −0.05; *P*=.02), indicating a small-to-moderate effect, with an estimated MD of −5.73 points after conversion to the original anxiety scale and moderate certainty (Table 1). However, moderate heterogeneity was observed among the studies (I^2=55%; *P*=.05). Subgroup analysis revealed no significant difference in the 0.3 mg/kg midazolam subgroup (SMD −0.11, 95% CI −0.47 to 0.25; *P*=.54), indicating a very small effect, with an estimated MD of −2.17 points. In contrast, a statistically significant effect was found in the 0.5 mg/kg midazolam subgroup (SMD −0.47, 95% CI −0.71 to −0.23; *P*<.001), indicating a moderate-to-large effect, equating to an estimated MD of −9.28 points. The heterogeneity test for subgroup differences indicated some variability (I^2=62.3%; *P*=.10) (Figure 3B).

XSL•FO
RenderX

Table . Result of assessment of certainty of evidence for all outcomes of the use of video games compared to midazolam in children in the operation room.

Outcomes	Relative effect (95% CI)[a]	No. of Participants (no. of studies)	Quality of the evidence (GRADE[b])
Anxiety T1[c]	SMD[d] −0.31 (−0.5 to −0.12)	437 (4 studies)	⊕⊕⊕⊕ high[e]
Anxiety T2[f]	SMD −0.29 (−0.52 to −0.05)	612 (6 studies)	⊕⊕⊕ moderate[gh]
Emergence delirium	MD[i] −2.01 (−4.62 to 0.59)	309 (3 studies)	⊕ very low[jkl]
Postoperative Behavior	SMD −0.35 (−0.62 to −0.09)	227 (3 studies)	⊕⊕⊕⊕ high
Length of stay in the PACU[m]	MD −19.43 minutes (−31.71 to −7.16)	309 (3 studies)	⊕⊕⊕⊕ high

[a]CI: confidence interval.

[b]Grading of Recommendations Assessment, Development, and Evaluation Working Group grades of evidence.

[c]Anxiety T1: anxiety at the time of parent separation.

[d]SMD: standardized mean difference.

[e]High quality: Further research is very unlikely to change our confidence in the estimate of effect.

[f]Anxiety T2: anxiety at the time of mask induction.

[g]Moderate quality: Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.

[h]Downgraded by 1 level due to moderate concerns about inconsistency, including notable heterogeneity in effect estimates across trials, reflecting variations in both the magnitude and direction of effect sizes (50%< I^2< 75%).

[i]MD: mean difference.

[j]Very low quality: We are very uncertain about the estimate.

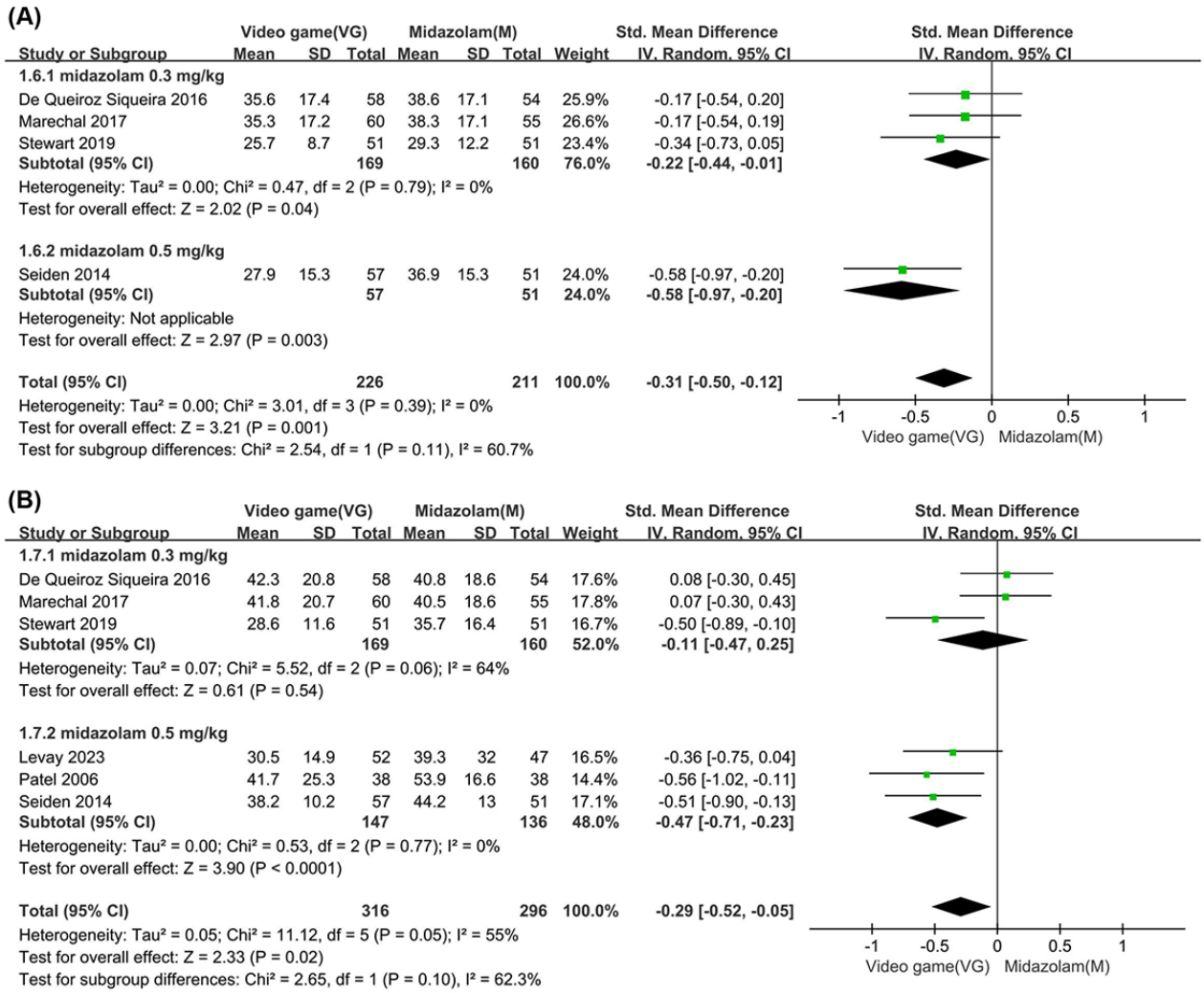
[k]Downgraded by 1 level due to significant concerns about imprecision, the confidence interval suggests the possibility of a null effect or benefit for either intervention.

[l]Downgraded by 2 levels due to serious concerns about inconsistency, with substantial heterogeneity in effect estimates across trials (I^2≥ 75%).

[m]PACU: postanesthesia care unit.

XSL•FO
RenderX

Figure 3. Forest plot comparing anxiety levels: video game intervention versus midazolam during (**A**) parent separation [23,30-32] and (**B**) mask induction [21,23,30-33].

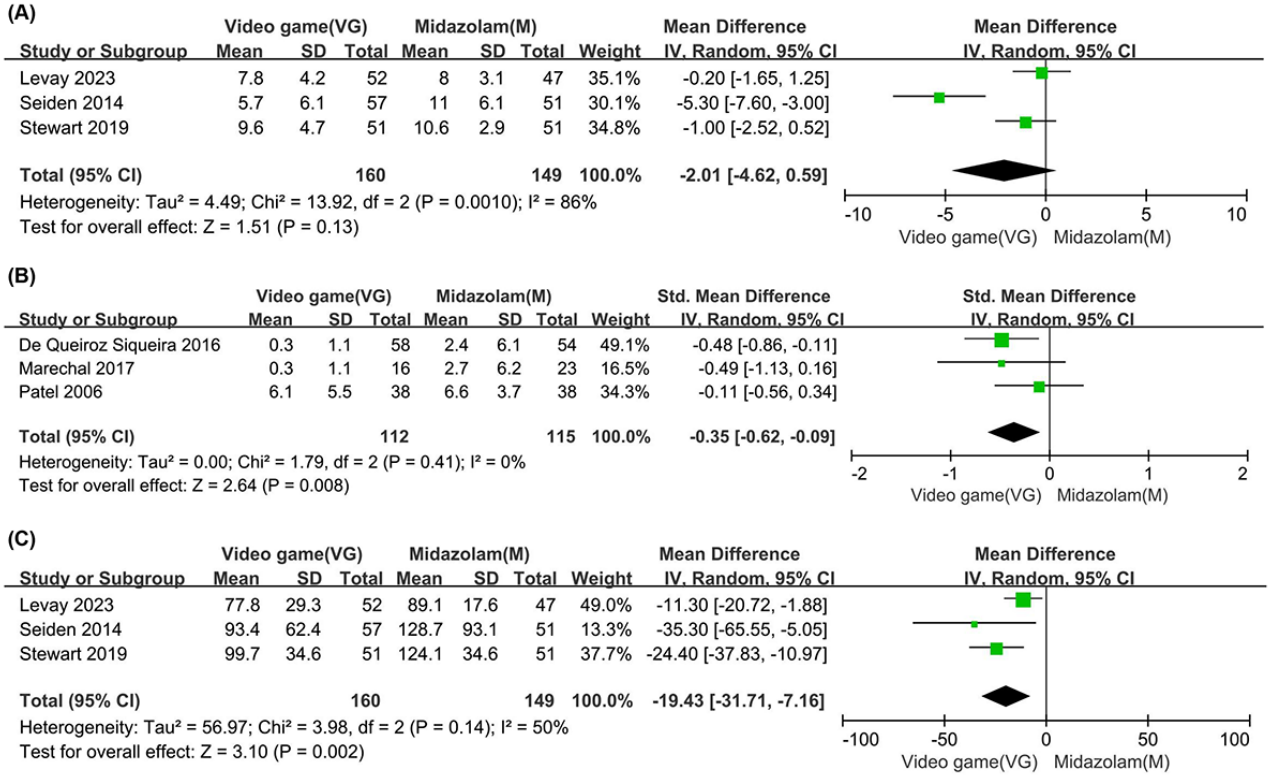


Emergence Delirium

Data on emergence delirium were collected from 3 studies, comprising a total of 160 participants [23,32,33]. The results showed no significant differences between children who played video games and those who received midazolam (MD −2.01, 95% CI −4.62 to 0.59; *P*=.13), with a very low certainty (Table 1). Additionally, substantial heterogeneity was observed (I^2=86%; *P*=.001) (Figure 4A).

XSL•FO
RenderX

Figure 4. Forest plot comparing video game intervention and midazolam: (**A**) emergence delirium [23,32,33], (**B**) postoperative behavior [21,30,31], and (**C**) length of stay in the postanesthesia care unit (PACU) [23,32,33].



Postoperative Behavior

Data on postoperative behavior were available from 3 studies, comprising a total of 112 participants [21,30,31]. A significant difference was found between the video game group and the midazolam group (SMD −0.35, 95% CI −0.62 to −0.09; *P*=.008), corresponding to a small-to-moderate effect size. After conversion to the original postoperative behavior scale, this equates to an estimated MD of −1.59 points. The quality of evidence was classified as high (Table 1). No substantial heterogeneity was observed (I^2=0%; *P*=.41) (Figure 4B).

Length of Stay in the PACU

Data on the length of stay in the PACU were available from 3 studies, totaling 160 participants [23,32,33]. Children who received video game interventions had significantly shorter PACU stays (MD −19.43 min, 95%CI −31.71 to −7.16 min; *P*=.002). The quality of the evidence was rated as high (Table 1), and moderate heterogeneity was observed across the studies (I^2=50%; *P*=.14) (Figure 4C).

Discussion

Principal Findings

This systematic review and meta-analysis assessed the effectiveness of video games compared to midazolam in reducing preoperative anxiety in pediatric patients. We identified 6 studies involving 612 children, which produced mixed outcomes. Our meta-analysis indicates that video games were more effective than midazolam in reducing anxiety during both parental separation and mask induction, critical moments of heightened stress for children undergoing surgery.

At the time of parental separation, the subgroup analysis demonstrated that video games outperformed midazolam at both the 0.3 mg/kg and 0.5 mg/kg doses, with a more pronounced effect observed in the 0.5 mg/kg group. This finding may be linked to the challenges of administering oral medications to pediatric patients, as well as midazolam's potential paradoxical effects, which can lead to increased agitation instead of calming the child. Additionally, factors such as the age and mood of the children may contribute to this variability [38]. One study found that approximately 14% of children who received oral midazolam prior to surgery still exhibited extreme anxiety and lack of compliance during anesthesia induction [39].

At the time of mask induction, a significant difference was found between video games and midazolam, suggesting that video games may be more effective in reducing anxiety during this phase. However, the subgroup analysis revealed no significant differences between the two interventions at the 0.3 mg/kg dose. This lack of difference may be due to the insufficient potency of this lower dose, which might not effectively reduce anxiety in children [40]. Additionally, the engaging nature of video games could provide enough distraction to manage anxiety, resulting in comparable outcomes to midazolam [19]. In contrast, at the 0.5 mg/kg dose, video games demonstrated superiority over midazolam. This superiority may be attributed to the potential for paradoxical reactions to midazolam, which can lead to increased agitation in some children, as well as the higher potency of this dose introducing variability in effectiveness [41]. These factors, along with individual differences in responses to medication, likely

XSL•FO
RenderX

contributed to the greater effectiveness of video games during this critical moment.

Significant differences were noted in postoperative behavior and length of stay in the PACU, suggesting that video games may offer an effective interactive distraction that helps children cope better in the postoperative period [19]. This engagement can serve to divert their attention from discomfort and anxiety associated with recovery, potentially leading to improved behavioral outcomes. The immersive nature of video games can facilitate a sense of control and agency, which is especially important for children facing medical procedures. Additionally, video games may promote relaxation and positive emotional states, further enhancing their ability to manage pain and anxiety [42]. Furthermore, the reduction in length of stay in the PACU observed in children receiving video game interventions suggests that these distractions may contribute to a more efficient recovery process. One possible mechanism is the attenuation of stress-related physiological responses, such as reduced sympathetic nervous system activation, which can facilitate faster stabilization of vital signs postoperatively. Moreover, by lowering perioperative anxiety and distress, video games may help mitigate the need for additional sedatives or analgesics, which could otherwise prolong PACU stays. Additionally, improved postoperative cooperation and reduced agitation may enable earlier discharge from the PACU, optimizing resource utilization in clinical settings. This highlights the potential for video game interventions not only to enhance patient experience but also to improve hospital workflow efficiency [43].

We found no significant differences between video games and midazolam regarding emergence delirium, indicating that nonpharmacological interventions may have limited effects on these outcomes. Several factors could contribute to this lack of difference, including the possibility that both interventions are similarly effective or that the nature of emergence delirium is such that it may not be easily alleviated by distractions alone [44]. Additionally, the context in which these interventions are applied may play a role, as factors such as individual patient characteristics [45] and the surgical environment [46] could influence outcomes. However, the evidence was classified as very low quality, and the substantial heterogeneity and inconsistency across studies weaken the strength of this conclusion, necessitating caution in interpretation. Further high-quality research is needed to confirm these findings and better understand the potential impact of nonpharmacological interventions on emergence delirium. Robust studies with larger sample sizes and standardized methodologies will be essential to draw more definitive conclusions.

Limitations

The credibility of our findings is limited by several factors. High levels of heterogeneity in the type, duration, and frequency of video games used in the studies made direct comparisons challenging. Bias was also a concern, as many studies lacked blinding, which may have skewed the results. Additionally, the small sample sizes in several studies likely affected statistical power, further limiting the generalizability of our findings. The subjectivity of various assessment scales is noteworthy; for instance, Seiden [23] reported mean baseline mYPAS values ranging from 23 to 45, while Levay [33] reported scores from 23 to 65. This variability may arise from the fact that mYPAS or other assessments were often administered by staff members, introducing potential bias.

Furthermore, the limited number of studies highlights the need for more high-quality research to strengthen our conclusions. Additionally, our review did not consider the potential impact of caregiver anxiety. A systematic review indicates that parent's and children's experiences are closely interconnected, with caregiver anxiety potentially exacerbating a child's anxiety and leading to long-term psychosocial effects, such as increased fear and guilt [47]. Future research should investigate the role of caregiver anxiety in pediatric surgical settings and explore effective interventions to alleviate it, ultimately enhancing the overall well-being of both children and their families during these critical experiences.

Conclusions

In summary, video games offer an accessible, low-cost, and well-tolerated intervention for reducing perioperative anxiety in pediatric patients undergoing general anesthesia, making them a promising nonpharmacological alternative to midazolam in certain contexts. However, to achieve a more comprehensive approach to pediatric perioperative care, a multimodal strategy should be implemented. This approach would integrate both pharmacological and nonpharmacological interventions tailored to each child's unique needs and specific clinical situation. Additionally, further high-quality, large-scale studies are essential to confirm these findings, reduce heterogeneity, and investigate the integration of caregiver anxiety management into pediatric perioperative protocols. By addressing the emotional needs of both children and their caregivers, we can enhance overall outcomes and better support families during surgical experiences.

Acknowledgments

This study was supported by the Noncommunicable Chronic Diseases-National Science and Technology Major Project (2023ZD0501801) and the Sichuan Province Health Research Project (ZH2025 -103).

Data Availability

All data generated or analyzed during this study are included in this published paper and its supplementary information files.

XSL•FO
RenderX

Authors' Contributions

Conceptualization: ZYL, SSD, TZ, GC
Data curation: ZYL, SSD
Formal analysis: ZYL, SSD, RHZ
Funding acquisition: ZYL, SSD, TZ, GC
Investigation: ZYL, SSD, RHZ
Methodology: ZYL, SSD, TZ, GC
Project administration: ZYL, SSD, TZ, GC
Resources: ZYL, SSD, TZ, GC
Supervision: ZYL, SSD, TZ, GC
Validation: ZYL, SSD
Visualization: ZYL, SSD
Writing – original draft: ZYL, SSD
Writing – review & editing: TZ, GC

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOCX File, 16 KB - games_v13i1e67007_app1.docx]

Multimedia Appendix 2
Characteristics of the studies included in the review.
[DOCX File, 16 KB - games_v13i1e67007_app2.docx]

Checklist 1
The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[PDF File, 104 KB - games_v13i1e67007_app3.pdf]

References

1. Mathias EG, Pai MS. Anxiety and pain in children undergoing surgery: a scoping review. J Perianesth Nurs 2022 Aug;37(4):545-550. [doi: 10.1016/j.jopan.2021.10.002] [Medline: 35279386]
2. Fischer S, Vinall J, Pavlova M, et al. Role of anxiety in young children's pain memory development after surgery. Pain 2019 Apr;160(4):965-972. [doi: 10.1097/j.pain.0000000000001473] [Medline: 30586022]
3. Kain ZN, Mayes LC, Caldwell-Andrews AA, Karas DE, McClain BC. Preoperative anxiety, postoperative pain, and behavioral recovery in young children undergoing surgery. Pediatrics 2006 Aug;118(2):651-658. [doi: 10.1542/peds.2005-2920] [Medline: 16882820]
4. Sveinsdóttir H, Zoëga S, Ingadóttir B, Blöndal K. Symptoms of anxiety and depression in surgical patients at the hospital, 6 weeks and 6 months postsurgery: a questionnaire study. Nurs Open 2021 Jan;8(1):210-223. [doi: 10.1002/nop2.620] [Medline: 33318829]
5. Kain ZN, Caldwell-Andrews AA, Maranets I, et al. Preoperative anxiety and emergence delirium and postoperative maladaptive behaviors. Anesth Analg 2004 Dec;99(6):1648-1654. [doi: 10.1213/01.ANE.0000136471.36680.97] [Medline: 15562048]
6. Cox RG, Nemish U, Ewen A, Crowe MJ. Evidence-based clinical update: does premedication with oral midazolam lead to improved behavioural outcomes in children? Can J Anaesth 2006 Dec;53(12):1213-1219. [doi: 10.1007/BF03021583] [Medline: 17142656]
7. Gropper MA. Miller's Anesthesia, 9th edition 2020.
8. McGraw T, Kendrick A. Oral midazolam premedication and postoperative behaviour in children. Paediatr Anaesth 1998;8(2):117-121. [doi: 10.1046/j.1460-9592.1998.00724.x] [Medline: 9549736]
9. Heikal S, Stuart G. Anxiolytic premedication for children. BJA Educ 2020 Jul;20(7):220-225. [doi: 10.1016/j.bjae.2020.02.006] [Medline: 33456954]
10. Dixon D. Respiratory depression following midazolam. Anaesthesia 1985 Sep;40(9):922-922. [doi: 10.1111/j.1365-2044.1985.tb11078.x] [Medline: 2932034]
11. Moon YE. Paradoxical reaction to midazolam in children. Korean J Anesthesiol 2013 Jul;65(1):2-3. [doi: 10.4097/kjae.2013.65.1.2] [Medline: 23904931]

XSL•FO
RenderX

12. Brosius KK, Bannister CF. Oral midazolam premedication in preadolescents and adolescents. Anesth Analg 2002 Jan;94(1):31-36. [doi: 10.1097/00000539-200201000-00006] [Medline: 11772796]
13. Millar K, Asbury AJ, Bowman AW, et al. A randomised placebo-controlled trial of the effects of midazolam premedication on children's postoperative cognition. Anaesthesia 2007 Sep;62(9):923-930. [doi: 10.1111/j.1365-2044.2007.05148.x] [Medline: 17697220]
14. Bradt J, Dileo C, Shim M. Music interventions for preoperative anxiety. Cochrane Database Syst Rev 2013 Jun 6;2013(6):CD006908. [doi: 10.1002/14651858.CD006908.pub2] [Medline: 23740695]
15. Manyande A, Cyna AM, Yip P, Chooi C, Middleton P. Non-pharmacological interventions for assisting the induction of anaesthesia in children. Cochrane Database Syst Rev 2015 Jul 14;2015(7):CD006447. [doi: 10.1002/14651858.CD006447.pub3] [Medline: 26171895]
16. Kim J, Chiesa N, Raazi M, Wright KD. A systematic review of technology-based preoperative preparation interventions for child and parent anxiety. Can J Anesth/J Can Anesth 2019 Aug;66(8):966-986. [doi: 10.1007/s12630-019-01387-8]
17. Simonetti V, Tomietto M, Comparcini D, Vankova N, Marcelli S, Cicolini G. Effectiveness of virtual reality in the management of paediatric anxiety during the peri-operative period: a systematic review and meta-analysis. Int J Nurs Stud 2022 Jan;125:104115. [doi: 10.1016/j.ijnurstu.2021.104115] [Medline: 34781118]
18. Rizzo MG Jr, Costello JP II, Luxenburg D, Cohen JL, Alberti N, Kaplan LD. Augmented reality for perioperative anxiety in patients undergoing surgery: a randomized clinical trial. JAMA Netw Open 2023 Aug 1;6(8):e2329310. [doi: 10.1001/jamanetworkopen.2023.29310] [Medline: 37589975]
19. Sajeev MF, Kelada L, Yahya Nur AB, et al. Interactive video games to reduce paediatric procedural pain and anxiety: a systematic review and meta-analysis. Br J Anaesth 2021 Oct;127(4):608-619. [doi: 10.1016/j.bja.2021.06.039] [Medline: 34340838]
20. Gates M, Hartling L, Shulhan-Kilroy J, et al. Digital technology distraction for acute pain in children: a meta-analysis. Pediatrics 2020 Feb;145(2):e20191139. [doi: 10.1542/peds.2019-1139] [Medline: 31969473]
21. Patel A, Schieble T, Davidson M, et al. Distraction with a hand-held video game reduces pediatric preoperative anxiety. Paediatr Anaesth 2006 Oct;16(10):1019-1027. [doi: 10.1111/j.1460-9592.2006.01914.x] [Medline: 16972829]
22. Kaos MD, Beauchamp MR, Bursick S, et al. Efficacy of online multi-player versus single-player exergames on adherence behaviors among children: a nonrandomized control trial. Ann Behav Med 2018 Sep 13;52(10):878-889. [doi: 10.1093/abm/kax061] [Medline: 30212848]
23. Seiden SC, McMullan S, Sequera-Ramos L, et al. Tablet-based interactive distraction (TBID) vs oral midazolam to minimize perioperative anxiety in pediatric patients: a noninferiority randomized trial. Paediatr Anaesth 2014 Dec;24(12):1217-1223. [doi: 10.1111/pan.12475] [Medline: 25040433]
24. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71. [doi: 10.1136/bmj.n71] [Medline: 33782057]
25. Sterne JAC, Savović J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019 Aug 28;366:l4898. [doi: 10.1136/bmj.l4898] [Medline: 31462531]
26. Wan X, Wang W, Liu J, Tong T. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. BMC Med Res Methodol 2014 Dec 19;14:135. [doi: 10.1186/1471-2288-14-135] [Medline: 25524443]
27. Patole S. Principles and Practice of Systematic Reviews and Meta-Analysis 2021.
28. McKenzie JE, Brennan SE, Ryan RE, Thomson HJ, Johnston RV. Chapter 9: summarizing study characteristics and preparing for synthesis. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al, editors. Cochrane Handbook for Systematic Reviews of Interventions Version 65: Cochrane; 2024. URL: www.training.cochrane.org/handbook
29. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 2008 Apr 26;336(7650):924-926. [doi: 10.1136/bmj.39489.470347.AD] [Medline: 18436948]
30. Queiroz Siqueira M, Tosseti S, Maréchal C, Cogniat B, Berthilier J, Chassard D. Abstract PR237: Effect of ipad tablet on children and parental anxiety during anesthesia. Anesthesia & Analgesia 2016;123(3S). [doi: 10.1213/01.ane.0000492635.96205.ee]
31. Marechal C, Berthiller J, Tosetti S, et al. Children and parental anxiolysis in paediatric ambulatory surgery: a randomized controlled study comparing 0.3 mg kg-1 midazolam to tablet computer based interactive distraction. Br J Anaesth 2017 Feb;118(2):247-253. [doi: 10.1093/bja/aew436] [Medline: 28100529]
32. Stewart B, Cazzell MA, Pearcy T. Single-blinded randomized controlled study on use of interactive distraction versus oral midazolam to reduce pediatric preoperative anxiety, emergence delirium, and postanesthesia length of stay. J Perianesth Nurs 2019 Jun;34(3):567-575. [doi: 10.1016/j.jopan.2018.08.004] [Medline: 30413359]
33. Levay MM, Sumser MK, Vargo KM, et al. The effect of active distraction compared to midazolam in preschool children in the perioperative setting: a randomized controlled trial. J Pediatr Nurs 2023;68:35-43. [doi: 10.1016/j.pedn.2022.09.021] [Medline: 36328912]
34. MacLaren JE, Thompson C, Weinberg M, et al. Prediction of preoperative anxiety in children: who is most accurate? Anesth Analg 2009 Jun;108(6):1777-1782. [doi: 10.1213/ane.0b013e31819e74de] [Medline: 19448201]

XSL•FO
RenderX

35. Jenkins BN, Kain ZN, Kaplan SH, et al. Revisiting a measure of child postoperative recovery: development of the post hospitalization behavior questionnaire for ambulatory surgery. Paediatr Anaesth 2015 Jul;25(7):738-745. [doi: 10.1111/pan.12678] [Medline: 25958978]
36. Sikich N, Lerman J. Development and psychometric evaluation of the pediatric anesthesia emergence delirium scale. Anesthesiology 2004 May;100(5):1138-1145. [doi: 10.1097/00000542-200405000-00015] [Medline: 15114210]
37. Cohen J. Statistical Power Analysis for the Behavioral Sciences: New York: Academic Press; 1969.
38. Jackson BF, Beck LA, Losek JD. Successful flumazenil reversal of paradoxical reaction to midazolam in a child. J Emerg Med 2015 Mar;48(3):e67-e72. [doi: 10.1016/j.jemermed.2014.09.062] [Medline: 25497845]
39. Kain ZN, MacLaren J, McClain BC, et al. Effects of age and emotionality on the effectiveness of midazolam administered preoperatively to children. Anesthesiology 2007 Oct;107(4):545-552. [doi: 10.1097/01.anes.0000281895.81168.c3] [Medline: 17893449]
40. Conway A, Rolley J, Sutherland JR. Midazolam for sedation before procedures. Cochrane Database Syst Rev 2016 May 20;2016(5):CD009491. [doi: 10.1002/14651858.CD009491.pub2] [Medline: 27198122]
41. Ashton H. Guidelines for the rational use of benzodiazepines. When and what to use. Drugs (Abingdon Engl) 1994 Jul;48(1):25-40. [doi: 10.2165/00003495-199448010-00004] [Medline: 7525193]
42. Cascella M, Cascella A, Monaco F, Shariff MN. Envisioning gamification in anesthesia, pain management, and critical care: basic principles, integration of artificial intelligence, and simulation strategies. J Anesth Analg Crit Care 2023 Sep 11;3(1):33. [doi: 10.1186/s44158-023-00118-2] [Medline: 37697415]
43. Griffiths M. Video games and health. BMJ 2005 Jul 16;331(7509):122-123. [doi: 10.1136/bmj.331.7509.122] [Medline: 16020833]
44. Wong DDL, Bailey CR. Emergence delirium in children. Anaesthesia 2015 Apr;70(4):383-387. [doi: 10.1111/anae.13043] [Medline: 25764401]
45. Chow CHT, Rizwan A, Xu R, et al. Association of temperament with preoperative anxiety in pediatric patients undergoing surgery: a systematic review and meta-analysis. JAMA Netw Open 2019 Jun 5;2(6):e195614. [doi: 10.1001/jamanetworkopen.2019.5614] [Medline: 31173131]
46. Vetter D, Barth J, Uyulmaz S, et al. Effects of art on surgical patients: a systematic review and meta-analysis. Ann Surg 2015 Nov;262(5):704-713. [doi: 10.1097/SLA.0000000000001480] [Medline: 26583656]
47. Gabriel MG, Wakefield CE, Vetsch J, et al. The psychosocial experiences and needs of children undergoing surgery and their parents: a systematic review. J Pediatr Health Care 2018;32(2):133-149. [doi: 10.1016/j.pedhc.2017.08.003] [Medline: 29066150]

Abbreviations

GRADE: Grading of Recommendations, Assessment, Development, and Evaluation
MD: mean difference
mYPAS: modified Yale Preoperative Anxiety Scale
PACU: postanesthesia care unit
PAED: pediatric anesthesia emergence delirium
PHBQ: posthospital behavior questionnaire
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RoB 2: version 2 of the Cochrane risk-of-bias tool for randomized trials
SMD: standardized mean difference

Edited by A Coristine; submitted 29.09.24; peer-reviewed by KEE Mills, T Baranowski; revised version received 16.02.25; accepted 17.02.25; published 10.03.25.

Please cite as:
Luo Z, Deng S, Zhou R, Ye L, Zhu T, Chen G
Comparative Efficacy of Video Games Versus Midazolam in Reducing Perioperative Anxiety in Pediatric Patients: Systematic Review and Meta-Analysis
JMIR Serious Games 2025;13:e67007
URL: https://games.jmir.org/2025/1/e67007
doi:10.2196/67007

© Ziyue Luo, Sisi Deng, Ruihao Zhou, Ling Ye, Tao Zhu, Guo Chen. Originally published in JMIR Serious Games (https://games.jmir.org), 10.3.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any

XSL•FO
RenderX

medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Acceptability of and Willingness to Use Virtual Reality Exergames for Weight Loss Among Young Adults With Overweight or Obesity in China: Qualitative Study

Yanya Chen[1,2*], MSN; Bingsheng Guan[3*], MD, PhD; Yaqi Zhang[4], MPH; Suen Chow Lee[2], MS; Jia-yu Liu[1], BSN; Sicun Li[2], MS; Ming Liu[2], MSc; Xiaoshen Zhang[5*], MD, PhD; Wai-kit Ming[2,6*], MD, PhD

[1]School of Nursing, Jinan University, Guangzhou, China

[2]Department of Infectious Diseases and Public Health, Jockey Club College of Veterinary Medicine and Life Sciences, City University of Hong Kong, To Yuen Building, 31 To Yuen Street, Hong Kong, China (Hong Kong)

[3]Department of Gastrointestinal Surgery, The First Affiliated Hospital, Jinan University, Guangzhou, China

[4]School of Public Health, University of Hong Kong, Hong Kong, China (Hong Kong)

[5]Department of Cardiovascular Surgery, The First Affiliated Hospital, Jinan University, Guangzhou, China

[6]Institute of Global Governance and Innovation for a Shared Future, City University of Hong Kong, Hong Kong, China (Hong Kong)

[*]these authors contributed equally

Corresponding Author:
Wai-kit Ming, MD, PhD
Department of Infectious Diseases and Public Health, Jockey Club College of Veterinary Medicine and Life Sciences, City University of Hong Kong, To Yuen Building, 31 To Yuen Street, Hong Kong, China (Hong Kong)

Abstract

Background: Overweight and obesity seriously affect physical and psychological health worldwide. They are common public health issues in young adults who are considered a “vulnerable group” prone to adopt unhealthy lifestyles that can lead to overweight and obesity. Virtual reality exergames could help increase balance performance among patients with Parkinson disease and improve depression and pain interference among individuals with chronic neck pain. Still, limited research has been conducted on the use of virtual reality exergames among young adults with overweight and obesity, and their willingness and acceptability remain unclear.

Objectives: This study aimed to assess the acceptability of and willingness to use virtual reality exergames for weight loss among young adults with overweight or obesity.

Methods: This was a qualitative study. Sixteen young adults with overweight or obesity were recruited in Guangzhou, China, and divided into 4 focus groups. They were interviewed between September and October 2023 through semistructured interviews. NVivo (version 14; Lumivero) was used to transcribe, code, and thematically analyze interviews.

Results: Four main themes and 8 subthemes emerged from the data. The main themes included perception of previous weight loss measures (regarded exercise and diet as the main measures for weight loss and difficulties in holding on to the weight loss methods), acceptability of virtual reality exergames (increasing opportunities for exercise, a powerful means for propelling doing exercise, positive impact on psychological well-being, and more appealing to office workers than students), willingness to use virtual reality exergames, and concerns (weight loss effect and other concerns).

Conclusions: Virtual reality exergames were seen as beneficial for maintaining exercise and promoting psychological well-being among young adults with overweight or obesity, despite concerns about effectiveness, cost, and privacy. Almost all young adults indicated their willingness to try these kinds of games if given the opportunity. These findings suggested that virtual reality exergames could be a promising tool for weight management in this population.

(JMIR Serious Games 2025;13:e66998) doi:10.2196/66998

KEYWORDS

acceptability; virtual reality; exergames; overweight; obesity

Introduction

The epidemic of obesity or overweight has gone beyond being a cosmetic concern to become a disease that poses a substantial threat to the well-being of individuals worldwide. Obesity raises the likelihood of developing a series of diseases, such as diabetes, hypertension, arteriosclerosis, hyperlipidemia, cardiovascular disease, and specific types of cancer [1].

XSL•FO
RenderX

Individuals with obesity also grapple with psychological disorders, including depression, lack of self-confidence, and low self-efficacy [2]. In 2025, more than 3041 million adults will be overweight (42% of the population), and 1249 million will be diagnosed with obesity (17% of the population) [3]. Especially, obesity prevalence among young adults ranged from 2.3% to 12%, and overweight was 28.8% [4]. Young adults (aged 18 - 25 years) are considered a "vulnerable group" prone to adopting unhealthy lifestyles that can lead to overweight and obesity [5,6]. This age range coincides with significant life transitions such as starting college, getting married, and beginning to build a family, which have been associated with weight gain [7]. Therefore, it is imperative to implement treatment and intervention programs targeting weight loss among young adults.

Engaging in enjoyable and sustainable physical activities that consistently motivate young adults is crucial for promoting weight loss and sustaining a healthy weight [8]. Nevertheless, conventional forms of physical exercise, such as jogging, yoga, and aerobic workouts, often fail to captivate individuals with obesity, leading to inconsistent and infrequent exercise habits. Moreover, young adults often struggle to find the time and energy to exercise regularly due to the demands of work and study [9,10]. The term "exergaming" (a portmanteau of "exercise" and "games") refers to technology-based physical activities that involve participants being physically active or exercising in order to play by using full-body movement as the primary means of interaction [11]. Despite potential drawbacks such as cybersickness, virtual reality (VR)-based exergames enhance involvement and immersion [5,12]. They also facilitate moderate-intensity physical activity, providing the convenience of independent use at home while overcoming common barriers such as distance and time constraints [13]. Evidence suggested that VR exergames could help increase balance performance among older patients with Parkinson disease [14] and improve depression and pain interference among individuals with chronic neck pain [15]. A study conducted in Taiwan, China, demonstrated that VR exergames could help reduce tension and induce calmness in patients with incomplete spinal cord injuries, thereby aiding their rehabilitation [16]. Among individuals with obesity, VR could effectively decrease the BMI levels of overweight middle-aged women and children, while also improving overall exercise performance [12,17]. It was also reported that Second Life (a social virtual world) could improve health self-efficacy for weight loss management among adults [18]. Still, limited research has been conducted on the use of VR exergames among young adults with overweight and obesity and their willingness and acceptability remain unclear. Consequently, in this study, the aim was to evaluate the willingness and acceptability to use VR exergames by young adults with overweight or obesity in weight loss settings through focus group interviews.

Methods

Participants and Study Design

This study adopted a qualitative approach using semistructured interviews to elucidate the viewpoints of the participants. Focus groups were selected as they could elicit open-ended responses, providing an in-depth understanding of individuals' perspectives, attitudes, beliefs, and opinions. Moreover, a group had the potential to foster synergies that transcended individual interviews as new insights could be generated through active listening, interactive engagement with peers, and reflective dialogue in focus groups [19].

The inclusion criteria were young adults aged 18 - 25 years, with a BMI of ≥24.0 kg/m^2 (overweight) or with a BMI of ≥28.0 kg/m^2 (obesity), did not have any prior firsthand experience with VR exergames, voluntary participation, living in Guangzhou, and have at least a high school education. Participants were recruited through posters at a university and a community located in Guangzhou, China. While sample sizes were typically not predetermined, it was common for qualitative studies to continue data collection until reaching saturation. First, we recruited 16 participants and divided them into 4 groups of 4 participants each. To ensure that the participants in a group comfortably talk as peers with other group members, the participants were divided into groups based on BMI. Two overweight groups enrolled participants with a BMI ranging from 24.0 to 27.9 kg/m^2, while 2 groups with obesity enrolled participants with a BMI of 28 kg/m^2. After interviewing these participants, we opted not to recruit other participants, as we determined that the data had reached saturation.

Interview Procedure

September to October 2023 was the collection period for the data. These were the procedures followed during the interviews: (1) Preparation: The focus group interviews were held in a quiet and isolated meeting room. The room used for the study measured approximately 20 m^2, providing sufficient space for comfortable seating of all participants. It was well lit with natural and artificial lighting to ensure optimal visibility. The main researcher, assistant, and field notetaker arrived at the meeting room 30 minutes before the participants arrived to arrange seating and prepare recording equipment. The main researcher involved in the evaluation holds a doctoral candidate in obesity fields and has more than 6 years of experience in conducting similar studies. The assistant has a master's degree and has participated in several research projects related to this topic. The field notetaker is an experienced researcher with a bachelor's degree and has been trained specifically for this study to ensure accurate and reliable data collection. Participants could sit comfortably in a circle and information might be transmitted through facial expressions and body posture. Each participant was seated approximately 1 - 1.5 m apart to ensure personal comfort and to maintain an inclusive and interactive environment. The main researcher, assistant researcher, and field notetaker were positioned within the circle but slightly offset to one side, allowing them to engage with participants without dominating the discussion. Before data collection, upon students' arrival, the researcher distributed sheets containing information about this study. Following this, the researchers described the study's purpose, methodology, and data-handling procedures in a comprehensive oral presentation. Subsequently, each participant signed the written consent form, ensuring that consent was obtained. The general characteristics of the

XSL•FO
RenderX

participants were collected, including age, gender, BMI, highest education, and work status. The BMI was determined through direct measurements. Each participant's height and weight were measured by trained research staff using the same equipment. These measurements were taken prior to the commencement of the study. (2) Introduction: The participants were explained that all obtained information would be used solely for research purposes. Once consent for recording was obtained, we turned on the portable recorder. (3) Discussion: The main researcher introduced topics based on the prepared interview-guiding questions and cultivated an environment where participants were allowed to discuss the topic freely (Textbox 1). After explaining how participants perceived their previous weight loss methods, a 7-minute Microsoft PowerPoint presentation was used to provide them with a brief overview regarding the concept and current state of VR exergames in the field of health to make sure that they had the same basic understanding. When no new contents or statements emerged, data saturation was identified; then data collection ended. Each interview took about 45 - 55 minutes. To stimulate the active participation of all participants during the interview, the main research also used nonverbal communication techniques, including nodding and smiling. (4) Based on participants' responses, the interview contents were reviewed. Each participant was provided a chance to add additional comments. After the interview, participants could receive a reward of 50 RMB (ie, US $6.83).

Textbox 1. Questions for focus groups.

1. Please share your feelings about your previous weight loss methods.
2. How do you feel about using virtual reality exergames to lose weight?
3. If virtual reality exercise games were used, how would they assist your weight loss?
4. What would your reaction be if you were suggested to exercise and lose weight using virtual reality exergames?
5. Is there anything you would like to add to the interview?

Ethical Considerations

Ethics approval was obtained through the institutional review board of Jinan University (reference number: JNUKY-2023 - 0128). Participant consent was obtained before interviews were conducted. This process included explaining the purpose of the research, methodology, and the procedures involved, and disclosure of compensation (participants received 50 RMB, ie, US $6.83, as reimbursement for their time). Participants could withdraw at any time without any setbacks. Furthermore, we implemented strict confidentiality measures to protect the privacy of our participants. Participants' information was anonymously coded during the transcription of the interviews and was securely stored and accessible only to authorized research team members.

Data Analysis

NVivo (version 14; Lumivero) software was used to code the data. Researchers systematically reviewed the transcripts, examining each sentence to gain a deeper understanding of the data and to ensure immersion before coding. The thematic analysis method described by Braun and Clarke [20] was adopted, including 6 steps totally, that is, familiarizing yourself with data, creating original codes, developing initial themes, reviewing themes, defining and naming themes, and preparing reports. Thus, we thematically analyzed the deidentified transcripts after coding. In addition, to enhance the validity of the analysis, member checking was used. After the initial coding and theme development, participants were invited to review the findings to ensure that their perspectives were accurately represented. Their feedback helped refine the themes and ensured the credibility of the interpretations. Two researchers (BG and SL) coded 2 transcripts to develop the initial coding framework. After discussing and revising the framework, 1 author (BG) coded all remaining transcripts and created categories. To ensure consistent coding across all transcripts, these researchers met regularly to further discuss the analysis, create codes, and categorize the transcripts. Subsequently, grounded in patterns of meaning in the data, emergent themes were developed and agreed upon by consensus within the main research team. To ensure transparency and reproducibility of our methods, the findings are presented in accordance with the COREQ (COnsolidated criteria for REporting Qualitative research) [21].

Results

Overview

A total of 16 participants (10 males and 6 females; aged 18 - 24 years), with an average BMI of 27.89 (SD 3.39), were included in this study (Table 1). For the focus group interviews, participants were divided into 4 groups, with a total recording time of 198 minutes. After an initial comprehensive review of the data, a total of 64 codes were developed and 8 subthemes were identified. These subthemes were then further refined and four main themes emerged, including (1) perception of previous weight loss measures, (2) acceptability of VR exergames, (3) willingness to use VR exercise games, and (4) concerns (Table 2).

XSL•FO
RenderX

Table . Sample characteristics.

Variable	Values
Age (years), mean (SD), n=16	21.00 (1.91)
Sex, n (%)	
Female	6 (38)
Male	10 (62)
BMI (kg/m^2), mean (SD), n=16	27.89 (3.39)
Highest education, n (%)	
Diploma or associate	4 (25)
Bachelor or higher	12 (75)
Work status, n (%)	
Employed	4 (25)
Unemployed[a]	12 (75)

[a]The work status of unemployed means the state of not having a job, including students and people who are looking for jobs.

Table . Main themes and subthemes from the focus group interviews.

Main themes	Subthemes
Perception of previous weight loss measures	• Regarded exercise and diet as the main measures for weight loss. • Difficulties in holding on to the weight loss methods.
Acceptability of virtual reality exergames	• Increasing opportunities for exercise. • A powerful means for propelling doing exercise. • Positive impact on psychological well-being. • More appealing to office workers than students.
Willingness to use virtual reality exergames	
Concerns	• Weight loss effect. • Other concerns.

Perception of Previous Weight Loss Measures

Regarded Exercise and Diet as the Main Measures for Weight Loss

Almost all participants (15 people) agreed that exercise and diet were the primary methods for weight loss, which they had used in the past. And these were the main ways they used to lose weight. Among them, 8 participants considered exercise and diet equally vital. Four reported losing weight through exercise without dietary control, while 2 noted weight loss through diet supplemented by exercise. In addition, participants also mentioned that they had used physiotherapy, acupuncture, and other crucial methods for losing weight.

> *Currently, I am controlling my weight through both diet and exercise. Previously, I have tried to lose weight through exercise or diet alone, but the result was not very significant. In an attempt to control my intake of sugar and fat, I once cooked for myself in the dormitory and ended up gaining weight. Perhaps… it's because my cooking is delicious.*
>
> *I don't diet much, but I exercise a lot. I play basketball quite often, and I also go swimming, but I don't control my eating at all.*

Difficulties in Holding on to the Weight Loss Methods

More than half of the participants found it challenging to stick to their previous weight loss methods. The main reasons cited by the participants for the difficulty in adhering to exercise included boredom (5 participants) and dislike of the negative experiences brought by exercise (7 participants). Negative experiences mentioned included sweating and muscle soreness. Moreover, work, study, and weather were also obstacles for participants in sticking to exercise. In terms of diet, participants indicated that lack of flavor in their meals was the issue.

> *Sweating makes me very uncomfortable! To clean myself, I need to walk back home and shower. But the*

XSL•FO
RenderX

price of walking home is too costly—I would walk a long way and sweat more!

Exercising alone doesn't seem very interesting... If I had to go by myself, I definitely wouldn't go out (to exercise).

Acceptability of VR Exergames

Increasing Opportunities for Exercise

Participants recognized that VR exergames provided them with more choices and opportunities to exercise. For example, they could exercise at home when it was difficult to exercise outside during cold or snowy winters. Furthermore, VR exergames offer an opportunity to exercise at home for those who might feel less comfortable to do exercise in public due to introversion or concerns about their body size.

During snowy or rainy conditions, people don't feel like going out… Yeah, just stays in one place and wraps himself in a blanket all day…. Virtual games (virtual reality exergames) can be played at home. There are simply more opportunities to exercise. Right?

I enjoy staying indoors by myself. I'm like a vampire who avoids sunlight, and my lifestyle habits are quite similar to those of a vampire. This (virtual reality exergames) may give me more exercise options.

During snowy or rainy conditions, people don't feel like going out…Yeah, just stays in one place and wraps himself in a blanket all day… Virtual games(virtual reality exergames) can be played at home. There are simply more opportunities to exercise. Right?

A Powerful Means for Propelling Doing Exercise

Many interviewees believed that VR exergames could help them maintain their exercise routines. On one hand, the fun aspect of VR exergames might motivate them to keep exercising. On the other hand, VR exergames could be played with family and friends, providing support and supervision for weight loss through exercise.

But the game can get people moving, and then you can lose weight while playing. That's what appeals to me the most, and it has the potential to make me stick to the exercise.

If I set a goal to do some exercises every day to lose weight, I might be able to stick to it on the first day. But on the second day, I might forget about it. Once I'm in bed or taking a shower, I just don't feel like exercising. There's nothing concrete to remind me. However, if there's a VR headset placed nearby, seeing it might remind me of something, like doing exercises. At the very least, it would serve as a reminder for me to do this task.

Positive Impact on Psychological Well-Being

Four participants mentioned that using VR exergames to lose weight might allow them to create an ideal image in the virtual world, giving them a feeling of "rebirth" and a new experience. They also felt that friendships established in the virtual world were not based on appearance, leading to increased confidence and more opportunities to make friends, which benefited their mental health. In addition, they felt less psychological burden when exercising at home in a place of relative privacy.

In reality, there are some people who don't like you because of your appearance, but in virtual reality, you can be reborn once again, you can shape your favorite image... I can play sports while playing games, and I have the opportunity to meet new friends through the game network, who will not discriminate against my appearance, and I will not feel inferior in the virtual world.

Exercising at home is causal and makes me feel relaxed. I dislike heat. But when I exercise at home using virtual reality exergames, I can basically only wear my underwear, without worrying about the eyes of others.

More Appealing to Office Workers Than Students

Two participants noted that VR exergames were suitable for office workers to lose weight but might be less suitable for students. Office workers, who were often busy and tired, might find it challenging to engage in outdoor sports. Interesting VR exergames could help them reduce pressure and lose weight.

We students have plenty of time, and there is a playground open 24 hours a day on our campus. So I prefer outdoor sports.

I usually get home very late and don't have the energy to go out (for outdoor activities). If I have the opportunity to exercise in my room, for example, through virtual reality (virtual reality exergames), I would still choose to use it for the sake of my health.

Willingness to Use VR Exercise Games

Almost all participants indicated their willingness to try using VR exergames for weight loss if given the opportunity. Six participants preferred to use VR for weight loss exercises at home rather than outdoors. Six participants considered VR exergames as an alternative to outdoor exercise. Among these 12 people, females prefer choosing dance-related sports in VR exergames, while males tend to prefer choosing aerobic and muscle strength exercises.

I would rather exercise at home through virtual reality exergames.

When I exercise outdoors, I can breathe the fresh air. Outdoor exercise is my first choice….

Concerns

Weight Loss Effect

Several participants expressed concerns that the intensity of VR exergames might not be sufficient, leading to unsatisfactory weight loss results.

Does this game really work? My friend played a Nintendo(Nintendo Switch®) game, and he told me that when he was playing badminton, he did not need

XSL•FO
RenderX

to swing his arm; he just needed to move his wrist with the handle, and the system would tell you that you had hit the ball. But in reality, it required us to swing our arms fully!

I go to the basketball court to play basketball in order to lose weight. Every time I end up sweating all over. I feel that virtual reality exergames can't achieve that level of physical activity.

Others Concerns

Some participants cited concerns beyond the effect of weight loss. Among them, 2 participants mentioned the high cost of devices, which might not be affordable for young people, including students. One participant expressed concern about the potential leakage of personal information when using VR exercise games. Another participant worried about becoming too addicted to VR exergames.

I once thought about buying the game suite: "Ring Fit Adventure for the Switch." But on second thought, I didn't buy the suite because it was too expensive.

Addiction, yes, many computer games allow you to just sit there without moving. I just manipulate a joystick, and sometimes I end up playing the whole day. This exercise game is still a game, and it's based on virtual reality. Could it be even more addictive?

Discussion

Principal Findings

The findings from this study indicate that most young adults with overweight or obesity reported difficulty maintaining their previously used weight control methods, such as exercise and diet. This finding is consistent with the study by Rana et al [22], which reported that nearly 91.7% of participants had attempted weight loss, yet almost 28% had rarely succeeded in maintaining their efforts. As demonstrated by Dabas et al [23], extreme temperatures and prolonged rainfall have the capacity to impact individual's travel, thereby significantly reducing the frequency of outdoor exercise and hindering the process of weight loss. As a result, some overweight people find it difficult to lose weight conventionally. In contrast to traditional exercise methods, VR games offer distinct advantages. They do not require a specific venue, and their flexible scheduling allows for easy adaptation to unexpected interruptions, such as inclement weather. Engaging in VR exergames can ensure the frequency and continuity of workouts, which helps the population with obesity adhere to their weight loss efforts. Furthermore, participants noted that negative experiences associated with exercise (muscle soreness, sweating, and fatigue) were also important factors that hindered their persistence in weight loss plans. Previous research has found that VR exergames were more engaging than standard exercise, as they can distract participants' attention, reduce the fatigue brought about by exercise to a certain extent, and allow participants to be more engaged in enjoying the process of exercise, ignoring the negative experiences brought about by the exercise itself and enjoying the process of exercise [24].

Along with physical symptoms, young adults with overweight or obesity often experience mental health issues, including depression, low self-esteem, bullying, social isolation, and body dissatisfaction, while individuals with normal weight are less likely to experience body dissatisfaction [25,26]. In the results of this qualitative study, some participants stated that VR exergames might have the potential to improve their psychological status, aligning with previous research [27]. This seemed a common result with some previous studies that had indicated a similar impact on depression when VR exergames were used among overweight middle-aged women [28], patients with cardiovascular disease, and patients with stroke [29,30]. Virtual reality allows users to experience situations indirectly that are hard to experience. These screen-based audiovisual stimuli could arouse interest, with the psychologically positive effect of the interaction of pleasure on depression [12]. Furthermore, some VR products allow users to create avatars in the virtual world according to their personal preferences, including Second Life and Nintendo's VR offering. Virtual reality exergames could assist young adults with obesity or overweight in improving their body satisfaction and self-efficacy [31]. Avatars in the virtual world were an opportunity to redress negative body image and perfect young adults' low self-esteem, low self-confidence, and other issues [32]. However, it was not easy to identify the underlying mechanism responsible for the improvement of depressive symptoms since it might involve neurobiological and psychological interactions [12].

In this study, young adults with overweight or obesity recognized that VR exergames were effective in engaging young adults with overweight or obesity in sustained physical exercise, aligning with the results of previous research [32]. Weight control was regarded as a big challenge for people with overweight or obesity who attempted to lose weight and it needed ongoing dietary, physical activities, and psychological support. One reason why young adults with overweight or obesity have a relatively high acceptability and willingness to use VR exergames might be that they considered that playing VR exergames socially with peers could increase enjoyment and intrinsic motivation for exercise, which was identified in similar research [33-35]. The degree of enjoyment derived from an activity has been recognized as a predictor of the efficacy of an exercise program, leading to interactive technology-based exercise becoming the all-time most favored approach for promoting physical activity [36]. Most VR exergames, including Nintendo Switch and Kinect, could be used in the home environment, and it was easy to configure [37] while providing both single-player mode and multiplayer mode. Virtual reality has already been defined as a motivating and enjoyable form of treatment, allowing them to interact more with family members, other peers, and patients [36]. Researchers found that playing VR exercise games with companions could improve social support and self-esteem, thereby motivating and maintaining physical activity [8]. When people engaged in VR exergames, they might gain satisfaction in performing well and competing with other players, which was related to the impact of VR exergames on the activation of the mesolimbic dopaminergic pathways and their influence on the reward system of the brain [8,38]. Another reason for young adults' acceptance of VR exergames might be the potential of VR exergames to

XSL•FO
RenderX

break through the obstacle of traditional outdoor sports, which could be limited by time and space. Virtual reality exercise games provide the possibility of indoor exercise, creating convenient conditions for patients to insist on exercise [13]. In particular, generation Z (born between 1997 and 2012), aligning with our study's focus on young adults, grew up in an environment of rapidly changing technologies and were highly adept at using digital technology [39], which influences their acceptability and willingness with VR exergames. Thus, VR exergames could become a potential way to increase physical exercise for people who were less likely to participate in regular aerobic exercise.

Although young adults with overweight or obesity thought that VR exergames could stimulate the intrinsic motivation to exercise and improve their psychological status, they still had some concerns, including the weight loss effect and the cost of the device. In the aspect of the weight loss effect of VR exercise games, Bosch et al [40] performed a study to evaluate the cardiorespiratory benefits of 30 minutes of boxing with 1 VR exergame (ie, Wii) compared with treadmill exercise in healthy young adults aged 23-27 years. It was concluded that VR exergames could provide sufficient aerobic activity for young adults. In addition, in line with previous findings [41,42], the acceptance of VR technology tended to be high among individuals with overweight or obesity as all participants in this study stated that they were willing to try to use VR exergames, although they did not have any experience of VR and had expressed worries about weight loss effect. This might be attributed to the playful nature of VR game exercises. Moreover, a previous study reported that only 1 out of 136 respondents with obesity had used VR or VR exergames as part of their obesity [42]. The low adoption rate of VR technology in obesity treatment is likely attributable to the high costs associated with VR hardware and software development [42]. However, as technology develops, the cost of VR may decrease, so the cost will not be one important factor influencing whether people use VR exergames in the future.

Limitations

This study had several limitations. The first limitation was that our participants did not have any prior experience with VR exergames, and thus, our focus group interviews were related to a hypothetical VR exergames that they had not experienced, although a brief overview regarding the concept and current state of VR exergames in the field of health was provided. After participating in an actual game, their perspectives would change and be different. However, previous research results indicated [42] that patients with obesity have limited knowledge about VR technology, with only 1 out of 136 participants considering VR technology as part of obesity treatment. That means this study's results can reflect most patients' views. In future research, opportunities can be provided for patients to attempt weight loss through VR exergames, and their perspectives can be explored through qualitative research. Another limitation is the limited generalization due to typically small and nonrandom samples. While samples for quantitative studies are selected from a population subgroup that is statistically representative of the population in order to generalize results, qualitative researchers recruit participants with experience of the phenomenon of interest as well as "information-rich cases" for in-depth analysis. Its focus is often on depth rather than breadth, limiting the ability to apply results universally. In future research, these limitations should be addressed by using a larger sample size and mixed methods, such as surveys and focus groups. However, this is the first study to investigate the acceptability of and willingness to use VR exergames by young adults with overweight or obesity in China. Our findings suggest that VR exergames are generally acceptable and have the potential to be used as a tool in weight loss interventions for young adults with overweight or obesity in China. Clinicians could consider incorporating these games into weight loss programs as a way to enhance engagement and motivation among young adults.

Conclusions

Compared with traditional weight loss methods, which were challenging to maintain, young adults with overweight or obesity considered VR exergames beneficial for maintaining regular exercise and promoting positive psychological well-being. Despite concerns regarding weight loss effectiveness, cost, and privacy issues, participants showed relatively high acceptance and willingness to engage with these games in weight loss settings. These results highlighted the potential of VR exergames as an engaging tool for weight management among young adults.

Acknowledgments

The authors would like to thank all participants for sharing their perspectives. This research was partially supported by SIRG-CityU Strategic Interdisciplinary Research (grant no. 7020093), Traditional Chinese Medicine Research Projects of Guangdong Provincial Administration of Traditional Chinese Medicine (no. 20241064), Guangdong Provincial Medical Research Fund (no. 2024313), and the University Youth Innovative Talent Project of Guangdong Province (no. 2023KQNCX006).

Disclaimer

We confirm that this work is original and has not been published elsewhere, nor is it currently under consideration for publication elsewhere.

Data Availability

The findings of this study are supported by data, which are accessible upon request from the corresponding author. The data cannot be publicly accessed due to privacy or ethical constraints.

XSL•FO
RenderX

Authors' Contributions

WM, XZ, BG, and YC led the study concept and design. YC and JL performed data collection. BG, SL, and ML performed data analysis. YC, YZ, BG, and SCL prepared the manuscript. The final draft was revised and approved by all authors. YC and BG made equal contributions as the co-first authors of this manuscript. WM and XZ, as the corresponding authors, made equal contributions to this manuscript.

Conflicts of Interest

None declared.

References

1. World Health Organization. Obesity and overweight. 2024 Mar. URL: https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight [accessed 2024-07-04]
2. Steptoe A, Frank P. Obesity and psychological distress. Philos Trans R Soc Lond B Biol Sci 2023 Oct 23;378(1888):20220225. [doi: 10.1098/rstb.2022.0225] [Medline: 37661745]
3. World Obesity Federation. Obesity atlas 2023. 2024. URL: https://data.worldobesity.org/publications/?cat=19 [accessed 2024-04-01]
4. Poobalan A, Aucott L. Obesity among young adults in developing countries: a systematic overview. Curr Obes Rep 2016 Mar;5(1):2-13. [doi: 10.1007/s13679-016-0187-x] [Medline: 26883372]
5. Xu W, Liang HN, Yu K, Baghaei N. Effect of gameplay uncertainty, display type, and age on virtual reality exergames. 2021 May 6 Presented at: CHI '21; May 8-13, 2021; Yokohama Japan URL: https://dl.acm.org/doi/proceedings/10.1145/3411764 [doi: 10.1145/3411764.3445801]
6. Chen Y, Guan L, Wu W, et al. Factors that influence young adults' preferences for virtual reality exergames in a weight control setting: qualitative study. J Med Internet Res 2024 Dec 30;26:e58422. [doi: 10.2196/58422] [Medline: 39753223]
7. Liu E, Chang SH. Self-esteem and weight status of young adults: findings from a pilot study. J Educ Health Promot 2022;11:263. [doi: 10.4103/jehp.jehp_1195_21] [Medline: 36325221]
8. Staiano AE, Abraham AA, Calvert SL. Motivating effects of cooperative exergame play for overweight and obese adolescents. J Diabetes Sci Technol 2012 Jul 1;6(4):812-819. [doi: 10.1177/193229681200600412] [Medline: 22920807]
9. Irandoust K, Taheri M, H'mida C, et al. Exergaming and aquatic exercises affect lung function and weight loss in obese children. Int J Sports Med 2021 Jun;42(6):566-572. [doi: 10.1055/a-1289-9307] [Medline: 33176381]
10. Hao Z, Liu K, Qi W, Zhang X, Zhou L, Chen P. Which exercise interventions are more helpful in treating primary obesity in young adults? A systematic review and Bayesian network meta-analysis. Arch Med Sci 2023;19(4):865-883. [doi: 10.5114/aoms/153479] [Medline: 37560714]
11. Abd-Alrazaq A, Alhuwail D, Al-Jafar E, et al. The effectiveness of serious games in improving memory among older adults with cognitive impairment: systematic review and meta-analysis. JMIR Serious Games 2022 Aug 9;10(3):e35202. [doi: 10.2196/35202] [Medline: 35943792]
12. Seo EY, Kim YS, Lee YJ, Hur MH. Virtual reality exercise program effects on body mass index, depression, exercise fun and exercise immersion in overweight middle-aged women: a randomized controlled trial. Int J Environ Res Public Health 2023 Jan 4;20(2):900. [doi: 10.3390/ijerph20020900] [Medline: 36673656]
13. Tatnell P, Atorkey P, Tzelepis F. The effectiveness of virtual reality interventions on smoking, nutrition, alcohol, physical activity and/or obesity risk factors: a systematic review. Int J Environ Res Public Health 2022 Aug 30;19(17):10821. [doi: 10.3390/ijerph191710821] [Medline: 36078528]
14. Maranesi E, Casoni E, Baldoni R, et al. The effect of non-immersive virtual reality exergames versus traditional physiotherapy in Parkinson's disease older patients: preliminary results from a randomized-controlled trial. Int J Environ Res Public Health 2022 Nov 10;19(22):14818. [doi: 10.3390/ijerph192214818] [Medline: 36429537]
15. Glavare M, Stålnacke BM, Häger CK, Löfgren M. Virtual reality exercises in an interdisciplinary rehabilitation programme for persons with chronic neck pain: a feasibility study. J Rehabil Med Clin Commun 2021;4:1000067. [doi: 10.2340/20030711-1000067] [Medline: 34527201]
16. Chen CH, Jeng MC, Fung CP, Doong JL, Chuang TY. Psychological benefits of virtual reality for patients in rehabilitation therapy. J Sport Rehabil 2009 May;18(2):258-268. [doi: 10.1123/jsr.18.2.258] [Medline: 19561368]
17. Polechoński J, Nierwińska K, Kalita B, Wodarski P. Can physical activity in immersive virtual reality be attractive and have sufficient intensity to meet health recommendations for obese children? A pilot study. Int J Environ Res Public Health 2020 Nov 1;17(21):8051. [doi: 10.3390/ijerph17218051] [Medline: 33139604]
18. Behm-Morawitz E, Lewallen J, Choi G. A second chance at health: how A 3D virtual world can improve health self-efficacy for weight loss management among adults. Cyberpsychol Behav Soc Netw 2016 Feb;19(2):74-79. [doi: 10.1089/cyber.2015.0317] [Medline: 26882324]
19. Chen Y, Li Z, Zou X, Yang Q, Guan B, Ming WK. Nursing students' perceptions of using virtual patients to prepare for clinical practice. J Nurs Educ 2022 Jul;61(7):398-402. [doi: 10.3928/01484834-20220613-04] [Medline: 35858137]

XSL•FO
RenderX

20. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
21. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
22. Rana ZH, Reed DB, Oldewage-Theron W, Lyford C, Colwell M, Dawson JA. Overweight or obesity onset and past attempts to lose/manage weight: a qualitative study. Obesities 2021;1(3):136-143. [doi: 10.3390/Obesities1030012]
23. Dabas J, Shunmukha Priya S, Alawani A, Budhrani P. What could be the reasons for not losing weight even after following a weight loss program? J Health Popul Nutr 2024 Mar 2;43(1):37. [doi: 10.1186/s41043-024-00516-4] [Medline: 38429842]
24. Faric N, Potts HWW, Hon A, et al. What players of virtual reality exercise games want: thematic analysis of web-based reviews. J Med Internet Res 2019 Sep 16;21(9):e13833. [doi: 10.2196/13833] [Medline: 31538951]
25. Wu Q, Han R, Li Z, et al. Effect of virtual reality-based exercise and physical exercise on adolescents with overweight and obesity: study protocol for a randomised controlled trial. BMJ Open 2023 Oct 11;13(10):e075332. [doi: 10.1136/bmjopen-2023-075332] [Medline: 37821136]
26. Gemesi K, Holzmann SL, Hochrein R, et al. Attitude of nutrition experts toward psychotherapy and virtual reality as part of obesity treatment-an online survey. Front Psychiatry 2022;13:787832. [doi: 10.3389/fpsyt.2022.787832] [Medline: 35546927]
27. Xu W, Liang HN, Baghaei N, et al. Effects of an immersive virtual reality exergame on university students' anxiety, depression, and perceived stress: pilot feasibility and usability study. JMIR Serious Games 2021 Nov 22;9(4):e29330. [doi: 10.2196/29330] [Medline: 34813487]
28. Kim M, Choi SW, Moon S, et al. Treatment effect of psychoeducation and training program using virtual reality technique in the patients with depressive symptoms. J Korean Neuropsychiatr Assoc 2020;59(1):51. [doi: 10.4306/jknpa.2020.59.1.51]
29. Chen Y, Cao L, Xu Y, Zhu M, Guan B, Ming WK. Effectiveness of virtual reality in cardiac rehabilitation: a systematic review and meta-analysis of randomized controlled trials. Int J Nurs Stud 2022 Sep;133:104323. [doi: 10.1016/j.ijnurstu.2022.104323] [Medline: 35870329]
30. Liu H, Cheng Z, Wang S, Jia Y. Effects of virtual reality-based intervention on depression in stroke patients: a meta-analysis. Sci Rep 2023 Mar 16;13(1):4381. [doi: 10.1038/s41598-023-31477-z] [Medline: 36928345]
31. Coons MJ, Roehrig M, Spring B. The potential of virtual reality technologies to improve adherence to weight loss behaviors. J Diabetes Sci Technol 2011 Mar 1;5(2):340-344. [doi: 10.1177/193229681100500221] [Medline: 21527103]
32. Navarro J, Cebolla A, Llorens R, Borrego A, Baños RM. Manipulating self-avatar body dimensions in virtual worlds to complement an internet-delivered intervention to increase physical activity in overweight women. Int J Environ Res Public Health 2020 Jun 5;17(11):4045. [doi: 10.3390/ijerph17114045] [Medline: 32517117]
33. Farič N, Smith L, Hon A, et al. A virtual reality exergame to engage adolescents in physical activity: mixed methods study describing the formative intervention development process. J Med Internet Res 2021 Feb 4;23(2):e18161. [doi: 10.2196/18161] [Medline: 33538697]
34. Hadi Mogavi R, Son J, Yang S, et al. The Jade Gateway to exergaming: how socio-cultural factors shape exergaming among East Asian older adults. Proc ACM Hum-Comput Interact 2024 Oct 14;8(CHI PLAY):1-34. [doi: 10.1145/3677106]
35. Xu W, Liang HN, Yu K, Wen S, Baghaei N, Tu H. Acceptance of virtual reality exergames among Chinese older adults. International Journal of Human–Computer Interaction 2023 Mar 16;39(5):1134-1148. [doi: 10.1080/10447318.2022.2098559]
36. García-Bravo S, Cuesta-Gómez A, Campuzano-Ruiz R, et al. Virtual reality and video games in cardiac rehabilitation programs. A systematic review. Disabil Rehabil 2021 Feb;43(4):448-457. [doi: 10.1080/09638288.2019.1631892] [Medline: 31258015]
37. Vieira Á, Melo C, Machado J, Gabriel J. Virtual reality exercise on a home-based phase III cardiac rehabilitation program, effect on executive function, quality of life and depression, anxiety and stress: a randomized controlled trial. Disabil Rehabil Assist Technol 2018 Feb;13(2):112-123. [doi: 10.1080/17483107.2017.1297858] [Medline: 28285574]
38. Klasen M, Weber R, Kircher TTJ, Mathiak KA, Mathiak K. Neural contributions to flow experience during video game playing. Soc Cogn Affect Neurosci 2012 Apr;7(4):485-495. [doi: 10.1093/scan/nsr021] [Medline: 21596764]
39. Yeh CH, Huang HM, Kuo CL, Huang CY, Cheng SF. Effectiveness of e-STORY App in clinical reasoning competency and self-directed learning among students in associate nursing program: a quasi experimental study. Nurse Educ Pract 2022 Oct;64:103456. [doi: 10.1016/j.nepr.2022.103456] [Medline: 36162302]
40. Bosch PR, Poloni J, Thornton A, Lynskey JV. The heart rate response to Nintendo Wii boxing in young adults. Cardiopulm Phys Ther J 2012 Jun;23(2):13-29. [Medline: 22833705]
41. Manghisi VM, Evangelista A, Semisa D, Latorre V, Uva AE. Evaluating the acceptance of cinematic virtual reality-based applications for rehabilitative interventions in schizophrenia. Games Health J 2022 Sep 8. [doi: 10.1089/g4h.2021.0223] [Medline: 36074091]
42. Luck-Sikorski C, Hochrein R, Döllinger N, et al. Digital communication and virtual reality for extending the behavioural treatment of obesity—the patients' perspective: results of an online survey in Germany. BMC Med Inform Decis Mak 2023 May 24;23(1):100. [doi: 10.1186/s12911-023-02197-1] [Medline: 37226164]

XSL•FO
RenderX

Abbreviations

COREQ: COnsolidated criteria for REporting Qualitative research
VR: virtual reality

Edited by F Spyridonis; submitted 28.09.24; peer-reviewed by A Rusyantia, HN Liang; revised version received 07.02.25; accepted 11.02.25; published 01.05.25.

Please cite as:
Chen Y, Guan B, Zhang Y, Lee SC, Liu JY, Li S, Liu M, Zhang X, Ming WK
Acceptability of and Willingness to Use Virtual Reality Exergames for Weight Loss Among Young Adults With Overweight or Obesity in China: Qualitative Study
JMIR Serious Games 2025;13:e66998
URL: https://games.jmir.org/2025/1/e66998
doi:10.2196/66998

© Yanya Chen, Bingsheng Guan, Yaqi Zhang, Suen Chow Lee, Jia-yu Liu, Sicun Li, Ming Liu, Xiaoshen Zhang, Wai-kit Ming. Originally published in JMIR Serious Games (https://games.jmir.org), 1.5.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Personal Protection Equipment Training as a Virtual Reality Game in Immersive Environments: Development Study and Pilot Randomized Controlled Trial

Liang Zhou[1,2], PhD; Haoyang Liu[2,3]; Mengjie Fan[1,2], MSc; Jiahao Liu[2,3]; Xingyao Yu[4], MSc; Xintian Zhao[5], MSc; Shaoxing Zhang[6], MD

[1]Institute of Medical Technology, Peking University Health Science Center, Beijing, China

[2]National Institute of Health Data Science, Peking University, Beijing, China

[3]Beijing Forestry University, Beijing, China

[4]Visualization Research Center (VISUS), University of Stuttgart, Stuttgart, Germany

[5]Department of Healthcare-Associated Infection Management, Peking University Third Hospital, Beijing, China

[6]Department of Otolaryngological, Peking University Third Hospital, 49 Huayuanbei Road, Beijing, China

Corresponding Author:
Shaoxing Zhang, MD
Department of Otolaryngological, Peking University Third Hospital, 49 Huayuanbei Road, Beijing, China

Abstract

Background: Proper donning and doffing of personal protection equipment (PPE) and hand hygiene in the correct spatial context of a health facility is important for the prevention and control of nosocomial infections. On-site training is difficult due to the potential infectious risks and shortages of PPE, whereas video-based training lacks immersion which is vital for the familiarization of the environment. Virtual reality (VR) training can support the repeated practice of PPE donning and doffing in an immersive environment that simulates a realistic configuration of a health facility.

Objective: This study aims to develop and evaluate a VR simulation focusing on the correct event order of PPE donning and doffing, that is, the item and hand hygiene order in the donning and doffing process but not the detailed steps of how to don and doff an item, in an immersive environment that replicates the spatial zoning of a hospital. The VR method should be generic and support customizable sequencing of PPE donning and doffing.

Methods: An immersive VR PPE training tool was developed by computer scientists and medical experts. The effectiveness of the immersive VR method versus video-based learning was tested in a pilot study as a randomized controlled trial (N=32: VR group, n=16; video-based training, n=16) using questionnaires on spatial-aware event order memorization questions, usability, and task workload. Trajectories of participants in the immersive environment were also recorded for behavior analysis and potential improvements of the real environment of the health facility.

Results: Comparable sequence memorization scores (VR mean 79.38, SD 12.90 vs video mean 74.38, SD 17.88; *P*=.37) as well as National Aeronautics and Space Administration Task Load Index scores (VR mean 42.9, SD 13.01 vs video mean 51.50, SD 20.44; *P*=.16) were observed. The VR group had an above-average usability in the System Usability Scale (mean 74.78>70.0) and was significantly better than the video group (VR mean 74.78, SD 13.58 vs video mean 57.73, SD 21.13; *P*=.009). The analysis and visualization of trajectories revealed a positive correlation between the length of trajectories and the completion time, but neither correlated to the accuracy of the memorization task. Further user feedback indicated a preference for the VR method over the video-based method. Limitations of and suggestions for improvements in the study were also identified.

Conclusions: A new immersive VR PPE training method was developed and evaluated against the video-based training. Results of the pilot study indicate that the VR method provides training quality comparable to video-based training and is more usable. In addition, the immersive experience of realistic settings and the flexibility of training configurations make the VR method a promising alternative to video instructions.

(JMIR Serious Games 2025;13:e69021) doi:10.2196/69021

KEYWORDS

virtual reality training; nosocomial infections control; visualization; human computer interaction; personal protection equipment; PPE

XSL•FO
RenderX

Introduction

Overview

Nosocomial infections are infections acquired within health care facilities. Such infections not only extend hospital stays and treatments of a patient but can also result in severe complications and even death [1,2]. Moreover, health care workers face an increased risk of infection due to cross-contamination that can escalate workload and disrupt health care services. Implementing preventive measures such as proper hand hygiene, disinfection protocols, appropriate personal protection equipment (PPE) use, and clean environment maintenance can significantly reduce the risk of such infections [3-5]. Notably, the availability, selection, and proper training of PPE is critical for reducing the risks of nosocomial infections [5,6]. The order of and the correct handling during donning and doffing of PPE is critical for reducing infections. It is known that a high risk of contamination exists during doffing [7].

On top of that, infection control protocols of a health care facility often involve spatial division for different protection and disinfection levels, for example, the operation room requires a higher PPE protection level than that of the outpatient building. In the case of a pandemic such as COVID-19, the “three areas and two corridors” division is used for health care facilities [8,9]. There, PPE donning and doffing have to be done in specific orders in specific regions to ensure the isolation of contaminated regions from clean regions.

PPE Training in Health Facilities

PPE training is typically organized by each health facility with in-person demonstrations and lecturing. A major benefit of in-person training in a health facility is that personnel can be trained in a real spatial context. However, training in the real environment has severe shortcomings: First, the risk of the trainee being infected and the contamination of the uninfected regions; second, the lack of repeated practice due to the potential shortage of PPE. To reduce differences in training and infectious risks, standardized PPE training videos are available. Participants passively acquire the information as the trainer presents lecture slides or demonstrates the procedure, and need to practice on their own with real PPE equipment that is sometimes in shortage. Simulation-based education is an alternative to lectures as it is effective for PPE training with reduced cognitive load [10,11].

Infection Control Training in Virtual Reality

Virtual reality (VR) and augmented reality are extensively used in medical training for a wide range of procedures, for example, surgery [12,13], needle insertion [14,15], and ultrasound [16]. Compared with augmented reality, VR provides a more immersive experience, saves training costs, and avoids exposure risks and resource consumption in real environments [17,18]. Therefore, we decide to devise our training method in VR.

VR methods are available for infection control training for various scenarios [19-24]. A VR infection control simulation program is available for training nurses to help children patients with COVID-19 [22]. There, donning and doffing of PPE, and the treatment of patients are simulated. However, the immersive environment does not focus on the spatial arrangements of zones of a real hospital; furthermore, the control group of this study did not receive video-based or conventional training but only did a questionnaire survey. For the management of patients with COVID-19, VR methods are developed for the infection control and diagnosis process [19] and proper case handling [20].

Recently, a VR simulator has been available for training detailed steps in the donning and doffing process of PPE [25]. The simulator is further extended by the use of reinforcement learning for training and interactive reinforcement learning for assessing the procedure of donning and doffing PPE [26].

Evidence is found for VR-based PPE donning and doffing training to be superior to traditional methods. Better performance is achieved with VR-based PPE donning and doffing training as higher performance scores than the traditional e-learning method are yielded [27]. Immersive 360-degree VR training is found to have comparable effectiveness to face-to-face training and is superior to video training [28].

VR is also used for infection control for operations in specific areas of a health care facility, such as preventing surgical-site infections in the operating room [23], and in neonatal intensive care units [21]. Hand hygiene and PPE donning are taught through 360-degree videos in VR for infection control training which is shown to be superior to traditional slideshow lecturing [24]. However, no work is currently available for interactive training of PPE donning and doffing embedded in an immersive environment with a spatial division for infection control that simulates a real health care facility.

Our VR Method for PPE Training in a Spatial Context

In this paper, we propose an immersive PPE training method in VR with simulated spatial division of a hospital for nosocomial infection prevention to address this issue. Our work is a close collaboration with medical professionals who provided health care support for the Winter Olympic Games 2022 during the COVID-19 pandemic. Our immersive training method involves the modeling of the workspace, the avatar, and PPE donning and doffing logic based on the actual setting of the supporting hospital of the Winter Olympic Games. In a pilot study of 32 participants, we compared our new method to traditional video-based instruction learning. In addition to traditional knowledge tests and subjective evaluations, the study also recorded the trajectories of participants in the immersive environment for behavior analysis and potential improvements of the simulation and the real health facility. Results demonstrate that our VR method has a comparable training quality in terms of memorizing PPE donning and doffing orders with better subjective evaluations.

The main benefit of our method is that it allows the practice of PPE donning and doffing according to the infection prevention and control protocols immersively in a virtual environment that resembles the actual zoning of a hospital without wasting PPE which can be potentially scarce. Another benefit is that our method allows for easy and flexible editing of the donning and doffing sequence with a spatial context that adapts to the actual situation of a health facility.

XSL•FO
RenderX

Methods

Immersive VR PPE Training Method

Overview

An immersive PPE training method as a mini VR game was developed as a collaboration between computer science experts and medical experts. Based on the requirement analysis, we designed the generic donning and doffing sequence model, the virtual environments, the avatar, and the training logic. Our VR training method was implemented by computer scientists in our group as a Unity application.

Requirement Analysis

The requirements were initially proposed by medical experts who took the PPE training in a support hospital for the Winter Olympic Games 2022, and finalized through iterative discussions between computer science experts and medical experts. The hospital was under a high infection prevention and control level to ensure the safety and high-quality health care services of the Games. Medical experts identified 2 challenges: first, PPE includes many items and they have to be donned and doffed properly in the correct order; second, the doffing has to be done in the correct zones that need to be familiarized. Specifically, in this case, PPE includes medical masks, N-95 masks, protection suits, gowns, inner gloves, outer gloves, goggles, face shields, and shoe covers. The hospital was spatially divided using the "three zones and two corridors" protocol and several rooms were allocated for different steps of doffing.

After discussions, both the medical and computer science experts agreed that an interactive VR method should be designed as it can support the repeated practice of donning and doffing and an immersive experience in an environment that simulates the hospital. The requirements of the VR method are as follows:

- Requirement 1 (R1): focus on the practice of PPE donning and doffing sequence.
- Requirement 2 (R2): Simulation of the spatial division of the hospital relevant to PPE donning and doffing.
- Requirement 3 (R3): configurable donning and doffing order for flexibility and extensibility.

Therefore, we conceptualized the VR method as a mini-game that guides users to learn the correct ordering. The trainee starts from the donning zone, and she or he has to follow the correct donning and hand sanitation order to get access to the work zone. The door unlocks only after the prescribed donning sequence is successfully performed without any error. Next, in the same fashion, the trainee has to correctly follow the doffing order with thorough hand sanitation to proceed to subsequent doffing zones to finish the game.

Generic Donning and Doffing Sequence Modeling

A generic donning and doffing sequence model is devised to enable flexible customization of PPE simulation training (R3). As shown in Figure 1, each PPE is associated with an eID, body part identifier bID, and region identifier sID. The donning and doffing event triples Ei are recorded in a list *L*:

$$L=(E_1,E_2,...,E_M),$$

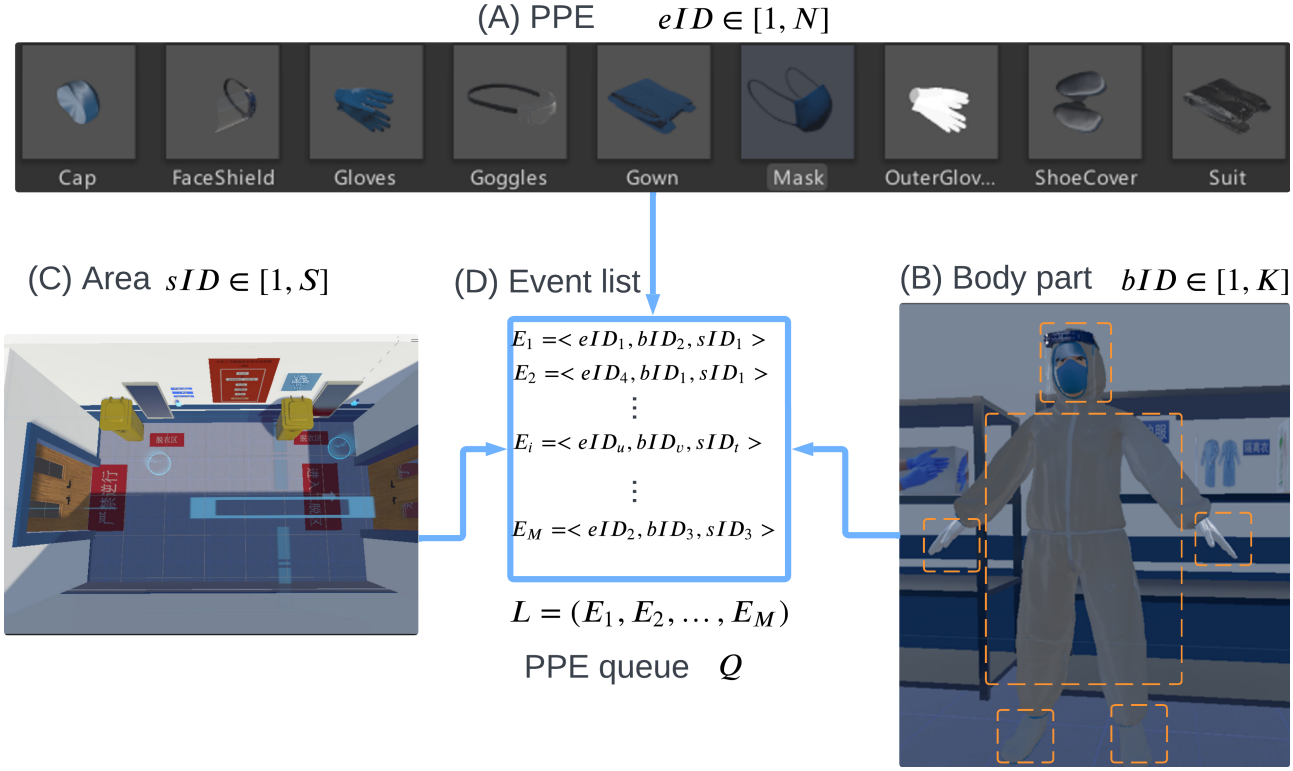
$$Ei=<eIDu,bIDv,sIDt>,$$

where $i\in[1,M],u\in[1,N],v\in[1,K],s\in[1,S]$.

Here, M,N,K, and S are the total steps of donning and doffing, the number of types of PPE, the number of body parts, and the number of zones, respectively. To ensure that all PPE is properly doffed, a queue Q is used to record the PPE status. If a piece of equipment is donned, it is enqueued into Q, and is dequeued if the equipment is doffed, and Q should be empty at the end of the training. The eID and bID in an event were prebounded so that the PPE can only be put on or removed when correctly corresponds to the body position and the collision detection has been passed. Therefore, the trainer only needs to set the sequence of events in corresponding areas to create a customized donning and doffing sequence according to the actual situation. The sequence in this paper follows the actual order in the supporting hospital during the 2022 Winter Olympics Games.

XSL•FO
RenderX

Figure 1. The generic donning and doffing sequence model. PPE: personal protection equipment.



Virtual Environment Modeling

Based on photos and videos of the environments of the actual hospital as shown in Figure 2, we create a virtual environment by reconstructing the rooms, items, and their arrangements (R1, R2). Four rooms are created: 1 for donning (Figure 3A) and 3 for doffing (Figure 3B-D), and the rooms are divided by doors. Other areas including the work zone of the hospital are omitted. In the donning room, PPE is placed on shelves in the order of the actual placement in the real hospital. Hand sanitizers, mirrors, and trash bins are placed at their corresponding locations. Indication signs (in Chinese), for example, floor signs "to the work zone" in the donning area (Figure 3A), "doffing zone" and instructions for doffing orders on the walls of doffing areas (Figure 3).

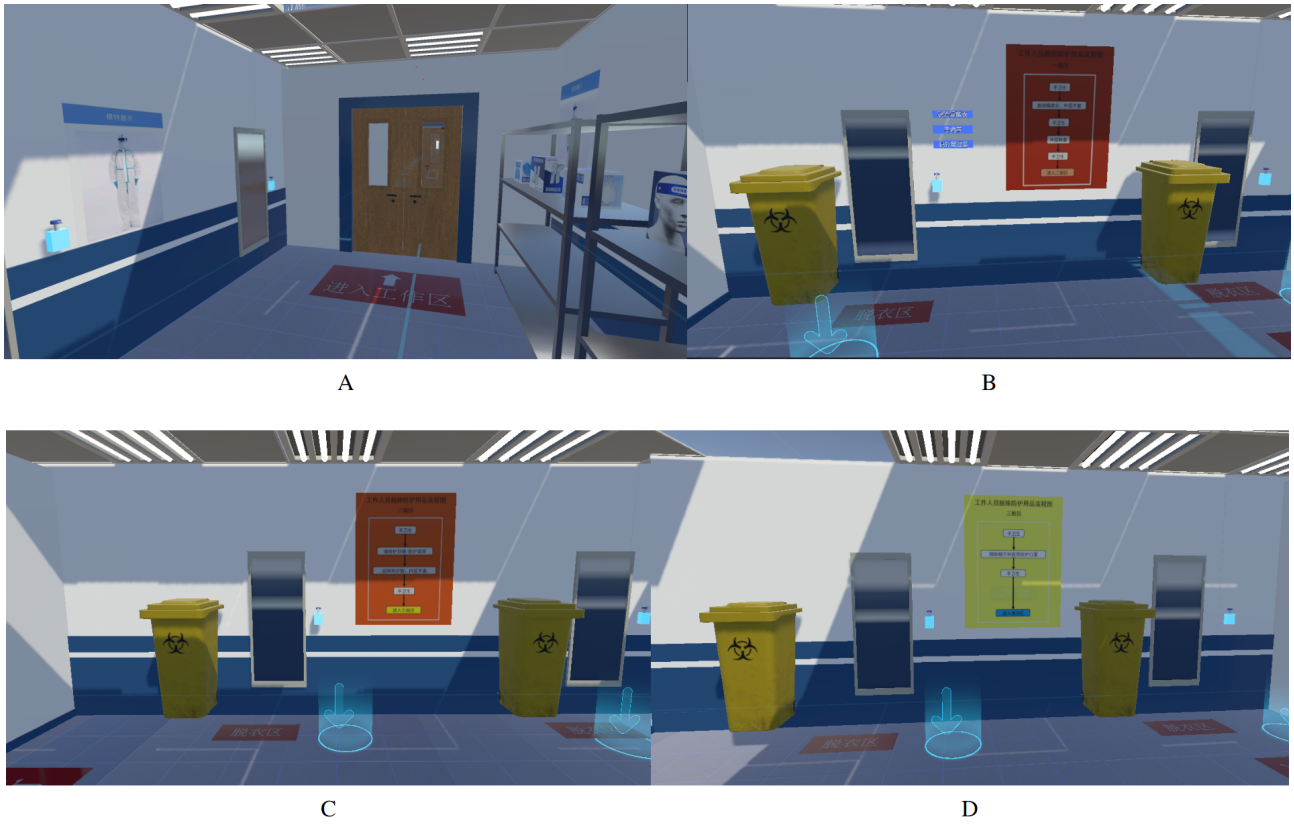
Teleportation zones are designated to facilitate moving to specific areas and prevent participants from accidentally exiting the room boundaries. To make the teleportation more effective, anchor points are placed in front of hand sanitizers and shelves so that users can conveniently reach target areas and items (semitransparent circles with arrows in blue off Figure 3).

XSL•FO
RenderX

Figure 2. Photos of the personal protection equipment (PPE) donning and doffing areas of the hospital. The overview (A) of the donning area with (B) details of the PPE on shelves. The 3 doffing areas are shown in (C-E).

XSL•FO
RenderX

Figure 3. Virtual scenes of (A) the donning area and (B-D) 3 doffing areas.



PPE Donning and Doffing Design

Mechanisms are designed for practicing the PPE donning and doffing ordering (R1) which can be conveniently configured according to actual situations of hospitals (R3).

Avatar

To increase the realism of the training, an avatar is designed (R2). The camera associated with the VR headset is aligned with the eye position of the avatar for a synchronized first-person perspective. Hand controllers are bound together with the hands of the avatar, facilitating an intuitive representation of hand movements. The avatar is skeletal rigged and inverse kinematics is used to simulate human movements, for example, the movements of the whole arms when PPE is picked up, being donned, or doffed. Upon equipping and removing each PPE item, the avatar is updated visually as shown in Figure 1B.

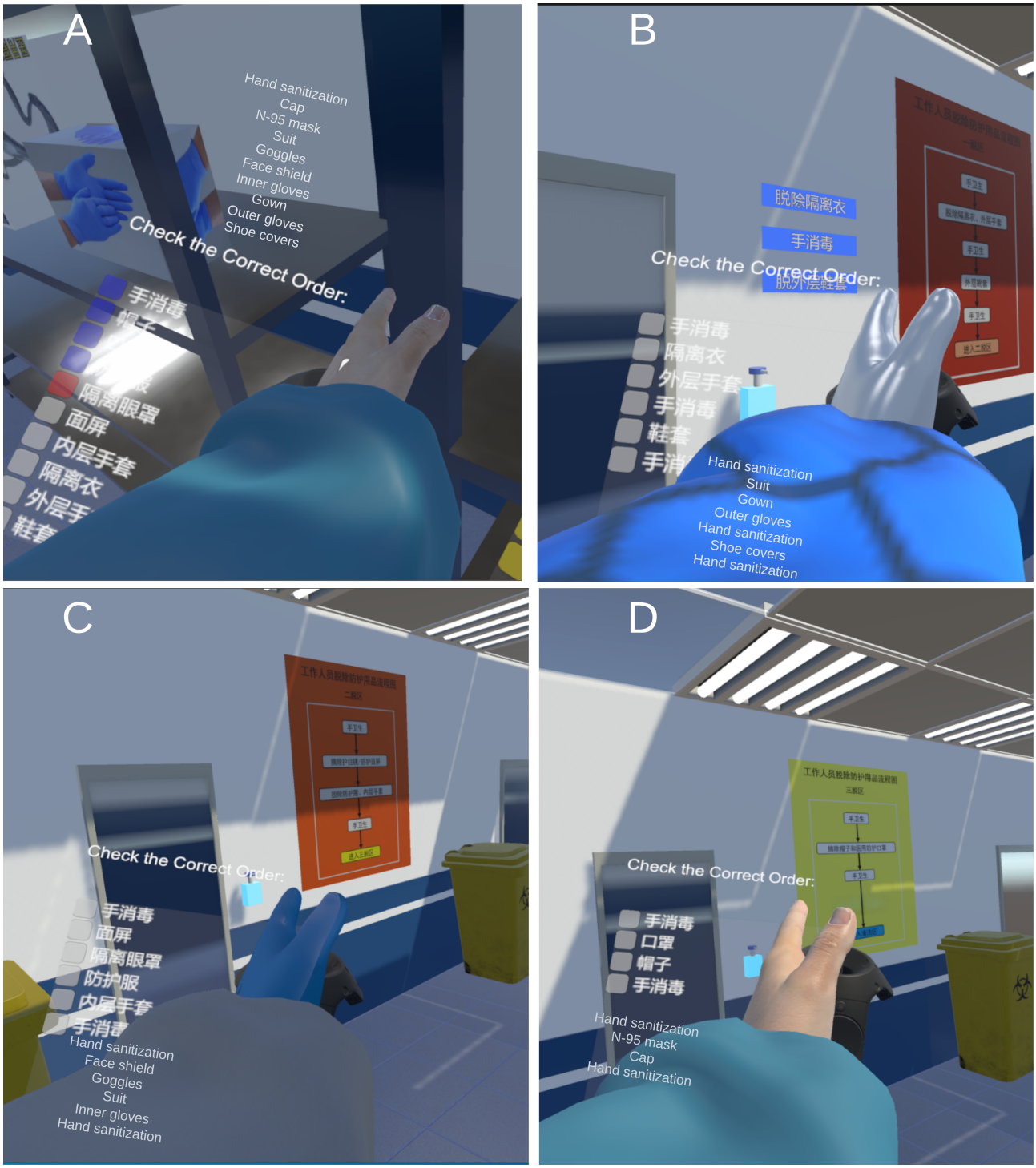
Donning

The donning order is shown as an item checklist in a user interface (UI) above the left-hand controller model (Figure 4). This UI displays the sequence for donning or doffing gear within this zone using checkboxes: the box in white that the action is pending, blue signifies a completed action, and red indicates that the action was performed incorrectly as shown in Figure 4A. In the donning zone, interactable boxes of PPE are distributed on shelves. The trainee uses the controller to touch a box and presses the trigger button to grasp the PPE from the box. PPE model must be brought close to the relevant body part to simulate the donning process, for example, goggles must be near the head, while a protective suit must be near the torso. When sufficiently close, the material of the protective item turns green, indicating correct placement. Releasing the trigger button causes the item to automatically adhere to the designated position on the body.

Following the correct donning, the indicator on the left-hand UI panel is updated to blue (Figure 4A). The right-hand UI tracks each action taken during the donning and doffing process, including any erroneous movements. After the entire donning process is complete, the door of the donning zone opens, allowing the user to proceed to the next section. Note that the work zone is omitted as it is irrelevant to our training.

XSL•FO
RenderX

Figure 4. The user interface (UI) instructions of (A) the donning area and (B-D) 3 doffing areas.



Doffing

Similar to the donning section, upon entering the doffing zones, the UI on the left controller shows the sequence for removing protective gear, along with indicative markers (Figure 4B-D). In this zone, users are required to follow the displayed instructions carefully, removing the protective items from specified parts of the body. For example, by approaching the head with the controller, the face shield can be grasped and lifted off; by approaching the torso, the protective suit can be taken off. The instruction flowchart placed on the wall within each doffing zone provides participants with guidance on the correct order for removing protective equipment as in the real hospital.

Trajectory Recording and Visualization

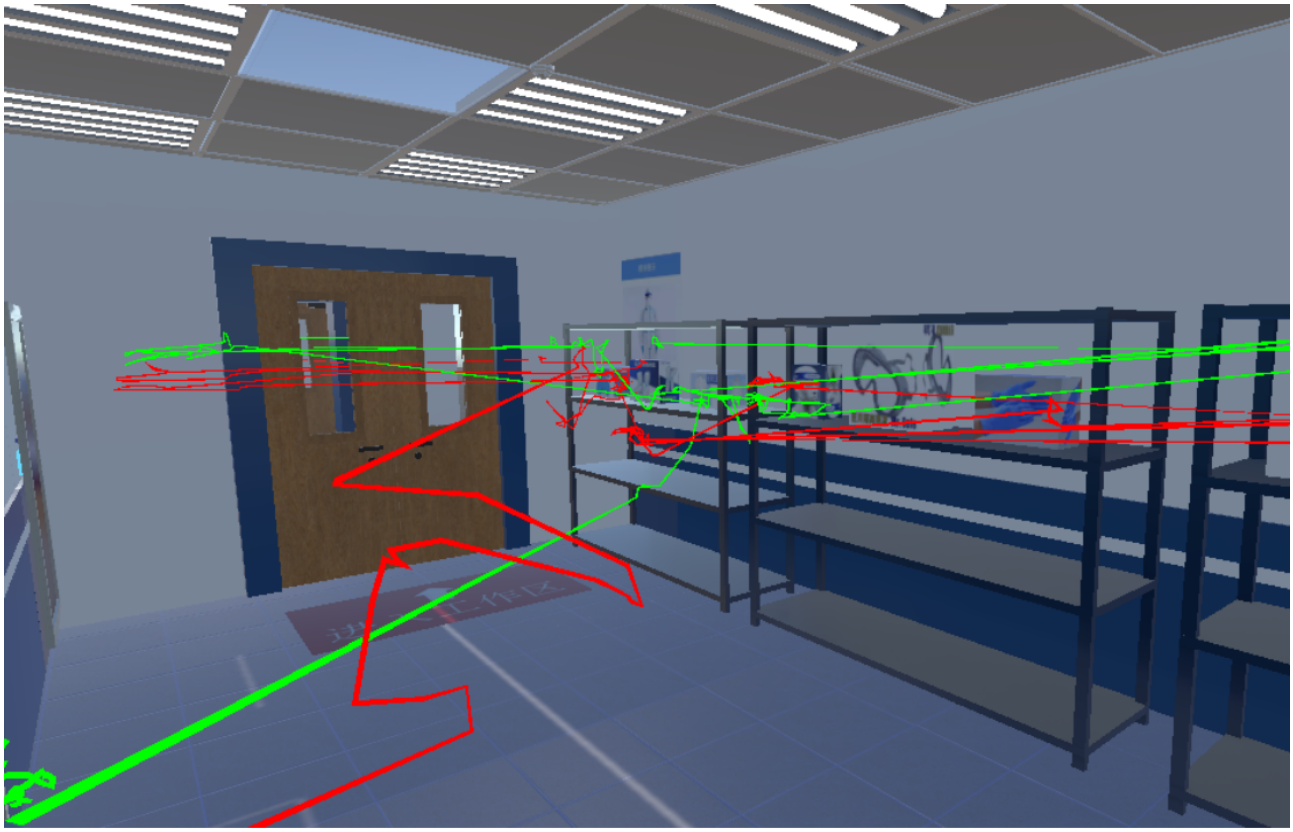
Overview

To support the analysis of the proficiency of trainees and also potential improvements in the hospital, our VR method supports the recording of trajectories and the visualization thereafter. Trajectories of the user's head, left hand, and right hand are recorded during the entire duration of use. Furthermore, the time spent by the user in each designated area, as well as the total time consumed, are also recorded.

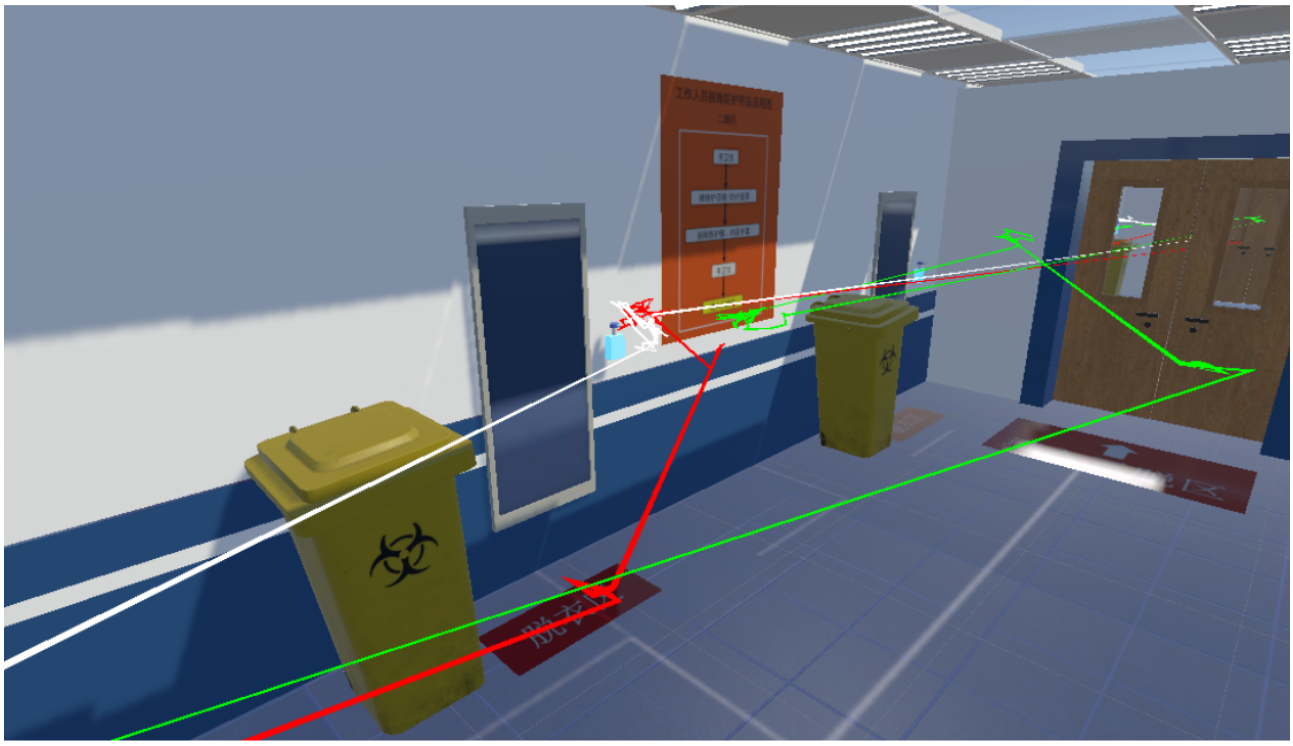
XSL•FO
RenderX

After the training session, the recorded trajectories of head positions can be visualized for analysis as shown in Figure 5. It can be seen that trajectories vary for different users quite a lot. This is probably due to their proficiency degrees and personal preferences.

Figure 5. Visualization of trajectories of users in (A) the donning zone and (B) the first doffing zone after a training session.



A



B

XSL•FO
RenderX

Pilot Study

We evaluated the effectiveness of our method in a controlled randomized trial.

Study Design

We used a between-subject design to test 2 training conditions: our immersive PPE training (VR group), and the instructional video training (video group). To compensate for the weight and experience of a head-mounted display (HMD), both training methods are tested through HMDs [29,30].

We hypothesize that "the VR group has a comparable accuracy of memorizing the PPE donning and doffing event orders to the video group, but the usability and subjective ratings of the VR group are higher than the video group."

Ethical Considerations

The study was approved by the ethics review board of Peking University under the approval number IRB00001052-23206. The ethics approval covered all aspects of the study, including participant recruitment, risk management, confidentiality, data collection, and analysis. This study did not measure any direct health outcomes and thus did not require registration.

Participants were required to fill out an informed consent form prior to the study. The form outlined the procedures, potential risks, and benefits of participation, and stated that participation was voluntary and that participants could withdraw from the study at any time. Data were stored anonymously in a local repository following the data security protocol. Participants received a monetary reward of RMB ¥30 (approximately US $4) regardless of whether they completed the study successfully.

Stimulus and Apparatus

We created an instructional video of approximately 5 minutes as no instructional video was readily available for the PPE training in the hospital of the 2022 Winter Olympic Games. We edited parts of a PPE donning and doffing training video released by the National Health Commission to the same order as the VR group; PPE doffing zoning was instructed using flowcharts that are identical to those on walls in the virtual and real environments. An HTC VIVE Pro Eye HMD and an accompanying PC with Intel i9 3.5GHz CPU, NVidia RTX 3080 GPU, and 64GB RAM were used for the study.

Tasks

The task for participants was to memorize the event orders of donning and doffing and answer a questionnaire about details of the ordering after the training with the assigned condition. Subjective evaluations of the test condition were acquired through the National Aeronautics and Space Administration Task Load Index (NASA-TLX) [31], and the System Usability Scale (SUS) [32] questionnaires.

To assess memory of the sequence, the questionnaire contains 10 multiple-choice questions on the order of PPE and hand hygiene operations as follows.

1. Which item of protective equipment should be donned first?
2. After donning the inner gloves, what is the next piece of equipment to wear?
3. What is the next step after donning the suit?
4. In which area does the process of doffing the suit commence?
5. At which stage in the donning process are goggles worn?
6. When are boot covers typically donned in the sequence?
7. What is the last PPE to be removed within the first doffing area?
8. Which piece of equipment is typically removed first in the first doffing area?
9. Which item is removed in the final doffing area?
10. After removing the goggles and the face shield in the second doffing area, which piece of equipment should be doffed next?

Each question is worth 10 points and has 4 choices: 1 correct answer and 3 distractors, and the total score of the memorization questionnaire is 100.

Participants

The number of participants, that is, the sample size n, of one group was determined using the 2-tailed power analysis:

$$n=2\sigma 2(Z1-\alpha+Z1-\beta)2\delta 2$$

Here, the test significance α is set to 0.05, and power 1-β is set to 80%. Therefore, $Z1-\alpha=Z0.95=1.96$, and $Z1-\beta=Z0.8=0.84$. We estimated that the VR group had a slightly higher mean value than the video group in tests but the variance in participants made both groups comparable. Therefore, we guessed that the scores are close, for example, 80/100 versus 70/100 making $\delta=(80-70)/100=0.1$, and an estimated variance of $\sigma=10/100=0.1$ was used. As a result, n is 15.68 which rounds to 16. Therefore, a total of 32 participants were needed for our experiment which is in line with typical VR or mixed reality studies in medical education [19,33,34].

Participants were recruited via posters published on web forums of the institute (Peking University Health Science Center) of the first author. Participants were required to be medical students who had no previous experience with PPE training and be able to perform actions as directed by the research team. Additionally, they were required to have normal or corrected-to-normal visual acuity and adequate head and body mobility, ensuring that physical conditions did not adversely affect the experimental process. We successfully recruited 32 volunteers (16 females) aged between 18 and 30 years (mean 22.53, SD 2.65) for the study.

Study Procedure

Before the study, each participant received a comprehensive overview of the project and provided informed consent. The participants had to sign the consent to confirm their willingness to participate in the experiment. Participants were told that they could quit the study at any time, but none of them did.

Upon agreeing to participate, each participant was randomly assigned to either the VR group or the video group. The study began with an introductory training session of approximately 10 minutes for participants to familiarize themselves with the necessary user interactions. Those in the video group were instructed on how to control video playback, pause, and adjust the progress bar using the controllers while wearing the HMD

XSL•FO
RenderX

(HTC Vive Eye Pro), ensuring they completed the entire instructional video. For participants in the VR group, initial guidance via a slideshow presentation was given, followed by hands-on training within the VR setting, where they learned to navigate spatially with the controllers and perform the motions of donning and doffing PPE. A researcher offered on-site direction to assist participants through the entire sequence of the experiment—from entering the donning area to exiting the doffing zones.

Next, a free practice session of 10 minutes began, and participants were told to practice and memorize the sequence of handling PPE as well as they could. The video learning group was granted the ability to rewatch instructional videos to reinforce their retention. In contrast, the VR group was engaged in interactive "think-aloud" exercises within the immersive environment, where they independently carried out the PPE donning and doffing, but could ask any questions to the researcher. Among these questions, most were about interactions in VR, for example, how to use the controllers to check the order list, how to don shoe covers, and a few were about the event order mechanism in the simulation, but none was about the donning and doffing process itself.

Next, participants were asked to complete 1 objective questionnaire and 2 subjective questionnaires to evaluate the effectiveness of the training. Finally, a brief interview session was conducted to elicit authentic feedback and valuable suggestions from the participants regarding their experience. The average completion time of the study by a participant was 30 minutes. Each participant was compensated with 30 RMB (approximately US $4).

Data Collection

We collected the trajectory data, questionnaires, and useful comments from the participants.

The trajectory data are potentially useful for analyzing the behavior of participants, and improving the arrangement of items and the allocation of areas for a more efficient workflow. The positional data of the HMD and both hand controllers of each user were recorded to construct trajectory profiles, at a sampling rate of 2 Hz to prevent latency during the VR training session. In addition to the questionnaire, we collected basic demographics of participants and their prior experience with VR.

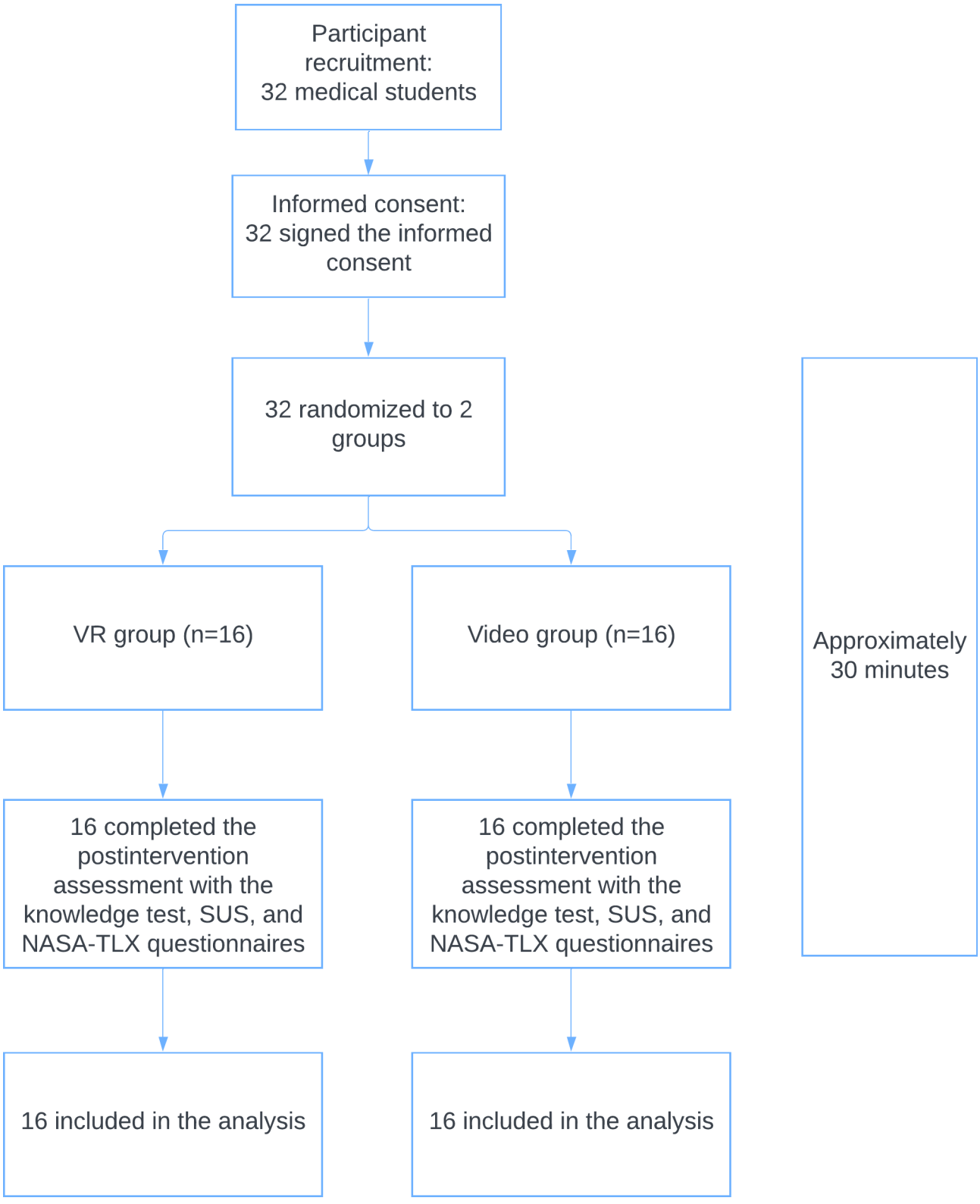
Results

Overview

The flow diagram of the trial is shown in Figure 6. The performance of the PPE event order memorization task, usability, task workload, and virtual trajectories are analyzed for the study. Results of the study are summarized in Table 1 and visualized as boxplots as shown in Figure 7.

XSL•FO
RenderX

Figure 6. Trial flow chart. NASA-TLX: National Aeronautics and Space Administration; SUS: System Usability Scale.



XSL•FO
RenderX

Table . Results of questionnaires of the pilot study. *P* values of Student *t* test between the 2 groups are also reported.

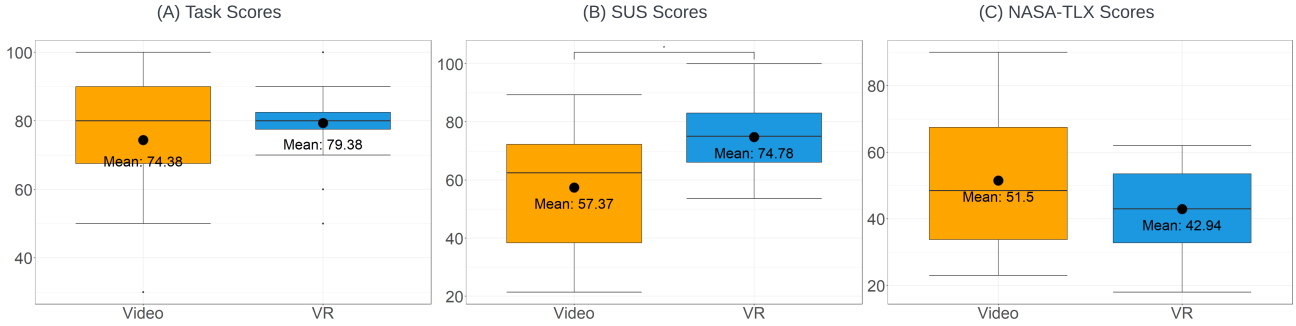
	VR[a] (N=16), mean (SD)	Video (N=16), mean (SD)	*P* value (*t* test)
Task scores	79.38 (12.89)	74.38 (17.88)	.37
SUS[b]	74.78 (13.58)	57.37 (21.13)	.009
NASA-TLX[c]	42.94 (13.01)	51.50 (20.44)	.17

[a]VR: virtual reality.

[b]SUS: System Usability Scale.

[c]NASA-TLX: National Aeronautics and Space Administration Task Load Index.

Figure 7. Boxplots of (A) the task scores, (B) SUS scores, and (C) NASA-TLX. NASA-TLX: National Aeronautics and Space Administration Task Load Index; SUS: System Usability Scale.



Event Order Memorization

Descriptive statistics of the event order memorization questionnaire are shown in the boxplot in Figure 7A. The VR group has a higher mean, equal median, and lower variance compared with the video group.

A Shapiro-Wilk normality test shows that both groups follow the assumption of normality (P=.17 for the VR group, P=.08 for the video group). Levene test ($F_{1, 30}$ =2.258, P=.14) shows that the equal variance assumption is met. The intergroup differences are then tested with a Student t test that shows no significance (P=.37).

System Usability Scale

The SUS scores of the VR group have a mean of 74.78 (SD 13.58) and a rather small variance. The boxplot of SUS scores can be seen in Figure 7B. The VR group has a higher mean and median, and a smaller variance, compared with the video group.

Both groups satisfy the normal distribution criteria with the Shapiro-Wilk test (W(16)=0.961, P=.67 for the VR group, and W(16)=0.926, P=.21 for the video group). Levene test demonstrates no significant disparity between the 2 groups ($F_{1, 30}$=2.990, P=.09), therefore, it is reasonable to deduce that variances are equivalent across the groups. A subsequent t test shows a statistical significance (P=.009) between the 2 groups.

NASA-TLX

The NASA-TLX result is shown in Figure 7C: the VR group has a lower mean and median, and a smaller variance compared with the video group. The NASA-TLX scores of both groups follow the normal distribution criteria after the Shapiro-Wilk test (W(16)=0.956, P=.60 and W(16)=0.962, P=.69, for VR and video groups, respectively), and show no significant differences in variance with the Levene test ($F_{1, 30}$=3.752, P=.06). No significant difference between the groups is found with the t test (P=.17).

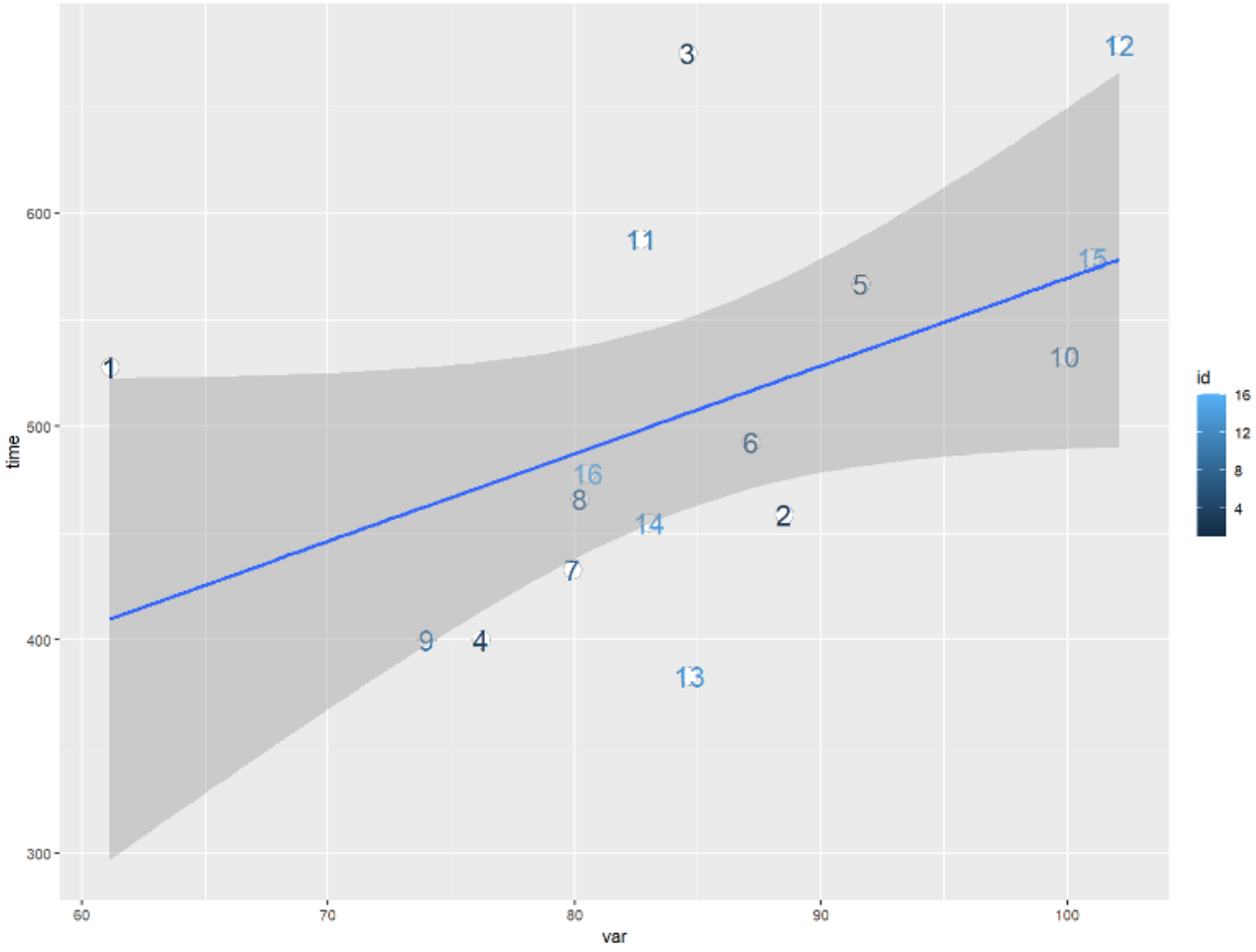
Trajectory Analysis

For quantitative analysis, we calculated the sum of the gradient magnitude of a trajectory as the total variance to describe the degree of fluctuation. A scatterplot of total variance versus task completion time is shown in Figure 8, where the user ID is drawn on data points. A positive correlation (Pearson correlation coefficient=0.47) is observed between the variance and time as shown in the blue fitted line in Figure 8.

Moreover, we visualize the actual trajectory of each user within the Unity scene, as shown in Figure 9. It can be seen that the trajectories are more cluttered in the donning area (Figure 9A) than in the doffing areas (Figure 9B-D). We identified some trajectories that are long and with few fluctuates (colored), and these correspond to participants who were fluent in VR.

XSL•FO
RenderX

Figure 8. The scatterplot of the total variance (var) of trajectories and task completion time (time) with the linearly fitted trend.



XSL•FO
RenderX

Figure 9. Visualization of trajectories of the participants as spaghetti plots.

XSL•FO
RenderX

Discussion

Main Findings

The VR group has a higher mean task score and a smaller variance than the video group. However, the *t* test on order task scores of the 2 groups is not significant.

One possible explanation for the similar learning outcomes is that participants in both groups used strategies to enhance their learning intentionally, as they were told to memorize the event order. Participants in the video group repeatedly watched clips of unfamiliar sequences, while those in the VR group read and memorized the order presented on the UI and the instructions displayed on the walls. Another factor is that the advantages of the immersiveness of the VR group and the familiarity of the real spatial settings thereof may not have been fully captured through the questionnaire.

For usability, it is important that the mean SUS score of the VR group (74.78, SD 13.58) exceeds the threshold of generally accepted systems [35] indicating above-average usability. This is because the mechanism and interactions of the simulation were iteratively designed and tested for usability.

The VR group also has significantly better usability than the video group. Our explanation is that in VR, the relatively small buttons of the video player had to be clicked remotely through rays cast from controllers require high accuracy. Compared with the ease of using a mouse or keyboard outside of VR, the usability is low. In addition, participants in the VR group were limited to simple interactions, which made them feel less engaged and resulted in lower subjective usability ratings. However, we also note that the simplicity of interactions in the video group makes its usability assessment of only secondary importance.

With the NASA-TLX scores, it can be inferred that the psychological workload or task load required by participants for both training methods was roughly equal for the case of memorizing donning and doffing event orders. The smaller variance suggests that the VR experience was more consistent across participants compared with the video group. The higher mean and larger variance observed in the video group can be attributed to their less engaging scenario and fewer interactions. Participants in this group focused only on memorizing the order information, perceiving the study mainly as a test. This focus may have induced more pressure compared with the VR group, which could lead to an increased perceived workload and, consequently, a higher mean. Additionally, the reactions of participants under pressure were more varied when compared with those in the VR group.

For the analysis of trajectories of the VR group, the positive correlation between total variance and time suggests that participants who tended to move around spent more time to complete. The visualization of trajectories reveals that participants took a relatively long time to get familiar with the environment, searched for PPE items and donned them, and moved back and forth to use the hand sanitizer. Unlike in the real environment, the simulation faced a number of design and implementation challenges: the sensitivity of collision detection required participants sometimes to make several attempts in the simulation to interact with PPE items; also, the mechanism does not support the participants to collect all PPE items at once and don them in one place in front of the mirror. We believe that all these factors contribute to the cluttering in the donning area.

Our trajectory visualization does not only provide insights into the simulation but also, more importantly, identifies potential issues in the real environment and helps nosocomial infection control experts make improvements. For instance, large labels of the donning order of each PPE, for example, 1 for the cap in our case, can be clearly placed on the shelf and the wall behind it in the donning area to reduce the searching time. Pictograms of the donning and doffing orders can be placed on walls in complement to text instructions to reduce the cognitive load.

Given the study results and the analysis, we conclude that the VR group has an above-average usability which is significantly better than the video group, and both groups have comparable donning and doffing event order memorization task scores, and task workload. Therefore, our hypothesis can be accepted.

Feedback

During the training, all participants noted that their initial focus was on familiarizing themselves with the interactions. It was only during the free practice phase that they shifted their attention to memorizing the sequence of donning and doffing. This suggests a need to reinforce operational guidance in the early training phase and to emphasize memory training exercises in the subsequent practice sessions. Most VR participants (14/16) indicated that they relied on the left-hand UI guidance bar to remember the sequence of steps. However, 2 participants suggested that providing a flow diagram in the donning area, similar to that in the doffing areas, could potentially improve the efficiency of memorization.

The majority of participants in the VR group reported a superior experience to those of the video group, highlighting increased engagement, satisfaction, and interest provided by this gamification method. They agreed that the spatial context in the virtual environment can improve their confidence in task performance and reduce the adaptation time in a real hospital. They also commented that the customizable donning and doffing order makes our VR method a generic and flexible tool that can adapt to a wide range of PPE training scenarios of varying prevention levels.

Limitations

Several limitations are identified for this study. The use of hand-held controllers instead of hand-tracking added to the overhead of the VR method and reduced its efficiency. The allocation to different groups may potentially introduce a bias in the satisfaction, and performance of participants.

The mean values of test scores and task load of the VR group were both better than those of the video group, but they were not of statistical significance suggesting that the selected sample size may be increased. A larger sample size would be helpful to show the existence or absence of significance of differences in task scores and task loads.

XSL•FO
RenderX

Our VR method does not currently support the creation of a customizable immersive space with room modules, for example, having fewer or more doffing areas. Simulating key steps during donning or doffing PPE with high-precision tracking of movements is not possible with our method. Finally, our method does not yet simulate emergent contamination situations for more comprehensive PPE training.

Conclusions and Future Work

We have presented a PPE training method in VR with a flexible donning and doffing ordering and spatial context. In immersive virtual environments that resemble the relevant configurations of an actual hospital, trainees can practice customized PPE donning and doffing in the corresponding zones to aid the memorization of the sequence and familiarize them with the hospital settings. With a pilot study (N=32), the method is compared against the video-based training in a randomized controlled trial. Results suggest that our new method yields a comparable memorization performance and cognitive load to the video training, but higher SUS scores, satisfaction, and engagement. Our method provides an immersive experience that is not possible with video training. It is important to note that our work does not intend to replace but to complement the existing training. Hands-on PPE donning and doffing in the real world is irreplaceable for successful infection prevention and control.

In the future, we will extend the functionalities of our VR method in several ways and conduct comparative experiments to study its effectiveness. We will extend our tool for the training of hazardous operations under personal protection for different roles in a hospital. We would like to cover more hospital areas and reconstruct realistic scenes captured using neural radiance field techniques [36] for improved realism. We also plan to support the training for detailed donning and doffing actions using real motions instead of controllers with the aid of an optical motion tracking system. For comparative studies in the future, we would like to include larger and more diverse participant groups.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (grant 62372012), the Clinical Medicine Plus X - Young Scholars Project of Peking University (PKU2024LCXQ028), and the Fundamental Research Funds for the Central Universities.

Conflicts of Interest

None declared.

Checklist 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File, 1070 KB - games_v13i1e69021_app1.pdf]

References

1. Kirkland KB, Briggs JP, Trivette SL, Wilkinson WE, Sexton DJ. The impact of surgical-site infections in the 1990s: attributable mortality, excess length of hospitalization, and extra costs. Infect Control Hosp Epidemiol 1999 Nov;20(11):725-730. [doi: 10.1086/501572] [Medline: 10580621]
2. Haque M, Sartelli M, McKimm J, Abu Bakar M. Health care-associated infections - an overview. Infect Drug Resist 2018;11:2321-2333. [doi: 10.2147/IDR.S177247] [Medline: 30532565]
3. Pittet D, Allegranzi B, Sax H, et al. Evidence-based model for hand transmission during patient care and the role of improved practices. Lancet Infect Dis 2006 Oct;6(10):641-652. [doi: 10.1016/S1473-3099(06)70600-4]
4. Stone PW, Pogorzelska-Maziarz M, Herzig CTA, et al. State of infection prevention in US hospitals enrolled in the National Health and Safety Network. Am J Infect Control 2014 Feb;42(2):94-99. [doi: 10.1016/j.ajic.2013.10.003]
5. John A, Tomas ME, Cadnum JL, et al. Are health care personnel trained in correct use of personal protective equipment? Am J Infect Control 2016 Jul 1;44(7):840-842. [doi: 10.1016/j.ajic.2016.03.031] [Medline: 27181222]
6. Haegdorens F, Franck E, Smith P, Bruyneel A, Monsieurs KG, Van Bogaert P. Sufficient personal protective equipment training can reduce COVID-19 related symptoms in healthcare workers: a prospective cohort study. Int J Nurs Stud 2022 Feb;126:104132. [doi: 10.1016/j.ijnurstu.2021.104132] [Medline: 34890835]
7. Saraswathy T, Nalliah S, Rosliza AM, et al. Applying interprofessional simulation to improve knowledge, attitude and practice in hospital- acquired infection control among health professionals. BMC Med Educ 2021 Dec;21(1):482. [doi: 10.1186/s12909-021-02907-1]
8. Notice on improving the prevention and control of infection in fever outpatient and medical institutions [Report in Chinese]. : National Health Commission of the People's Republic of China; 2020 URL: http://www.nhc.gov.cn/yzygj/s3573d/202006/4e456696ceef482996a5bd2c3fb4c3db.shtml [accessed 2025-03-17]
9. Chang Y, Ai Z, Wargocki P, Liu Y, Hu Y. Design of convertible patient care unit for both non-pandemic and pandemic times: prototype, building spatial layout, and ventilation design. Build Environ 2024 Jun;258:111597. [doi: 10.1016/j.buildenv.2024.111597]

XSL•FO
RenderX

10. Mumma JM, Durso FT, Ferguson AN, et al. Human factors risk analyses of a doffing protocol for ebola-level personal protective equipment: mapping errors to contamination. Clin Infect Dis 2018 Mar 5;66(6):950-958. [doi: 10.1093/cid/cix957] [Medline: 29471368]
11. Díaz-Guio DA, Ricardo-Zapata A, Ospina-Velez J, Gómez-Candamil G, Mora-Martinez S, Rodriguez-Morales AJ. Cognitive load and performance of health care professionals in donning and doffing PPE before and after a simulation-based educational intervention and its implications during the COVID-19 pandemic for biosafety. Infez Med 2020;28(suppl 1):111-117. [Medline: 32532947]
12. Yu K, Roth D, Strak R, et al. Mixed reality 3D teleconsultation for emergency decompressive craniotomy: an evaluation with medical residents. Presented at: 2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR); Oct 16-20, 2023; Sydney, Australia. [doi: 10.1109/ISMAR59233.2023.00081]
13. Chheang V, Schott D, Saalfeld P, et al. Advanced liver surgery training in collaborative VR environments. Comput Graph 2024 Apr;119:103879. [doi: 10.1016/j.cag.2024.01.006]
14. Heinrich F, Joeres F, Lawonn K, Hansen C. Comparison of projective augmented reality concepts to support medical needle insertion. IEEE Trans Visual Comput Graphics 2019;25(6):2157-2167. [doi: 10.1109/TVCG.2019.2903942]
15. Heinrich F, Schwenderling L, Joeres F, Lawonn K, Hansen C. Comparison of augmented reality display techniques to support medical needle insertion. IEEE Trans Visual Comput Graphics 2020;26(12):3568-3575. [doi: 10.1109/TVCG.2020.3023637]
16. Allgaier M, Huettl F, Hanke L, et al. LiVRSono - virtual reality training with haptics for intraoperative ultrasound. Presented at: 2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR); Oct 16-20, 2023; Sydney, Australia.
17. Zhao G, Fan M, Yuan Y, Zhao F, Huang H. The comparison of teaching efficiency between virtual reality and traditional education in medical education: a systematic review and meta-analysis. Ann Transl Med 2021 Feb;9(3):252. [doi: 10.21037/atm-20-2785] [Medline: 33708879]
18. Kavanagh S, Luxton-Reilly A, Wuensche B, Plimmer B. A systematic review of virtual reality in education. Theme Sci Technol Edu 2017;10:85-119 [FREE Full text]
19. Birrenbach T, Zbinden J, Papagiannakis G, et al. Effectiveness and utility of virtual reality simulation as an educational tool for safe performance of COVID-19 diagnostics: prospective, randomized pilot trial. JMIR Serious Games 2021 Oct 8;9(4):e29586. [doi: 10.2196/29586] [Medline: 34623315]
20. Buyego P, Katwesigye E, Kebirungi G, et al. Feasibility of virtual reality based training for optimising COVID-19 case handling in Uganda. BMC Med Educ 2022 Dec;22(1):274. [doi: 10.1186/s12909-022-03294-x]
21. Yu M, Yang M, Ku B, Mann JS. Effects of virtual reality simulation program regarding high-risk neonatal infection control on nursing students. Asian Nurs Res (Korean Soc Nurs Sci) 2021 Aug;15(3):189-196. [doi: 10.1016/j.anr.2021.03.002]
22. Yu M, Yang MR. Effectiveness and utility of virtual reality infection control simulation for children with COVID-19: quasi-experimental study. JMIR Serious Games 2022;10(2):e36707. [doi: 10.2196/36707]
23. Masson C, Birgand G, Castro-Sánchez E, et al. Is virtual reality effective to teach prevention of surgical site infections in the operating room? Study protocol for a randomised controlled multicentre trial entitled VIP room study. BMJ Open 2020 Jun 21;10(6):e037299. [doi: 10.1136/bmjopen-2020-037299] [Medline: 32565477]
24. Omori K, Shigemoto N, Kitagawa H, et al. Virtual reality as a learning tool for improving infection control procedures. Am J Infect Control 2023 Feb;51(2):129-134. [doi: 10.1016/j.ajic.2022.05.023] [Medline: 35659561]
25. Kakdas YC, Demirel D, Barker JR, et al. Virtual reality-based donning and doffing simulator. Presented at: 2024 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW); Mar 16-21, 2024; Orlando, FL, USA. [doi: 10.1109/VRW62533.2024.00043]
26. Kakdas YC, Kockara S, Halic T, Demirel D. Enhancing medical training through learning from mistakes by interacting with an ill-trained reinforcement learning agent. IEEE Trans Learn Technol 2024;17:1248-1260. [doi: 10.1109/tlt.2024.3372508] [Medline: 39431279]
27. Kravitz MB, Dadario NB, Arif A, et al. The comparative effectiveness of virtual reality versus e-module on the training of donning and doffing personal protective equipment: a randomized, simulation-based educational study. Cureus 2022 Mar;14(3):e23655. [doi: 10.7759/cureus.23655] [Medline: 35510011]
28. Tsukada K, Yasui Y, Miyata S, et al. Effectiveness of virtual reality training in teaching personal protective equipment skills. JAMA Netw Open 2024 Feb 5;7(2):e2355358. [doi: 10.1001/jamanetworkopen.2023.55358]
29. Skreinig LR, Kalkofen D, Stanescu A, et al. guitARhero: interactive augmented reality guitar tutorials. IEEE Trans Visual Comput Graphics 2023;29(11):4676-4685. [doi: 10.1109/TVCG.2023.3320266]
30. Haltner P, Goddy-Worlu R, Forren J, Nicholas C, Reilly D. A comparative evaluation of AR embodiments vs. videos and figures for learning bead weaving. Presented at: 2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR); Oct 16-20, 2023; Sydney, Australia.
31. Brooke J. SUS: A “quick and dirty” usability scale. In: Usability Evaluation In Industry: CRC Press; 1996:189-194.
32. Hart SG, Staveland LE. Development of NASA-TLX (task load index): results of empirical and theoretical research. In: Hancock PA, Meshkati N, editors. Human Mental Workload: North-Holland; 1988, Vol. 52:139-183. [doi: 10.1016/S0166-4115(08)62386-9]

XSL•FO
RenderX

33. Bogomolova K, van der Ham IJM, Dankbaar MEW, et al. The effect of stereoscopic augmented reality visualization on learning anatomy and the modifying effect of visual‐spatial abilities: a double‐center randomized controlled trial. Anatomical Sciences Ed 2020 Sep;13(5):558-567. [doi: 10.1002/ase.1941]
34. Ho S, Liu P, Palombo DJ, Handy TC, Krebs C. The role of spatial ability in mixed reality learning with the HoloLens. Anat Sci Educ 2022 Nov;15(6):1074-1085. [doi: 10.1002/ase.2146] [Medline: 34694737]
35. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. Int J Hum Comput Interact 2008 Jul 29;24(6):574-594. [doi: 10.1080/10447310802205776]
36. Mildenhall B, Srinivasan PP, Tancik M, Barron JT, Ramamoorthi R, Ng R. NeRF: representing scenes as neural radiance fields for view synthesis. Presented at: Computer Vision – ECCV 2020: 16th European Conference; Aug 23-28, 2020; Glasgow, United Kingdom.

Abbreviations

HMD: head-mounted display
NASA-TLX: National Aeronautics and Space Administration Task Load Index
PPE: personal protection equipment
R1: requirement 1
R2: requirement 2
R3: requirement 3
SUS: System Usability Scale
UI: user interface
VR: virtual reality

Edited by A Coristine, K Erenli; submitted 20.11.24; peer-reviewed by D Demirel, X Cheng; revised version received 12.02.25; accepted 20.02.25; published 20.03.25.

Please cite as:
Zhou L, Liu H, Fan M, Liu J, Yu X, Zhao X, Zhang S
Personal Protection Equipment Training as a Virtual Reality Game in Immersive Environments: Development Study and Pilot Randomized Controlled Trial
JMIR Serious Games 2025;13:e69021
URL: https://games.jmir.org/2025/1/e69021
doi:10.2196/69021

© Liang Zhou, Haoyang Liu, Mengjie Fan, Jiahao Liu, Xingyao Yu, Xintian Zhao, Shaoxing Zhang. Originally published in JMIR Serious Games (https://games.jmir.org), 20.3.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Impact of a 3-Month Recall Using High-Fidelity Simulation or Screen-Based Simulation on Learning Retention During Neonatal Resuscitation Training for Residents in Anesthesia and Intensive Care: Randomized Controlled Trial

Anne-Claire Louvel[1*], MD; Cécile Dopff[1], MD; Gauthier Loron[2,3], Prof Dr Med; Daphne Michelet[4,5*], Prof Dr Med

[1]Department of Anesthesia and Intensive Care, Hôpital Bichat-Claude-Bernard, Paris, France

[2]Department of Anesthesia and Intensive Care, Centre Hospitalier Universitaire de Reims, Reims, France

[3]CReSTIC EA 3804, Université de Reims Champagne-Ardenne, Reims, France

[4]Department of Anesthesia and Intensive Care, Centre Hospitalier Universitaire de Reims, 45 rue Cognacq-Jay, Reims, France

[5]C2S laboratory, Université de Reims Champagne-Ardenne, Reims, France

[*]these authors contributed equally

Corresponding Author:
Daphne Michelet, Prof Dr Med
Department of Anesthesia and Intensive Care, Centre Hospitalier Universitaire de Reims, 45 rue Cognacq-Jay, Reims, France

Abstract

Background: Retention capacities are dependent on the learning context. The optimal interval between two learning sessions to maintain a learner's knowledge is often a subject of discussion, along with the methodology being used. Screen-based simulation could represent an easy alternative for retraining in neonatal resuscitation.

Objective: The aim of the study was to evaluate the benefits of a 3-month recall session using high-fidelity simulation or screen-based simulation, assessed 6 months after an initial neonatal resuscitation training session among anesthesia and intensive care residents.

Methods: All participating anesthesia and intensive care residents were volunteers, and they underwent training in the same session, which included a theoretical course and high-fidelity simulation. The attendees were then randomized into three groups: one with no 3-month recall, one with a high-fidelity simulation recall, and one with a screen-based simulation recall. To reassess the skills of each participant, a high-fidelity simulation was performed at 6 months. The primary outcomes included expert assessment of technical skills using the Neonatal Resuscitation Performance Evaluation score and nontechnical skills assessed by the Anesthesia Non-Technical Skills score. Secondary outcomes included a knowledge quiz and self-assessment of confidence. We compared the results between groups and analyzed intragroup progressions.

Results: Twenty-eight participants were included in the study. No significant differences were observed between groups at the 6-month evaluation. However, we observed a significant improvement in theoretical knowledge and self-confidence among students over time. Regarding nontechnical skills, as evaluated by the Anesthesia Non-Technical Skills score, there was significant improvement between the initial training and the 6-month session in both recall groups (16 vs 12.8, P=.01 in the high-fidelity group; 16 vs 13.9, P=.05 in the simulation group; 14.7 vs 15.1, P=.50 in the control group). For technical skills assessed by the Neonatal Resuscitation Performance Evaluation score, a nonsignificant trend toward improvement was observed in the two recall groups, while a regression was observed in the control group (all Ps>.05). The increase in students' self-confidence was significant across all groups but remained higher in the two 3-month recall groups.

Conclusions: Initial neonatal resuscitation training for anesthesia and intensive care residents leads to improved knowledge and self-confidence that persist at 6 months. A 3-month recall session, whether through high-fidelity simulation or screen-based simulation, improves nontechnical skills (eg, situation management and team communication) and technical skills. Screen-based simulation, which saves time and resources, appears to be an effective educational method for recall after initial training. The study outcomes justify the need for further studies with larger sample sizes to confirm the promising role of serious games in educational programs for medical students.

(JMIR Serious Games 2025;13:e57057) doi:10.2196/57057

XSL•FO
RenderX

KEYWORDS

screen-based simulation; high-fidelity simulation; neonatal resuscitation; pediatric; infant; neonatal; newborns; emergency; urgent; simulation; resuscitation; intensive care; medical education; anesthesia; anesthesiology; high fidelity; educational; student; resident; knowledge retention; learner; teaching; intensive care unit; ICU

Introduction

Neonatal resuscitation is a potentially critical situation that requires training. Nearly 10% of newborns and 80% of infants weighing less than 1500 g require resuscitation at birth, and the quality of care provided during the first minute of life is directly related to the prognosis [1-3]. Theoretical knowledge based on current recommendations and practical training are key for ensuring optimal neonatal resuscitation. Anesthesia and intensive care physicians may need to intervene when a pediatrician is not immediately available, for instance, to assist with intubation. Therefore, we decided to train and prepare our residents for such scenarios.

Screen-based simulation is an emerging simulation-based training tool for health care professionals, ensuring comparable learning effectiveness to traditional learning methods [4]. Recent developments in computer science have allowed the creation of highly realistic medical digital simulators, improving the acquisition of both knowledge and nontechnical skills, as well as technical skills [5,6]. A screen-based simulator (NRP eSim), designed by Laerdal Medical in collaboration with the American Academy of Pediatrics is included in the Neonatal Resuscitation Program as one of the six educational components of the 7th edition of the Neonatal Resuscitation Program training program [7]. In the neonatal resuscitation context, technical skills refer to diagnostic and therapeutic actions, whereas nontechnical skills include leadership, teamwork, communication, and task management.

Few studies have already assessed knowledge retention following screen-based simulation training. While results are relatively positive when evaluated within one or two months after simulation [8,9], they suggest lower effectiveness compared to traditional learning methods when assessed after 6 months [10].

The question of knowledge retention is still debated, and the ideal interval for refresher training remains uncertain, regardless of the method [11]. Recent studies indicate the need for regular training for complex procedures, such as the management of cardiac arrest with intervals of 3 to 6 months [12], or even monthly [13].

The objective of our study is to analyze the retention of knowledge and skills at 6 months after an initial training in neonatal resuscitation for anesthesia and intensive care residents, with or without a 3-month recall training session using either screen-based simulation or high-fidelity (HF) simulation.

Methods

Ethical Considerations

This randomized controlled simulation study was conducted from February 2021 to November 2021 at the University Hospital of Reims, France. The study was approved by the CERAR (Ethics Committee for Research in Anesthesia Resuscitation; approval number: 00010254-2021-048). All students provided written informed consent (Multimedia Appendix 1). All data were anonymized using research numbers. No compensation was provided to the residents who volunteered to participate in the study. Consent was provided for the video recording and analysis of primary and secondary outcomes. The CONSORT-eHealth checklist is provided (Checklist 1) [14].

Participants

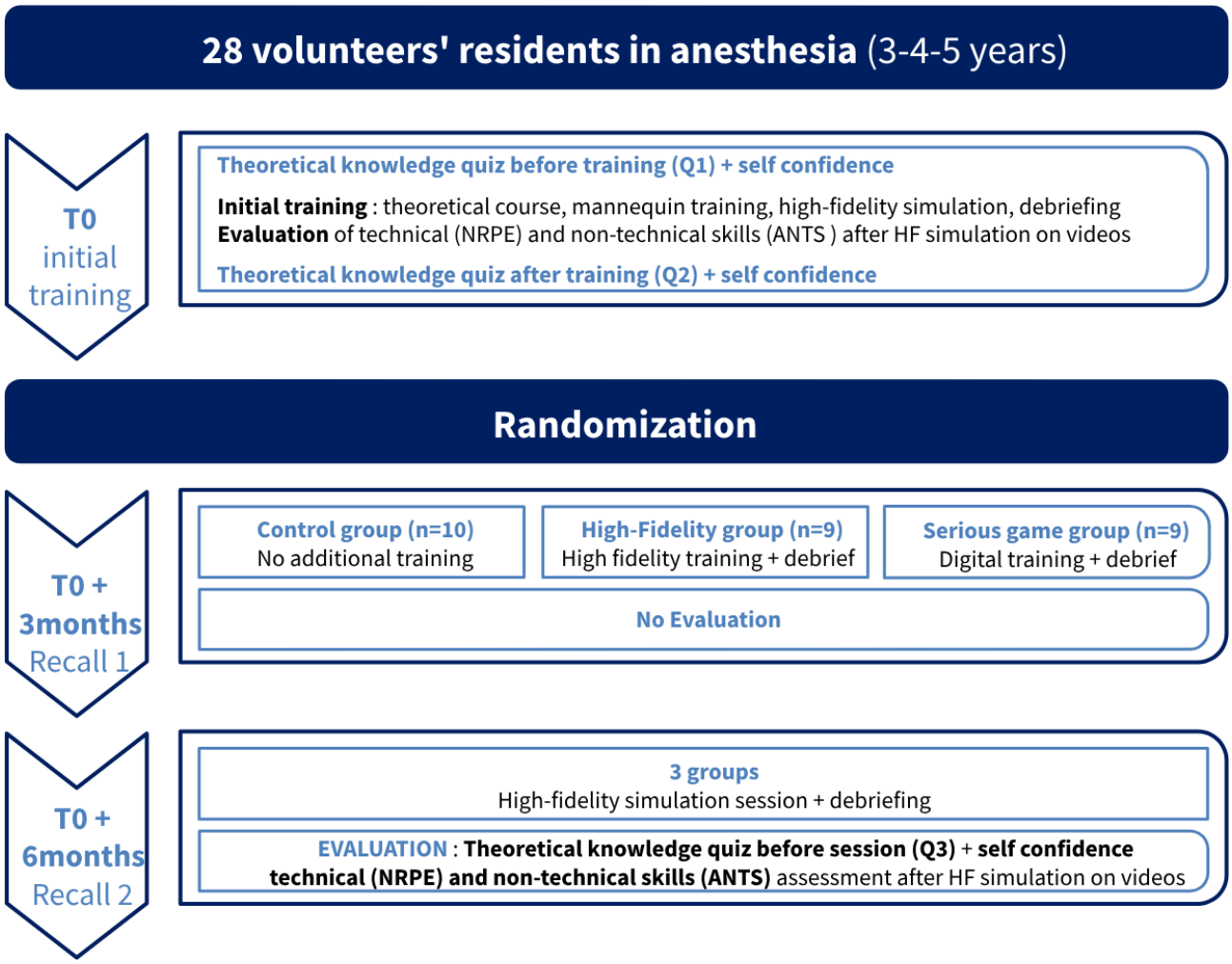
Volunteer participants were recruited from third-, fourth-, and fifth-year anesthesiology and intensive care residency classes in Reims. None had previous courses or training in neonatal resuscitation. They were informed about the study via email from the teaching supervisor (DM). To confirm their participation, they had to ensure that they could be present at all times during the study (ie, 3 and 6 months).

Design of the Study

The experimental design is summarized in Figure 1.

XSL•FO
RenderX

Figure 1. Experimental design. ANTS: Anesthesia Non-Technical Skills; HF: high-fidelity; NRPE: Neonatal Resuscitation Performance Evaluation; T0: time zero.



Initial Training

The initial training was conducted by a pediatric intensivist and an anesthetist. It consisted of a 45-minute course on the basics and essential knowledge required for neonatal resuscitation based on the International Liaison Committee on Resuscitation (ILCOR) 2020 recommendations [15]. Subsequently, the interns practiced ventilation using a low-fidelity mannequin using the Neopuff (Fisher & Paykel Healthcare), intubation, and cardiac massage. Subsequently, they individually participated in an HF simulation that included a presentation of the material, briefing, scenario execution, and debriefing (see scenario details in Multimedia Appendix 2). The simulation was conducted in a real-life delivery room setting at Reims University Hospital using the SimNewBorn mannequin (Laerdal Medical). The simulation sessions were recorded for later analysis.

Residents were randomized into three parallel groups (randomization.com) in a 1:1:1 ratio. The control group did not receive a recall session at 3 months. The two intervention groups received a recall session at 3 months after the initial session using either serious game (SG) screen-based simulation or HF simulation.

Three-Month Recall

HF Simulation

Each participant individually participated in an HF simulation using the SimNewBorn mannequin in the delivery room. The scenario is provided in Multimedia Appendix 2. The session lasted approximately 45 minutes, including briefing, scenario execution, and debriefing with one of the study trainers. The material used was identical to that of the initial training, and participants were given a 5-minute refresher before the simulation.

Screen-Based Simulation

Each participant performed an individual screen-based simulation consisting of a 10-minute tutorial explaining the various possible actions and a scenario. The screen-based simulation, PerinatSims, was designed by Medusims. It features a 3D virtual environment of a delivery room, with a newborn placed on a neonatal resuscitation table (Figure 2). The simulation used a point-and-click interface with a first-person point of view.

The scenario was similar to that used for the HF group (see Multimedia Appendix 2) and was followed by a digital debriefing. The session lasted 45 minutes and was conducted in the presence of a trainer to prevent technical issues from

XSL•FO
RenderX

affecting the research. Participants logged onto the digital simulator on a computer at the hospital, with access granted for the duration of the study.

Figure 2. Screenshot of interface and virtual environment of the PerinatSims screen-based simulation.



Six-Month Evaluation

Overview

The final evaluation for each participant of the study involved an HF simulation, initiated with a briefing and presentation of the material, followed by scenario execution using the SimNewBaby Mannequin, and debriefing. The evaluation scenario is provided in Multimedia Appendix 2. Simulation sessions were recorded for later analysis.

The potential exposure of each participant to real-life neonatal resuscitation cases or additional training during the 6-month study period was controlled and monitored.

Primary End Points: Comparison of Nontechnical and Technical Skills Retention

Two independent, blinded anesthetist raters retrospectively evaluated the technical and nontechnical skills by analyzing video recordings from the initial and final sessions.

Technical skills were assessed by the Neonatal Resuscitation Performance Evaluation (NRPE) scoring system (20 points) adapted for each scenario [16], and nontechnical skills using the Anesthesia Non-Technical Skills (ANTS) observation system, which includes four categories (ie, situation awareness, task management, team work, and decision-making) [17]. ANTS is a validated tool used to assess nontechnical skills in various simulation situations, ranging from emergency training for medical students [18] to neonatal resuscitation for midwives using a specific modified ANTS version [19]. The ANTS scores were recorded as the overall category scores on a scale of 1 - 4 (poor performance: 1 to good performance: 4). The global score (out of 16 points) is presented as a 20-point scale in the Results section, as described by the authors of the ANTS. Interrater reliability calculations were performed.

Secondary End Points

Comparison of Knowledge Retention

Knowledge was assessed by a validated questionnaire containing 25 single or multiple-choice questions based on ILCOR recommendations [15]. Assessments were performed at Q1: beginning of the initial training; Q2: end of the initial training; and Q3: beginning of the 6-month evaluation.

Comparison of the Self-Confidence Auto-Evaluation

The self-confidence question assessed residents' perception of their performance, based on the question, "How much are you confident in your capability to organize and execute a neonatal resuscitation?" The responses were recorded using a 6-point Likert scale, ranging from "not at all confident" (scored as 0) to "very confident" (scored as 5) [20]. Self-confidence was assessed at the end of the initial training and the 6-month evaluation session.

XSL•FO
RenderX

Data Collection

Data were collected directly on paper for the knowledge and self-confidence questionnaires, as well as for clinical and demographic information. The primary outcome measures were assessed via expert evaluation of the video recordings a posteriori.

Statistical Analysis

Based on a previous study, a baseline ANTS score of approximately 12/20 was expected [21], with a target improvement to 16/20 through recall interventions. With an alpha risk of 5% and 80% power, the number of participants required per group was 9.

The statistical hypothesis tested whether SG and HF simulation were equivalent in reactivating skills and knowledge, with both differing from the control group. Data are presented as median (Q1-Q3) for continuous data, given the small sample size. Interrater reliability calculations were performed for both evaluations, showing good agreement between the two raters (kw>0.66; *P*=.01) [22]. The statistical analyses were conducted using the Friedman test and Wilcoxon test for paired samples, and the Kruskal-Wallis test for intergroup comparisons. All tests were two-tailed, with statistical significance considered at *P*<.05. Statistical analyses were performed using SPSS software (version 25.0; IBM Corp).

Results

Participant Characteristics

Twenty-eight participants were included in the study; 10 were randomly assigned to the control group, 9 to the HF simulation 6-month recall group (HF), and 9 to the screen-based simulation 6-month recall group (SG). All were anesthesiology and intensive care residents in their third, fourth, or fifth year of a 5-year training program in France. Seventeen participants (60%) were men, with a median age of 28 years (IQR 27-31). Baseline demographic and clinical characteristics are presented in Table 1. All participants completed the study (CONSORT [Consolidated Standards of Reporting Trials] flow diagram is provided) (Figure 3).

Seventy-five percent (21/28) of the participants had previously participated in training involving simulation. Twenty-one participants were familiar with HF simulation, and 15 had prior experience with a SG, though none had training in neonatal resuscitation. No participants had encountered or participated in a real neonatal resuscitation during the 6-month study period or benefited from other training. The findings are presented in Table 2.

Table . Baseline demographic and clinical characteristics of each group.

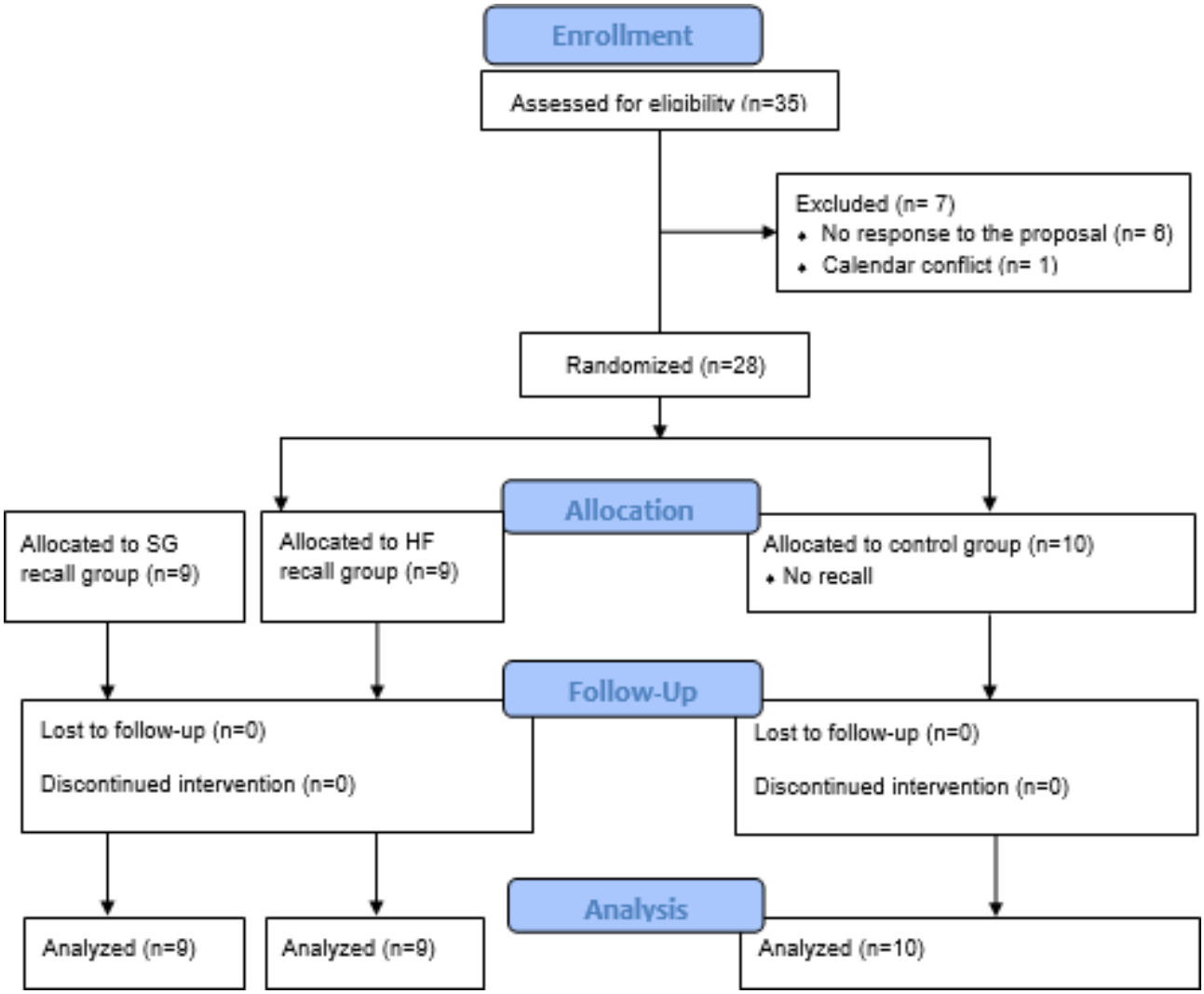
Variables	HF[a] group (n=9)	SG[b] group (n=9)	Control group (n=10)
Age (years), median (IQR)	27.5 (27-29)	30 (29-31)	28 (27-32)
Year of anesthesia training (out of 5), median (IQR)	3.5 (3-5)	4.5 (3-5)	4 (3-5)
Sexes, n (%)			
Female	5 (56)	2 (22)	4 (40)
Male	4 (44)	7 (78)	6 (60)
Previous experience in HF[a] simulation, n (%)	5 (56)	8 (89)	8 (80)
Previous experience in SG[b] simulation, n (%)	4 (44)	5 (56)	6 (60)

[a]HF: High-fidelity.

[b]SG: Serious game screen-based.

XSL•FO
RenderX

Figure 3. CONSORT flowchart of participant enrollment, allocation, follow up, and analysis. CONSORT: Consolidated Standards of Reporting Trials; HF: high-fidelity; SG: serious games.



XSL•FO
RenderX

Table . Comparison of the study parameters within groups using the Friedman and Wilcoxon tests for paired samples.

Participants and end points	Preinitial training, median (IQR)	Postinitial training, median (IQR)	6-month evaluation, median (IQR)	*P* value
Overall (N=28)				
Quiz (20 points)	10 (8-11)	16 (14-16)	13.5 (11.25-15)	.001
Confidence (5 points)	0 (0-1)	3 (2-3)	2 (2-3)	<.001
NRPE[a] (20 points)	—[b]	14.6 (12.6-15.9)	14.4 (12.4-16)	.86
ANTS[c] (20 points)	—[b]	14.5 (12.8-15.7)	15.8 (14.6-17)	.005
HF[d] group (n=9)				
Quiz (20 points)	9 (7-11.5)	16 (15-17.5)	12.5 (11-14.75)	.001
Confidence (5 points)	0 (0-1)	3 (2-3.5)	3 (2-3)	.004
NRPE (20 points)	—	13.3 (12.3-15.9)	14.9 (11.9-17.2)	.50
ANTS (20 points)	—	12.8 (11.9-15.6)	16 (14.8-17.7)	.01
SG[e] group (n=9)				
Quiz (20 points)	9 (8-11)	16 (13.5-17.5)	13 (10.5-15)	.001
Confidence (5 points)	0 (0-1.5)	3 (1.5-3)	2.5 (2-3.75)	.002
NRPE (20 points)	—	14.5 (12.1-15.3)	15 (14.1-15.9)	.33
ANTS (20 points)	—	13.9 (12.8-15.8)	16 (15.4-17.1)	.05
Control group (n=10)				
Quiz (20 points)	10 (8.75-11.25)	15 (14-15)	14 (13.25-14.75)	.001
Confidence (5 points)	0 (0-1)	2.5 (1-3)	2 (1-2.75)	.003
NRPE (20 points)	—	15.7 (14.7-16.5)	12.6 (11.9-13.9)	.21
ANTS (20 points)	—	15.1 (14.7-15.6)	14.7 (13.4-15.6)	.50

[a]NRPE: Neonatal Resuscitation Performance Evaluation.

[b]Not applicable.

[c]ANTS: Anesthesia Non-Technical Skills.

[d]HF: High-fidelity.

[e]SG: Serious game screen-based simulation.

Primary End Points: Comparison of Technical and Nontechnical Skills Retention

Nontechnical Skills (ANTS)

Nontechnical skills, assessed by the ANTS score, showed a significant improvement between the initial training and 6-month session in the two groups that received recall training session.

No significant differences were observed between the groups (*P*=.15).

Technical Skills (NRPE)

Technical skills assessed by the NRPE score indicated a nonsignificant trend toward improvement in the two recall groups, whereas a decline was observed in the control group.

No significant differences were observed between the groups (*P*=.19).

Secondary End Points

Comparison of Knowledge Retention

Knowledge scores improved across all groups between the beginning of the study and the end of the initial training. At 6 months, knowledge decreased in all groups but remained higher than the initial level.

No significant differences were observed between the groups (*P*=.12).

Comparison of Self-Confidence Evaluation

The increase in students' self-confidence was significant in all groups between the initial training and 6-month recall; the increase was more pronounced in the two groups that received a 3-month recall session.

Discussion

Principal Findings

This study highlights the benefits of a recall session 3 months after an initial neonatal resuscitation training for novice

XSL•FO
RenderX

anesthesia and intensive care residents. Regardless of the recall technique—whether an SG or HF simulation—learning retention improved at 6 months.

The importance of knowledge and skill maintenance has been demonstrated several times in medical education, although no ideal interval or optimal method has been defined. A significant loss of skills over time is observed in the absence of continuous training [23,24]. Short, systematic, and regularly repeated training sessions help to optimize the retention of skills, as the consolidation of learning is based on repetition [25,26]. Retention is a marker that is regularly used to assess the quality of learning. The literature demonstrates a significant variability in the duration of retention depending on the type of skills taught and the learner population. In an HF simulation study on the cricothyroidotomy technique involving already qualified anesthetists, retention lasted up to a year after training [27]. This may be attributed to the complexity of the taught task and the high expertise of the learners. However, for a group of novices training in new technical skills, retention can be significantly shorter [28]. Therefore, the teaching format, learners, and the type of skill taught all have an impact on the duration of retention.

Screen-based simulation serves as an alternative to HF simulation or an interesting addition to training programs, as it represents an innovative contribution in line with the expectations of newer generations. It offers flexibility in terms of usage and learning methods—whether individual or group-based, remote or in-person—while preserving a safe environment for immersive and realistic training that is conducive to making mistakes. Additionally, screen-based simulation is less expensive than HF simulation [29]. A recent study examined its implementation to facilitate the development of medical teaching programs using screen-based simulation [30].

Few studies have assessed retention of learning after screen-based simulation training in the medical field. In aviation, one study found better knowledge retention after immersive digital simulator training compared to conventional training tools after one week [31]. In a medical study, the contribution of screen-based simulation was evaluated in comparison to traditional teaching for head and neck physiology and anatomy for second-year medical students. Knowledge retention, assessed through a questionnaire at 6 months, was similar between groups; the traditional teaching group showed progression over time, whereas the screen-based simulation group did not [10]. Another study involving nursing students compared knowledge retention after a 45-minute course on managing respiratory distress, followed by either screen-based simulation or low-fidelity simulation. Screen-based simulation resulted in superior knowledge retention at two months [32]. Additionally, a recent study demonstrated the potential of an online SG as an effective learning tool for improving and retaining knowledge related to the diagnosis and treatment planning of oral lesions, with good knowledge retention observed at one week [33].

Our study is the first to analyze the impact of a 3-month recall session on the retention of learning at 6 months and to compare two recall techniques: screen-based simulation and HF simulation. It is important to emphasize that both recall groups benefitted from identical training durations at 3 months (45 min in total), allowing for a direct comparison of the training methods. Our findings confirm the benefits of a 3-month recall session, with improvement in learning in both recall groups, unlike the control group without recall, in which skills declined.

A 3-month recall appears to enable an improvement in knowledge, particularly in technical and nontechnical skills and participants' self-confidence. The only element that progressed similarly in the control group was the knowledge questionnaire score, though this may be the least reliable criterion, as learners had the opportunity to review the ILCOR algorithm before participating in the 6-month session. Assessing skills is a more discriminating measure of learning retention.

Limitations of the Study

The main limitation of our study is the sample size. The study involved voluntary third-, fourth-, and fifth-year anesthesia and intensive care residents, resulting in a limited number of eligible participants (28 out of 35), which corresponds to the sample size calculation but did not allow us to detect any differences between the groups. Notably, the baseline level of the ANTS score was higher than expected, particularly in the control group. However, our population was homogeneous, consisting of neonatal resuscitation novices, and we were able to follow them over the 6 months of the study.

Although this study is a good reflection of a typical educational program, our findings remain to be confirmed using larger cohorts. Additionally they should be validated using other pedagogical themes to refine the role of screen-based simulation in educational pathways and to assess the effect of different methods on the long-term retention of learning.

Conclusion

A 3-month recall after initial neonatal resuscitation training for novice residents results in improved learning retention at 6 months, regardless of the teaching technique used—numerical or HF simulation. While our findings should be confirmed by future studies with larger sample sizes, they support the value and feasibility of incorporating screen-based simulation into training programs.

Acknowledgments

The authors would like to thank the pediatric intensivists who actively contributed to the study, the delivery room staff who facilitated the high-fidelity simulations, and all voluntary participants.

XSL•FO
RenderX

Data Availability

The datasets used or analyzed during the current study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: GL, DM
Data curation: DM
Formal analysis: ACL, CD, DM
Investigation: ACL, CD, DM
Methodology: GL, DM
Project administration: DM
Resources: DM
Software: DM
Supervision: DM
Validation: ACL, CD
Visualization: DM
Writing – original draft: ACL, CD, DM
Writing – review & editing: ACL, CD, GL, DM

Conflicts of Interest

None declared.

Multimedia Appendix 1
Informed written consent (in French).
[DOC File, 43 KB - games_v13i1e57057_app1.doc]

Multimedia Appendix 2
Simulation scenario.
[DOCX File, 13 KB - games_v13i1e57057_app2.docx]

Checklist 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File, 637 KB - games_v13i1e57057_app3.pdf]

References

1. Ryan CA, Clark LM, Malone A, Ahmed S. The effect of a structured neonatal resuscitation program on delivery room practices. Neonatal Netw 1999 Feb;18(1):25-30. [doi: 10.1891/0730-0832.18.1.25] [Medline: 10205469]
2. Curran VR, Aziz K, O'Young S, Bessell C. Evaluation of the effect of a computerized training simulator (ANAKIN) on the retention of neonatal resuscitation skills. Teach Learn Med 2004;16(2):157-164. [doi: 10.1207/s15328015tlm1602_7] [Medline: 15276892]
3. Pearlman SA, Zern SC, Blackson T, Ciarlo JA, Mackley AB, Locke RG. Use of neonatal simulation models to assess competency in bag-mask ventilation. J Perinatol 2016 Mar;36(3):242-246. [doi: 10.1038/jp.2015.175] [Medline: 26583941]
4. Gentry SV, Gauthier A, L'Estrade Ehrstrom B, et al. Serious gaming and gamification education in health professions: systematic review. J Med Internet Res 2019 Mar 28;21(3):e12994. [doi: 10.2196/12994] [Medline: 30920375]
5. Barré J, Michelet D, Job A, et al. Does repeated exposure to critical situations in a screen-based simulation improve the self-assessment of non-technical skills in postpartum hemorrhage management? Simul Gaming 2019 Apr;50(2):102-123. [doi: 10.1177/1046878119827324]
6. Aster A, Laupichler MC, Zimmer S, Raupach T. Game design elements of serious games in the education of medical and healthcare professions: a mixed-methods systematic review of underlying theories and teaching effectiveness. Adv Health Sci Educ Theory Pract 2024 Nov;29(5):1825-1848. [doi: 10.1007/s10459-024-10327-1] [Medline: 38563873]
7. Sawyer T, Umoren RA, Gray MM. Neonatal resuscitation: advances in training and practice. Adv Med Educ Pract 2017;8:11-19. [doi: 10.2147/AMEP.S109099] [Medline: 28096704]
8. Kanthan R, Senger JL. The impact of specially designed digital games-based learning in undergraduate pathology and medical education. Arch Pathol Lab Med 2011 Jan;135(1):135-142. [doi: 10.5858/2009-0698-OAR1.1] [Medline: 21204720]
9. Barré J, Michelet D, Truchot J, Cabon P, Tesniere A. Midwifery students' retention of learning after screen-based simulation training on neonatal resuscitation: a pilot study. BMJ Simul Technol Enhanc Learn 2021;7(1):31-34. [doi: 10.1136/bmjstel-2019-000525] [Medline: 35521074]

XSL•FO
RenderX

10. Rondon S, Sassi FC, Furquim de Andrade CR. Computer game-based and traditional learning method: a comparison regarding students' knowledge retention. BMC Med Educ 2013 Feb 25;13:30. [doi: 10.1186/1472-6920-13-30] [Medline: 23442203]
11. Au K, Lam D, Garg N, et al. Improving skills retention after advanced structured resuscitation training: a systematic review of randomized controlled trials. Resuscitation 2019 May;138:284-296. [doi: 10.1016/j.resuscitation.2019.03.031] [Medline: 30928503]
12. Monsieurs KG, Nolan JP, Bossaert LL, et al. European Resuscitation Council Guidelines for Resuscitation 2015: Section 1. Executive summary. Resuscitation 2015 Oct;95:1-80. [doi: 10.1016/j.resuscitation.2015.07.038] [Medline: 26477410]
13. Anderson R, Sebaldt A, Lin Y, Cheng A. Optimal training frequency for acquisition and retention of high-quality CPR skills: a randomized trial. Resuscitation 2019 Feb;135:153-161. [doi: 10.1016/j.resuscitation.2018.10.033] [Medline: 30391370]
14. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126. [doi: 10.2196/jmir.1923] [Medline: 22209829]
15. Réanimation néonatale en salle de naissance. Réseaux de santé de Champagne-Ardenne. 2023 Jan 18. URL: https://reseaux-sante-ca.org/IMG/pdf/2021-01-copege_algo_rea_ilcor_2020_avec_aide_et_leader_-_modif_couleurs_avec_num_rpca.pdf [accessed 2025-03-10]
16. van der Heide PA, van Toledo-Eppinga L, van der Heide M, van der Lee JH. Assessment of neonatal resuscitation skills: a reliable and valid scoring system. Resuscitation 2006 Nov;71(2):212-221. [doi: 10.1016/j.resuscitation.2006.04.009] [Medline: 16987590]
17. Fletcher G, Flin R, McGeorge P, Glavin R, Maran N, Patey R. Anaesthetists' Non-Technical Skills (ANTS): evaluation of a behavioural marker system. Br J Anaesth 2003 May;90(5):580-588. [doi: 10.1093/bja/aeg112] [Medline: 12697584]
18. Hagemann V, Herbstreit F, Kehren C, et al. Does teaching non-technical skills to medical students improve those skills and simulated patient outcome? Int J Med Educ 2017 Mar 29;8:101-113. [doi: 10.5116/ijme.58c1.9f0d] [Medline: 28355594]
19. Cavicchiolo ME, Cavallin F, Staffler A, et al. Decision making and situational awareness in neonatal resuscitation in low resource settings. Resuscitation 2019 Jan;134:41-48. [doi: 10.1016/j.resuscitation.2018.10.034] [Medline: 30391367]
20. Roh YS, Lee WS, Chung HS, Park YM. The effects of simulation-based resuscitation training on nurses' self-efficacy and satisfaction. Nurse Educ Today 2013 Feb;33(2):123-128. [doi: 10.1016/j.nedt.2011.11.008] [Medline: 22153054]
21. Michelet D, Barre J, Truchot J, Piot MA, Cabon P, Tesniere A. Effect of computer debriefing on acquisition and retention of learning after screen-based simulation of neonatal resuscitation: randomized controlled trial. JMIR Serious Games 2020 Aug 11;8(3):e18633. [doi: 10.2196/18633] [Medline: 32780021]
22. Randolph JJ. Online kappa calculator [computer software]. Justus Randolph. 2008. URL: http://justusrandolph.net/kappa/ [accessed 2025-03-10]
23. Torrelles Nadal C, Paris Mañas G, Sabrià Bernadó B, Alsinet Mora C. Assessing teamwork competence. Psicothema 2015;27(4):354-361. [doi: 10.7334/psicothema2014.284] [Medline: 26493573]
24. Granry JC. État de l'art (National et International) En Matière de Simulation Dans Le Domaine de La Santé Dans Le Cadre Du Développement Professionnel Continu (DPC) et de La Prévention Des Risques Associés Aux Soins: HAS; 2012. URL: https://www.has-sante.fr/upload/docs/application/pdf/2012-01/simulation_en_sante_-_rapport.pdf [accessed 2025-03-10]
25. McGaugh JL. The perseveration-consolidation hypothesis: Mueller and Pilzecker, 1900. Brain Res Bull 1999;50(5-6):445-446. [doi: 10.1016/s0361-9230(99)00126-4] [Medline: 10643476]
26. McGaugh JL. Memory--a century of consolidation. Science 2000 Jan 14;287(5451):248-251. [doi: 10.1126/science.287.5451.248] [Medline: 10634773]
27. Boet S, Borges BCR, Naik VN, et al. Complex procedural skills are retained for a minimum of 1 yr after a single high-fidelity simulation training session. Br J Anaesth 2011 Oct;107(4):533-539. [doi: 10.1093/bja/aer160] [Medline: 21659406]
28. Moulton CAE, Dubrowski A, Macrae H, Graham B, Grober E, Reznick R. Teaching surgical skills: what kind of practice makes perfect?: a randomized, controlled trial. Ann Surg 2006 Sep;244(3):400-409. [doi: 10.1097/01.sla.0000234808.85789.6a] [Medline: 16926566]
29. Haerling KA. Cost-utility analysis of virtual and mannequin-based simulation. Simul Healthc 2018 Feb;13(1):33-40. [doi: 10.1097/SIH.0000000000000280] [Medline: 29373382]
30. Ebm C, Del Pozo C, Barbarello A, Poli G, Brusa S. Unleashing excellence: using a project management approach to effectively implement a simulation curriculum to improve residents' preparedness. BMC Med Educ 2024 Mar 4;24(1):234. [doi: 10.1186/s12909-024-05166-y] [Medline: 38438940]
31. Chittaro L, Buttussi F. Assessing knowledge retention of an immersive serious game vs. a traditional education method in aviation safety. IEEE Trans Vis Comput Graph 2015 Apr;21(4):529-538. [doi: 10.1109/TVCG.2015.2391853] [Medline: 26357103]
32. Padilha JM, Machado PP, Ribeiro A, Ramos J, Costa P. Clinical virtual simulation in nursing education: randomized controlled trial. J Med Internet Res 2019 Mar 18;21(3):e11529. [doi: 10.2196/11529] [Medline: 30882355]

XSL•FO
RenderX

33. Buajeeb W, Chokpipatkun J, Achalanan N, Kriwattanawong N, Sipiyaruk K. The development of an online serious game for oral diagnosis and treatment planning: evaluation of knowledge acquisition and retention. BMC Med Educ 2023 Nov 3;23(1):830. [doi: 10.1186/s12909-023-04789-x] [Medline: 37924052]

Abbreviations

ANTS: Anesthesia Non-Technical Skills
CONSORT: Consolidated Standards of Reporting Trials
HF: high-fidelity
ILCOR: International Liaison Committee on Resuscitation
NRPE: Neonatal Resuscitation Performance Evaluation
SG: serious game

Edited by A McDaniel; submitted 03.02.24; peer-reviewed by A Dwijayanti, C Lejus, M Liaqat; revised version received 09.10.24; accepted 01.11.24; published 21.03.25.

Please cite as:
Louvel AC, Dopff C, Loron G, Michelet D
Impact of a 3-Month Recall Using High-Fidelity Simulation or Screen-Based Simulation on Learning Retention During Neonatal Resuscitation Training for Residents in Anesthesia and Intensive Care: Randomized Controlled Trial
JMIR Serious Games 2025;13:e57057
URL: https://games.jmir.org/2025/1/e57057
doi:10.2196/57057

© Anne-Claire Louvel, Cécile Dopff, Gauthier Loron, Daphne Michelet. Originally published in JMIR Serious Games (https://games.jmir.org), 21.3.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Perceptions Toward an Attentional Bias Modification Mobile Game Among Individuals With Low Socioeconomic Status Who Smoke: Qualitative Study

Michael Wakeman[1], BS; Lydia Tesfaye[1], BS; Gunnar Baskin[2], BS; Tim Gregory[2], BA; Greg Gruse[2], BS; Erin Leahy[2], BA; Brandon Kendrick[2], MA; Sherine El-Toukhy[1], MA, PhD

[1]Division of Intramural Research, National Institute on Minority Health and Health Disparities, National Institutes of Health, 11545 Rockville Pike, Rockville, MD, United States

[2]ICF, Reston, VA, United States

Corresponding Author:
Sherine El-Toukhy, MA, PhD
Division of Intramural Research, National Institute on Minority Health and Health Disparities, National Institutes of Health, 11545 Rockville Pike, Rockville, MD, United States

Abstract

Background: Attentional bias modification (ABM) can help address implicit attention from individuals who smoke toward smoking cues, which trigger cravings and lapses that impede smoking cessation. We developed an ABM mobile game, Fruit Squish, to support individuals who smoke and are quitting as part of a multicomponent smoking cessation mobile app, Quit Journey. Users advance in the game by tapping on neutral (ie, fruit) rather than smoking-related (eg, cigarette pack) imagery that they are presented with, essentially training them to avoid focusing on smoking cues.

Objective: This study aimed to gauge acceptance of an ABM smoking cues mobile game among young adults who smoked and were socioeconomically disadvantaged.

Methods: We recruited 38 individuals who smoked cigarettes, aged 18 - 29 years, who were neither 4-year college graduates nor enrollees in 4-year colleges to participate in 12 semistructured digital focus groups. Sessions were audio recorded and transcribed verbatim. We used ATLAS.ti software to code the transcripts for salient themes based on the Second Unified Theory of Acceptance and Use of Technology constructs (ie, effort expectancy, facilitating conditions, hedonic motivation, performance expectancy, and social influence) and sentiment (ie, negative, neutral, and positive).

Results: Performance expectancy of the mobile game was the dominant technology acceptance construct discussed (34/110, 30.90%). Perceived usefulness of the game was mixed in sentiment owing to perceptions that the game aimed to distract individuals who smoke during cravings and concerns that cue imagery in the game could trigger cravings. Hedonic motivation was the second most discussed technology acceptance construct (17/110, 15.45%), with participants describing the game as neither fun nor engaging. Participants referenced their past experiences with mobile games and mobile device characteristics as facilitating conditions for using the game (10/110, 9.09%). Although effort expectancy was minimally discussed (6/110, 5.45%), the game was characterized as easy to use. To improve the game, participants suggested adding new levels with increasing difficulty (eg, increase stimuli speed and limit session time) and new game elements (eg, leaderboard). Other suggestions included improving game graphics and renaming the game to capture its relation to smoking cessation.

Conclusions: Young adults with low socioeconomic status who smoke had mixed reactions to a mobile smoking cues ABM game. Results suggest the need to communicate the rationale underlying ABM games to users and their potential positive effects on smoking cessation. To promote the uptake and sustained use of ABM mobile games, they need to be on par with commercially available entertainment mobile apps. Research is needed to explore the efficacy of gamified ABM on cognitive biases in real-life settings.

(JMIR Serious Games 2025;13:e59515) doi:10.2196/59515

KEYWORDS

serious games; focus groups; smoking cessation; attentional bias modification; mobile interventions; young adults

XSL•FO
RenderX

Introduction

Implicit cognitive processes play a critical role in the maintenance of addictive behaviors, including tobacco use disorders [1,2]. Individuals who smoke have unconsciously enhanced attention toward smoking cues, which are stimuli associated with smoking behaviors ranging from physical objects to smells and observed behaviors [3-5]. Attentional biases result from positive rewards associated with smoking that are conditioned over time. Accordingly, attentional bias toward smoking cues can trigger cravings and have negative affect, which in turn cause smoking behaviors or lapses [6]. Furthermore, attentional biases can persist among individuals who formerly smoked and are associated with a heightened risk of relapse [7-9]. Attentional biases have been documented through both self-reports (eg, self-reported cravings after exposure to cues) and objective measures (eg, faster reaction times when responding to smoking cues) [8,10]. Several theories explain the development of attentional biases, including the incentive-sensitization theory, the elaborative intrusion theory of desire, and the theory of current concerns [1,2]. Collectively, these models suggest that substance cues do not merely trigger cravings but that the two exist as part of a bidirectional relationship and positive feedback cycle [1,2].

Clinical guidelines for treating nicotine dependency recommend that individuals who smoke avoid or reduce exposure to smoking cues in their environments [11]. Individuals who smoke can follow these recommendations for smoking cues present in their immediate environments (eg, ashtrays in their homes). However, these cues can be pervasive and beyond their control in other instances (eg, media and advertisements) and are, thus, more difficult to avoid [5,12]. This is particularly true for disadvantaged individuals who smoke, are exposed to a high volume of tobacco advertising, are targets of tobacco marketing tactics (eg, coupons), and reside in areas with a higher density of tobacco retailers [13-15]. For instance, retail marketing (eg, tobacco price promotions) disproportionately targets African Americans [16]. Similarly, tobacco retailer density is higher in African American, Hispanic, and lower-income neighborhoods [15]. These strategies contribute to the saliency of tobacco-related imagery, which triggers smoking behaviors and lapses, and are a potential mechanism for higher smoking prevalence and difficulty quitting among populations at the receiving end of these strategies [17].

Cognitive bias modifications (CBMs) are interventions that target and restructure implicit cognitive biases related to disorders, such as substance use, using varied cognitive tasks [18]. Attentional bias modification (ABM) is a CBM intervention approach where individuals are trained to refocus their attention away from substance cues, thereby severing previously formed associations [2]. One approach involves a visual-probe task where both neutral and substance-related stimuli are presented on a screen followed by a probe (eg, a small dot) to redirect the user's attention to the neutral stimuli [19]. ABM has previously been explored for addictive behaviors, including tobacco and alcohol use, with evidence of their effects on substance use being mixed. For example, a review of meta-analyses found that ABM can be effective in shifting attentional biases, but the effect sizes were short-lived and generally modest for appetitive stimuli (eg, smoking) [18]. Another meta-analysis focusing on tobacco and alcohol found that CBM has a small reductive effect on cognitive biases but did not reduce substance use [20]. These results could be attributed to several factors, mainly the limited number of training sessions [18]. However, there is evidence that multiple (vs single) ABM sessions can decrease attentional bias over longer time spans [21,22].

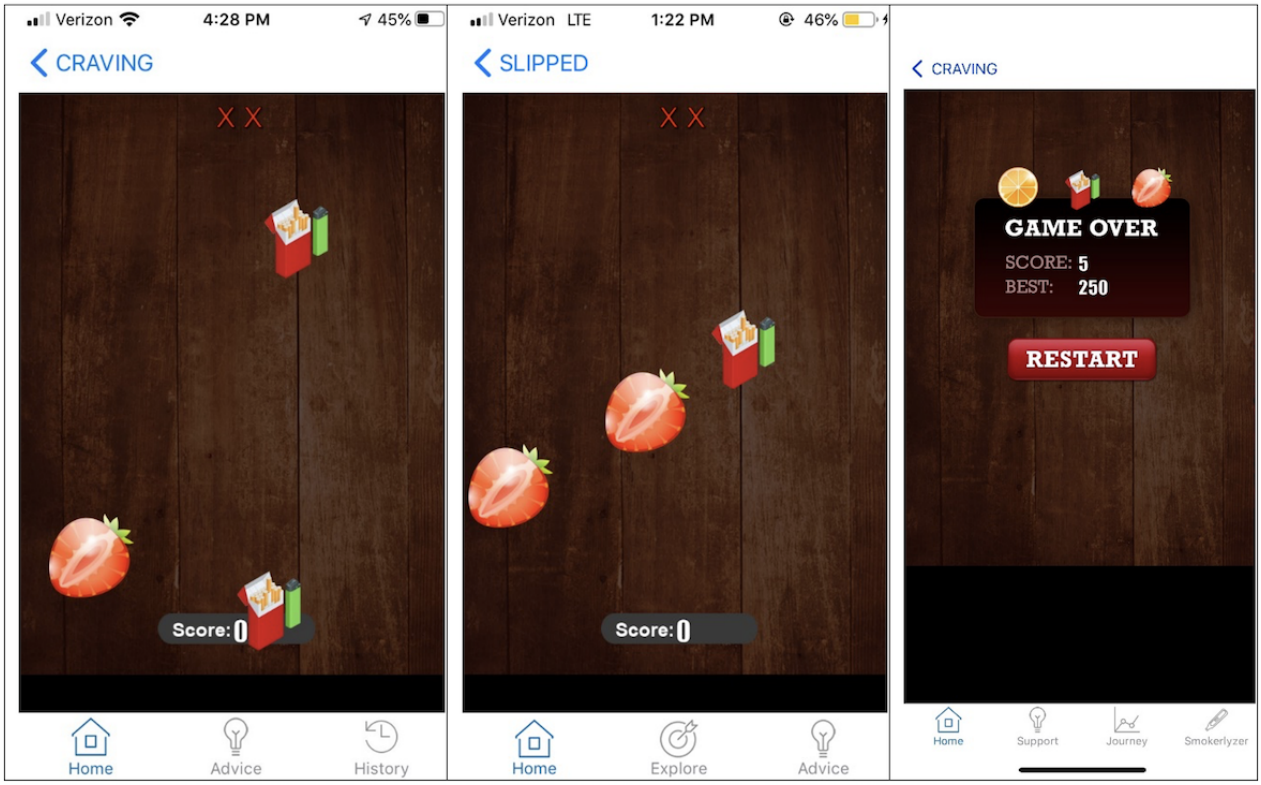
Smartphones, with their high use, market penetration, and advanced features [23], can conveniently deliver multisession, self-administered ABM interventions to individuals who smoke in their natural environments. More importantly, advancements in mobile technologies afford opportunities to gamify ABM interventions to engage target populations and increase their adherence to ABM interventions over multiple sessions. The use of mobile apps to deliver CBM for substance use disorders (eg, alcohol use) has been promising [24,25]. For example, a review of trials using apps to deliver CBM interventions found that 7 of 8 trials effectively modified existing biases or reduced related symptoms (eg, anxiety and alcohol consumption). Nearly all of these apps relied on the visual-probe task [26]. In the context of smoking cessation, few attempts have been made to translate ABM techniques outside laboratory settings using digital technologies and platforms (eg, personal digital assistants and web-based interventions) [2]. Further, a review identified only 3 commercially available CBM apps for tobacco use (ie, Quitty, StopSmoking-Quit Smoking, Quit Smoking: Simple and Quick), none of which were supported by published literature [26]. Although limited in number, studies examining smoking-related ABM delivered via smartphones have primarily used modified visual-probe tasks and had mixed effects on attentional biases and no effects on symptoms (eg, cravings and smoking behaviors) [27-29].

To our knowledge, gamifying mobile ABM interventions has been limited to substance use disorders (ie, alcohol and opioid use) primarily using modified visual-probe tasks. Gamified ABM interventions have yet to be implemented in smoking cessation apps as part of multicomponent interventions [30]. To that effect, we developed a smoking cues ABM mobile game called Fruit Squish. The game aims to help individuals who smoke reduce or avoid attention paid to smoking cues, a behavioral change technique that facilitates smoking cessation [31,32]. Briefly, the game mimics popular mobile games (eg, Fruit Ninja) where players must slice select stimuli (eg, pieces of fruit) rather than undesired stimuli (eg, bombs) to advance. We translated the visual-probe task typically used in attentional bias training to a Fruit Ninja game template by substituting images of bombs with smoking-related images (eg, cigarette packs). In Fruit Squish, players must tap on the fruit (rather than smoking-related images) to receive points and advance in the game (Figure 1). The game is part of Quit Journey, a novel smoking cessation mobile app targeting individuals with low socioeconomic status who smoke. Briefly, Quit Journey is a multicomponent smoking cessation intervention that includes features common to smoking cessation apps (eg, educational content). Additionally, the app includes 4 novel features: carbon monoxide physiological tracking, performance-based coupon

XSL•FO
RenderX

incentives, an ABM game (ie, Fruit Squish), and a user-generated message library with smoking-related messages (feature under development) that provides access to on-demand messages and to just-in-time support messages that are sent when app users are at risk of smoking [33, Tesfaye et al, 2024; Wakeman et al, 2024, unpublished data]. Due to the novelty of ABM games in the context of smoking cessation, understanding users' perceptions of them is important. Prior research has established that user perceptions of new technologies are essential for their acceptance and use [34,35]. In this study, we gauge perceptions of individuals with low socioeconomic status who smoke toward a smoking cues ABM mobile game, Fruit Squish, to further improve the game as a feature of Quit Journey. The study falls under the preparation phase of the multiphase optimization strategy for developing and evaluating behavioral interventions [36,37].

Figure 1. Screenshots from Fruit Squish.



Methods

Ethical Considerations

The National Institutes of Health Institutional Review Board deemed the study exempt on October 11, 2019. ICF's Institutional Review Board deemed the study exempt on November 19, 2019 and approved an amendment on February 26, 2020. All participants received information about the study and its procedures and verbally consented to participate in the study before the audio recording began. Userworks, Inc (Silver Spring, MD), a user experience and usability research firm, had access to personally identifiable information for participant recruitment and screening purposes, whereas all authors did not. UserWorks assigned enrolled participants identification numbers (eg, P1), which were used during the focus group discussions, and subsequently in the transcripts and data analysis. Participants received US $150 gift cards as an incentive.

Participants and Recruitment

We partnered with Userworks to recruit a convenience sample of 38 participants through its research panels and other commercial platforms (eg, Craigslist). UserWorks contacted candidate participants via email and screened those interested over the phone for enrollment eligibility (Note S1 in Multimedia Appendix 1). Participants were eligible for inclusion in the study if they self-reported being 18 - 29 years old, neither 4-year college educated nor enrolled in a 4-year college as an indicator of low socioeconomic status [38-41], current cigarette smokers who smoked at least 100 cigarettes in their lifetime and smoked every day or some days, willing to quit smoking within 6 months, not using smoking cessation aids, not using noncigarette combustible tobacco products (eg, cigars), smartphone owners, and able to speak English. Recruitment occurred between January and April 2020 and no participants withdrew after enrollment. Recruitment was not based on data saturation. However, the sample size was deemed sufficient to achieve data saturation [42]. The 32-item consolidated criteria for reporting qualitative research checklist appears in Table S1 in Multimedia Appendix 1 [43].

Procedures

We held 12 semistructured focus groups to explore perceptions of a newly developed smoking cessation mobile intervention, Quit Journey, and its features, including the ABM mobile game described here, among young adults who smoked and were socioeconomically disadvantaged. The game was presented as

XSL•FO
RenderX

a concept in 8 predesign focus group sessions and as a prototype version in 4 postdesign sessions. A total of 12 participants were involved in both pre and postdesign focus group sessions. Discussions lasted ≈1.5 hours and were held on GoTo Meeting, audio recorded, and transcribed verbatim. Transcripts were autogenerated by GoTo Meeting, verified by 3 members of our team, and were not returned to participants for comment.

Discussions followed a moderation guide based on the technology acceptance models (Note S2 in Multimedia Appendix 1) [34,35]. TG, a user experience strategist and researcher (male), was the primary moderator, while EL, a strategic communications and marketing project director (female), was the backup moderator and note-taker. TG had training and certification in user research, including focus groups. The moderators had no prior relationship with participants who were informed that the moderators were unaffiliated with the research group that commissioned the study. Only the participants, TG, EL, and SEL, were present during the focus group sessions. Before conducting the focus groups, we conducted dry runs.

Analysis

Following a deductive thematic analysis [44,45], we developed a codebook and corresponding themes using 5 constructs from the Second Unified Theory of Acceptance and Use of Technology: effort expectancy, facilitating conditions, hedonic motivation, performance expectancy, and social influence [34,35]. These constructs were defined as follows: (1) effort expectancy referred to "perceived ease of use or effortfulness with which one can navigate mobile applications and their features and seamlessly integrate them in one's life"; (2) facilitating conditions were defined as "factors that can aid or impede the uptake or use of mobile applications or their features. These include individual-related (eg, skills, predispositions, prior experiences) and technical-related (eg, infrastructure) factors". (3) Hedonic motivation referred to "perceived fun, pleasure, or enjoyment (or lack thereof) associated with the use of mobile applications and their features". (4) Performance expectancy referred to "perceived usefulness or helpfulness of smoking cessation mobile applications and their features in achieving desired health goals and behaviors". (5) Social influence referred to "perceived importance of significant others' recommendations and approval of using mobile applications and their features" [33,46,47, Wakeman M et al, 2024; Tesfaye L et al, 2024, unpublished data]. Quotes that did not fit under a technology acceptance construct were coded as "not applicable". Additionally, we included sentiment (ie, negative, neutral, and positive), design concepts, and users' suggestions. Positive and negative sentiment captured statements that indicated "a sense of approval, praise, or certainty" or "a sense of disapproval, criticism, or skepticism about any aspect of smoking cessation mobile applications and their features such as their worthiness, utility and impact, time and effort investment, and compatibility with one's life," respectively. Neutral sentiment captured statements "that were neither positive or negative in tone, contained an equal number of positive and negative remarks, or were conditional (ie, positive in nature but depended on the presence or absence of another factor" [33,46,47, Tesfaye et al, 2024; Wakeman et al, 2024, unpublished data]. The design concepts code captured statements related to "user interface, content input and delivery, information structure, navigation, typography, and aesthetics of smoking cessation applications or their features" [33,46,47, Tesfaye et al, 2024; Wakeman et al, 2024, unpublished data]. Finally, the suggestions code captured statements that were "concerned with improvements, modifications, or additions to smoking cessation applications or their features aimed to improve their functionality or design" [33,46,47, Tesfaye et al, 2024; Wakeman et al, 2024, unpublished data]. We created additional codes for emerging themes (ie, intent/willingness to use the game, novelty). We coded the transcripts for salient themes applying only one theme from each semantic domain to a quote. When multiple themes were present, we selected the predominant or underpinning theme in conditional quotes (eg, a quote was coded as hedonic motivation if the participant considered the game to be useful if it was fun and engaging). We coded the transcripts for all Quit Journey app features, but only present here results related to the smoking cues mobile game. Our research on the use of mobile technologies for smoking cessation and on Quit Journey and its features has been published elsewhere or is currently under review [33,46,47, Tesfaye et al, 2024; Wakeman et al, 2024, unpublished data].

A member of our team extracted quotes for exploratory analysis using Excel (Microsoft Corporation). Two authors (MW and LT) independently coded focus group transcripts using ATLAS.ti (version 8; ATLAS.ti Scientific Software Development GmbH), which was also used to calculate intercoder agreement [48]. We calculated the Krippendorff c-α, a measure of separating relevant and irrelevant content whereby coders code texts of similar locations and lengths, and the Krippendorff cu-α, a measure of semantic domain reliability whereby coders code a given text for the presence or absence of a semantic domain [48]. Discrepancies were resolved through discussions between MW, LT, and SEL. Qualitative results were not returned to participants for feedback.

Results

Patient Characteristics

The Krippendorff c-α was 0.82, whereas the Krippendorff cu-α was 0.66 for technology acceptance, 0.70 for sentiment, and 0.94 for Quit Journey app features. Sample characteristics appear in Table 1, and detailed participant characteristics in Table 2.

XSL•FO
RenderX

Table . Participant characteristics (N=38) [33,46,47].[a]

Characteristic	Values, n (%)
Sex	
Female	20 (52.63)
Male	18 (47.36)
Race and ethnicity	
NH[b] American Indian or Alaska Native	1 (2.63)
NH Asian, Native Hawaiian or Pacific Islander	3 (7.89)
NH Black or African American	11 (28.94)
Hispanic or Latino	6 (15.78)
NH White	16 (42.10)
Mixed	1 (2.63)
Highest level of education	
Less than high school	3 (7.89)
High school graduate	10 (26.31)
High school equivalent	3 (7.89)
Some college, no degree	18 (47.36)
2-year associate degree	4 (10.52)
Smoking frequency	
Every day	30 (78.94)
Some days	8 (21.05)
Quit timeframe	
7 days	11 (28.94)
30 days	22 (57.89)
6 months	5 (13.15)
Smartphone operating system	
Android	21 (55.26)
iOS	17 (44.73)

[a]Includes unpublished data from Tesfaye et al, 2024; Wakeman et al, 2024.

[b]NH: Non-Hispanic.

XSL•FO
RenderX

Table . Detailed participant characteristics [33,46,47].[a] Participants who identified as Hispanics or Latinos were considered as such regardless of race.

ID	Sex	Race and ethnicity	Highest level of education	Smoking frequency	Quit timeframe	Smartphone operating system
P01[b]	Female	NH[c] White	Some college, ND[d]	Every day	30 days	Android
P02[b]	Male	Hispanic/Latino	Some college, ND	Every day	30 days	Android
P03	Female	Hispanic/Latino	HS[e] incomplete	Every day	7 days	Android
P04[b]	Female	NH White	Some college, ND	Some days	30 days	Android
P05	Male	NH White	Some college, ND	Every day	7 days	Android
P06	Male	NH Black/AA[f]	Some college, ND	Every day	30 days	Android
P07	Male	NH Asian, NHPI[g]	HS equivalent	Every day	30 days	iOS
P08[b]	Female	Hispanic/Latino	Some college, ND	Every day	30 days	iOS
P09	Male	NH White	HS equivalent	Every day	30 days	iOS
P10[b]	Male	NH White	HS graduate	Every day	7 days	Android
P11[b]	Female	Hispanic/Latino	Some college, ND	Every day	30 days	iOS
P12[b]	Male	NH White	Some college, ND	Every day	7 days	Android
P13[b]	Female	NH White	HS graduate	Every day	30 days	Android
P14[b]	Female	NH White	2-year AD[h]	Some days	30 days	iOS
P15[b]	Female	NH Black/AA	HS incomplete	Some days	30 days	Android
P16[b]	Female	AIAN[i]	HS incomplete	Every day	30 days	iOS
P17[b]	Male	NH Black/AA	HS graduate	Every day	30 days	Android
P18	Female	NH Black/AA	Some college, ND	Every day	7 days	iOS
P19	Female	NH White	HS equivalent	Every day	7 days	Android
P20	Female	NH White	HS graduate	Every day	7 days	Android
P21	Female	NH White	Some college, ND	Every day	7 days	iOS
P22	Female	NH White	Some college, ND	Every day	30 days	iOS
P23	Male	NH White	Some college, ND	Every day	7 days	iOS
P24	Female	NH Black/AA	Some college, ND	Every day	30 days	Android
P25	Female	NH Black/AA	HS graduate	Some days	30 days	Android
P26	Male	NH Black/AA	Some college, ND	Every day	30 days	iOS
P27	Male	Hispanic/Latino	Some college, ND	Some days	30 days	Android
P28	Female	NH White	2-year AD	Every day	30 days	Android
P29	Female	NH White	Some college, ND	Every day	30 days	iOS
P30	Male	NH Asian, NHPI	HS graduate	Some days	6 months	Android
P31	Male	NH Black/AA	Some college, ND	Every day	6 months	Android
P32	Female	NH Asian, NHPI	HS graduate	Every day	30 days	iOS
P33	Male	Hispanic/Latino	2-year AD	Some days	6 months	iOS
P34	Male	NH Black/AA	2-year AD	Every day	6 months	Android
P35	Female	NH Black/AA	HS graduate	Every day	30 days	iOS
P36	Male	NH White	Some college, ND	Every day	7 days	Android
P37	Male	NH Black/AA	HS graduate	Every day	6 months	iOS
P38	Male	NH Mixed	HS graduate	Some days	7 days	iOS

XSL•FO
RenderX

[a]Includes unpublished data from Tesfaye et al, 2024; Wakeman et al, 2024.

[b]Participated in 2 focus groups.

[c]NH: Non-Hispanic.

[d]ND: no degree.

[e]HS: high school.

[f]AA: African American.

[g]NHPI: Native Hawaiian/Pacific Islander.

[h]AD: associate degree.

[i]AIAN: American Indian/Alaska Native.

We extracted 143 quotes across all semantic domains. A total of 110 quotes captured technology acceptance constructs and sentiment toward the smoking cues mobile game, whereas 33 quotes included suggestions to improve the game. Sentiment toward the smoking cues game was mixed, with negative (45/110, 40.90%) and neutral (36/110, 32.72%) quotes outweighing those with positive sentiment (29/110, 26.36%) (Table 3). Most quotes centered on performance expectancy (34/110, 30.90%), followed by hedonic motivation (17/110, 15.45%), facilitating conditions (10/110, 9.09%), and effort expectancy (6/110, 5.45%). Social influences on game acceptance and use were not discussed. Below we present illustrative quotes, with participant ID attribution, first on performance expectancy, then followed by all other technology acceptance themes as they were discussed to a lesser extent (Table S2 in Multimedia Appendix 1). Last, we present participants' suggestions to improve the game's functionality and design (Table S3 in Multimedia Appendix 1).

Table . Distribution of the number of quotes on smoking cues attentional bias modification mobile game by technology acceptance constructs and sentiment.

Themes	Sentiment			Total, n (%)
	Negative, n (%)	Neutral, n (%)	Positive, n (%)	
Effort expectancy	0 (0)	2 (33.33)	4 (66.67)	6 (5.45)
Facilitating conditions	4 (40.00)	5 (50.00)	1 (10.00)	10 (9.09)
Hedonic motivation	8 (47.05)	5 (29.41)	4 (23.52)	17 (15.45)
Performance expectancy	9 (26.47)	12 (35.29)	13 (38.23)	34 (30.90)
Social influence	0 (0)	0 (0)	0 (0)	0 (0)
Not applicable	24 (55.81)	12 (27.90)	7 (16.27)	43 (39.09)
Total	45 (40.90)	36 (32.72)	29 (26.36)	110 (100.00)

Performance Expectancy

Perceptions of the smoking cues game's usefulness were mixed in sentiment with 26.47% (9/34) of the quotes having a negative sentiment, 35.29% (12/34) neutral, and 38.23% (13/34) positive. Some participants attributed the usefulness of the game to its potential role as a distraction tool from cravings, which was rooted in participants' familiarity with mobile games and their prior use of them as distraction tools.

> *The distraction would be good and just having another thing to do and having my mind just getting distracted and not focusing on smoking. That'll be helpful, really beneficial, really useful.* [P06]

> *I feel like it'd be very useful because I too kind of play games to distract myself, from negative thoughts, or habits.* [P10]

Few participants who understood that the true purpose of the game was to modify unconscious attention to smoking cues emphasized the novelty of the concept and its potential usefulness in facilitating cessation.

> *You're kinda … psychologically … switching … I guess [the game] makes you less likely to focus on those things in real life. But I never really thought about that. I think that's pretty cool.* [P30]

> *You're using [the game] as a way to relocate your mind from a cigarette, basically, at that time that you need it the most.* [P20]

Some participants expressed concerns over the smoking cues imagery, which could trigger cravings and be counterproductive to their quit attempts.

> *I like the idea of a game, but I feel like it makes it kind of tough because … what if I'm playing this game and I see packs of cigarettes and stuff, and I'm like, oh, well, I could go for a cigarette now … Like, maybe trigger something that I wouldn't already do … I like the idea of a game but maybe not necessarily having like smoking-related images.* [P21]

> *I almost feel like [the game is] trying to utilize … like cognitive behavioral therapy … where it's like exposure therapy. But I think for people who are trying to not think about cigarettes, to think about them doesn't necessarily help them.* [P36]

XSL•FO
RenderX

Regardless of what participants thought the purpose of the game was, they wanted evidence of the effectiveness of the game in helping them quit smoking.

> *I would have to see the game … in action I guess … to see … how enticing it is … and also know that it's supposed to be proven to really help like psychologically and really impacts your brain functions as far as wanting to quit.* [P27]

> *The pictures [of the game] are … kinda lame. And I would want to like see … how that helps, because I think if that was just … a game, I would be like, what the heck is this. I wouldn't understand without someone … explaining that it's gonna actually help.* [P11]

Other Technology Acceptance Themes

Hedonic Motivation

Participants characterized the game as outdated, boring, and not fun, with 47.05% (8/17) of the quotes being negative in sentiment.

> *I feel like the game seems just kinda boring.* [P10]

The game fared worse in comparison to other mobile games that participants were accustomed to playing, especially as they highlighted enjoyability as an important user engagement factor.

> *If I want to play a game and distract myself, I might just do that of my own accord. I just have a feeling that … the game within the app is probably not going to be as much fun as … whatever I was gonna play anyway, which is gonna distract me anyway.* [P12]

> *I think [the game] would have to be fun, but … not stressful either, because … I think a lot of times when some people play games, they want to smoke cause sometimes it can be a little bit stressful or boring, so it'd definitely have to be … very engaging and fun to actually want to play it and use that, and actually help.* [P13]

Facilitating Conditions

Past experiences and attitudes toward mobile gaming, both positive and negative, affected participants' perceptions of the game.

> *[The game] reminds me of Fruit Ninja, but without flashing things, and I never really liked that game. That game frustrates me. So, this one probably would frustrate me too, which is the opposite of what I'm trying to do.* [P28]

Additionally, compatibility of the game with their smart devices and the time demands to play the game were highlighted as possible barriers.

> *If it's not too … over the top, like, some people might not have … good enough phones to run [the game] … But as long as it doesn't crash and is … decent then, yeah [I think the game would be easy to use].* [P13]

> *I think it just depends on the audience, because if you're not a gamer and you're more into social media, and that's you're downtime … it just depends on how determined you are and if you're game savvy and things like that.* [P16]

Effort Expectancy

Six quotes referencing effort expectancy were generally positive in sentiment about the ease of playing the game (66.67%).

> *I feel like [the game] would be a pretty easy to use feature.* [P23]

Suggestions

A total of 33 quotes included suggestions to improve the smoking cues game (Table S3 in Multimedia Appendix 1). First, participants suggested renaming the game to reference smoking.

> *Definitely [the name of the game should reference smoking], if it's going to be on a smoking app.* [P09]

Because of its role in helping them quit, some participants suggested that the game playing time and scoring system should be linked to their quitting journey. For example, one participant suggested that players could score more points if they played the game when craving cigarettes.

> *It might be cool, like, after a week of your set trigger times, like, you would only get points if you did … the little games at those trigger times … Instead of smoking a cigarette, you did a game. Instead of just someone playing games all day, getting points.* [P21]

Because of concerns over cravings when seeing smoking-related images in the game, some suggested changing the premise of the game to enable players to "destroy" smoking stimuli instead of neutral stimuli.

> *I think the game should be about … destroying cigarettes … instead of … choosing fruit over the cigarette, because … when you see cigarettes and you see the lighter … it's … kind of giving me a craving, seeing the picture of cigarettes. So, I think it should be like something to destroy cigarettes, to show that it's no good.* [P25]

Drawing on their experiences with mobile games, participants suggested adding different levels to the game with varying difficulty (eg, increasing speed and time limits), competitive elements (eg, leaderboards), and the ability to invite friends to play the game as in multiplayer games.

> *Maybe make it … like a high score kind of thing like old arcade games where you … have like a leaderboard and try to get the highest score and stuff. Just make it a … little competition, like friendly competition. Give it a little bit more of a distraction.* [P05]

Other participants suggested adding other games to our cessation mobile app (eg, mindfulness games) to help individuals who smoke deal with smoking antecedents such as stress.

> *Mindfulness might be better [for the game]. I think it would be maybe cool if … you picked like your stress,*

XSL•FO
RenderX

then you could have like a meditation. Or … if you're like sad, you could have, like, a really cute game. Or, like something that would correlate to your mood to personalize it and … be like as helpful as possible. [P11]

Discussion

Principal Findings

This is one of the first studies, to our knowledge, that examines perceptions of a gamified ABM module within a multicomponent smoking cessation mobile app targeting individuals with low socioeconomic status who smoke [27,30]. As a novel concept, participants were willing to try the game provided there was evidence of its efficacy as a tool in aiding them to quit. However, only a few grasped the true purpose of the game as an attentional bias training tool, whereas most viewed the game as a distraction tool during cravings. Skepticism about the usefulness of the game was attributed to the possibility of triggering cravings upon the presentation of smoking cue imagery and the availability of other mobile games, more sophisticated in nature, that could distract individuals who smoke during cravings. Opinions were mixed concerning the game's enjoyability but were largely positive for ease of use. Our findings underscore that the primary challenge to maximizing user acceptance is to communicate the purpose and benefit of an ABM game and to improve the game's functionality and design elements to meet users' expectations and withstand competition from commercially available entertainment games. This is particularly true if an ABM game were to engage users for an extended time that is needed to allow for the delivery of multiple training sessions to positively affect cessation. Beyond our ABM smoking cues game, these results could inform the development and user acceptability of gamified self-help ABM and CBM interventions for other substance use disorders and tobacco products. This is particularly relevant as CBM- and ABM-based mobile games and apps have generally had mixed effects on behavioral change [30]. Finally, our results highlight the importance of user-centered approaches for intervention development. User feedback will help us further improve our ABM mobile game by incorporating participants' suggestions prior to evaluating its usability and efficacy in field studies.

Comparison to Prior Work

Our approach to embedding an ABM mobile game within a multicomponent app for smoking cessation is novel. Past attempts to translate ABM techniques to mobile platforms as games have been limited to substance use disorders other than tobacco dependency (eg, opioid use) [49]. Few smartphone-based ABM interventions for smoking cessation have been standalone interventions. Furthermore, these interventions usually deliver the traditional visual probe test on a mobile platform rather than transforming the technique to embrace smartphones as a medium [27-29]. Our approach would allow for user-initiation of the training module at their own will and in their natural environments, eliminating one potential cause of the underperformance of traditional ABM interventions. This could be particularly beneficial to populations that cannot avoid or reduce exposure to smoking-related imagery per clinical guidelines for treating tobacco dependency [11]. Given the scarcity of evidence on ABM mobile games for smoking cessation in general and with disadvantaged groups in particular, research is needed to examine their efficacy in reducing smoking and its antecedents (eg, cravings and reactivity to smoking cues).

Additional strengths of our game-based approach to smoking cues ABM include the accessibility and familiarity of smartphones and mobile games, particularly among young and middle-aged adults [23,50]. Indeed, most participants in our study agreed the game was easy to use and referenced prior experiences with mobile games. The downside to this approach is that it excludes by default individuals with certain predispositions, thus limiting the efficacy of smoking cessation intervention with an ABM component. For example, less digitally literate and older individuals might not be willing or able to use mobile health games [50]. More importantly, high user expectations of ABM game functionality and design present a fundamental challenge. Mobile health games must compete with commercially available, professionally developed entertainment games for users' attention and engagement, especially as most mobile apps rapidly lose users after their first use [51]. While smoking cessation interventions are on average 6 weeks long, sustaining engagement during this period is a challenge. However, limited evidence suggests that the use of a game was positively associated with the number of smoking cessation app sessions and time spent on the app [52,53].

Indeed, enjoyability emerged as a main factor in user acceptance of the ABM game. Past work has established the important role of enjoyability in users' adoption of new technologies as well as its impact on user engagement for serious games [35,54]. User engagement, in turn, has been associated with abstinence in pilot testing [55]. Accordingly, negative sentiments in discussing the game's level of fun and enjoyment are concerning. Some remarked that the game appeared outdated, which may be on account of seeing an early prototype or on account of having experiences with increasingly sophisticated mobile games. Partnerships between academics and mobile game developers may be a beneficial approach to elevate health games [52]. Our game's level of enjoyability could be improved by implementing user suggestions, which primarily focused on adding new elements or approaches commonplace in other games. Several of the suggested additions (eg, leaderboards, levels, and rewards) have been desired by individuals who smoke in other work and successfully implemented in both smoking cessation and ABM apps [30,56-58].

Our findings show that communicating the intent and benefit of an ABM is a primary requisite to user acceptance. Skepticism of the game's performance stemmed primarily from concerns that the presentation of cue imagery could trigger cravings and from perceptions of the game as a distraction tool, in which case preferable alternatives are readily available. Suggestions to redesign the game to allow users to "destroy" smoking-related images as well as the common use of mobile apps and games for distractions echo our participants' perceptions of the ABM game [57,59]. Previous research on an ABM app for substance use disorder documented similar results [56]. Moreover, cravings in response to cues are well-documented among individuals

XSL•FO
RenderX

who smoke and provide credence to participants' concerns [60]. However, the game requires focusing one's attention on neutral images (ie, fruit slices) to score. For the ABM to be effective, it is necessary to use stimuli that are relevant to and associated with users' attentional biases [61]. Past qualitative research has identified the need for users to understand the rationale behind CBM and ABM [62,63], which is critical for engagement [54]. Thus, it is imperative to identify appropriate means to communicate the rationale behind the game to users and its potential benefit in helping them quit. This could potentially be accomplished by providing a written rationale, written instructions, or informational tooltips, as some participants suggested. Future research is needed to field test the effects of an ABM game on intended (eg, modifying attentional biases) and unintended (eg, triggering cravings) outcomes.

Strengths and Limitations

This study has several strengths, including the recruitment of a diverse sample of participants with low socioeconomic status with roughly ≤50% of the participants from one race or sex. Furthermore, our qualitative methodology provides a wealth of information on user perceptions of the ABM mobile game under investigation. Nonetheless, this study has several limitations. First, although the moderator assured participants that there were no right or wrong answers, our results could have been influenced by social desirability bias [64]. Second, our focus groups were held during the COVID-19 pandemic, which resulted in participants seeing web-based mock pages of the game rather than using it had the discussions been conducted in person. Participants were shown an early prototype of the smoking cues mobile game, which may have affected their perceptions. Our results may have been impacted by the high exposure to and familiarity with mobile games that young adults have, which may not be representative of all individuals who smoke [50]. However, we will field test our smoking cessation app to determine whether the ABM game significantly contributes to quit rates among adults older than 18 years and to shed light on their experiences using the game in their natural environment, which will complement this formative work. Last, although the intercoder agreement for technology acceptance and sentiment domains was low, all inconsistencies were resolved through discussions between MW, LT, and SEL.

Conclusions

This study is one of the first to characterize users' perceptions of an ABM mobile game for smoking cessation. Participants had mixed perceptions of the game's usefulness and enjoyability but were willing to use it provided evidence of its efficacy in aiding smoking cessation. These results will help guide the development and improvement of the game to support adults with low socioeconomic status who smoke. Further work should explore the real-life use and effects of gamified ABM techniques in general and for smoking cessation in particular both independently and as part of multicomponent behavioral interventions.

Acknowledgments

ICF services for this manuscript were performed under the support of the National Institute on Minority Health and Health Disparities through contract #75N91019F00132 with the National Cancer Institute. The National Cancer Institute had no input on the development of Quit Journey smoking cessation mobile app or its features nor on the study conceptualization, data analysis, or interpretation. The efforts of MW, LT, and SEL have been supported by the Division of Intramural Research of The National Institute on Minority Health and Health Disparities. Charmaine Chan, Kristyn Kamke, and Zahra Ansari reviewed the transcripts against audio files. Rahaf Zaza extracted quotes in an exploratory phase of the analysis. This work was supported by the Division of Intramural Research, National Institute on Minority Health and Health Disparities (ZIA MD000011).

Disclaimer

The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the National Institute on Minority Health and Health Disparities; the National Institutes of Health; or the US Department of Health and Human Services.

Data Availability

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Authors' Contributions

SEL was responsible for conceptualization, methodology, funding acquisition, and project administration. Data curation was carried out by TG, EL, and SEL. MW and LT conducted the formal analysis. GB and GG developed the software. MW drafted the original manuscript. LT, GB, GG, TG, EL, BK, and SEL reviewed and edited the manuscript. BK also contributed to project administration.

Conflicts of Interest

GB, TG, GG, EL, and BK are all employed by ICF. All other authors have no conflicts of interest, financial or otherwise, to declare.

Multimedia Appendix 1

XSL•FO
RenderX

Supplementary files.
[DOCX File, 65 KB - games_v13i1e59515_app1.docx]

References

1. Field M, Cox WM. Attentional bias in addictive behaviors: a review of its development, causes, and consequences. Drug Alcohol Depend 2008 Sep 1;97(1-2):1-20. [doi: 10.1016/j.drugalcdep.2008.03.030] [Medline: 18479844]
2. Tan YZ, Zhang MWB, Choo ACC. Perspectives on modifying attentional biases amongst individuals with tobacco use disorder using technology: a review. Int J Environ Res Public Health 2019 Jul 24;16(15):2644. [doi: 10.3390/ijerph16152644] [Medline: 31344921]
3. Cortese BM, Uhde TW, LaRowe SD, et al. Olfactory cue reactivity in nicotine-dependent adult smokers. Psychol Addict Behav 2015 Mar;29(1):91-96. [doi: 10.1037/adb0000018] [Medline: 25180553]
4. Shiffman S, Dunbar M, Kirchner T, et al. Smoker reactivity to cues: effects on craving and on smoking behavior. J Abnorm Psychol 2013 Feb;122(1):264-280. [doi: 10.1037/a0028339] [Medline: 22708884]
5. Kang Y, Cappella JN, Strasser AA, Lerman C. The effect of smoking cues in antismoking advertisements on smoking urge and psychophysiological reactions. Nicotine Tob Res 2009 Mar;11(3):254-261. [doi: 10.1093/ntr/ntn033] [Medline: 19251767]
6. Hutchison KE, Niaura R, Swift R. Smoking cues decrease prepulse inhibition of the startle response and increase subjective craving in humans. Exp Clin Psychopharmacol 1999 Aug;7(3):250-256. [doi: 10.1037//1064-1297.7.3.250] [Medline: 10472513]
7. Rehme AK, Bey K, Frommann I, et al. Selective attention to smoking cues in former smokers. Eur Neuropsychopharmacol 2018 Feb;28(2):276-284. [doi: 10.1016/j.euroneuro.2017.12.003] [Medline: 29371023]
8. Masiero M, Lucchiari C, Maisonneuve P, Pravettoni G, Veronesi G, Mazzocco K. The attentional bias in current and former smokers. Front Behav Neurosci 2019;13:154. [doi: 10.3389/fnbeh.2019.00154] [Medline: 31354446]
9. Powell J, Dawkins L, West R, Powell J, Pickering A. Relapse to smoking during unaided cessation: clinical, cognitive and motivational predictors. Psychopharmacology (Berl) 2010 Dec;212(4):537-549. [doi: 10.1007/s00213-010-1975-8] [Medline: 20703450]
10. Attwood AS, O'Sullivan H, Leonards U, Mackintosh B, Munafò MR. Attentional bias training and cue reactivity in cigarette smokers. Addiction 2008 Nov;103(11):1875-1882. [doi: 10.1111/j.1360-0443.2008.02335.x] [Medline: 19032536]
11. Fiore MC, Jaén CR, Baker TB, Bailey WC, Benowitz NL, Curry SJ. Treating Tobacco Use and Dependence: 2008 Update Clinical Practice Guideline: U.S. Department of Health and Human Services, Public Health Service; 2008. URL: https://www.ncbi.nlm.nih.gov/books/NBK63952/ [accessed 2025-05-06]
12. Lochbuehler K, Voogd H, Scholte RHJ, Engels R. Attentional bias in smokers: exposure to dynamic smoking cues in contemporary movies. J Psychopharmacol 2011 Apr;25(4):514-519. [doi: 10.1177/0269881110388325] [Medline: 21098549]
13. El-Toukhy SM, Choi K. Magazine hyped: trends in tobacco advertising and readership characteristics, 2010-2014. Prev Med 2016 Oct;91:132-137. [doi: 10.1016/j.ypmed.2016.08.017] [Medline: 27519170]
14. Hamilton-Moseley KR, McNeel TS, Choi K. Cumulative cigarette discount coupon exposure and trajectories of cigarette smoking: a longitudinal analysis in US adults. Tob Control 2024 Sep 25;33(5):565-569. [doi: 10.1136/tc-2022-057801] [Medline: 37015744]
15. Mills SD, Kong AY, Reimold AE, Baggett CD, Wiesen CA, Golden SD. Sociodemographic disparities in tobacco retailer density in the United States, 2000-2017. Nicotine Tob Res 2022 Jul 13;24(8):1291-1299. [doi: 10.1093/ntr/ntac020] [Medline: 35079790]
16. Ribisl KM, D'Angelo H, Feld AL, et al. Disparities in tobacco marketing and product availability at the point of sale: results of a national study. Prev Med 2017 Dec;105:381-388. [doi: 10.1016/j.ypmed.2017.04.010] [Medline: 28392252]
17. US Department of Health and Human Services. Smoking Cessation: A Report of the Surgeon General: Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health; 2020. URL: https://www.hhs.gov/sites/default/files/2020-cessation-sgr-full-report.pdf [accessed 2025-05-06]
18. Jones EB, Sharpe L. Cognitive bias modification: a review of meta-analyses. J Affect Disord 2017 Dec 1;223:175-183. [doi: 10.1016/j.jad.2017.07.034] [Medline: 28759865]
19. MacLeod C, Rutherford E, Campbell L, Ebsworthy G, Holker L. Selective attention and emotional vulnerability: assessing the causal basis of their association through the experimental manipulation of attentional bias. J Abnorm Psychol 2002 Feb;111(1):107-123. [Medline: 11866165]
20. Boffo M, Zerhouni O, Gronau QF, et al. Cognitive bias modification for behavior change in alcohol and smoking addiction: Bayesian meta-analysis of individual participant data. Neuropsychol Rev 2019 Mar;29(1):52-78. [doi: 10.1007/s11065-018-9386-4] [Medline: 30644025]
21. Lopes FM, Pires AV, Bizarro L. Attentional bias modification in smokers trying to quit: a longitudinal study about the effects of number of sessions. J Subst Abuse Treat 2014 Jul;47(1):50-57. [doi: 10.1016/j.jsat.2014.03.002] [Medline: 24666812]

XSL•FO
RenderX

22. Heitmann J, Bennik EC, van Hemel-Ruiter ME, de Jong PJ. The effectiveness of attentional bias modification for substance use disorder symptoms in adults: a systematic review. Syst Rev 2018 Oct 13;7(1):160. [doi: 10.1186/s13643-018-0822-6] [Medline: 30316302]
23. Mobile fact sheet. Pew Research Center. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2024-02-06]
24. Cox M, Intrilligator J, Hillier C. Chimpshop and alcohol reduction - using technology to change behaviour. Perspect Public Health 2015 May;135(3):126-127. [doi: 10.1177/1757913915580926] [Medline: 25925306]
25. Garnett C, Crane D, West R, Brown J, Michie S. The development of Drink Less: an alcohol reduction smartphone app for excessive drinkers. Transl Behav Med 2019 Mar 1;9(2):296-307. [doi: 10.1093/tbm/iby043] [Medline: 29733406]
26. Zhang M, Ying J, Song G, Fung DS, Smith H. Attention and cognitive bias modification apps: review of the literature and of commercially available apps. JMIR Mhealth Uhealth 2018 May 24;6(5):e10034. [doi: 10.2196/10034] [Medline: 29793899]
27. Choo CC, Tan YZ, Zhang MWB. A smartphone app for attentional bias retraining in smokers: mixed methods pilot study. JMIR Form Res 2022 Jan 3;6(1):e22582. [doi: 10.2196/22582] [Medline: 34982037]
28. Robinson JD, Cui Y, Linares Abrego P, et al. Sustained reduction of attentional bias to smoking cues by smartphone-delivered attentional bias modification training for smokers. Psychol Addict Behav 2022 Nov;36(7):906-919. [doi: 10.1037/adb0000805] [Medline: 35025555]
29. Forray A, Gunter-Riley RG, Maltz C, Waters AJ. Pilot study of attentional retraining for postpartum smoking relapse. Front Psychiatry 2023;14:1231702. [doi: 10.3389/fpsyt.2023.1231702] [Medline: 37900298]
30. Zhang M, Ying J, Song G, Fung DS, Smith H. Gamified cognitive bias modification interventions for psychiatric disorders: review. JMIR Ment Health 2018 Oct 25;5(4):e11640. [doi: 10.2196/11640] [Medline: 30361194]
31. Michie S, Richardson M, Johnston M, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
32. Michie S, Hyder N, Walia A, West R. Development of a taxonomy of behaviour change techniques used in individual behavioural support for smoking cessation. Addict Behav 2011 Apr;36(4):315-319. [doi: 10.1016/j.addbeh.2010.11.016] [Medline: 21215528]
33. Tesfaye L, Wakeman M, Baskin G, et al. A feature-based qualitative assessment of smoking cessation mobile applications. PLOS Digit Health 2024 Nov;3(11):e0000658. [doi: 10.1371/journal.pdig.0000658] [Medline: 39571041]
34. Venkatesh V, Morris MG, Davis BG, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003;27(3):425-478. [doi: 10.2307/30036540]
35. Venkatesh V, Thong JYL, Xu X. Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. MIS Q 2012;36(1):157-178. [doi: 10.2307/41410412]
36. Landoll RR, Vargas SE, Samardzic KB, Clark MF, Guastaferro K. The preparation phase in the multiphase optimization strategy (MOST): a systematic review and introduction of a reporting checklist. Transl Behav Med 2022 Feb 16;12(2):291-303. [doi: 10.1093/tbm/ibab146] [Medline: 34850214]
37. Collins LM, Baker TB, Mermelstein RJ, et al. The multiphase optimization strategy for engineering effective tobacco use interventions. Ann Behav Med 2011 Apr;41(2):208-226. [doi: 10.1007/s12160-010-9253-x] [Medline: 21132416]
38. Galobardes B, Shaw M, Lawlor DA, Lynch JW, Davey Smith G. Indicators of socioeconomic position (part 1). J Epidemiol Community Health 2006 Jan;60(1):7-12. [doi: 10.1136/jech.2004.023531] [Medline: 16361448]
39. Cao P, Jeon J, Tam J, et al. Smoking disparities by level of educational attainment and birth cohort in the U.S. Am J Prev Med 2023 Apr;64(4 Suppl 1):S22-S31. [doi: 10.1016/j.amepre.2022.06.021] [Medline: 36935129]
40. Adler NE, Newman K. Socioeconomic disparities in health: pathways and policies. Health Aff (Millwood) 2002;21(2):60-76. [doi: 10.1377/hlthaff.21.2.60] [Medline: 11900187]
41. Shavers VL. Measurement of socioeconomic status in health disparities research. J Natl Med Assoc 2007 Sep;99(9):1013-1023. [Medline: 17913111]
42. Hennink M, Kaiser BN. Sample sizes for saturation in qualitative research: a systematic review of empirical tests. Soc Sci Med 2022 Jan;292:114523. [doi: 10.1016/j.socscimed.2021.114523] [Medline: 34785096]
43. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
44. Braun V, Clarke V. Thematic analysis. In: Cooper PM, Camic PM, Long DL, Panter AT, Rindskopf D, Sher KJ, editors. APA Handbook of Research Methods in Psychology, Vol 2: Research Designs: Quantitative, Qualitative, Neuropsychological, and Biological American Psychological Association: American Psychological Association; 2012:57-71. [doi: 10.1037/13620-004]
45. Miles MB, Huberman AM, Saldana J. Qualitative Data Analysis: A Methods Sourcebook, 3rd edition: SAGE Publications; 2013. URL: https://collegepublishing.sagepub.com/products/qualitative-data-analysis-4-246128 [accessed 2025-05-06]
46. Tesfaye L, Wakeman M, Gregory T, et al. Acceptability of a smart lighter for tracking cigarette smoking: a focus group study. Digit Health 2025;11:20552076251323998. [doi: 10.1177/20552076251323998] [Medline: 40151637]

XSL•FO
RenderX

47. Wakeman M, Tesfaye L, Gregory T, Leahy E, Kendrick B, El-Toukhy S. Perceptions of the use of mobile technologies for smoking cessation: focus group study with individuals of low socioeconomic status who smoke. JMIR Form Res 2024 Oct 11;8:e58221. [doi: 10.2196/58221] [Medline: 39392684]
48. Friese S. ATLAS.ti 8 windows user manual. : ATLASTi Scientific Software Development GmbH; 2020 URL: https://downloads.atlasti.com/docs/manual/atlasti_v8_manual_en.pdf [accessed 2025-04-29]
49. Zhang MWB, Heng S, Song G, Fung DSS, Smith HE. Development of a co-designed gamified attention bias modification app for individuals with opioid use disorder. Technol Health Care 2021;29(3):521-526. [doi: 10.3233/THC-202522] [Medline: 33074203]
50. Clement J. Share of adults in the United States who play games on a smartphone or tablet once a week or more as of March 2024, by age group. Statista. 2024. URL: https://www.statista.com/statistics/248892/frequency-of-using-smartphones-to-play-games-in-the-us/#:~:text=An%20October%202022%20survey%20found,with%20aged%20between%2035%2D54 [accessed 2024-02-13]
51. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11. [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
52. White JS, Salem MK, Toussaert S, et al. Developing a game (Inner Dragon) within a leading smartphone app for smoking cessation: design and feasibility evaluation study. JMIR Serious Games 2023 Aug 11;11:e46602. [doi: 10.2196/46602] [Medline: 37566442]
53. Bindoff I, Ling TR, Gee P, Geelan B, Ferguson SG, Peterson GM. Effects of a mobile app called Quittr, which utilizes premium currency and games features, on improving engagement with smoking cessation intervention: pilot randomized controlled trial. JMIR Serious Games 2020 Dec 14;8(4):e23734. [doi: 10.2196/23734] [Medline: 33315016]
54. El-Hilly AA, Iqbal SS, Ahmed M, et al. Game on? Smoking cessation through the gamification of mHealth: a longitudinal qualitative study. JMIR Serious Games 2016 Oct 24;4(2):e18. [doi: 10.2196/games.5678] [Medline: 27777216]
55. Goodwin S, Nastasi JA, Newman ST, Rapoza D, Raiff BR. Feasibility and acceptability of a mobile game to support smoking cessation: repeated measures study. JMIR Serious Games 2024 Aug 21;12:e54684. [doi: 10.2196/54684] [Medline: 39167443]
56. Zhang M, Heng S, Song G, Fung DS, Smith HE. Co-designing a mobile gamified attention bias modification intervention for substance use disorders: participatory research study. JMIR Mhealth Uhealth 2019 Oct 3;7(10):e15871. [doi: 10.2196/15871] [Medline: 31584003]
57. Zhang M, Wolters M, O'Connor S, Wang Y, Doi L. Smokers' user experience of smoking cessation apps: a systematic review. Int J Med Inform 2023 Jul;175:105069. [doi: 10.1016/j.ijmedinf.2023.105069] [Medline: 37084673]
58. Derksen ME, van Strijp S, Kunst AE, Daams JG, Jaspers MWM, Fransen MP. Serious games for smoking prevention and cessation: a systematic review of game elements and game effects. J Am Med Inform Assoc 2020 May 1;27(5):818-833. [doi: 10.1093/jamia/ocaa013] [Medline: 32330255]
59. Vilardaga R, Casellas-Pujol E, McClernon JF, Garrison KA. Mobile applications for the treatment of tobacco use and dependence. Curr Addict Rep 2019 Jun;6(2):86-97. [doi: 10.1007/s40429-019-00248-0] [Medline: 32010548]
60. Betts JM, Dowd AN, Forney M, Hetelekides E, Tiffany ST. A meta-analysis of cue reactivity in tobacco cigarette smokers. Nicotine Tob Res 2021 Jan 22;23(2):249-258. [doi: 10.1093/ntr/ntaa147] [Medline: 32772094]
61. Field M, Christiansen P. Commentary on, "Internal reliability of measures of substance-related cognitive bias". Drug Alcohol Depend 2012 Aug 1;124(3):189-190. [doi: 10.1016/j.drugalcdep.2012.02.009] [Medline: 22386685]
62. Beard C, Weisberg RB, Primack J. Socially anxious primary care patients' attitudes toward cognitive bias modification (CBM): a qualitative study. Behav Cogn Psychother 2012 Oct;40(5):618-633. [doi: 10.1017/S1352465811000671] [Medline: 22127022]
63. Kuckertz JM, Schofield CA, Clerkin EM, et al. Attentional bias modification for social anxiety disorder: what do patients think and why does it matter? Behav Cogn Psychother 2019 Jan;47(1):16-38. [doi: 10.1017/S1352465818000231] [Medline: 29729676]
64. Bergen N, Labonté R. "Everything is perfect, and we have no problems": detecting and limiting social desirability bias in qualitative research. Qual Health Res 2020 Apr;30(5):783-792. [doi: 10.1177/1049732319889354] [Medline: 31830860]

Abbreviations

ABM: attentional bias modification
CBM: cognitive bias modification

XSL•FO
RenderX

Edited by A Coristine, T Leung; submitted 15.04.24; peer-reviewed by A Eisingerich; revised version received 05.02.25; accepted 05.02.25; published 09.05.25.

Please cite as:
Wakeman M, Tesfaye L, Baskin G, Gregory T, Gruse G, Leahy E, Kendrick B, El-Toukhy S
Perceptions Toward an Attentional Bias Modification Mobile Game Among Individuals With Low Socioeconomic Status Who Smoke: Qualitative Study
JMIR Serious Games 2025;13:e59515
URL: https://games.jmir.org/2025/1/e59515
doi:10.2196/59515

© Michael Wakeman, Lydia Tesfaye, Gunnar Baskin, Tim Gregory, Greg Gruse, Erin Leahy, Brandon Kendrick, Sherine El-Toukhy. Originally published in JMIR Serious Games (https://games.jmir.org), 9.5.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Exergame (ExerG)-Based Physical-Cognitive Training for Rehabilitation in Adults With Motor and Balance Impairments: Usability Study

Silvia Herren[1*], MSc; Barbara Seebacher[2,3,4*], PhD; Sarah Mildner[2], MSc; Yanick Riederer[5], MSc; Ulrike Pachmann[6], MBA, MSc; Nija Sonja Böckler[7], MA; Stephan Niedecken[5], Dipl; Sabrina Alicia Sgandurra[8], MA; Leo Bonati[1,9,10], MD; Isabella Hotz[2], MSc; Alexandra Schättin[5], PhD; Roman Jurt[7], BA; Christian Brenneis[4,11], MD; Katharina Lenfert[2], MA; Frank Behrendt[1,12], PhD; Stefan Schmidlin[7], MA, MSc; Lennart Nacke[13], PhD; Corina Schuster-Amft[1,12,14*], PhD; Anna Lisa Martin-Niedecken[5,7*], PhD

[1]Research Department, Reha Rheinfelden, Rheinfelden, Switzerland
[10]Department of Clinical Research, University of Basel, Basel, Switzerland
[11]Department of Neurology, Clinic for Rehabilitation Muenster, Muenster, Austria
[12]School of Engineering and Computer Science, Bern University of Applied Sciences, Burgdorf, Switzerland
[13]HCI Games Group, Stratford School of Interaction Design and Business, University of Waterloo, Waterloo, ON, Canada
[14]Department for Sport, Exercise and Health, University of Basel, Basel, Switzerland
[2]Department of Rehabilitation Science, Clinic for Rehabilitation Muenster, Gröben 700, Muenster, Austria
[3]Clinical Department of Neurology, Medical University of Innsbruck, Innsbruck, Austria
[4]Karl Landsteiner Institute for Interdisciplinary Rehabilitation Research, Muenster, Austria
[5]Sphery Ltd, Zurich, Switzerland
[6]VASCage – Centre on Clinical Stroke Research, Innsbruck, Austria
[7]Department of Design, Institute for Design Research, Zurich University of the Arts, Zurich, Switzerland
[8]English Language and Literature, Faculty of Arts, University of Waterloo, Waterloo, ON, Canada
[9]Stroke Center and Department of Neurology, University Hospital Basel, Basel, Switzerland
[*]these authors contributed equally

Corresponding Author:
Barbara Seebacher, PhD
Department of Rehabilitation Science, Clinic for Rehabilitation Muenster, Gröben 700, Muenster, Austria

Related Article:

This is a corrected version. See correction statement: https://games.jmir.org/2025/1/e73405

Abstract

Background: Exergames are increasingly used in rehabilitation, yet their usability and user experience for patients and therapists, particularly for functional model systems, are underresearched. The diverse needs and preferences of users make conducting usability studies challenging, emphasizing the need for further investigation in real-world settings.

Objective: This study aimed to evaluate the usability, safety, and user experience of a novel exergame functional model, the ExerG, from the perspectives of patients and therapists in a rehabilitation setting.

Methods: In this mixed methods study, 15 patients undergoing rehabilitation (primary end users [PEUs]) and 20 therapists (secondary end users [SEUs]) from 2 rehabilitation centers in Switzerland and Austria participated in exercising and observation sessions with the ExerG. SEUs received training on system use and technical issue management, enabling them to fulfill their therapist roles while treating patients or mock patients. Rapid Iterative Testing and Evaluation was used and the training software adjusted based on participant feedback. Usability was assessed with questionnaires, semistructured interviews, and through observations during the ExerG testing. System acceptability was evaluated using specific quantitative thresholds based on PEU performance and feedback. An observation protocol tracked SEUs' correct use, errors, hesitations, task completion time, and needed assistance across scenarios.

Results: Patients and therapists reported overall good usability and positive experiences with the exergame. PEUs rated 23/29 (79%) instructions as acceptable, showed good-to-very-good exercise performance in 19/29 (65%) tasks, and completed 28/29

XSL•FO
RenderX

(97%) tasks. Patients reported no adverse events, showing improved performance and enjoyment across ExerG exercising rounds, with 79/90 (88%) expressing positive emotions and reporting median scores of 9 (IQR 7.5 - 10) on a 1 - 10 user satisfaction scale. Patients were willing to continue using the device if the graphic design was improved (5/15), tracking systems and projector quality were enhanced (each 3/15), instructions clarified (12/15), and the game variety increased (2/15). PEUs felt secure in the safety harness (15/15) but recommended swivel arm movement enhancements (5/15). SEUs effectively executed scenarios, with hesitation and difficulties observed in only 14/41 tasks and 2/41 tasks, across all 20 therapists, accounting for 1.7% and 0.2% of the 820 total task cases, respectively. Therapists' quantitative usability ratings were high (median System Usability Scale score 82.5, IQR 65 - 95). All SEUs expressed their willingness to use the ExerG (20/20) and reported being able to operate the system using the user handbook (20/20). They emphasized the motivation-enhancing effect of video-game based training (12/20) and considered the activities supportive for physical and cognitive skills (20/20). They suggested incorporating daily living task simulations (13/20), more customizable options (6/20), more targeted motivational feedback (9/20), clearer performance ratings (9/20), and more concise activity instructions (6/20).

Conclusions: The interdisciplinary, iterative ExerG development approach shows promise. The findings will inform future optimizations. Future work will assess long-term impact.

Trial Registration: ClinicalTrials.gov NCT05967078; https://clinicaltrials.gov/study/NCT05967078 and OSF Registries OSFCQ9AT; https://osf.io/cq9at

(JMIR Serious Games 2025;13:e66515) doi:10.2196/66515

KEYWORDS

exergame; rehabilitation; user-centered design; usability testing; mixed-methods; interdisciplinary research; concept functional model proofs; exercise; cognitive training; technology acceptance; motor; cognitive impairment; safety; user experience; balance impairments; balance

Introduction

The rapid advancement of digital technologies has revolutionized various aspects of health care, including rehabilitation. Among these innovations, exergames—video games that require physical activity to play—have emerged as promising tools for engaging patients in rehabilitative exercises. Exergames use physical activity and brain challenges to boost patient motivation, adherence, and results in rehabilitation settings [1,2].

Despite the growing interest in exergames for rehabilitation, there is a lack of research on their usability and user experience, particularly for functional model systems [3,4]. A functional model is a representation that describes how a system, device, product, or process operates in terms of its functions and the interactions between its components [5]. It focuses on the purposes of different parts and how they contribute to the overall objectives of the system [5]. Usability, which refers to the effectiveness, efficiency, and satisfaction with which users can achieve specific goals in a particular environment [6,7], is a critical factor in the adoption and success of rehabilitation technologies. Poor usability can lead to frustration, reduced motivation, and even abandonment of the technology [8]. Similarly, user experience, which encompasses the user's perceptions, emotions, and responses to the technology [9], can significantly influence patient engagement and adherence to rehabilitative interventions [8,10].

Evaluating the usability and user experience of exergame functional models is essential for iterative design and optimization. Functional model testing allows developers to identify and address usability issues early in the development process, reducing the risk of costly redesigns and improving the final product's quality [11]. However, conducting usability studies with rehabilitation populations presents challenges due to the diverse needs, abilities, and preferences of both patients and therapists [12]. Key considerations include safety, training goals, individuality, game environment, social interactions, and physical and technical overload for patients, as well as facets like meaningfulness, distractions from the game environment, safety, gamification elements, and the availability and accessibility of the exergame for therapists [13]. Therefore, further investigation into the usability and user experience of exergame functional models in real-world rehabilitation settings is essential.

To address this gap, we developed a novel exergame functional model called ExerG, designed for use in rehabilitation. The ExerG incorporates a range of physical-cognitive training elements, including balance, coordination, and dual-tasking exercises, which are crucial for rehabilitation [14]. The development of the ExerG followed an interdisciplinary, iterative approach, involving collaboration between game designers, rehabilitation experts, and primary end users (PEUs; patients) and secondary end users (SEUs; therapists) [15]. This approach allows for the integration of clinical knowledge, user feedback, and technical expertise to create a user-centered design that meets the specific needs of the rehabilitation context [16,17].

The primary aim of this study was to evaluate the usability, safety, performance, and user experience of the ExerG functional model in a rehabilitation setting. We hypothesized that ExerG would demonstrate good overall usability and positive user experiences for both patients and therapists, despite the challenges associated with its functional model status. Additionally, we aimed to identify areas for optimization and inform future iterations of the ExerG system [12,18].

XSL•FO
RenderX

Methods

Study Design

To achieve this study's objectives, we conducted a mixed methods study involving patients (PEUs) and therapists (SEUs) from 2 rehabilitation centers in Switzerland and Austria. This summative usability study employed a convergent mixed methods design to comprehensively evaluate the usability, safety, performance, and user experience of ExerG, a novel training solution for rehabilitation in adults with motor and balance impairments [12,18]. Summative usability testing is conducted to assess whether intended users can use the device safely and effectively in its intended environments [19]. The Good Reporting of a Mixed Methods Study (GRAMMS) framework [20] was adhered to (Multimedia Appendix 1). Our testing procedures adhered to the International Organization for Standardization's guidelines for usability engineering of medical devices (IEC 62366-1:2015 and IEC TR 62366 - 2). We used the Rapid Iterative Testing and Evaluation Method [21], which enabled swift adjustments to the training software based on participant feedback. This approach enabled continuous refinement of the system throughout this study, ensuring responsiveness to user needs and enhancing the ExerG's overall usability.

Exercise and Test Items

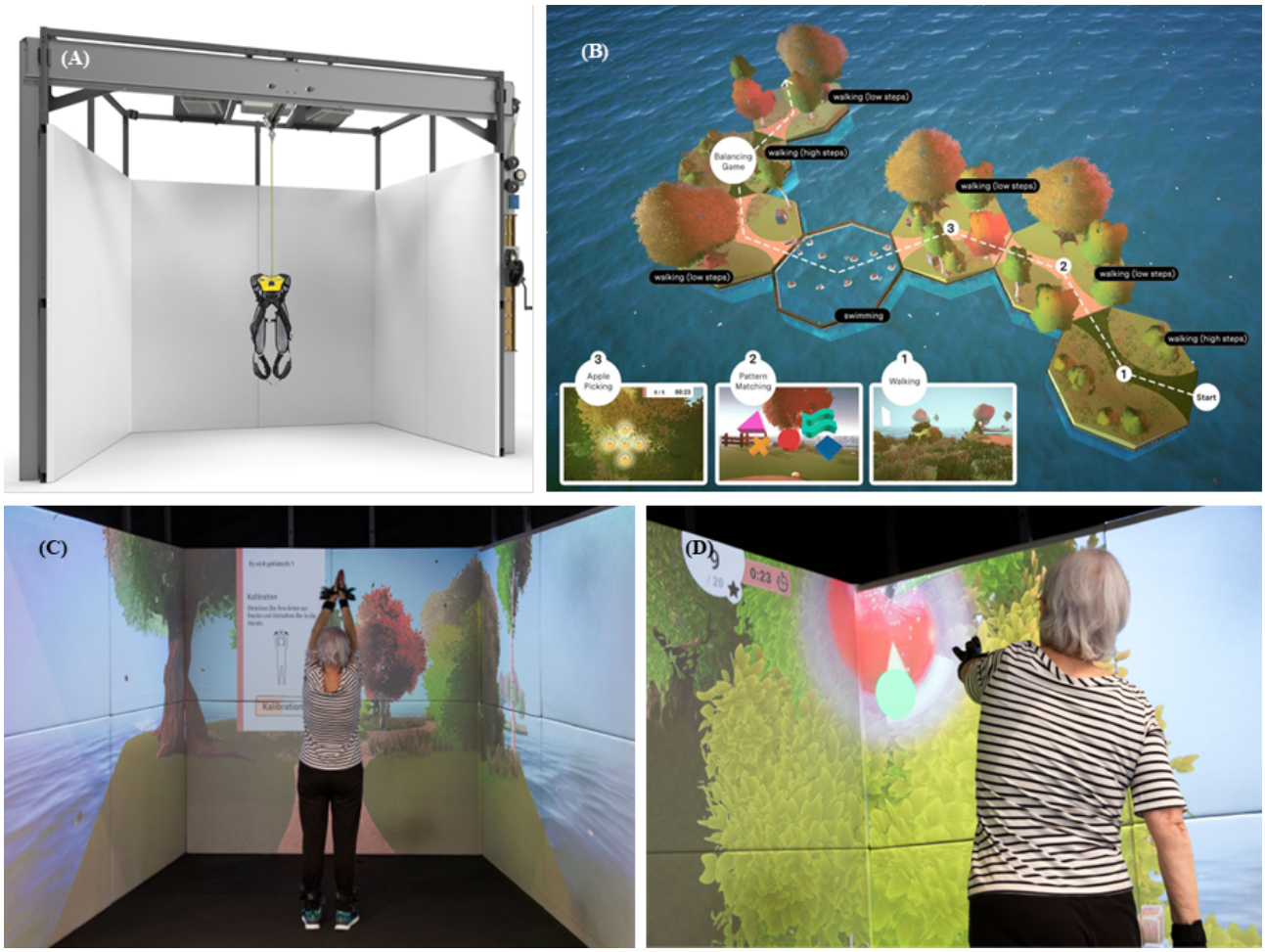
The new training solution comprises the ExerG training software for patients, the ExerCube supporting material, and the safety harness for the ExerG. The ExerG training solution is intended to be operated by professionals (therapists/SEUs) and used by patients (PEUs). The ExerG exercise program integrates a variety of physical-cognitive training elements, including balance, coordination, and dual-tasking, through engaging mini-games such as apple picking, pattern matching, rowing, trunk jumping, and balloon catching. These activities are designed to enhance coordinative skills by refining precise spatial and temporal movements of the legs and arms, as well as improving the control and regulation of motor activity in the central nervous system. They also promote conditional abilities by requiring dynamic balance during fast leg movements to help prevent falls. Therapists can select exercises, personalize settings before each session, and make real-time adjustments to tailor the program to each individual's needs. Training intensities vary from easy to hard.

The exercise program features a range of activities that progress in difficulty. Walking variations include low to high leg lifts at different speeds and dual-task walking with complex arm movements. Side steps incorporate arm movements and responses to stimuli, while front steps and lunges vary from small to large movements, and squat depths increase. Jumping exercises advance to include jumps that break floor contact, and the swaying tree pose becomes more challenging. Balance exercises include single-leg stances and full leg lifts, with symmetrical and asymmetrical arm tasks, along with body rotations. Swimming arm movements speed up, and challenges involve reaching for wall stimuli. Balloon-catching activities vary in speed, while obstacle jumps progress from galloping to standard jumps, and fruit sorting speeds change from slow to fast (Multimedia Appendix 2).

The ExerG training software and the safety harness are complementary and adapted technologies of the ExerCube, an already existing immersive fitness training device [22] (Multimedia Appendix 2). In combination with the safety harness (Petzl; Figure 1A), its purpose is to enable training in a safe environment and mitigate the risk of patients falling (Figure 1B-D).

XSL•FO
RenderX

Figure 1. (A) Overview of the ExerG hardware, (B) scenarios overview, (C) trackers calibration of the system for the user, and (D) the apple picking scenario.



Patients and Participants

Participant Recruitment

Recruitment took place from May 16, 2023, to October 31, 2023. Using convenience sampling, our study involved 2 user groups: patients (PEUs) and therapists (SEUs, participants), from rehabilitation centers in Switzerland (Reha Rheinfelden) and Austria (Clinic for Rehabilitation Muenster).

Primary End Users

Patients from inpatient, outpatient, and Neuro Day Care units were considered at 1 of the 2 centers. Recruitment involved screening entry lists, reviewing treated patients, distributing flyers, emailing outpatient physiotherapists, and engaging therapists. Medical records were screened for basic eligibility, and eligible patients received verbal and written study information. They had at least 24 hours to decide and sign the consent form if they were interested. See Textbox 1 for eligibility criteria. For more details on the Mini Mental State Examination (MMSE) and Berg Balance Scale, see Multimedia Appendix 3.

XSL•FO
RenderX

Textbox 1. Eligibility criteria of primary end users.

Inclusion criteria

- Aged ≥18 years
- Undergoing inpatient or outpatient rehabilitation
- Motor impairment of the upper or lower extremities, gait disorder, balance disorder, visual-spatial disorder, or cognitive disorder due to any disease
- Ability to speak and comprehend German and understand digitally transmitted training instructions
- No prior training experience with the device
- Body height ranging from 160 cm to 200 cm, in accordance with hardware specifications
- Body weight not exceeding 120 kg, as per the safety harness specifications
- Capable of performing the movements required for video game-based training. This includes independently transitioning from a sitting to a standing position (scoring ≥3 on the Berg Balance Scale [BBS] [23] item 1, "sitting to standing") and maintaining the standing position without assistance (scoring ≥3 on the BBS item 2, "standing unsupported")
- Capable of participating in the entire clinical investigation, as determined by this study's principal investigator

Exclusion criteria

- Moderate-to-severe cognitive impairment as defined by a Mini Mental State Examination (MMSE) [24] score of ≤18 [25]
- Known cybersickness
- Severe visual, neurological, cardiorespiratory, psychiatric, or orthopedic impairments that reduce a person's ability to follow instructions or play the games
- Epileptic seizures within the past 3 months
- Recent surgery, fractures, joint replacement, or malignancy within the past 3 months
- Impairment of hearing resulting in the inability to engage in verbal communication
- Severe movement pain exceeding 5 on the 11-point pain Numeric Rating Scale (NRS) [26]
- Physical conditions that prevent the proper wearing of the safety harness or could result in pain or health complications, such as skin lesions or open wounds (eg, severe osteoporosis)
- Joint contractures (eg, in the shoulder, knee, or hip) that may result in limitations during end user evaluation of the ExerG
- Severe neurological conditions (eg, severe epilepsy, advanced Parkinson disease, or condition after severe stroke)
- Severe psychiatric conditions (eg, pronounced paranoid states or severe depression)
- Terminal illness (estimated life expectancy <12 months)

Secondary End Users

SEU recruitment included physiotherapists, occupational therapists, training therapists, and sports scientists from the rehabilitation centers and external clinics. This diverse sampling approach ensured a comprehensive evaluation of the ExerG across various settings and professional perspectives. SEUs were required to be at least aged 21 years and hold a therapist qualification or a relevant bachelor's degree.

Sample Size

This usability study sample included 15 PEUs and 20 SEUs. The SEU group emphasized diverse recruitment, engaging staff from various therapeutic disciplines with distinct specializations. Further, 5 therapists tested the ExerG in real-world scenarios with inpatients and outpatients. To balance sample sizes, an additional 10 PEUs were enrolled. Saturation was evaluated using code meaning, which entailed reviewing each interview to document identified codes and checking subsequent interviews for any new aspects, dimensions, or nuances of those codes until saturation was reached with no new information [27].

Study Procedures

Overview

PEUs participated in an initial eligibility evaluation and baseline data collection, followed by a single session with 2 exercising rounds to test the ExerG and safety harness. SEUs, who were already familiar with technology due to their daily work, received training on system usage and technical issue management. This training included detailed explanations of their responsibilities while treating patients or mock patients during the evaluation session, with real individuals acting as stand-ins for actual patients (see the figure in the Results section for a flow chart).

Evaluation Session for Primary End Users

PEUs attended a single study visit with 2 training rounds. At the start of each training round, participants received an overview of the training content and procedures. Researchers

XSL•FO
RenderX

collected demographic and disease-specific data and conducted baseline assessments. Patients were then briefed on the gaming activities and fitted with safety harnesses and wrist and ankle motion sensors. Once secured, the first training round began, lasting approximately 10 - 15 minutes, depending on the patient's physical and cognitive capabilities.

Evaluation Session for Secondary End Users

SEUs from both centers participated in a study visit that included ExerG usage training and 2 observation rounds. Before the training, SEUs received the system manual and were advised to read it thoroughly. During the session, they attended an introduction and training on system usage, following a standardized protocol for explanation and demonstration, and practiced procedures for memorization.

In the usage observation rounds 1 and 2, SEUs set up the system, designed a physical-cognitive training session using the software's exergames, prepared a patient with training materials, and guided them through the session. This study's team observed and documented the process. At the conclusion, SEUs completed rating scales and questionnaires, with the key difference being that 1 center used mock patients during the observation rounds.

Data Collection

Overview

Data were collected during and after the PEU and SEU study visits using an observation protocol, questionnaires, and interviews. Quantitative data on usability, user experience, and acceptance were gathered with standardized rating scales. Qualitative measures, including semistructured interviews and specific questions about ExerG training scenarios, provided a more comprehensive assessment.

Primary Outcomes

This usability study focused on system interaction and safety, encompassing both software and hardware as the primary outcome. Adverse events were monitored throughout this study. To identify difficulties, problems, or errors during use, the team followed a detailed observation protocol with predefined categories. Both end user groups provided objective measurements (scale ratings) and subjective feedback as part of the protocol. Performance was assessed by calculating the proportion of participants who completed or met the criteria for each exercise. The test moderator evaluated exercise performance using an observation sheet with the following ratings: very good, good, half-half, poor, and very poor. Performance was categorized as "very good" if the sum of "very good" and "good" ratings was ≥85%. Tasks were aligned with the training program flow, assessing instruction clarity, exercise performance, task completion, and exercise intensity. The system's acceptability was evaluated using specific quantitative thresholds based on PEU performance and feedback. The system was deemed acceptable by meeting the following criteria: (1) training completion: at least 85% of PEUs successfully completed the first training round; (2) exercise skipping: fewer than 20% of exercises were skipped by PEUs; (3) instruction comprehensibility: at least 85% of PEUs rated their understanding of the training instructions as "well" or "very well"; and (4) performance: at least 85% of PEUs performed the physical and cognitive activities with "good" to "very good" performance ratings.

For SEUs, an observation protocol recorded correct use, errors, hesitations, task completion time, and needed assistance across 8 scenarios: hardware preparation, safety system preparation or attachment, fall simulation, initiating or supporting training, compiling training, ending training, removing safety system, and shutting down hardware. "Hesitation" implied uncertainty or a lack of confidence, while "difficulties" represented objective barriers (Multimedia Appendices 4 and 5).

Secondary Outcomes for Primary End Users

Following each training round, PEUs evaluated perceived physical exertion and mental effort using the Borg CR10 (Borg Category Ratio Scale) [28] and the Paas Mental Effort Rating Scale [29]. They assessed player experience using the Player Experience Inventory (PXI) [30,31], focusing on both functional and psychosocial aspects of the exergame (Multimedia Appendix 3). Semistructured interviews and observations were also conducted for deeper insights into exergame functionality.

At the end of their study visit, PEUs participated in a brief semistructured interview featuring both structured and open-ended questions to gather insights into their perceptions of game control and comfort. Throughout the exergaming training, this study's team used an observation sheet with predetermined categories for flow, motivation, and motion sickness, documenting all training activities in sequence. This sheet included checkboxes for ratings on:

- Level (easy, medium, or hard)
- Execution (yes, no, or skipped)
- Clarity of instructions (very good, good, half-half, poor, or very poor)
- Ability to execute the required movements (very good, good, half-half, poor, or very poor)
- Training intensity (too difficult, difficult, optimal, easy, or too easy)

Additionally, a study team member evaluated the perceived helpfulness of the exergame feedback using a rating scale (very good to very poor). A trained member also systematically observed and noted the patient's emotions during gameplay [32]. Any emotions and difficulties observed or reported by the patient during or after the training were recorded and discussed to analyze their causes and potential solutions [32] (Multimedia Appendix 3).

Secondary Outcomes for Secondary End Users

SEUs evaluated the ExerG using the System Usability Scale (SUS) [33], a common questionnaire for assessing technology usability. The SUS includes 10 items with alternating positive and negative statements, each rated on a 5-point Likert scale, and is widely used in human-computer interaction research. According to Bangor et al [34], SUS scores of 70 - 79 indicate good or acceptable usability, while scores of 80 - 89 indicate excellent usability.

After completing all use scenarios, any difficulties encountered were discussed with SEUs. A semistructured interview, featuring

XSL•FO
RenderX

both structured and open-ended questions, was conducted after the second ExerG usage observation round to evaluate user-friendliness and comprehensively assess SEUs' experience with the device, focusing on exercise content, relevance to everyday life, and usability.

Data Analyses

Statistical Data Analysis

Demographic and usability metrics were analyzed using descriptive statistics. Absolute and relative frequencies were calculated for count and nominal data. Ordinal data (MMSE, Borg CR10, Paas Mental Effort Rating Scale, and SUS) are presented with medians and IQRs, which reflect the range between the 25th and 75th percentiles of the data. Continuous data are reported with means and SDs. The usability problem discovery rate was calculated by dividing the number of identified problems by the total number of possible problems and then multiplying by 100 to obtain a percentage. Statistical analysis was performed using IBM SPSS software (version 28.0; IBM Corp) and GraphPad Prism 9 (GraphPad Software).

Qualitative Data Analysis

The patients' responses from the brief interview were recorded and subsequently entered into an Excel (Microsoft Corp) spreadsheet for further analysis. Aligning with the principles of qualitative content analysis as outlined by Mayring [35,36], and following familiarization with the material, a coding scheme was developed inductively, based on the data and the research question. This process resulted in categories that were unidimensional, mutually exclusive, and exhaustive. Clear coding rules were established to represent recurring themes, ideas, and concepts within the texts. Each condensed phrase was assigned to one or more codes from the coding scheme, ensuring a thorough classification of segments. Through an iterative process, the codes and their assignments were reviewed and discussed by 2 researchers to enhance intersubjective conformability and ensure reliability [37]. The codes were applied to relevant segments of text, ensuring that each segment was accurately classified according to the established categories [37]. The frequencies of the codes were tallied, and the coded data were synthesized to create overarching themes that captured the key insights from the analysis. Themes were then described and illustrated with quotations accompanied by corresponding participant IDs.

Ethical Considerations

This study was approved by the Research Ethics Committee of Northwestern and Central Switzerland (IRB approval: amendment3_2022 - 00559), and informed consent was obtained from participants, who were able to opt out at any time. Participants' data were anonymized. In Austria, this study did not require Ethics Committee approval as it involved only therapists and focused on hardware and software handling. No compensation was provided to participants in this research study. This study was prospectively registered (ClinicalTrials.gov ID NCT05967078; OSF Preregistration OSF.IO/CQ9AT).

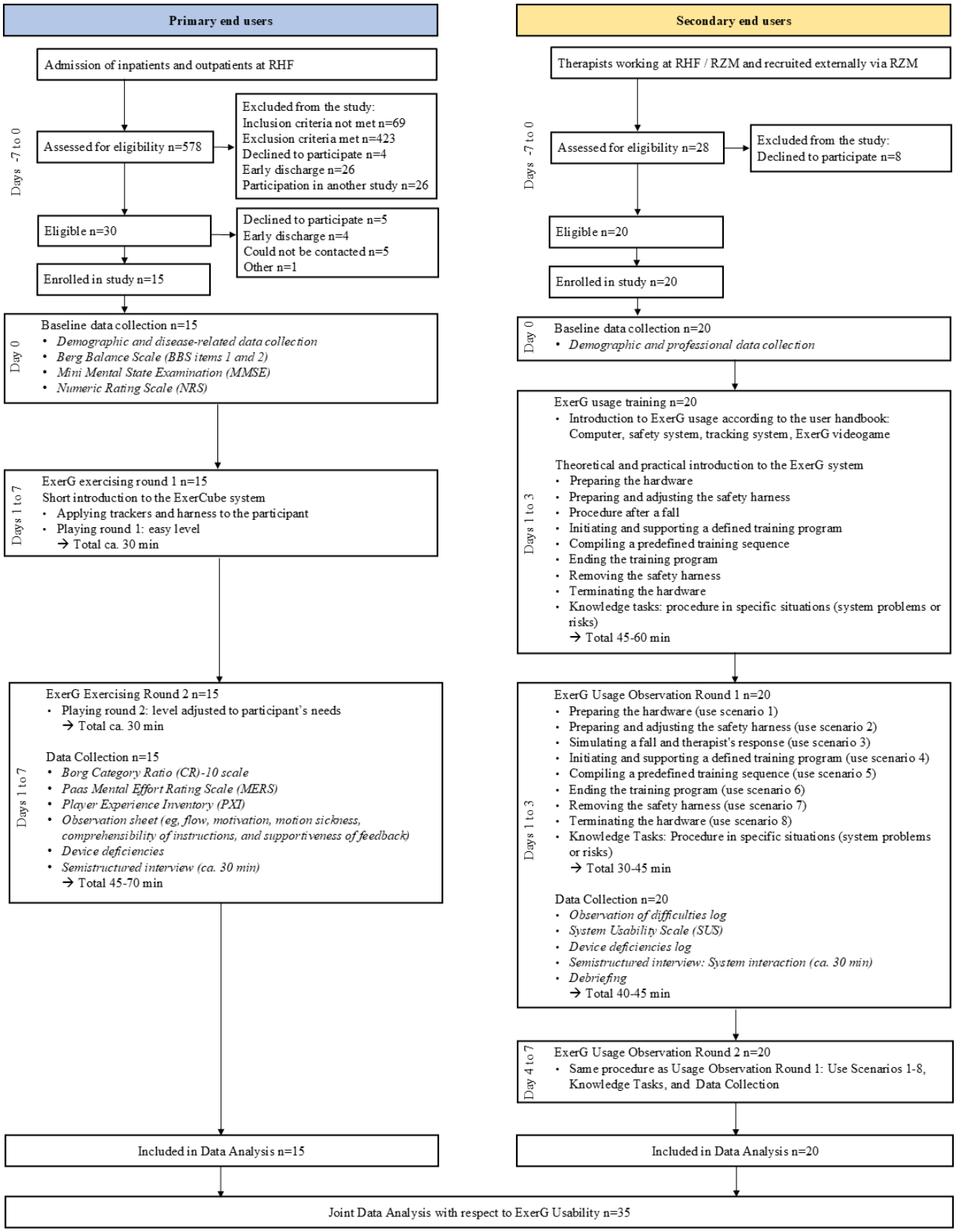
Results

Patients' and Participants' Characteristics

A total of 15 PEUs and 20 SEUs were recruited across both study centers (Figure 2).

XSL•FO
RenderX

Figure 2. Study flow diagram. RHF: Reha Rheinfelden; RZM: Reha Zentrum Muenster.



Primary End User Group

In total, 15 patients, with neurologist-assessed motor and balance impairment, exhibiting a mean age of 57.8 (SD 12.25) years, completed this study (Table 1). This study's population comprised 3 patients using walking aids and 5 with a history of falls. Among the PEUs, 11 (73%) had prior experience with technology-based training, primarily familiar with exergames and stationary weight-relieving systems. Only 1 (7%) PEU was acquainted with both mobile body-weight-supporting systems and virtual reality.

XSL•FO
RenderX

Table . Primary end user characteristics.

	Values (n=15)
Age (years), mean (SD; range)	57.8 (12.25; 40 - 78)
Sex, n (%)	
Male	9 (60)
Female	6 (40)
Mobility aids, n (%)	
None	12 (80)
Walking sticks	1 (6.7)
Walker	2 (13.3)
Fallers, n (%)	5 (33.3)
Nonfallers, n (%)	10 (66.7)
Falls in the previous 5 months, n (%)	
None	10 (66.7)
1 time	4 (26.7)
1 - 5 times	1 (6.7)
MMSE[a] score, median (IQR; range)	29 (28 - 30; 22 - 30)
Primary diagnoses, n (%)	
Multiple sclerosis	4 (26.7)
Lacunar thalamo-capsular stroke	2 (13.3)
Middle cerebral artery stroke	1 (6.7)
Brainstem stroke	1 (6.7)
Cerebellar stroke	1 (6.7)
Meningoencephalitis	1 (6.7)
Orthostatic tremor	1 (6.7)
Pneumococcal sepsis	1 (6.7)
Progressive gait disorder	1 (6.7)
Dermatomyositis	1 (6.7)
Guillain-Barré syndrome	1 (6.7)
Secondary diagnoses (multiple), n (%)	
Arterial hypertonia	3 (20)
Atherosclerosis	1 (6.7)
Chronic renal insufficiency	1 (6.7)
Diabetes mellitus	1 (6.7)
Dyslipidemia	1 (6.7)
Dyslipoproteinemia	1 (6.7)
Episodic migraine	1 (6.7)
Meningoencephalitis	1 (6.7)
Mild obesity	1 (6.7)
Multiple sclerosis	1 (6.7)
Progressive gait ataxia	1 (6.7)
Recurrent depressive mood disorder	1 (6.7)
Sensorimotor polyneuropathy	1 (6.7)

XSL•FO
RenderX

[a]MMSE: Mini Mental State Examination.

Secondary End User Group

A total of 20 therapists and sports scientists participated in and completed this study (Table 2).

Table . Secondary end user characteristics.

	Values (n=20)
Age (years), mean (SD; range)	33.61 (9.4; 23.7 - 59.3)
Sex, n (%)	
Male	6 (30)
Female	14 (70)
Professional experience (years), mean (SD; range)	8.9 (8.9; 1 - 38)
Profession, n (%)	
Physiotherapist	14 (70)
Occupational therapist	3 (15)
Sport scientist	3 (15)
Professional qualification, n (%)	
Vocational diploma	4 (20)
Bachelor's degree	7 (35)
Master's degree	7 (35)
Doctoral degree	2 (10)
Professional field (multiple), n (%)	
Neurological rehabilitation	16 (80)
Orthopedic rehabilitation	3 (15)
Geriatric rehabilitation	5 (25)
Medical fitness	2 (10)
Internal medicine, oncological rehabilitation	4 (20)
Psychomotor rehabilitation	3 (15)
Mixed	3 (15)

Primary Outcomes

Primary End User Group

No adverse events were reported during this study. Table 3 summarizes the activities per exercising round, detailing the clarity of instructions, PEUs' exercise performance, and completion rates, including activity intensity for ExerG exercising round 2. PEUs became familiar with the tasks in round 1, leading to improved performance in round 2. However, newly introduced activities continued to pose challenges, with some usability issues noted. Overall, 23 of 29 (79%) instructions were rated acceptable, good-to-very-good exercise performance was observed in 19 of 29 (65%) tasks, and 28 of 29 (97%) tasks were completed.

XSL•FO
RenderX

Table . Primary end user—ExerG software interaction.

Session and task	Clarity of instructions	Exercise performance	Task completion	Exercise intensity	
	Good to very good	Good to very good	Yes	Optimal	(Very) easy
ExerG exercise round 1					
System calibration	93.3[a]	93.3[a]	93.3[a]	—[b]	—
Walking	73.3	80	100[a]	—	—
Apple picking	93.3[a]	86.7[a]	93.3[a]	—	—
Walking	86.7[a]	80	100[a]	—	—
Pattern matching	86.7[a]	86.7[a]	100[a]	—	—
Walking	93.3	86.7[a]	100[a]	—	—
Rowing	73.3	60	80	—	—
Tall grass walking	66.7	73.3	100[a]	—	—
Apple picking	100[a]	60	100[a]	—	—
Walking	100[a]	100[a]	100[a]	—	—
Pattern matching	100[a]	93.3[a]	100[a]	—	—
Walking	100[a]	100[a]	100[a]	—	—
ExerG exercise round 2					
System calibration	93.3[a]	93.3[a]	93.3[a]	0	93.3[a]
Walking	100[a]	93.3[a]	100[a]	20	80
Apple picking	93.3[a]	80	100[a]	13.3	86.7[a]
Walking	100[a]	100[a]	100[a]	20	80
Trunk jumping	80	86.7[a]	93.3[a]	40	33.3
Walking	100[a]	100[a]	100[a]	13.3	86.7[a]
Rowing	100[a]	93.3[a]	100[a]	20	80
Walking	100[a]	100[a]	100[a]	13.3	80
Pick apples, avoid nuts	80	60	100[a]	40	26.7
Walking	100[a]	100[a]	100[a]	13.3	86.7[a]
Tall grass walking	93.3[a]	86.7[a]	100[a]	26.7	53.3
Walking	100[a]	100[a]	100[a]	13.3	86.7[a]
Time-limited pattern matching	100[a]	80	100[a]	33.3	66.7
Walking	100[a]	100[a]	100[a]	13.3	86.7[a]
Catching balloons	60	80	93.3[a]	40	26.7
Walking	100[a]	100[a]	100[a]	13.3	86.7[a]
Multitasking in apple picking	93.3[a]	60	100[a]	40	46.7

[a]These numbers indicate values above the 85% acceptance criteria.

[b]Not applicable.

Figure S2 in Multimedia Appendix 4 illustrates the emotions observed during PEU exercise activities. Enjoyment, curiosity, attention, and amazement were categorized as positive emotions, accounting for 79 of 90 (88%) of the expressed emotions during

XSL•FO
RenderX

training. Some PEUs exhibited negative emotions, such as nervousness, boredom, and confusion, but none expressed anger or resistance. Perceived physical and mental effort increased with higher exercise levels, with physical exertion generally viewed as greater than mental effort (Figure S3 in Multimedia Appendix 4).

Secondary End User Group

Analysis of the observation protocol revealed that most SEUs successfully executed the use scenarios. Minor difficulties or hesitation were noted in 14 of 41 tasks and difficulties in 2 of 41 tasks among the 20 therapists accounting for 1.7% and 0.2% of the total 820 task cases, respectively. Assistance from this study's personnel was required in 9 instances across all scenarios (1.1%), with no use errors detected. Independent, correct, and effective use was observed for all tasks in fall simulation (3), supporting training (4), compiling training (5), removing the safety system (7), and shutting down hardware (8). The mean time to complete tasks was 2.6 (SD 3.8) minutes, indicating the clarity and effectiveness of the ExerG user handbook and ExerG usage training (Tables 4 and 5).

XSL•FO
RenderX

Table . Secondary end user—ExerG hardware and software interaction across use scenarios 1 - 4 (n=20).

Use scenarios and tasks for secondary end users	Use[a]	Support[a]
Preparing the hardware (use scenario 1)[b]		
Turn on main power switch	✓	No
Turn on computer	✓	No
Turn on touchscreen or check if it is on	Difficulties 1	Yes 1
Turn on the keypad or check if it is on	✓	No
Turn on the 3 projectors using the remote	Hesitation 1	No
Select correct HDMI channels for projectors with remote	Hesitation 2	Yes 1
Enter the computer password	✓	No
Check internet connection	✓	No
Preparing and attaching the safety system (use scenario 2)[c]		
Check proper rope position and ensure it is undamaged	Hesitation 2	No
Perform the end stop test	Hesitation 2	No
Properly don the Petzl Newton Fast harness (correct size)	✓	No
Correctly connect the Petzl harness system	Hesitation 1	No
Tighten the rope if necessary	✓	No
Adjust the drop height correctly using the crank handle	✓	No
Simulating a fall scenario (use scenario 3)[d]		
If the patient falls and is caught by the safety system:		
Move to chair to help them stand	✓	No
Lower to the floor to stand independently	✓	No
Initiating a predefined training program (use scenario 4A)[e]		
Remove the trackers from the charging station	✓	No
Attach trackers securely to person’s ankles and wrists	Hesitation. 1	Yes 1
Turn on the trackers (green light).	Hesitation 1	Yes 1
Select the ExerG training program from the menu	✓	No
Choose the predefined training sequence from the menu	Hesitation 2	Yes 2
Start the selected training program	✓	No
Properly guide the (mock) patient through calibration	✓	No
Supporting a predefined training program (use scenario 4B)[f]		
Accompany (mock) patient during exergaming training	✓	No

[a]With columns “use” and “support,” the numbers represent the number of therapists, who showed one of the following behaviors: independent, correct, and effective use indicated by a checkmark, and hesitation or experiencing difficulties.

[b]Mean total duration 1.6 (SD 1.8) minutes.

XSL•FO
RenderX

[c]Mean total duration 2.4 (SD 0.8) minutes.

[d]Mean total duration 0.9 (SD 0.5) minutes.

[e]Mean total duration 1.9 (SD 1) minutes.

[f]Mean total duration 12.7 (SD 2.7) minutes.

Table . Secondary end user—ExerG hardware and software interaction across use scenarios 5 - 8 (n=20).

Use scenarios and tasks for secondary end users	Use[a]	Support[a]
Compilation of a predefined training sequence (use scenario 5)[b]		
Select the ExerG training program in the launcher	✓	No
Open a new training sequence	✓	No
Start editing mode	✓	No
Choose appropriate exercises	✓	No
Name the training sequence with the specified patient ID	✓	No
Save the training sequence	✓	No
Start the new training sequence	✓	No
Ending a training program (use scenario 6)[c]		
End the ExerG training program	Hesitation 2	Yes 1
Exit the user interface (launcher)	✓	No
Removing the safety system (use scenario 7)[d]		
Remove the wrist and ankle trackers	✓	No
Attach the 4 trackers to the charging station	✓	No
Crank down to release rope tension	✓	No
Release the rope from the Petzl harness system	✓	No
Loosen and remove the Petzl Newton Fast harness	✓	No
Terminating the hardware (use scenario 8)[e]		
Turn off the 3 projectors with the remote	Difficulties 1	Yes 2
Shut down the computer	✓	No
Switch off the main power switch of the system	✓	No

[a]With columns “use” and “support,” the numbers represent the number of therapists, who showed one of the following behaviors: independent, correct, and effective use indicated by a checkmark, and hesitation or experiencing difficulties.

[b]Mean total duration 1.9 (SD 0.6) minutes.

[c]Mean total duration 0.2 (SD 0.08) minutes.

[d]Mean total duration 1.5 (SD 0.6) minutes.

[e]Mean total duration 0.4 (SD 0.3) minutes.

SEUs demonstrated mastery of the knowledge tasks, reflecting clear user handbook instructions. These tasks included summarizing the product’s purpose, adjusting the video game volume, troubleshooting software issues (eg, user interface not opening or incidents during training), and managing ongoing activities (eg, skipping activities or restarting programs). Hardware tasks involved resolving issues with HTC Vive (HTC Corporation) cameras and trackers, ensuring proper recognition at startup, and correcting screen projections. Backup system tasks required knowledge of troubleshooting rope positioning and resistance during the end-stop test.

Secondary Outcomes

Primary End User Group

Ratings of patient player experience using the PXI indicated a positive or very positive experience. Functional consequences, such as audiovisual appeal or ease-of-control, were generally rated high. In contrast, related to psychosocial consequences, “autonomy” received the lowest score and showed a large IQR,

XSL•FO
RenderX

suggesting that PEUs felt they lacked control over their engagement in the game. Conversely, perceived mastery was rated highly (Figure 3 and Multimedia Appendix 4), closely linked to the aspect of autonomy.

Figure 3. Ratings of patient player experience. PXI: Player Experience Inventory.

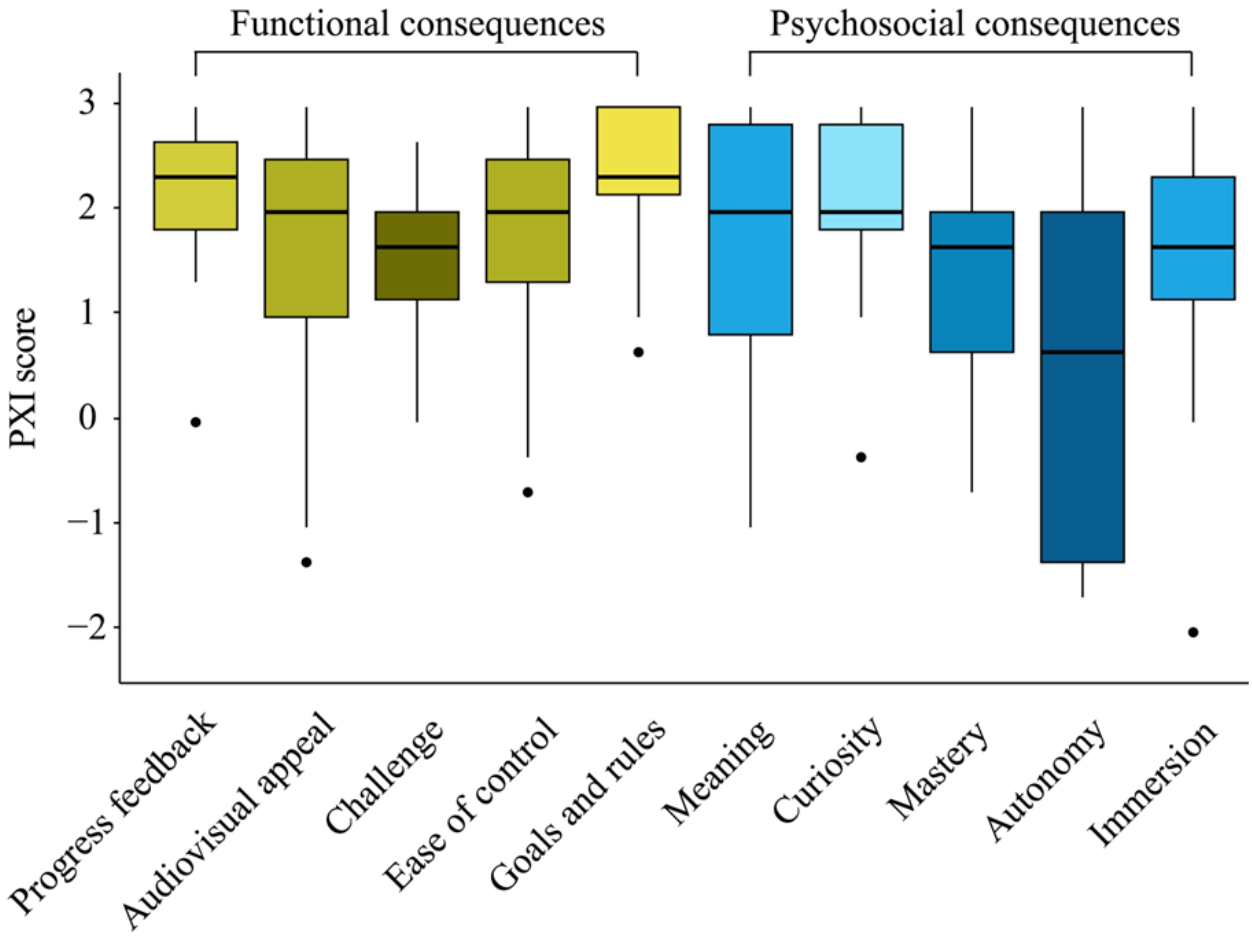


Figure 3 shows a boxplot of player experience across patients as assessed using the PXI, grouped by functional and psychosocial consequences. The boxes represent the IQR, with a median line, while whiskers extend to the minimum and maximum within 1.5 times the IQR, and outliers are indicated as points beyond this range.

Figure S4 in Multimedia Appendix 4 presents findings from the structured portion of the semistructured interview with PEUs. Most patients expressed enjoyment of the ExerG training, with positive feedback on the software and hardware. For the first question, PEUs rated their overall satisfaction at a median of 9 (IQR 7.5 - 10) on a 1 - 10 scale.

Saturation was achieved after 15 patient interviews, which revealed three themes through content analysis of the qualitative data: (1) enjoyment and acceptance, (2) live interaction, and (3) safety and comfort. Most patients enjoyed the training and would continue using the device (each n=12, 80%) if certain improvements were made (n=13, 87%), such as better graphic design (n=5, 33%), upgraded tracking systems (n=3, 20%), refined swivel arms (n=8, 53%), and enhanced projector quality (n=3, 20%). More variety in game activities and difficulty levels would also enrich the experience (n=2, 13%). While most users could understand the activity instructions (n=8, 53%) and feedback texts (n=9, 60%), they preferred shorter, more concise instructions with clear task-related wording and adequate reading time (n=12, 80%). Patients felt safe (n=15, 100%) and comfortable using the safety harness (n=13, 87%) but recommended improving the safety system, particularly the swivel arm's movements, for smoother balance control (n=5, 33%). Many also noted limitations with exploring the game environment (n=11, 73%) and discomfort with the rope pulley during specific activities (n=3, 20%). Table S1 in Multimedia Appendix 6 presents theme descriptions and quotes.

Secondary End User Group

In SEUs, a median SUS score of 82.5 (IQR 18.7) was reported, indicating excellent usability of the ExerG [34]. The structured interview analysis showed that therapists were satisfied with the system (median score of 8, IQR 7 - 8 on a scale of 1 - 10), were willing to use it (n=20, 100%), and found it highly user-friendly (n=20, 100%). Less than half of the therapists found the feedback helpful (n=8, 40%), and one-fifth considered the video feedback to be comprehensive (n=4, 20%). SEUs indicated that the training content does not or only partly resembles everyday tasks (Figure S5 in Multimedia Appendix 4).

XSL•FO
RenderX

Saturation was achieved after 19 therapist interviews—content analysis of the qualitative data identified five themes: (1) acceptance and motivation, (2) meaningful activities, (3) training feedback, (4) individualization, and (5) safety and autonomy. Therapists noted that video-game based training could boost patient motivation through its engaging design and surprising elements, making it more enjoyable than standard therapy (n=12, 60%). They found the activities beneficial for physical and cognitive functions (n=20, 100%) but recommended incorporating more exercises that simulate daily living tasks for improved relevance (n=13, 65%). Audiovisual feedback and star awards were seen as helpful (n=8, 40%), with a preference for more specific motivational feedback (n=9, 45%) and clearer performance summaries (n=8, 40%). Performance ratings should be simple and understandable (n=9, 45%), with varying levels to sustain motivation (n=6, 30%). The system's innovative approach to addressing cognitive and physical functions was well-received (n=9, 45%), and therapists requested a wider range of customizable exercises to cater to individual patient needs (n=10, 50%). The safety harness was valued for creating secure training environments (n=4, 20%). For independent exercising, therapists suggested clearer, shorter activity instructions that remain visible until acknowledged (n=6, 30%), underscoring the importance of considering the patient's cognitive and physical capabilities for independent use (n=16, 80%). Theme descriptions and quotes are available in Table S1 in Multimedia Appendix 7.

Both User Groups

PEUs and SEUs have identified opportunities to enhance system functionality and user experience. Key improvements include reducing inertia and swing of the swivel arm for better training comfort, making the safety system operable for therapists of all heights, and ensuring accurate tracking. Adjustments for color-blind users and projector light issues will enhance visual clarity. Clear instructional texts and synchronized movements will improve gameplay fluidity. Additionally, automatic screen arrangement correction, seamless activity transitions, and reliable HTC Vive tracker recognition will boost overall system performance.

Discussion

Principal Results and Comparison With Prior Work

Patients and therapists reported positive experiences and good usability with the ExerG exergame. PEUs deemed 79% of instructions as acceptable, achieved good exercise performance in 65% of tasks, and completed 97% of tasks without adverse events. Patients expressed satisfaction, with 88% reporting positive emotions and a median user satisfaction score of 9. Many patients indicated a willingness to continue using the device, contingent upon improvements in graphic design, tracking systems, projector quality, clarity of instructions, and game variety. PEUs felt secure in the safety harness, though some suggested enhancements for swivel arm movement. SEUs demonstrated effective task execution, with minimal hesitation reported. Therapist usability ratings were high, with all willing to use the ExerG and being able to operate it based on the user handbook. They noted the motivating effects of video-game-based training and its support for physical skills. Suggestions for improvement included incorporating daily living task simulations, more customization, targeted feedback, clearer performance ratings, and concise instructions.

This usability study with patients and therapists from 2 European rehabilitation centers identified specific usability challenges and strengths, informing design recommendations. Despite initial challenges inherent to the functional model, patients reported increased engagement and enjoyment across exercising rounds. Positive emotions, particularly heightened attention and enjoyment, characterized their experiences. While feedback on game mechanics was mostly favorable, some patients expressed a desire for greater autonomy. These findings underscore ExerG's potential for user engagement and highlight key areas for refinement in future versions.

Our findings support the OPTIMAL (Optimizing Performance Through Intrinsic Motivation and Attention for Learning) theory of motor learning [38], which emphasizes the importance of learner autonomy in improving motor learning. Research shows that people are more motivated when they have control over their actions and their impact on the environment, even if it requires more effort. This indicates an inherent reward in exercising control. Eitam et al [39] demonstrated that motivation increases when individuals feel their actions have meaningful effects, underscoring the role of autonomy in fostering intrinsic motivation. Most patients had prior experience with rehabilitation technologies, but the exergame in this study was more complex due to its multisensory and multimodal features, unique movement concepts, and cognitive challenges. This complexity may have affected patients' autonomy, particularly with in-game instructions that were always clear. However, exergames are known to promote autonomy and competence through playful engagement once users become familiar with the technology [40]. Patients reported an increase in perceived physical effort from the first to the second ExerG exercising round, while mental effort was perceived as lower. Overall, patients had a very good player experience, suggesting strong potential for ExerG training despite initial challenges.

In the quantitative assessment, semistructured interviews with PEUs showed that most patients enjoyed training with the ExerG software and had positive responses to its accessories, with no adverse events reported across study centers. SEUs found that the setup and closure of ExerG usage training and observation rounds were generally well-received. No adverse events were reported across study centers.

Potential optimizations for the ExerG system include reducing inertia and swing effects in the swivel arm for patient comfort, ensuring accurate tracking without offset, making accommodations for color-blind individuals, addressing projector light issues affecting color perception, and ensuring proper setup of ground patterns, room darkening, and fully charged trackers.

From a motor learning perspective, these adjustments are important for several reasons: First, ExerG training involves physical-cognitive dual-task training for patients with motor or cognitive disorders, where cognitive-motor interference leads to decreased performance [41]. Second, aging [42] and

XSL•FO
RenderX

neurological disorders [43] increase susceptibility to this interference. Third, both aging [44] and neurological disorders [43,45] impair off-line gains in motor skill learning and heighten susceptibility to contextual interference. This interference occurs when practicing multiple skills in 1 session, with high contextual interference aiding retention and transfer despite lower immediate performance [44]. Therefore, minimizing external factors, such as lighting, noise, or the ExerG's safety harness, is essential to optimize performance.

Older individuals are primarily motivated to engage with exergames by generativity, peer recommendations, self-improvement, and curiosity [46]. Retention is enhanced by achieving goals, immersion, and staying active, supported by structured schedules and adjustable difficulty levels. However, barriers such as stagnant progress, fatigue, or discomfort can hinder participation. Thus, it is essential to adapt the ExerG training software and accessories based on these psychological and user experience insights.

From an SEU perspective, clear instructions, accurate alignment of object movements with player actions, automatic screen arrangement, smooth transitions, easy tracker recognition and calibration, user-friendly safety systems, and prompt projector remote responses were highlighted. Our findings align with a systematic review indicating that ease-of-use, comfort, learnability, usefulness, and perceived efficacy are crucial usability aspects for health care professionals [47]. These insights underscore the need for close collaboration between the game design and development team and SEUs to incorporate SEUs' domain knowledge into the game design. Overall, this study's findings suggest that the ExerG training software and accessories can be further iterated and optimized to better meet patient and therapist requirements.

Strengths and Limitations

This study adds to the literature on exergames for rehabilitation by delivering insights into the usability, safety, performance, and user experience of a novel exergame functional model in a real-world setting. Despite the challenges, the ExerG showed overall good usability and positive experiences for both patients and therapists. Patients reported increased enjoyment and improved performance, with 88% of their emotions being positive, while therapists successfully executed training scenarios and identified areas for optimization, including clearer instructions, more accurate tracking, and refined feedback mechanisms. Importantly, no adverse events occurred during this study, indicating the safety of the ExerG system.

Our research underscores the importance of an interdisciplinary, iterative development approach for creating user-centered designs in rehabilitation technologies. By involving game and industrial designers, programmers, rehabilitation experts, and end users throughout the process, we developed a functional model that meets specific rehabilitation needs. Insights from this usability study will guide future iterations of the ExerG system to better serve patients and therapists. This approach can serve as a model for developing other rehabilitation technologies, emphasizing the significance of user involvement, iterative design, and real-world testing. Our study also has some limitations. While 20 SEUs were recruited from 2 rehabilitation centers, only patients from 1 center participated, resulting in a smaller sample size of 15 patients. Saturation was achieved in the qualitative strand, and prior research suggests that even with 15 users, a detection rate of 97% (minimum 90) is feasible [48]. The PEU group was heterogeneous, reflecting diverse diagnoses and symptoms, which may increase variability in outcomes but also enriches feedback by providing a broader range of perspectives. This diversity enhances the ExerG's applicability and improves this study's ecological validity, aligning findings more closely with real-world usage scenarios.

One significant limitation of our study is that, although our eligibility criteria permitted the inclusion of patients with an MMSE score of 19 or greater, only 1 patient with mild cognitive impairment participated, and there were no participants with moderate cognitive impairment. This lack of diversity in cognitive impairment levels may restrict the generalizability of our findings. Furthermore, motor and balance impairments were assessed by neurologists at the start of rehabilitation for this study's participants. This study's population included 3 patients who required walking aids and 5 others with a history of falls. However, aside from the 2 Berg Balance Scale items, no objective balance assessments or formal motor function assessments were conducted. Additionally, the presence of researchers in our usability study may have influenced participants' responses and behavior. While patients were unaware if their study therapist was a researcher, some may have felt compelled to provide positive feedback or altered their behavior due to the observers' presence. Although we recognize the potential for observer bias, it is a common concern in usability studies. Finally, assessing the exergame in its functional model stage may have affected usability and experience. However, since the development process followed an iterative approach, the findings can directly inform design optimizations for the exergame.

Conclusions

The newly developed exergame (ExerG training software and accessories) demonstrated good usability and positive effects on both therapists and patients, despite challenges related to its functional model status. The interdisciplinary, iterative approach used in its development facilitated direct optimization based on study findings. Future iterations hold promise, including follow-up studies such as randomized controlled trials to assess long-term training effects and user experience with the finalized ExerG system. In conclusion, our study offers a comprehensive evaluation of the usability and user experience of a novel exergame in a rehabilitation setting, enhancing the understanding of how exergames can be designed to support rehabilitation outcomes while identifying key challenges and improvement opportunities.

XSL•FO
RenderX

Acknowledgments

We would like to thank all the patients and therapists who took part in our study. We truly appreciate the time and effort everyone dedicated during the training and evaluation sessions, as their valuable insights into the ExerG training device were essential for the completion of this research. The ExerGetic project is cofunded by the Active Assisted Living Programme (project AAL-2020-7-48-CP) and the national authorities in Austria (Austrian Research Promotion Agency, FFG) and Switzerland (Innosuisse). The funding agency was not involved in designing this study, sampling and recruitment, data collection, data interpretation, or writing this paper.

Data Availability

All data generated or analyzed during this study are included in this paper itself and its supporting material. The ExerG is manufactured by Sphery Ltd, CH. Interested researchers can contact Sphery Ltd for information sharing.

Authors' Contributions

ALMN, AS, CSA, SN, LN, CB, and BS conceptualized this study and secured funding. SN, ALMN, AS, YR, RJ, SS, and NSB handled software and hardware adaptations and provided technical support. UP, CB, BS, and CSA served as study coordinators and supervisors. BS, CSA, LB, CB, SM, KL, IH, and SH were involved in participant screening and recruitment, while BS, CSA, SH, KL, IH, and SM were responsible for data collection and curation. Formal data analysis, interpretation, and visualization were carried out by BS, CSA, FB, SH, KL, IH, and SM. Manuscript preparation was managed by SH, BS, CSA, FB, LN, SAS, and ALMN. All authors reviewed, read, and approved the final paper.

Conflicts of Interest

AS, YR, and SN are or have been employees at Sphery Ltd. ALMN and SN are cofounders of the Zuercher Hochschule der Kuenste (ZHdK) spinoff Sphery Ltd, which developed the ExerCube based on the results of ALMNs' previous research projects. AS was employed by Sphery as Senior Research and Development Manager from November 2019 to February 2023. The remaining authors have no conflicts of interest to declare. No revenue was paid or promised to be paid directly to ALMN, AS, YR, SN, Sphery Ltd, or any of the authors and research institutions.

Multimedia Appendix 1
Good Reporting of a Mixed Methods Study (GRAMMS) checklist.
[PDF File, 116 KB - games_v13i1e66515_app1.pdf]

Multimedia Appendix 2
Study material–exercise and test item details.
[PDF File, 217 KB - games_v13i1e66515_app2.pdf]

Multimedia Appendix 3
Details on screening instruments and secondary outcome measures for primary end users.
[DOCX File, 39 KB - games_v13i1e66515_app3.docx]

Multimedia Appendix 4
Case report forms of primary and secondary end users.
[PDF File, 2277 KB - games_v13i1e66515_app4.pdf]

Multimedia Appendix 5
Further illustration of outcomes in both user groups.
[PDF File, 234 KB - games_v13i1e66515_app5.pdf]

Multimedia Appendix 6
Primary end user theme description.
[PDF File, 127 KB - games_v13i1e66515_app6.pdf]

Multimedia Appendix 7
Secondary end user theme description.
[PDF File, 138 KB - games_v13i1e66515_app7.pdf]

XSL•FO
RenderX

References

1. Luimula M, Ailio P, Botha-Ravyse C, et al. Gaming for health across various areas of life. Presented at: 2018 IEEE 9th International Conference on Cognitive Infocommunications (CogInfoCom); Aug 22-24, 2018; Budapest, Hungary. [doi: 10.1109/CogInfoCom.2018.8639955]
2. Fernandes CS, Magalhães B, Lima A, Nóbrega P, Silva M, Santos C. Impact of exergames on the mental health of older adults: a systematic review and GRADE evidence synthesis. Games Health J 2022 Jul 11;11(6):355-368. [doi: 10.1089/g4h.2021.0229] [Medline: 35819413]
3. Wiemeyer J, Deutsch J, Malone LA, et al. Recommendations for the optimal design of exergame interventions for persons with disabilities: challenges, best practices, and future research. Games Health J 2015 Feb;4(1):58-62. [doi: 10.1089/g4h.2014.0078] [Medline: 26181682]
4. Tobaigy A, Alshehri MA, Timmons S, Helal OF. The feasibility of using exergames as a rehabilitation tool: the attitudes, awareness, opinions and experiences of physiotherapists, and older people towards exergames. J Phys Ther Sci 2018 Apr;30(4):555-562. [doi: 10.1589/jpts.30.555] [Medline: 29706705]
5. Erden MS, Komoto H, van Beek TJ, D'Amelio V, Echavarria E, Tomiyama T. A review of function modeling: approaches and applications. AIEDAM 2008;22(2):147-169. [doi: 10.1017/S0890060408000103]
6. McNeill MDJ, Charles DK, Burke JW, Crosbie JH, McDonough SM. Evaluating user experiences in rehabilitation games. J Assist Technol 2012 Sep 7;6(3):173-181. [doi: 10.1108/17549451211261290]
7. Tsai TH, Tseng KC, Wong AM, Chang HJ. A study exploring the usability of an exergaming platform for senior fitness testing. Health Informatics J 2020 Jun;26(2):963-980. [doi: 10.1177/1460458219853369] [Medline: 31264499]
8. Pramuka M, van Roosmalen L. Telerehabilitation technologies: accessibility and usability. Int J Telerehabil 2009;1(1):85-98. [doi: 10.5195/ijt.2009.6016] [Medline: 25945165]
9. Mitchell J, Shirota C, Clanchy K. Factors that influence the adoption of rehabilitation technologies: a multi-disciplinary qualitative exploration. J Neuroeng Rehabil 2023 Jun 20;20(1):80. [doi: 10.1186/s12984-023-01194-9] [Medline: 37340496]
10. Karsh BT. Beyond usability: designing effective technology implementation systems to promote patient safety. Qual Saf Health Care 2004 Oct;13(5):388-394. [doi: 10.1136/qhc.13.5.388] [Medline: 15465944]
11. Van Buskirk R, Moroney BW. Extending prototyping. IBM Syst J 2003;42(4):613-623. [doi: 10.1147/sj.424.0613]
12. Alhasani R, George N, Radman D, Auger C, Ahmed S. Methodologies for evaluating the usability of rehabilitation technologies aimed at supporting shared decision-making: scoping review. JMIR Rehabil Assist Technol 2023 Aug 15;10:e41359. [doi: 10.2196/41359] [Medline: 37581911]
13. Ringgenberg N, Mildner S, Hapig M, et al. ExerG: adapting an exergame training solution to the needs of older adults using focus group and expert interviews. J Neuroeng Rehabil 2022 Aug 16;19(1):89. [doi: 10.1186/s12984-022-01063-x] [Medline: 35974409]
14. Swinnen N, Vandenbulcke M, de Bruin ED, Akkerman R, Stubbs B, Vancampfort D. Exergaming for people with major neurocognitive disorder: a qualitative study. Disabil Rehabil 2022 May;44(10):2044-2052. [doi: 10.1080/09638288.2020.1822934] [Medline: 32962436]
15. Pirovano M, Surer E, Mainetti R, Lanzi PL, Alberto Borghese N. Exergaming and rehabilitation: a methodology for the design of effective and safe therapeutic exergames. Entertain Comput 2016 May;14:55-65. [doi: 10.1016/j.entcom.2015.10.002]
16. Siriaraya P, Visch V, Boffo M, et al. Game design in mental health care: case study-based framework for integrating game design into therapeutic content. JMIR Serious Games 2021 Dec 1;9(4):e27953. [doi: 10.2196/27953] [Medline: 34855611]
17. Ekbia HR, Lee J, Wiley S. Rehab games as components of workflow: a case study. Games Health J 2014 Aug;3(4):215-226. [doi: 10.1089/g4h.2014.0039] [Medline: 26192370]
18. Medina JLP, Acosta-Vargas P, Rybarczyk Y. A systematic review of usability and accessibility in tele-rehabilitation systems. In: Rybarczyk Y, editor. Assistive and Rehabilitation Engineering: IntechOpen; 2019:31-50. [doi: 10.5772/intechopen.85869]
19. Human factors studies and related clinical study considerations in combination product design and development: draft guidance for industry and FDA staff. US Food and Drug Administration. 2016. URL: https://www.fda.gov/media/96018/download [accessed 2025-01-29]
20. O'Cathain A, Murphy E, Nicholl J. The quality of mixed methods studies in health services research. J Health Serv Res Policy 2008 Apr;13(2):92-98. [doi: 10.1258/jhsrp.2007.007074] [Medline: 18416914]
21. Medlock MC. The Rapid Iterative Test and Evaluation method (RITE). In: Games User Research: Oxford University Press; 2018:203-216 URL: https://doi.org/10.1093/oso/9780198794844.003.0013 [accessed 2025-01-29]
22. Martin-Niedecken AL, Mahrer A, Rogers K, de Bruin ED, Schättin A. "HIIT" the ExerCube: comparing the effectiveness of functional high-intensity interval training in conventional vs. exergame-based training. Front Comput Sci 2020;2:33. [doi: 10.3389/fcomp.2020.00033]
23. Berg KO, Wood-Dauphinee SL, Williams JI, Maki B. Measuring balance in the elderly: validation of an instrument. Can J Public Health 1992;83 Suppl 2:S7-11. [Medline: 1468055]
24. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975 Nov;12(3):189-198. [doi: 10.1016/0022-3956(75)90026-6] [Medline: 1202204]

XSL•FO
RenderX

25. Creavin ST, Wisniewski S, Noel-Storr AH, et al. Mini-Mental State Examination (MMSE) for the detection of dementia in clinically unevaluated people aged 65 and over in community and primary care populations. Cochrane Database Syst Rev 2016 Jan 13;2016(1):CD011145. [doi: 10.1002/14651858.CD011145.pub2] [Medline: 26760674]
26. Hawker GA, Mian S, Kendzerska T, French M. Measures of adult pain: Visual Analog Scale for Pain (VAS Pain), Numeric Rating Scale for Pain (NRS Pain), McGill Pain Questionnaire (MPQ), Short-Form McGill Pain Questionnaire (SF-MPQ), Chronic Pain Grade Scale (CPGS), Short Form-36 Bodily Pain Scale (SF-36 BPS), and Measure of Intermittent and Constant Osteoarthritis Pain (ICOAP). Arthritis Care Res (Hoboken) 2011 Nov;63 Suppl 11(S11):S240-S252. [doi: 10.1002/acr.20543] [Medline: 22588748]
27. Hennink M, Kaiser BN. Sample sizes for saturation in qualitative research: a systematic review of empirical tests. Soc Sci Med 2022 Jan;292:114523. [doi: 10.1016/j.socscimed.2021.114523] [Medline: 34785096]
28. Borg GAV. Psychophysical bases of perceived exertion. Med Sci Sports Exerc 1982 May;14(5):377. [doi: 10.1249/00005768-198205000-00012]
29. Paas F. Training strategies for attaining transfer of problem-solving skill in statistics: a cognitive-load approach. J Educ Psychol 1992;84(4):429-434. [doi: 10.1037//0022-0663.84.4.429]
30. Abeele VV, Spiel K, Nacke L, Johnson D, Gerling K. Development and validation of the player experience inventory: a scale to measure player experiences at the level of functional and psychosocial consequences. Int J Hum Comput Stud 2020 Mar;135:102370. [doi: 10.1016/j.ijhcs.2019.102370]
31. Graf L, Altmeyer M, Emmerich K, Herrlich M, Krekhov A, Spiel K. Development and validation of a German version of the player experience inventory (PXI). 2022 Sep 4 Presented at: MuC '22; Sep 4-7, 2022; Darmstadt Germany p. 265-275 URL: https://dl.acm.org/doi/proceedings/10.1145/3543758 [accessed 2025-01-29] [doi: 10.1145/3543758.3543763]
32. Pekrun R, Linnenbrink-Garcia L, editors. International Handbook of Emotions in Education: Routledge/Taylor & Francis Group; 2014. [doi: 10.4324/9780203148211]
33. Brooke JB. SUS: a "quick and dirty" usability scale. In: Jordan PW, Thomas B, McClelland IL, Weerdmeester B, editors. Usability Evaluation In Industry, 1st edition: Taylor & Francis - CRC Press; 1996, Vol. 189:180-189 URL: https://www.crcpress.com/product/isbn/9780748404605 [accessed 2025-02-05]
34. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the System Usability Scale. Int J Hum Comput Interact 2008 Jul 29;24(6):574-594. [doi: 10.1080/10447310802205776]
35. Mayring P. Qualitative Inhaltsanalyse Grundlagen Und Techniken: Beltz; 2015.
36. Mayring P. Qualitative content analysis: theoretical background and procedures. In: Bikner-Ahsbahs A, Knipping C, Presmeg N, editors. Approaches to Qualitative Research in Mathematics Education: Examples of Methodology and Methods: Springer Netherlands; 2015:365-380. [doi: 10.1007/978-94-017-9181-6_13]
37. Noble H, Smith J. Issues of validity and reliability in qualitative research. Evid Based Nurs 2015 Apr;18(2):34-35. [doi: 10.1136/eb-2015-102054] [Medline: 25653237]
38. Wulf G, Lewthwaite R. Optimizing Performance Through Intrinsic Motivation and Attention for Learning: The OPTIMAL theory of motor learning. Psychon Bull Rev 2016 Oct;23(5):1382-1414. [doi: 10.3758/s13423-015-0999-9] [Medline: 26833314]
39. Eitam B, Kennedy PM, Higgins ET. Motivation from control. Exp Brain Res 2013 Sep;229(3):475-484. [doi: 10.1007/s00221-012-3370-7] [Medline: 23288323]
40. Osorio G, Moffat DC, Sykes J. Exergaming, exercise, and gaming: sharing motivations. Games Health J 2012 Jun;1(3):205-210. [doi: 10.1089/g4h.2011.0025] [Medline: 26193438]
41. Wang X, Pi Y, Chen P, Liu Y, Wang R, Chan C. Cognitive motor interference for preventing falls in older adults: a systematic review and meta-analysis of randomised controlled trials. Age Ageing 2015 Mar 1;44(2):205-212. [doi: 10.1093/ageing/afu175]
42. Gallou-Guyot M, Mandigout S, Bherer L, Perrochon A. Effects of exergames and cognitive-motor dual-task training on cognitive, physical and dual-task functions in cognitively healthy older adults: an overview. Ageing Res Rev 2020 Nov;63:101135. [doi: 10.1016/j.arr.2020.101135] [Medline: 32768594]
43. McIsaac TL, Fritz NE, Quinn L, Muratori LM. Cognitive-motor interference in neurodegenerative disease: a narrative review and implications for clinical management. Front Psychol 2018;9:2061. [doi: 10.3389/fpsyg.2018.02061] [Medline: 30425673]
44. Roig M, Ritterband-Rosenbaum A, Lundbye-Jensen J, Nielsen JB. Aging increases the susceptibility to motor memory interference and reduces off-line gains in motor skill learning. Neurobiol Aging 2014 Aug;35(8):1892-1900. [doi: 10.1016/j.neurobiolaging.2014.02.022] [Medline: 24680325]
45. Schweighofer N, Lee JY, Goh HT, et al. Mechanisms of the contextual interference effect in individuals poststroke. J Neurophysiol 2011 Nov;106(5):2632-2641. [doi: 10.1152/jn.00399.2011] [Medline: 21832031]
46. Crane BM, Drazich BF, Taylor JL, et al. Older adults and three-dimensional exergaming: motivators and barriers to participation and retention. Games Health J 2023 Apr;12(2):150-157. [doi: 10.1089/g4h.2022.0131] [Medline: 36706426]
47. Zanatta F, Giardini A, Pierobon A, D'Addario M, Steca P. A systematic review on the usability of robotic and virtual reality devices in neuromotor rehabilitation: patients' and healthcare professionals' perspective. BMC Health Serv Res 2022 Apr 20;22(1):523. [doi: 10.1186/s12913-022-07821-w] [Medline: 35443710]

XSL•FO
RenderX

48. Alroobaea R, Mayhew PJ. How many participants are really enough for usability studies? Presented at: 2014 Science and Information Conference (SAI); Aug 27-29, 2014; London, UK. [doi: 10.1109/SAI.2014.6918171]

Abbreviations

CR10: Borg Category Ratio Scale
MMSE: Mini Mental State Examination
OPTIMAL: Optimizing Performance Through Intrinsic Motivation and Attention for Learning
PEU: primary end user
PXI: Player Experience Inventory
SEU: secondary end user
SUS: System Usability Scale

Edited by A Coristine; submitted 15.09.24; peer-reviewed by E Dove, S Quaglini; revised version received 25.12.24; accepted 21.01.25; published 14.02.25.

Please cite as:
Herren S, Seebacher B, Mildner S, Riederer Y, Pachmann U, Böckler NS, Niedecken S, Sgandurra SA, Bonati L, Hotz I, Schättin A, Jurt R, Brenneis C, Lenfert K, Behrendt F, Schmidlin S, Nacke L, Schuster-Amft C, Martin-Niedecken AL
Exergame (ExerG)-Based Physical-Cognitive Training for Rehabilitation in Adults With Motor and Balance Impairments: Usability Study
JMIR Serious Games 2025;13:e66515
URL: https://games.jmir.org/2025/1/e66515
doi:10.2196/66515

© Silvia Herren, Barbara Seebacher, Sarah Mildner, Yanick Riederer, Ulrike Pachmann, Nija Sonja Böckler, Stephan Niedecken, Sabrina Alicia Sgandurra, Leo Bonati, Isabella Hotz, Alexandra Schättin, Roman Jurt, Christian Brenneis, Katharina Lenfert, Frank Behrendt, Stefan Schmidlin, Lennart Nacke, Corina Schuster-Amft, Anna Lisa Martin-Niedecken. Originally published in JMIR Serious Games (https://games.jmir.org), 14.2.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Transformative Gamified Binocular Therapy for Unilateral Amblyopia in Young Children: Pilot Prospective Efficacy and Safety Study

Wenqing Zhu[1,2], MD, PhD; Shuneng Gu[3], MSc; Jian Li[3], PhD; Jin Lin[1], BSc; Chanling Hu[1], BSc; Rui Liu[1,2,4], MD, PhD

[1]Department of Ophthalmology, Eye and ENT Hospital of Fudan University, No.83 Fenyang Road, Xuhui District, Shanghai, China

[2]Key Laboratory of Myopia and Related Eye Diseases, Chinese Academy of Medical Sciences, Fudan University, Shanghai, China

[3]BOKE Digital Health Research Institute, BOKE Medical Technology Co, Ltd, Shanghai, China

[4]State Key Laboratory of Medical Neurobiology and MOE Frontiers Center for Brain Science, Fudan University, Shanghai, China

Corresponding Author:
Rui Liu, MD, PhD
Department of Ophthalmology, Eye and ENT Hospital of Fudan University, No.83 Fenyang Road, Xuhui District, Shanghai, China

Abstract

Background: Amblyopia is a common cause of visual impairment in children. Compliance with traditional treatments for amblyopia is challenging due to negative psychosocial impacts. Recent shifts in amblyopia treatment have moved from suppressing the dominant eye to enhancing binocular visual function. Binocular digital therapy has become a promising approach.

Objective: The aim of this study was to evaluate the effects of binocular gamified digital therapy on visual acuity and stereoacuity (SA) in children with unilateral amblyopia.

Methods: This pilot prospective study enrolled 11 children aged 4-6 years with unilateral amblyopia. Following at least 8 weeks of refractive correction, participants underwent binocular gamified digital therapy for 60 minutes per day, 5 days a week. The therapy used a roguelike shooting game delivered under binocular conditions through two independent channels with a real-time artificial intelligence visual engine. Assessments of distance visual acuity (DVA), near visual acuity (NVA), and SA were conducted at baseline and again at 4, 8, and 12 weeks.

Results: At 12 weeks, the following significant improvements were noted: amblyopic eye DVA improved by 1.0 line (P=.01; d=0.77), binocular DVA improved by 0.7 lines (P=.006; d=1.00), and SA improved by 0.3 logarithm (log) arcseconds (P=.01; d=0.97). At 8 weeks, improvements included amblyopic eye DVA by 0.9 lines (P=.046; d=1.00) and SA by 0.28 log arcseconds (P=.02; d=0.90). No significant adverse events were observed, although one participant developed progressive esotropia.

Conclusions: Binocular gamified digital therapy is effective and safe for improving visual outcomes in children aged 4-6 years with unilateral amblyopia.

Trial Registration: Chinese Clinical Trial Registry ChiCTR2300072066; https://www.chictr.org.cn/showproj.html?proj=198625

(JMIR Serious Games 2025;13:e63384) doi:10.2196/63384

KEYWORDS

amblyopia; binocular treatment; digital therapy; game; stereoacuity; visual acuity

Introduction

Amblyopia refers to a unilateral or bilateral decrease of vision caused by abnormal vision development in childhood. It is a common cause of visual impairment in children, significantly affecting their social environment and quality of life. Studies indicate that amblyopia affects 0.7%-2.6% of children aged 30‐71 months and 1%-5.5% of older children [1-3]. Anisometropia is linked to a higher risk of developing amblyopia, with about one-third of children with amblyopia displaying this condition [4]. Additionally, 19%-50% of cases of unilateral amblyopia are associated with strabismus, which increases the risk of amblyopia by 2.7-18 fold [5]. Traditional treatments include refractive correction [6,7] and clinical therapies like part-time patching or atropine penalization of the dominant eye [8,9]. However, compliance with these treatments is challenging, with only about 50% of children adhering to patching regimens due to negative psychosocial impacts [10]. Patching also fails to enhance binocular vision effectively and may disrupt motion processing [11], reading fluency [12], fixation [13], and visual perception [14].

XSL•FO
RenderX

Recent shifts in amblyopia treatment have moved from suppressing the dominant eye with monocular interventions to enhancing binocular visual function. Binocular therapy has thus become a promising approach [15,16]. Many children with amblyopia have a functional binocular visual system and cerebral cortex that can facilitate vision in the amblyopic eye by reducing the contrast sensitivity in the dominant eye. This concept underpins the clinical use of binocular vision therapy [17,18]. Gamification has been reported to support adherence to therapeutic amblyopia treatment and improve visual-perceptual defects in children [19,20]. The Vision Planet Visual Training System in this study, a gamified home-based medical software, treats amblyopia under binocular conditions by presenting separate visual stimuli to each eye. The software features an in-game joystick and a shoot button that allow the user to control a character, aiming to improve gaze and binocular fusion through interactive gameplay.

In this study, we present an initial evaluation of a novel binocular digital therapy that employs active viewing with feedback, designed to strengthen the amblyopic eye without the participants' awareness. The objective of this prospective pilot study was to assess the efficacy and safety of this therapy in children with unilateral amblyopia.

Methods

Participants

The study was a pilot, prospective, single-arm proof-of-concept trial conducted at the Eye and ENT Hospital of Fudan University in Shanghai, China, from June 2023 to January 2024. A total of 11 participants were enrolled.

The primary inclusion criteria were as follows: (1) children aged between 4-7 years; (2) diagnosed with anisometropic, strabismic, or combined mechanism amblyopia; (3) visual acuity (VA) in the amblyopic eye ranging from 20/32 (0.2 base 10 logarithm of the minimum angle of resolution [logMAR]) to 20/200 (1.0 logMAR); (4) best corrected visual acuity (BCVA) in the dominant eye of 20/40 (0.3 logMAR) or better for children aged 4-5 years, and 20/32 (0.2 logMAR) or better for children 5 years or older; (5) an interocular difference of ≥2 logMAR lines (≥10 letters); (6) consistent spectacle wear for at least 8 weeks, if required; (7) a pause in patching or any binocular therapy for at least 2 weeks; (8) partially accommodative esotropia with ≤10 prism diopters (PDs) of near deviation postrefractive correction or well-controlled intermittent exotropia. Major exclusion criteria included (1) a spherical equivalent of ≤–0.5 diopter in either eye, (2) any other condition potentially causing reduced BCVA, (3) severe developmental delays affecting participation in treatment or evaluation, (4) noncompliance with spectacle wear, (5) a history of light-induced seizures, (6) use of rigid gas permeable contact lenses, (7) any reported anatomical ocular anomalies, and (8) prior intraocular or strabismus surgery.

Ethical Considerations

This study involving human subjects received ethical approval from the institutional review board (IRB) at the Eye and ENT Hospital of Fudan University under approval number 2023061. The research adhered to all relevant ethical guidelines for studies involving humans, medical records, patient information, and secondary data analyses. No exemptions or waivers were applied for this study, and the ethics approval was granted in accordance with institutional requirements. Informed consent was obtained from all participants prior to their inclusion in the study. Participants were informed about the purpose, procedures, potential risks, and their right to withdraw at any point without consequence. For secondary analyses of existing data, the original informed consent or IRB approval covered the use of this data for further analysis without requiring additional consent. All data collected in this study was anonymized to ensure the privacy and confidentiality of the participants. Personal identifiers were removed, and data was stored securely in compliance with institutional privacy protocols. In cases where data could potentially be linked to individuals, appropriate measures, including encryption and access control, were in place to protect participant information. Participants in this study received compensation for their time and involvement. The amount and type of compensation were clearly communicated to all participants prior to their participation that they will receive a transportation allowance of approximately US $80 in total (about $20 per visit). Details of compensation were outlined in the informed consent form, ensuring transparency and fairness in the compensation process. No identifiable features of research participants are visible in any images included in the manuscript or supplementary materials.

Gamified Binocular Therapy

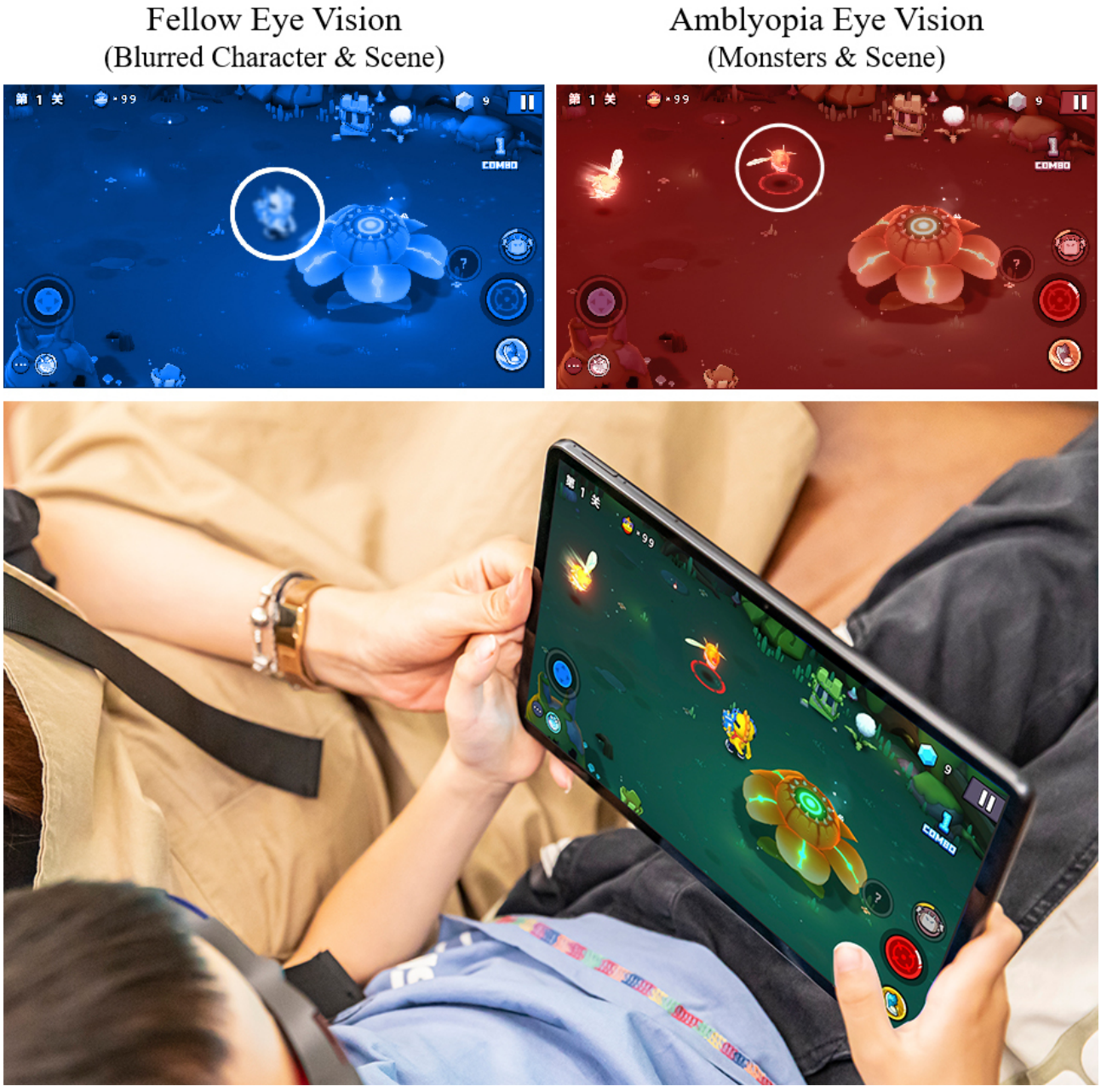
The Vision Planet Visual Training System is a gamified medical software developed using Unity version 2020.3.48 (Unity Technologies) and features a roguelike shooting game. Roguelike games are characterized by minimalistic graphics and random generation of contents, such as items, enemies, and level layouts. The system employed the classic mechanics, dynamics, and aesthetics model as a framework, which involves operating characters to shoot and defeat monsters. The dynamic model involved players combining different equipment and buffs to construct personalized play styles to battle with monsters in various planets, which enabled players to experience the aesthetics of discovery when uncovering unknown elements, challenges when facing powerful bosses, and immersion in the storyline through fantasy. The system used the Berlin Interpretation to instruct the roguelike mechanism, including factors like tactical challenge, random environment generation, permadeath, and resource management. It employed a novel approach to treat amblyopia by delivering visual stimuli through two independent channels, presenting them separately to each eye. Vision Planet was a newly developed serious game and certified as a Class II medical device by the National Medical Products Administration in mainland China (identifier: zhexiezhuzhun20212210379).

The treatment setup included (1) an 11-inch Huawei MatePad (Model DBR-W00) that operated the software and presented the visual stimuli, (2) anaglyph glasses that separated the visual stimuli and applied a monocular Gaussian blur to the fellow eye channel, and (3) the Vision Planet gamified medical software. The software automatically adjusted the magnitude of the blur based on the baseline VA difference between the

XSL•FO
RenderX

two eyes, with a greater disparity resulting in a higher blur intensity. During the sessions, character visuals were directed to the fellow eye and monster visuals to the amblyopic eye. Participants controlled the character's movements and attacks on the monsters using a virtual joystick and shoot button, aiming to achieve simultaneous gaze and binocular fusion, as illustrated in Figure 1.

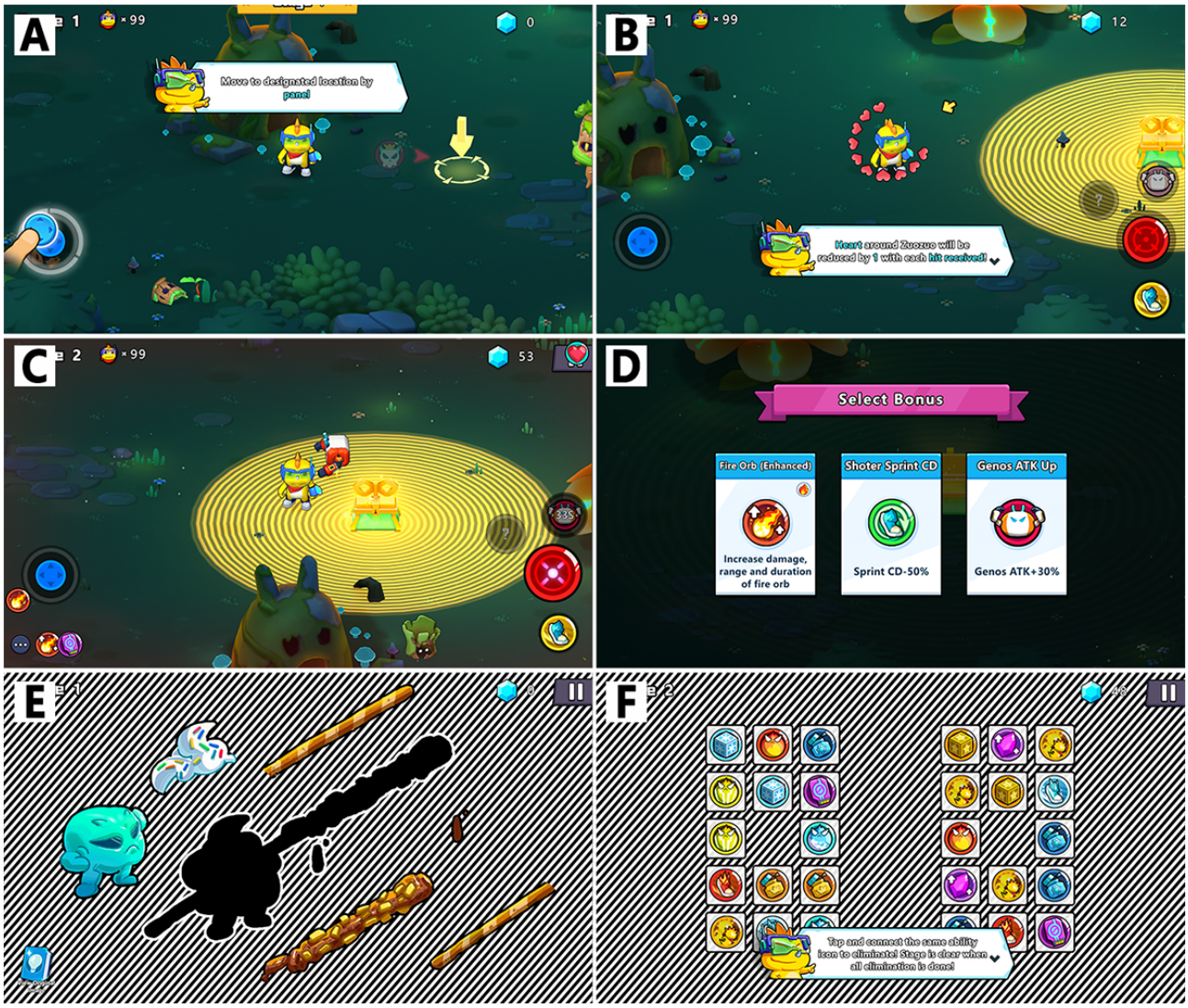
Figure 1. Binocular treatment setup. The top illustration shows a red-blue dichoptic game frame with a blurred character visual intended for the fellow eye and a monster visual for the amblyopic eye. The bottom image is a participant during a treatment session.



During the first Vision Planet training session, users were guided through a tutorial level that familiarized them with basic operational procedures, including movement, attack mechanics, injury responses, level completion strategies, and enemy attack patterns. In the primary game levels of Vision Planet, users should have eliminated all monsters to complete the level. Upon successful clearance, a treasure box was awarded which contained 3 buffs. The interfaces of the tutorial level, main game levels, and bonus levels are illustrated in Figure 2.

XSL•FO
RenderX

Figure 2. Training journey. The training journey in Vision Planet includes a tutorial level, main game levels, and bonus levels. (A and B) A structured tutorial level was designed to acquaint users with the basic operational procedures at the first training session. (C and D) In main game levels, users were rewarded with a treasure box upon successfully clearing a level, which contained a selection of 3 buffs. (E and F) Users progressed to a bonus level after completing 3 regular levels. The bonus gameplay was distinct and less challenging to offer a rhythmic break from the main game's intensity.



Vision Planet comprised 5 worlds that could be unlocked by meeting daily training goals, offering new elements and motivating users to continue their training. Boss level completion was rewarded with energy stones for card draws in the gacha system. Acquired cards could be strategically deployed for buffs in battles. Gems earned from defeating monsters could be used to purchase unique weapons from the store. Users had the option to select and equip their preferred weapon in the preparation interface (Figure 3).

XSL•FO
RenderX

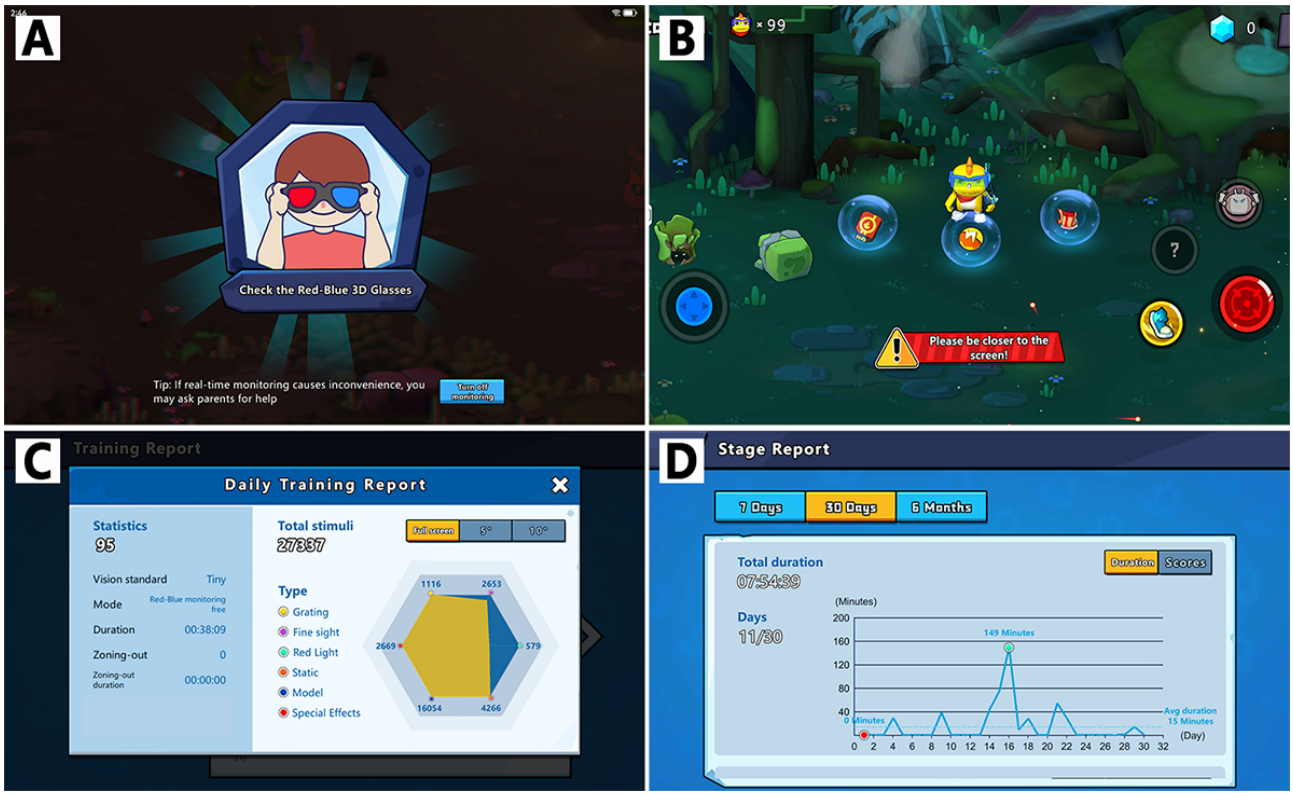
Figure 3. Incentive system. The incentive system mainly consists of unlockable levels, a gacha system, and an in-game store. (A and B) Vision Planet comprised 5 worlds with only the first available from the start, and others unlocked after meeting daily training time requirements for set days. (C and D) Users earned energy stones by defeating bosses for gacha draws, getting cards that boosted battles. (E and F) Users defeated monsters to earn gems for purchasing unique weapons, which could be equipped in the preparation interface to customize their training experience.



Using a real-time artificial intelligence visual engine, Vision Planet was capable of continuously monitoring and reminding participants to keep their eyes parallel to the screen at a distance of about 40 cm, while ensuring that participants wore the anaglyph glasses properly throughout the training session to guarantee the standard conduction of the intervention. Images were captured in real time by the built-in camera of the training tablet, and processed through the visual engine (algorithm) of the Vision Planet software. Vision Planet also calculated daily training metrics, including the duration of play and the scores, which could be shown in reports for various training cycles (Figure 4).

XSL•FO
RenderX

Figure 4. Back-end monitoring included supervised training and a summary report. (A and B) The visual engine monitored and prompted users to maintain a 40 cm distance and wear anaglyph glasses correctly for effective training. (C and D) A parent center and developer back end offered daily training metrics to help parents and providers monitor progress and prevent gaming addictions.



Study Schedule

After a minimum of 8 weeks of refractive correction, participants underwent binocular digital therapy for 60 minutes daily, 5 days a week. The primary outcomes measured were DVA, NVA, and SA. DVA was assessed using an electronic VA system with current refractive correction as necessary, following the Pediatric Eye Disease Investigator Group Amblyopia Treatment Study protocols, which use Lea symbols with crowded optotypes. NVA was tested at a distance of 40 cm using the Amblyopia Treatment Study's near acuity test, which included flip cards featuring single-surrounded Lea symbol charts [21]. SA was evaluated using the Titmus SA chart (by Stereo Optical) with current spectacle corrections as required. Ocular alignment was determined through the simultaneous prism and cover test. All assessments were performed by the same study-certified examiner and were scheduled at 4, 8, and 12 weeks. Patient compliance and satisfaction with the binocular digital therapy were evaluated at the week 8 visit using the Vision Planet Questionnaire (Multimedia Appendix 1).

Statistical Analysis

Data was analyzed using the SPSS version 25.0 software (IBM Corp). Continuous variables were presented as mean and SD, and categorical variables were reported as frequency and percentage. Stereopsis measurements were transformed to a natural logarithmic scale for statistical analyses. Differences in amblyopic eye and binocular DVA, NVA, and SA between the baseline and follow-up visits at 4, 8, and 12 weeks were evaluated using paired 2-tailed *t* tests. The effect size was calculated using a Cohen *d*, determined by dividing the mean difference by the SD of the differences. *P* values <.05 were considered statistically significant.

Results

Participant Characteristics

The study involved 11 children with amblyopia aged 4 - 6 (mean 5.6, SD 0.7) years. The cohort included 9 participants with anisometropia and 2 with combined mechanism amblyopia. All participants had received refractive correction at least 8 weeks prior to the study. In total, 8 participants had previously undergone patching and paused at least 2 weeks before the study. No participants had prior exposure to any other binocular therapies. Baseline characteristics of the participants are detailed in Table 1.

XSL•FO
RenderX

Table . Baseline characteristics of the participants.

Participant number	Age (years)	Sex	Amblyopic eye	Amblyopic eye VA[a] (logMAR[b])	Fellow eye VA (logMAR)	Spherical refraction (amblyopic eye)	Spherical refraction (fellow eye)	Type of amblyopia	Tropia	Refractive correction duration (years)	Previous patching treatment	Initial stereoacuity (arcseconds)
1	4.8	F[c]	OS	0.3	0.1	+5.25	+2.50	C[d]	10 E(T)[e]	2.5	Yes	800
2	5.9	F	OS	0.4	0	+2.75	+1.25	A[f]	—[g]	1.5	Yes	400
3	6.1	M[h]	OD	0.3	0	+8.00	+1.50	A	—	2.5	Yes	400
4	4.4	F	OS	0.2	0	+4.75	+1.25	C	30 X(T)[i]	1.5	Yes	800
5	6.9	M	OD	0.6	–0.1	+2.75	+1.00	A	—	1	Yes	400
6	4.9	F	OD	0.3	0	+6.50	+2.75	A	—	1.5	No	200
7	5.8	F	OS	0.4	0.1	+5.50	+4.00	A	—	0.2	No	60
8	6.3	M	OS	0.4	–0.1	+5.75	+2.00	A	—	0.2	No	60
9	5.9	M	OD	0.2	0	+6.00	+4.00	A	—	1.5	Yes	80
10	5.7	M	OS	0.4	0	+7.25	+3.25	A	—	1	Yes	400
11	5.3	M	OS	0.2	0	+4.50	+3.25	A	—	0.5	Yes	50

[a]VA: visual acuity.

[b]logMAR: base 10 logarithm of the minimum angle of resolution.

[c]F: female.

[d]C: combined.

[e]E(T): intermittent esotropia.

[f]A: anisometropic.

[g]Em dash indicates that participant did not have tropia.

[h]M: male.

[i]X(T): intermittent exotropia.

Visual Acuity and Stereoacuity

Before the intervention, the average BCVA of the amblyopic eye was 0.34 (SD 0.12) logMAR for DVA and 0.33 (SD 0.17) logMAR for NVA. The average binocular BCVA was 0 (SD 0.06) logMAR for DVA and 0.02 (SD 0.08) logMAR for NVA. The average SA was 2.34 (SD 0.46) log arcseconds. The DVA for the amblyopic eye improved to 0.31 (SD 0.13) logMAR at 4 weeks, 0.25 (SD 0.10) logMAR at 8 weeks, and 0.24 (SD 0.09) logMAR at 12 weeks. Statistically significant improvements were noted in the DVA of the amblyopic eye from baseline, with an increase of 0.9 lines at 8 weeks (P=.046; d=1.00) and 1.0 line at 12 weeks (P=.01; d=0.77), while no significant change was observed at 4 weeks (P=.34). Binocular BCVA for DVA improved to –0.1 (SD 0.08) logMAR at 4 weeks, 0 (SD 0.07) logMAR at 8 weeks, and −0.07 (SD 0.07) logMAR at 12 weeks, corresponding to a significant enhancement of 0.7 lines at 12 weeks (P=.006; d=1.00), with no significant changes at 4 weeks (P=.10) or 8 weeks (P=.19). The SA improved to 2.16 (SD 0.46) log arcseconds at 4 weeks, 2.06 (SD 0.41) log arcseconds at 8 weeks, and 2.04 (SD 0.39) log arcseconds at 12 weeks. Statistically significant improvements were noted in SA from baseline, with improvements of 0.28 log arcseconds at 8 weeks (P=.02; d=0.90) and 0.3 log arcseconds at 12 weeks (P=.01; d=0.97), while no significant change was observed at 4 weeks (P=.06). For the NVA, no statistically significant changes were observed for either the amblyopic eye or binocularly at any time point. The NVA for the amblyopic eye improved to 0.32 (SD 0.13) logMAR at 4 weeks (P=.68), 0.28 (SD 0.13) logMAR at 8 weeks (P=.27), and 0.24 (SD 0.09) logMAR at 12 weeks (P=.06). Binocular BCVA for NVA was recorded as 0.05 (SD 0.07) logMAR at 4 weeks (P=.27), 0 (SD 0.05) logMAR at 8 weeks (P=.34), and 0 (SD 0.06) logMAR at 12 weeks (P=.17).

Statistical results for VA and SA across all time points are shown in Table 2. Improvements in VA and SA at all time points are depicted in Figure 5. The progress of each participant in BCVA for the amblyopic eye, binocular BCVA, and SA at the week 8 follow-up is illustrated in Figure 6.

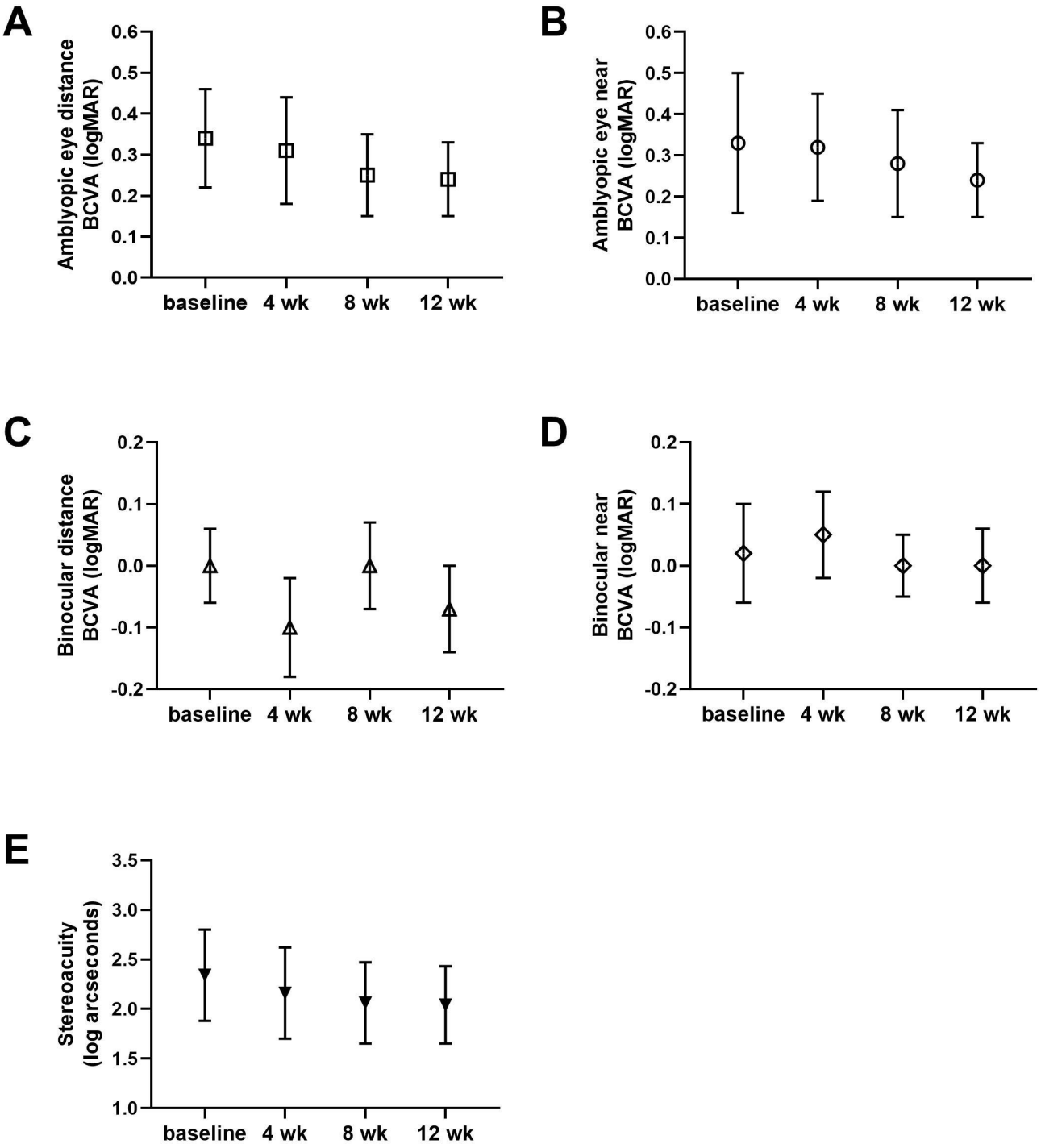
XSL•FO
RenderX

Table . Pairwise comparisons and Cohen *d*.

	Mean difference	Standardized mean difference (95% CI)	Cohen *d*	*t* value	*P* value
Distance visual acuity of the amblyopic eye					
Baseline vs 4 wks	0.02	0.06 (–0.02 to 0.06)	0.50	1.00	.34
Baseline vs 8 wks	0.06	0.09 (0-0.13)	1.00	2.28	.046
Baseline vs 12 wks	0.12	0.13 (0.03-0.21)	0.77	2.95	.01
Near visual acuity of the amblyopic eye					
Baseline vs 4 wks	0.01	0.07 (–0.04 to 0.06)	0.14	0.43	.68
Baseline vs 8 wks	0.05	0.13 (–0.04 to 0.13)	0.38	1.17	.27
Baseline vs 12 wks	0.08	0.13 (0-0.17)	0.69	2.17	.06
Binocular distance visual acuity					
Baseline vs 4 wks	0.04	0.07 (–0.01 to 0.08)	0.43	1.79	.10
Baseline vs 8 wks	0.03	0.06 (–0.02 to 0.07)	0.50	1.40	.19
Baseline vs 12 wks	0.05	0.05 (0.02-0.09)	1.00	3.46	.006
Binocular near visual acuity					
Baseline vs 4 wks	–0.04	0.10 (–0.11 to 0.03)	0.30	–1.17	.27
Baseline vs 8 wks	0.02	0.06 (–0.02 to 0.06)	0.50	1.00	.34
Baseline vs 12 wks	0.02	0.04 (–0.01 to 0.05)	0.50	1.49	.17
Stereoacuity					
Baseline vs 4 wks	0.17	0.26 (–0.01 to 0.35)	0.69	2.16	.06
Baseline vs 8 wks	0.27	0.31 (0.06-0.48)	0.90	2.89	.02
Baseline vs 12 wks	0.30	0.31 (0.09-0.51)	0.97	3.18	.01

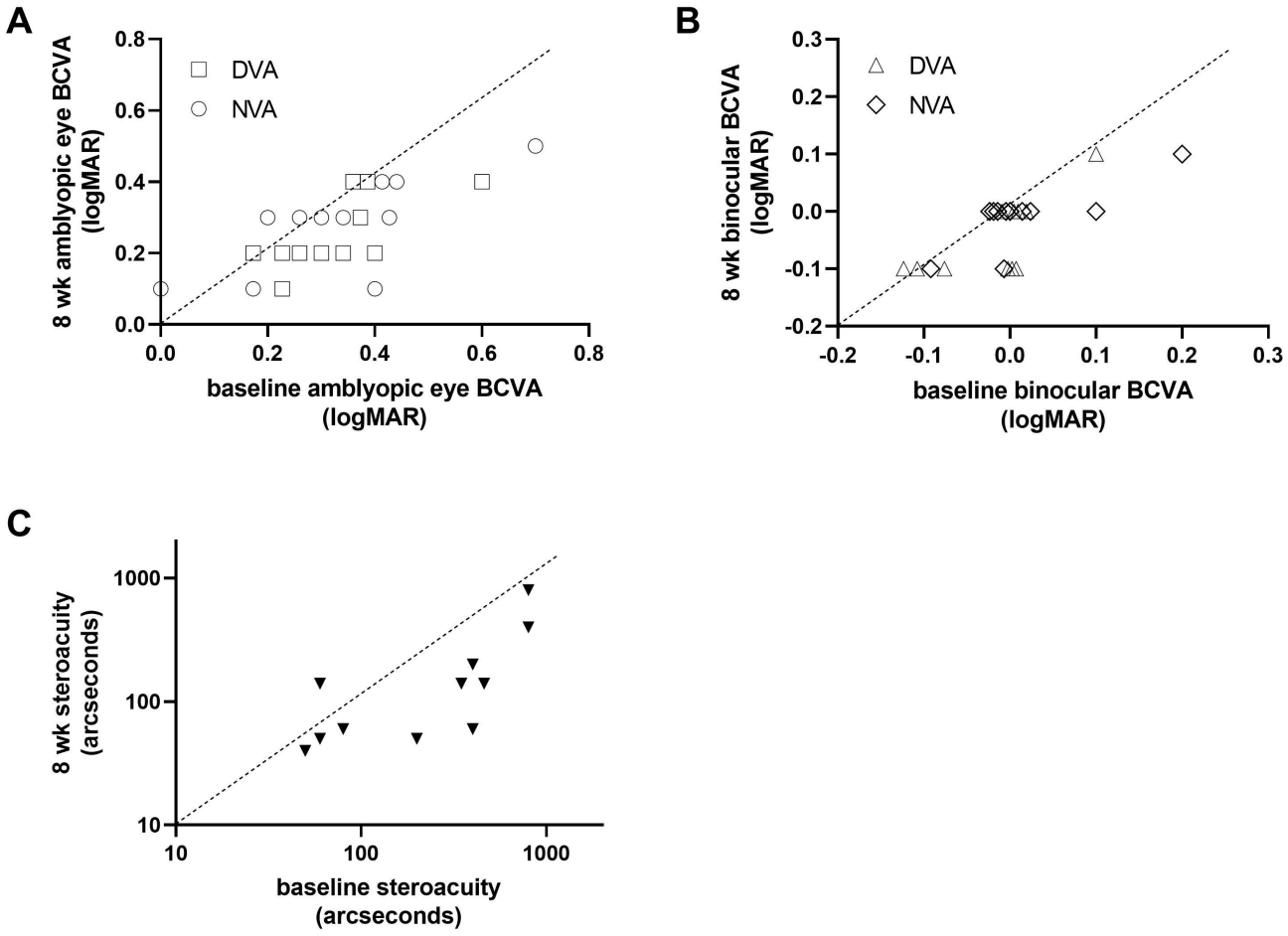
XSL•FO
RenderX

Figure 5. Improvement in visual acuity and stereoacuity. Best corrected visual acuity (BCVA) and stereoacuity were examined at 4 study time points—baseline, 4 weeks, 8 weeks, and 12 weeks. These are presented for (A) distance amblyopic eye BCVA, (B) near amblyopic eye BCVA, (C) distance binocular BCVA, (D) near binocular BCVA, and (E) stereoacuity. Error bars represent SD. log: logarithm; logMAR: base 10 logarithm of the minimum angle of resolution.



XSL•FO
RenderX

Figure 6. Improvement in BCVA and stereoacuity at the week 8 follow-up visit, as a function of baseline measurements. Individual data is depicted for the measurements of (A) distance and near amblyopic eye BCVA, (B) distance and near binocular BCVA, and (C) stereoacuity. Data under the diagonal unity line indicated amelioration. A value of 800 arcseconds was arbitrarily assigned if subjects failed the Titmus test. BCVA: best corrected visual acuity; DVA: distance visual acuity; NVA: near visual acuity.



Satisfaction Questionnaires and Compliance

The exit questionnaire revealed that 91% (10/11) of participants found Vision Planet easy or very easy to use and 64% (7/11) would prefer Vision Planet over patching for future amblyopia treatments. Detailed satisfaction data are available in Multimedia Appendix 1. The overall training completion rate for all participants at the week 8 follow-up was 98.5% (range: 93-114.5%), with 82% (9/11) of participants completing more than 90% of the prescribed training time. Detailed compliance data are shown in Multimedia Appendix 2.

Adverse Events

Participants were consistently monitored for adverse events such as diplopia, new or worsening eye heterotropia, deteriorative VA, or unexpected adverse events. No significant adverse events were observed, although one participant developed progressive esotropia. This participant had a near and far deviation of 10 PD esotropia with glasses at baseline, which increased to 20 PD esotropia with glasses by 12 weeks. The amblyopic eye DVA remained stable at 0.3 logMAR from baseline to 12 weeks. Stereoacuity improved from 800 arcseconds at baseline to 400 arcseconds at 12 weeks. Deviations remained stable at 4 and 8 weeks, but progressive esotropia was noted at the last follow-up. A recycloplegic refraction was immediately performed and full hyperopia spectacles were prescribed.

Discussion

Principal Findings

This pilot study confirmed the efficacy and safety of binocular digital therapy for treating anisometropic amblyopia in children aged 4-6 years. We observed improvements in amblyopic eye DVA and SA following this treatment. Apart from one case of progressive esotropia, no significant adverse events were reported.

Previous research on binocular therapy has documented VA enhancements through the use of dichoptic video games, contrast-adjusted films, and shutter glasses during video or interactive tasks [22,23]. Our findings add to the existing literature by demonstrating significant improvements in both DVA of the amblyopic eye and binocular DVA using the Vision Planet binocular treatment. However, another study involving children aged 7 - 12 years did not show any improvement in amblyopic eye VA at 16 weeks [24], possibly due to the older age of the participants. In contrast, a multicenter randomized controlled trial reported a 1.8-line improvement in amblyopic eye VA with digital treatment compared to a 0.8-line improvement with continued spectacle use alone at 12 weeks,

XSL•FO
RenderX

in children aged 4 - 7 years [25]. Recently, Wygnanski-Jaffe et al [26] found a 2.8-line improvement in BCVA of the amblyopic eye in children aged 4 - 9 years using a binocular eye tracking–based home treatment (CureSight) over 16 weeks, showing noninferior results compared to patching treatment, which achieved a 2.3-line improvement. Compared to the 1.0-line improvement in our study, the significant results in their study could be attributed to the longer duration of training (90 min/d for 16 weeks). Additionally, our prior research analyzed the dosing schedule of CureSight binocular treatment, revealing that more frequent treatment (5 days per week) was more effective than less frequent sessions (3 days per week), with improvements of 1.8 lines versus 1.5 lines for DVA and 2.9 lines versus 1.7 lines for NVA, respectively [27]. This suggests that a longer duration for binocular treatment may be more effective in treating amblyopia and the treatment in our study (60 min/d, 5 d/wk, for 12 weeks) may not have been sufficient to achieve optimal results.

Numerous studies have supported the improvement in SA with binocular treatment for amblyopia [18,28,29], while others have reported no significant changes [24,25,30]. In this study, we observed statistically significant enhancements in SA from baseline by 0.28 log arcseconds at 8 weeks and by 0.3 log arcseconds at 12 weeks with Vision Planet treatment. Similarly, our prior study with CureSight treatment showed significant improvement in SA by 0.38 log arcseconds at 12 weeks. Research indicates that binocular therapy could substantially benefit the restoration of binocular function. SA is closely linked with the depth of suppression but does not correlate with decreases in VA [31]. By addressing suppression and enhancing contrast balance, binocular digital therapy has shown potential as an effective alternative for treating amblyopia [30]. The success of binocular digital treatments may be closely related to the age of the participants and the engaging nature of the treatment content. Young children may benefit more from SA improvement due to cortical plasticity. Importantly, the content of binocular treatments should be engaging enough to maintain children's attention throughout the training period.

Comparison to Other Gamification Therapies

The falling blocks game is a commonly used gamification therapy for amblyopia. Participants score points in the game by moving the falling blocks to form complete lines of blocks. Holmes et al [32] reported that the mean amblyopic eye VA improved by 1.05 lines in the binocular falling blocks game group and 1.35 lines in the patching group at 16 weeks among children aged 5 to <13 years, which was similar to the DVA improvement observed at 12 weeks in our study (1.0 line). Gao et al [33] used a home-based binocular falling blocks game to treat amblyopia, reporting that the mean amblyopic eye VA improved by 0.6 lines from baseline in the binocular game group and 0.7 lines in the placebo game group at 6 weeks, with no statistical improvement noted. The limited effect of the gamified binocular therapy may be attributed to the short duration of the treatment.

Another gamification therapy for amblyopia is the Dig Rush game, an action-oriented adventure game with 42 levels in which miners dig for gold. Kelly et al [30] reported a mean improvement of 1.5 lines with 2 weeks of binocular Dig Rush game treatment in children aged 4-10 years, which was greater than the improvement observed in those treated with part-time patching (0.7 lines). However, the Pediatric Eye Disease Investigator Group found no benefit to VA or SA in amblyopia from binocular Dig Rush game treatment for children aged 7-12 years at 8 weeks, which might be attributed to the older age of the participants [34].

Finally, Gambacorta and colleagues [35] used the mirror stereoscope to achieve alignment and dichoptic games to treat amblyopia. The children were instructed to move around, collect health points, and tag robot opponents by aligning the fixation scope onto the center of the robot and clicking the mouse to activate the pointer tool. Following 20 hours of training, the VA of the amblyopic eye improved by 1.4 lines for the dichoptic group and by 0.6 lines for the monocular group. However, this therapy needed to be conducted in the lab, which increased the time and cost.

Compared to the falling blocks game and the Dig Rush game, the Vision Planet system incorporates a roguelike shooting game element, adding an entertaining dimension to the therapeutic process and potentially increasing adherence and motivation among patients undergoing amblyopia treatment. Roguelike games possess inherent factors such as random environment generation, nonmodality, complexity, and resource management. Additionally, roguelike mechanics have a high degree of compatibility with action shooting games. By leveraging these factors, different needs of various patients could be met and the game would achieve a high replayability, further improving the adherence. These characteristics align well with the cognitive abilities of patients with amblyopia and the repetitive training requirements of amblyopia treatment. This system not only adheres to the therapeutic principles of binocular treatment but also enhances patient engagement through the integration of gamified mechanisms.

Adverse Events

In the safety profile of the Vision Planet binocular treatment, one participant developed progressive esotropia. The study included 2 participants with combined mechanism amblyopia. One child, diagnosed with intermittent exotropia, displayed a stable deviation posttreatment. The other child with esotropia experienced worsening deviations of 10 PDs in both near and far assessments, possibly due to overaccommodation induced by intensive and close visual training. The child had a consistent deviation of 10 PDs at all distances pretreatment. This was based on their stable alignment profile during the study, which we documented as E(T) in the table to acknowledge the intermittent nature of the deviation. It was presumed that a microtropia with anomalous retinal correspondence might have decompensated to reveal their true angle posttreatment. The improvement in SA for this child was modest, with a change of approximately one octave. Based on established literature, a change of around 2 octaves in the SA threshold is generally required to surpass the test-retest variability for most SA assessments [36]. Given that the change observed in this case was less than 2 octaves, we acknowledge that this improvement could fall within the bounds of normal test-retest variability. As such, while the child

XSL•FO
RenderX

showed some improvements, we cannot definitively state that it reflects a true clinical gain beyond this variability. Further research is needed to explore the impact of binocular treatment on eye alignment.

Limitations

This study has several limitations, including its small sample size, modest follow-up duration, lack of randomization, and absence of a control group. To more definitively evaluate the effectiveness of binocular treatment compared to other amblyopia treatments, a larger, multicenter randomized clinical trial using Vision Planet treatment is necessary. Another limitation is the use of the Titmus test for assessing SA, which includes monocular cues. It is worth noting, however, that participants who experience VA improvement are likely to become better at recognizing and interpreting these monocular cues compared to their initial performance. Despite these limitations, this study serves as an initial assessment of a new approach to treating amblyopia. Future research will need to include more comprehensive and structured clinical evaluations to further explore this treatment's effectiveness. Additional studies and clinical trials are essential to confirm the efficacy and long-term benefits of this innovative treatment method.

Conclusion

The Vision Planet treatment has shown positive effects on both DVA and SA in children aged 4-6 years with amblyopia, indicating the potential effectiveness of binocular digital therapy as a treatment option for unilateral amblyopia.

Acknowledgments

This research was supported by the Shanghai Municipal Health Commission Program (201940485) and Shanghai Hospital Development Center Program (SHDC2023CRD013). We thank BOKE Medical Technology (Shanghai) Co, Ltd for providing the facilities that supported this study. The authors also extend gratitude to the participants and the staff for their invaluable contributions to this research.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

WZ was responsible for study conceptualization, methodology, drafting, and editing the manuscript. RL was responsible for study conceptualization, methodology, reviewing, and correspondence. SG and J Li were responsible for the software and methodology. J Lin and CH were responsible for the recruitment of the participants and data collection.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Vision Planet Questionnaire.
[DOCX File, 15 KB - games_v13i1e63384_app1.docx]

Multimedia Appendix 2
Compliance in 8 weeks.
[DOCX File, 14 KB - games_v13i1e63384_app2.docx]

References

1. Multi-ethnic Pediatric Eye Disease Study Group. Prevalence of amblyopia and strabismus in African American and Hispanic children ages 6 to 72 months the multi-ethnic pediatric eye disease study. Ophthalmology 2008 Jul;115(7):1229-1236. [doi: 10.1016/j.ophtha.2007.08.001] [Medline: 17953989]
2. Friedman DS, Repka MX, Katz J, et al. Prevalence of amblyopia and strabismus in White and African American children aged 6 through 71 months the Baltimore Pediatric Eye Disease Study. Ophthalmology 2009 Nov;116(11):2128-2134. [doi: 10.1016/j.ophtha.2009.04.034] [Medline: 19762084]
3. McKean-Cowdin R, Cotter SA, Tarczy-Hornoch K, et al. Prevalence of amblyopia or strabismus in Asian and non-Hispanic White preschool children: multi-ethnic pediatric eye disease study. Ophthalmology 2013 Oct;120(10):2117-2124. [doi: 10.1016/j.ophtha.2013.03.001] [Medline: 23697956]
4. Tarczy-Hornoch K, Varma R, Cotter SA, et al. Risk factors for decreased visual acuity in preschool children: the multi-ethnic pediatric eye disease and Baltimore pediatric eye disease studies. Ophthalmology 2011 Nov;118(11):2262-2273. [doi: 10.1016/j.ophtha.2011.06.033] [Medline: 21856014]
5. Birch EE, Stager DR. Monocular acuity and stereopsis in infantile esotropia. Invest Ophthalmol Vis Sci 1985 Nov;26(11):1624-1630. [Medline: 4055294]

XSL•FO
RenderX

6. Chen PL, Chen JT, Tai MC, Fu JJ, Chang CC, Lu DW. Anisometropic amblyopia treated with spectacle correction alone: possible factors predicting success and time to start patching. Am J Ophthalmol 2007 Jan;143(1):54-60. [doi: 10.1016/j.ajo.2006.09.027] [Medline: 17113556]
7. Cotter SA, Pediatric Eye Disease Investigator Group. Treatment of anisometropic amblyopia in children with refractive correction. Ophthalmology 2006 Jun;113(6):895-903. [doi: 10.1016/j.ophtha.2006.01.068] [Medline: 16751032]
8. Holmes JM, Kraker RT, Beck RW, et al. A randomized trial of prescribed patching regimens for treatment of severe amblyopia in children. Ophthalmology 2003 Nov;110(11):2075-2087. [doi: 10.1016/j.ophtha.2003.08.001] [Medline: 14597512]
9. Wang S, Wen W, Zhu W, et al. Effect of combined atropine and patching vs patching alone for treatment of severe amblyopia in children aged 3 to 12 years: a randomized clinical trial. JAMA Ophthalmol 2021 Sep 1;139(9):990-996. [doi: 10.1001/jamaophthalmol.2021.2413] [Medline: 34264296]
10. Hussein MAW, Coats DK, Muthialu A, Cohen E, Paysse EA. Risk factors for treatment failure of anisometropic amblyopia. J AAPOS 2004 Oct;8(5):429-434. [doi: 10.1016/j.jaapos.2003.09.005] [Medline: 15492734]
11. Webber AL, Wood JM, Gole GA, Brown B. The effect of amblyopia on fine motor skills in children. Invest Ophthalmol Vis Sci 2008 Feb;49(2):594-603. [doi: 10.1167/iovs.07-0869] [Medline: 18235004]
12. Kanonidou E, Proudlock FA, Gottlob I. Reading strategies in mild to moderate strabismic amblyopia: an eye movement investigation. Invest Ophthalmol Vis Sci 2010 Jul;51(7):3502-3508. [doi: 10.1167/iovs.09-4236] [Medline: 20207968]
13. Kelly KR, Cheng-Patel CS, Jost RM, Wang YZ, Birch EE. Fixation instability during binocular viewing in anisometropic and strabismic children. Exp Eye Res 2019 Jun;183:29-37. [doi: 10.1016/j.exer.2018.07.013] [Medline: 30006273]
14. Webber AL, Wood JM, Gole GA, Brown B. Effect of amblyopia on self-esteem in children. Optom Vis Sci 2008 Nov;85(11):1074-1081. [doi: 10.1097/OPX.0b013e31818b9911] [Medline: 18981922]
15. Hess RF, Mansouri B, Thompson B. A new binocular approach to the treatment of amblyopia in adults well beyond the critical period of visual development. Restor Neurol Neurosci 2010;28(6):793-802. [doi: 10.3233/RNN-2010-0550] [Medline: 21209494]
16. Knox PJ, Simmers AJ, Gray LS, Cleary M. An exploratory study: prolonged periods of binocular stimulation can provide an effective treatment for childhood amblyopia. Invest Ophthalmol Vis Sci 2012 Feb 21;53(2):817-824. [doi: 10.1167/iovs.11-8219] [Medline: 22169103]
17. Baker DH, Meese TS, Mansouri B, Hess RF. Binocular summation of contrast remains intact in strabismic amblyopia. Invest Ophthalmol Vis Sci 2007 Nov;48(11):5332-5338. [doi: 10.1167/iovs.07-0194] [Medline: 17962490]
18. Hess RF, Thompson B. Amblyopia and the binocular approach to its therapy. Vis Res 2015 Sep;114:4-16. [doi: 10.1016/j.visres.2015.02.009] [Medline: 25906685]
19. Bocqué C, Wang J, Rickmann A, Julich-Haertel H, Kaempf U, Januschowski K. Gamification to support adherence to a therapeutic ambylopia treatment for children: retrospective study using a focal ambient visual acuity stimulation game. JMIR Pediatr Parent 2023 Feb 1;6:e32282. [doi: 10.2196/32282] [Medline: 36724007]
20. Wu WL, Huang YL, Liang JM, Chen CH, Wang CC, Ho WH. Interactive digital game for improving visual-perceptual defects in children with a developmental disability: randomized controlled trial. JMIR Serious Games 2022 Apr 15;10(2):e34756. [doi: 10.2196/34756] [Medline: 35436215]
21. Repka MX, Cotter SA, Beck RW, et al. A randomized trial of atropine regimens for treatment of moderate amblyopia in children. Ophthalmology 2004 Nov;111(11):2076-2085. [doi: 10.1016/j.ophtha.2004.04.032] [Medline: 15522375]
22. Birch EE, Jost RM, De La Cruz A, et al. Binocular amblyopia treatment with contrast-rebalanced movies. J AAPOS 2019 Jun;23(3):160. [doi: 10.1016/j.jaapos.2019.02.007] [Medline: 31103562]
23. Herbison N, MacKeith D, Vivian A, et al. Randomised controlled trial of video clips and interactive games to improve vision in children with amblyopia using the I-BiT system. Br J Ophthalmol 2016 Nov;100(11):1511-1516. [doi: 10.1136/bjophthalmol-2015-307798] [Medline: 26951772]
24. Elhusseiny AM, Bishop K, Staffa SJ, Zurakowski D, Hunter DG, Mantagos IS. Virtual reality prototype for binocular therapy in older children and adults with amblyopia. J AAPOS 2021 Aug;25(4):217. [doi: 10.1016/j.jaapos.2021.03.008] [Medline: 34246761]
25. Xiao S, Angjeli E, Wu HC, et al. Randomized controlled trial of a dichoptic digital therapeutic for amblyopia. Ophthalmology 2022 Jan;129(1):77-85. [doi: 10.1016/j.ophtha.2021.09.001] [Medline: 34534556]
26. Wygnanski-Jaffe T, Kushner BJ, Moshkovitz A, Belkin M, Yehezkel O, CureSight Pivotal Trial Group. An eye-tracking-based dichoptic home treatment for amblyopia: a multicenter randomized clinical trial. Ophthalmology 2023 Mar;130(3):274-285. [doi: 10.1016/j.ophtha.2022.10.020] [Medline: 36306974]
27. Zhu W, Tian T, Yehezkel O, et al. A prospective trial to assess the efficacy of eye-tracking-based binocular treatment versus patching for children's amblyopia: a pilot study. Semin Ophthalmol 2023 Nov;38(8):761-767. [doi: 10.1080/08820538.2023.2223275] [Medline: 37339068]
28. Hess RF, Mansouri B, Thompson B. A binocular approach to treating amblyopia: antisuppression therapy. Optom Vis Sci 2010 Sep;87(9):697-704. [doi: 10.1097/OPX.0b013e3181ea18e9] [Medline: 20622704]
29. Stewart CE, Wallace MP, Stephens DA, Fielder AR, Moseley MJ, MOTAS Cooperative. The effect of amblyopia treatment on stereoacuity. J AAPOS 2013 Apr;17(2):166-173. [doi: 10.1016/j.jaapos.2012.10.021] [Medline: 23622448]

XSL•FO
RenderX

30. Kelly KR, Jost RM, Dao L, Beauchamp CL, Leffler JN, Birch EE. Binocular iPad game vs patching for treatment of amblyopia in children: a randomized clinical trial. JAMA Ophthalmol 2016 Dec 1;134(12):1402-1408. [doi: 10.1001/jamaophthalmol.2016.4224] [Medline: 27832248]
31. Webber AL, Schmid KL, Baldwin AS, Hess RF. Suppression rather than visual acuity loss limits stereoacuity in amblyopia. Invest Ophthalmol Vis Sci 2020 Jun 3;61(6):50. [doi: 10.1167/iovs.61.6.50] [Medline: 32579677]
32. Holmes JM, Manh VM, Lazar EL, et al. Effect of a binocular iPad game vs part-time patching in children aged 5 to 12 years with amblyopia: a randomized clinical trial. JAMA Ophthalmol 2016 Dec 1;134(12):1391-1400. [doi: 10.1001/jamaophthalmol.2016.4262] [Medline: 27812703]
33. Gao TY, Guo CX, Babu RJ, et al. Effectiveness of a binocular video game vs placebo video game for improving visual functions in older children, teenagers, and adults with amblyopia: a randomized clinical trial. JAMA Ophthalmol 2018 Feb 1;136(2):172-181. [doi: 10.1001/jamaophthalmol.2017.6090] [Medline: 29302694]
34. Pediatric Eye Disease Investigator Group, Holmes JM, Manny RE, et al. A randomized trial of binocular Dig Rush game treatment for amblyopia in children aged 7 to 12 years. Ophthalmology 2019 Mar;126(3):456-466. [doi: 10.1016/j.ophtha.2018.10.032] [Medline: 30352226]
35. Gambacorta C, Nahum M, Vedamurthy I, et al. An action video game for the treatment of amblyopia in children: a feasibility study. Vis Res 2018 Jul;148:1-14. [doi: 10.1016/j.visres.2018.04.005] [Medline: 29709618]
36. Adams WE, Leske DA, Hatt SR, Holmes JM. Defining real change in measures of stereoacuity. Ophthalmology 2009 Feb;116(2):281-285. [doi: 10.1016/j.ophtha.2008.09.012] [Medline: 19091410]

Abbreviations

BCVA: best corrected visual acuity
DVA: distance visual acuity
log: logarithm
logMAR: base 10 logarithm of the minimum angle of resolution
NVA: near visual acuity
PD: prism diopter
SA: stereoacuity
VA: visual acuity

Edited by R Baranyi; submitted 18.06.24; peer-reviewed by JY Yao, M Piano, P Haring; revised version received 03.11.24; accepted 08.11.24; published 16.01.25.

Please cite as:
Zhu W, Gu S, Li J, Lin J, Hu C, Liu R
Transformative Gamified Binocular Therapy for Unilateral Amblyopia in Young Children: Pilot Prospective Efficacy and Safety Study
JMIR Serious Games 2025;13:e63384
URL: https://games.jmir.org/2025/1/e63384
doi:10.2196/63384

© Wenqing Zhu, Shuneng Gu, Jian Li, Jin Lin, Chanling Hu, Rui Liu. Originally published in JMIR Serious Games (https://games.jmir.org), 16.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

User Experience of a Semi-Immersive Musical Serious Game to Stimulate Cognitive Functions in Hospitalized Older Patients: Questionnaire Study

Laurent Samson[1*]; Lena Carcreff[2*], PhD; Frédéric Noublanche[2,3,4], PhD; Sophie Noublanche[1], Dr med; Hélène Vermersch-Leiber[1], Dr med; Cédric Annweiler[1,2,3,4,5], MD, PhD, Prof Dr med

[1]Department of Medical and Rehabilitation Care, Angers University Hospital, Angers, France

[2]Research Center on Autonomy and Longevity, Angers Living Lab in Hospital Geriatrics (Allegro), Department of Geriatric Medicine and Memory Clinic, Angers University Hospital, 4 rue Larey, Angers, France

[3]Laboratory of Psychology of Pays de la Loire, EA 4638, Angers University, Angers, France

[4]Gérontopôle Autonomie Longévité des Pays de la Loire, Nantes, France

[5]Robarts Research Institute, Department of Medical Biophysics, Schulich School of Medicine and Dentistry, The University of Western Ontario, London, ON, Canada

[*]these authors contributed equally

Corresponding Author:
Lena Carcreff, PhD
Research Center on Autonomy and Longevity, Angers Living Lab in Hospital Geriatrics (Allegro), Department of Geriatric Medicine and Memory Clinic, Angers University Hospital, 4 rue Larey, Angers, France

Abstract

Background: Reminiscence therapy through music is a psychosocial intervention with benefits for older patients with neurocognitive disorders. Therapies using virtual or augmented reality are efficient in ecologically assessing, and eventually training, episodic memory in older populations. We designed a semi-immersive musical game called "A Life in Songs," which invites patients to immerse themselves in a past era through visuals and songs from that time period. The game aspires to become a playful, easy-to-use, and complete tool for the assessment, rehabilitation, and prevention of neurocognitive decline associated with aging.

Objective: This study aimed to assess the user experience (UX) associated with the newly designed serious game.

Methods: After one or several sessions of the game guided by the therapist, patients of the geriatric wards were asked to answer questions selected from 2 widely known UX scales (AttrakDiff and meCUE [modular evaluation of the components of user experience]) with the therapist's help. The internal consistency of the UX dimensions was assessed through Cronbach α to verify the validity of the dimensions. The level of engagement of the patient throughout the experimental session was also assessed following an internally developed scale, which included 5 levels (interactive, constructive, active, passive, and disengaged behaviors). UX mean scores were computed and presented graphically. Verbal feedbacks were reported to support the quantitative results.

Results: Overall, 60 inpatients with a mean age of 84.2 (SD 5.5) years, the majority of whom were women (41/60, 68%), were included. Their score on the Mini-Mental State Examination (MMSE) ranged between 12 and 29. A majority of patients (27/56, 48%) had no major neurocognitive disorder (MNCD), 22/56 (39%) had mild MNCD, and 7/56 (13%) had moderate MNCD. The results revealed very positive UX with mean values beyond the neutral values for every UX dimension of both scales. The overall mean (SD) judgment was rated 3.92 (SD 0.87) (on a scale of −5 to 5). Internal consistency was acceptable to good for the emotional dimensions of the meCUE. Questionable to unacceptable consistency was found for the other UX dimensions. Participants were mostly active (23/60, 38%) and constructive (21/60, 35%).

Conclusions: These findings demonstrated a very good appreciation of the game by geriatric inpatients. Participants' and health care professionals' verbal comments strongly aligned with the quantitative results. The poor internal consistency in the UX dimensions reflected the high heterogeneity among the included patients. Further studies are needed to evaluate the potential benefits of clinical factors such as neurocognitive functions, mood, depression, or quality of life.

(JMIR Serious Games 2025;13:e57030) doi:10.2196/57030

KEYWORDS

virtual reality; geriatrics; reminiscence; episodic memory; serious game; neurocognitive disorders; older adults; user experience

XSL•FO
RenderX

Introduction

Physical frailty and cognitive impairment are common age-related alterations. While the prevalence of cognitive frailty in the community-dwelling older population is low (<5%), it considerably increases in clinical settings [1] and will also increase due to the current demographic transition [2]. Among cognitive impairments, memory loss is one of the earliest signs before the onset of major neurocognitive disorder (MNCD) (ie, dementia) [3]. MNCD, as described by the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* (*DSM-5*) [4], refers to a significant decline in cognitive functioning, affecting one or more cognitive domains such as attention, memory, and language. The deficits interfere with the person's ability to live independently, requiring assistance in daily activities.

Episodic (or autobiographical) memory is defined as the memory of personal life knowledge and personally experienced events [5]. It has a primordial role in personal identity, maintaining the feeling of time continuity [6]. In addition to the mental images that punctuate one's life, emotion is the privileged mediator of episodic memory. This form of memory is particularly vulnerable to age-linked changes, and its alteration is seen as a hallmark of early mild cognitive impairment and Alzheimer disease [7].

Aside from pharmacological interventions, psychosocial interventions have been developed and adopted with the objective of maintaining or improving the functions, relationships, and well-being of people with cognitive impairments [8]. Among these psychosocial interventions, reminiscence therapy is seen as a credible and efficient intervention. It consists of discussing past events or experiences with the patient, classically with the help of tangible prompts such as pictures, familiar items, music, or sounds [9]. It has been demonstrated in a recent meta-analysis that reminiscence therapy has positive effects on the quality of life, cognition, communication, and mood of people living with MNCD [10].

Virtual reality (VR) and, more recently, augmented reality (AR) (superimposition of VR elements onto the real-world environment) have become popular in the medical field [11]. They were found efficient not only in surgery education and training, but also in pain management, in motor and cognitive rehabilitation after stroke, and in mental health conditions such as anxiety and depression [11]. The advantage of using such supports is to stimulate both cognitive and motor functions [12]. The use of VR and AR has also been investigated in older populations to ecologically assess, and eventually train, episodic memory [7,13], thanks to the high level of immersion. The findings were in favor of positive effects on well-being as well as cognitive function improvements [14,15]. Depending on the immersion level, the technology can be categorized as immersive, semi-immersive, or nonimmersive; semi-immersive being certainly the most appropriate for older patients with cognitive impairments [14]. Nonimmersive VR was found to positively influence the rehabilitation of the most common geriatric syndromes [15].

In this context, a semi-immersive musical game called "A Life in Songs" ("*Une vie en Chanson*") has been designed. It immerses the user in a past decade (from 1950 to 2020), thanks to remarkable songs and events of this decade and thanks to a visual of a decade-related period living room (Figure 1). The game stimulates autobiographical memory without interpretative purpose, in a playful way while allowing to mentally relive autobiographical and semantic events of the user's life. "A Life in Songs" aspires to become a stand-alone tool for the assessment, rehabilitation, and prevention of neurocognitive decline associated with aging, thus promoting autonomy.

The design of the game required vigilance, in particular regarding ease of use, ergonomics, aesthetics, reliability, and adaptability. A good user experience (UX) is crucial for the success of any digital product. UX has recently emerged in the field of human-computer interaction as an extension of the concept of usability. It helps to consider the whole factors beyond the usefulness of a product, thus considering not only the pragmatic qualities but also the hedonic qualities, the emotions, and the intention to use [16]. This approach has been considered to include humans, their context, and their needs in the creative process. Before conducting a full-scale trial to assess the game's effectiveness, it is essential to first evaluate its UX.

This study thus aimed to evaluate the UX of the game among a geriatric hospitalized population. A secondary objective was to compare the UX between patients with and without MNCD. We hypothesized that overall UX would be found favorable (above the mean neutral value of the UX scales) and that UX would be similar for both groups of patients.

XSL•FO
RenderX

Figure 1. Overview of the game. (A) Virtual living room appearance of the decade 1960 - 1970. (B) The virtual reality (VR) device on the floor and the patient wearing VR glasses. (C) Examples of remarkable French singers and events of the decade 1960 - 1970. (D) Opened dialogue between the patient and the therapist.



Methods

Study Design

This study was a unicentric nonrandomized cross-sectional UX study using a "task-based experiment" method [17]. It was conducted in the Geriatric Department of Angers University Hospital, supported by the investigative team of its living lab [18], between October 2022 and May 2023.

Ethical Considerations

The protocol was approved by the ethical committee of Angers University Hospital (2022 - 024 and 2023 - 038), and it was declared to the National Commission for Information Technology and Civil Liberties (ar21-0159v0). Oral nonopposition to participation was obtained in accordance with French legal requirements for category 3 clinical research [19]. All data were deidentified. No compensation was provided to participants, as the study was integrated into their care program. Consent for publication was obtained from identifiable individuals featured in Figure 1 .

Description of the Serious Game

"A Life in Songs" ("*Une vie en Chanson*") is a musical game created in 2017 by LS, the first author of this paper and an art therapist, within the university hospital. The game was initially built on Microsoft PowerPoint support, based on the Goose Game principle where each square represented a year, starting from 1950 until 2020. A remarkable French song and French or international event were associated to each year (square), for instance, *La Javanaise* from Serge Gainsbourg and the assassination of John F Kennedy in 1963. In order to be more attractive and exploitable by other therapists, the game has been redesigned and incorporated into a plug-and-play semi-immersive VR solution (CADWall, Imagin-VR) in 2021. This solution consists of the 3D projection of a virtual living room on a screen wall (Figure 1B). The users wear VR glasses to explore the environment by moving their heads. The game is divided into 7 sections, corresponding to 7 decades (from 1950 to 2020). For each decade, the appearance of the virtual living room changes (decoration and objects according to the period) (Figure 1A). The therapist plays 5 songs from this decade and asks the participant to remember the singer's name. The 5 corresponding singer faces appear on the virtual wall for help. A total of 5 pictures related to 5 remarkable events of the decade appear on the wall as well (Figure 1C). The game is completely customizable to meet the needs of the patient, therapist, or caregiver. In summary, "A Life in Songs" in VR was designed to immerse patients in the past in an innovative, playful, and pleasant way, aiming to open the dialogue and stimulate episodic memory (Figure 1D). Immersion in the virtual environment promotes natural and instinctive interactions in real time via sensorimotor interfaces. This innovation was also thought to provide health care professionals with an interactive, collaborative, and turnkey tool tailored to their patients and offer the possibility to create their own training protocols. One advantage of the CADWall solution is that it is easily

XSL•FO
RenderX

transportable and simple to set up, thus usable in several units of the institution. At the time of the study, the game was not registered with the agency for the protection of programs. It is co-owned by Angers University Hospital and Imagin-VR.

Recruitment

Patients hospitalized in a geriatric care unit were informed about the study by the medical doctor if they were aged 75 years or older and if they were able to respond to a questionnaire (UX assessment). Interested patients were then approached by the art therapist to give oral and written detailed information about the study and the protocol. Patients under legal protection, as well as non-francophone patients, were not included.

Protocol

The patients were taken to a specific art therapy room within the geriatric unit by nursing assistants. The patient was allowed to participate on his or her own or accompanied by a relative or with other patients. The material and the game were presented, and the patient was given VR glasses. The investigator (the art therapist) stayed visible to the patient while wearing the glasses. After a few minutes of getting used to the device, the patient was asked to choose a decade to start the game. The investigator launched the corresponding decade: the appropriate living room 3D visual appeared on the virtual wall. A total of 5 songs (from 5 different years of the decade) were played and the patient was asked if he or she knew the singer after each of them. Additionally, after each song, a picture of a remarkable national or international event that happened during the same year appeared on the wall and the patient was asked to talk about this event if he or she remembered it. The investigator facilitated the session in order to guide the patients as much as possible and let them talk about their evoked memories. On the patient's willing basis, the session could be prolonged by playing with another decade or two.

At the end of the game session, the patient was asked to answer several questions about his or her UX. Finally, the investigator rated the engagement of the patient. Throughout the entire experimental session, the investigator manually recorded verbal feedback. At a later time, health care professionals' opinions were collected through brief, unrecorded interviews.

UX Assessment

UX was assessed through 2 self-administered (with the therapist's assistance if needed) questionnaires inspired from 2 standardized and well-known questionnaires: the AttrakDiff [20] and the modular evaluation of the components of user experience (meCUE) [21] in their French versions [22,23]. The AttrakDiff is composed of 28 items assessing 4 UX dimensions: pragmatic quality (PQ), hedonic quality identification (HQI), hedonic quality stimulation (HQS), and attractiveness (ATT). Each item is presented as a Likert scale semantic differential that represents opposites (eg, "simple-complicated"). The rates range between −3 and 3. To avoid the tendency of acquiescence, the valence of the items was mixed: words on the left of the Likert scale were sometimes positive, sometimes negative [22]. The meCUE is composed of 30 items, structured in 4 modules (product perceptions, emotions, consequences, and global assessment). Each item is presented as a sentence to which the user agrees or disagrees on a Likert scale (from 1 "strongly disagree" to 7 "strongly agree," except for the item of global assessment which is from −5 "bad" to 5 "good"). The items of the meCUE are categorized into 10 components (usefulness, usability, visual aesthetics, status, commitment, positive and negative emotions, product loyalty, intention to use, and overall evaluation) [24].

Some items have been removed from the standardized questionnaires due to the expected difficulties of understanding the patients with neurocognitive disorders. In total, 12 items were kept from the AttrakDiff (3 of each dimension), and 17 were kept from the meCUE (6 in the module "product perception," 8 in the module "emotions," 2 in the module "consequences," and 1 item of "global assessment").

Engagement Behavior

The level of engagement during the whole game session was rated by the experimenter using an internally designed scale called "ICAPD" (interactive, constructive, active, passive, and disengaged), which was inspired from the ICAP (interactive, constructive, active, and passive) hypothesis proposed by Chi et al [25,26]. The ICAPD scale differentiated five engagement behaviors as follows: (1) interactive—the patient discussed, questioned, and debated with the experimenter; (2) constructive—the patient asked questions, constructed his answers, justified them, and offered ideas to the experimenter; (3) active—the patient followed the experimental procedure and answered to the questions; (4) passive—the patient participated summarily in the experimental procedure and answered summarily to the experimenter's questions; (5) disengaged—the patient did not follow the experimental procedure. The scale is provided in Multimedia Appendix 1. The investigator attributed the level of engagement according to the patient's dominant behavior during the whole game session.

Data Collection

The age and sex of the participants, the 29 UX scores, and the engagement levels were collected. The level of neurocognitive disorders assessed through the Mini-Mental State Examination (MMSE) [27] was collected from the computer-based patient record. The MMSE provided a brief screening test that quantitatively assessed the severity of the cognitive disorder. The cutoff value of ≤26 out of 30 was used to dichotomize the cohort into 2 groups: without MNCD (MMSE≥26) and with MNCD (MMSE<26) [28], for the purpose of the secondary objective of this study.

Data Analysis

The scoring for the items of the AttrakDiff was adapted in order to report negative values for negative valence and positive values for positive valence. In addition, the scoring for the items of the meCUE relative to negative emotions were inverted so that a higher score in this dimension corresponded to fewer negative emotions, and thus better UX. Quantitative variables were presented as mean (SD) or median (IQR), according to the data distribution. Qualitative data were described in terms of numbers and percentages. The UX scores were presented graphically as means and SDs as recommended in a previous study [16].

XSL•FO
RenderX

Comparison of the UX scores between the patients with and without MNCD was performed through Wilcoxon unpaired tests given the low number of participants in each group. Patients with missing MMSE scores were excluded from this analysis. The distribution of each engagement behavior was reported. All descriptive statistics and graphical representations were performed using R (v4.1.1) and the RStudio interface (v1.4) [29]. The internal consistency of the UX dimensions was assessed through Cronbach α when possible (number of items >2 per dimension). Internal consistency was interpreted as excellent if $\alpha \geq 0.9$, good if $0.8 \leq \alpha < 0.9$, acceptable if $0.7 \leq \alpha < 0.8$, questionable if $0.6 \leq \alpha < 0.7$, poor if $0.5 \leq \alpha < 0.6$, and unacceptable if $\alpha < 0.5$ [30]. The artwork was created in Inkscape.

The analyses and reporting of the results followed the CONSORT-eHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) guidelines [31].

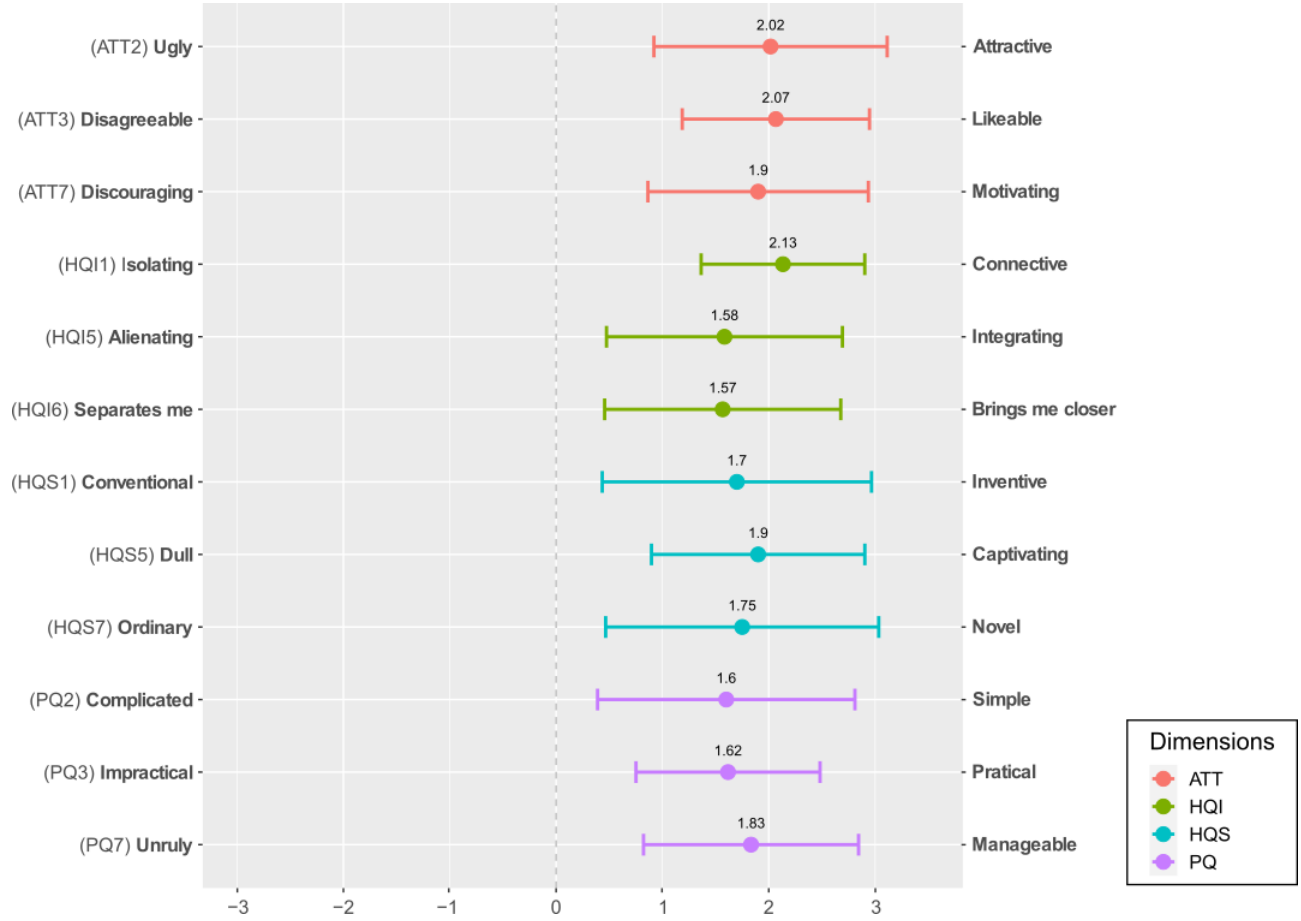
Results

UX results are presented in Figures 2 and 3. The scores to the AttrakDiff items were all positive (Figure 2) with a mean score of 1.81 (SD 0.2) across all items. The mean (SD) scores obtained for each dimension were as follows: 1.99 (1.01) for ATT, 1.76 (1.04) for HQI, 1.78 (1.19) for HQS, and 1.68 (1.04) for PQ. The scores to modules I, II, and III of the meCUE were also above the neutral value of 4, with a mean score of 5.58 (SD 0.44) across all items of these modules (Figure 3). The lower bounds of the error bars (SDs) for 3 of the 16 meCUE items extended beyond the neutral value 4. Global assessment (module IV) was rated 3.92 (0.87). The details of the scores for each item and each dimension are available in Multimedia Appendix 2.

A total of 60 patients, 41 women and 19 men, with a mean age of 84.5 (SD 5.5) years were included in the study between October 2022 and May 2023. Out of these 60 patients, 4 patients did not have an MMSE score recorded in their computer-based records. Among the 56 patients with available MMSE scores, the mean score was 24.4 (SD 3.9), with a minimum of 12 and a maximum of 29. Using the cutoff value of 26/30, 27 out of 56 (48%) were categorized without MNCD and 29 out of 56 (52%) with MNCD. The main reasons for hospitalization were repeated falls, fractures, memory loss, associated anxiety-depressive disorders, poststroke rehabilitation, Parkinson disease, neurological pathologies, and chronic pain.

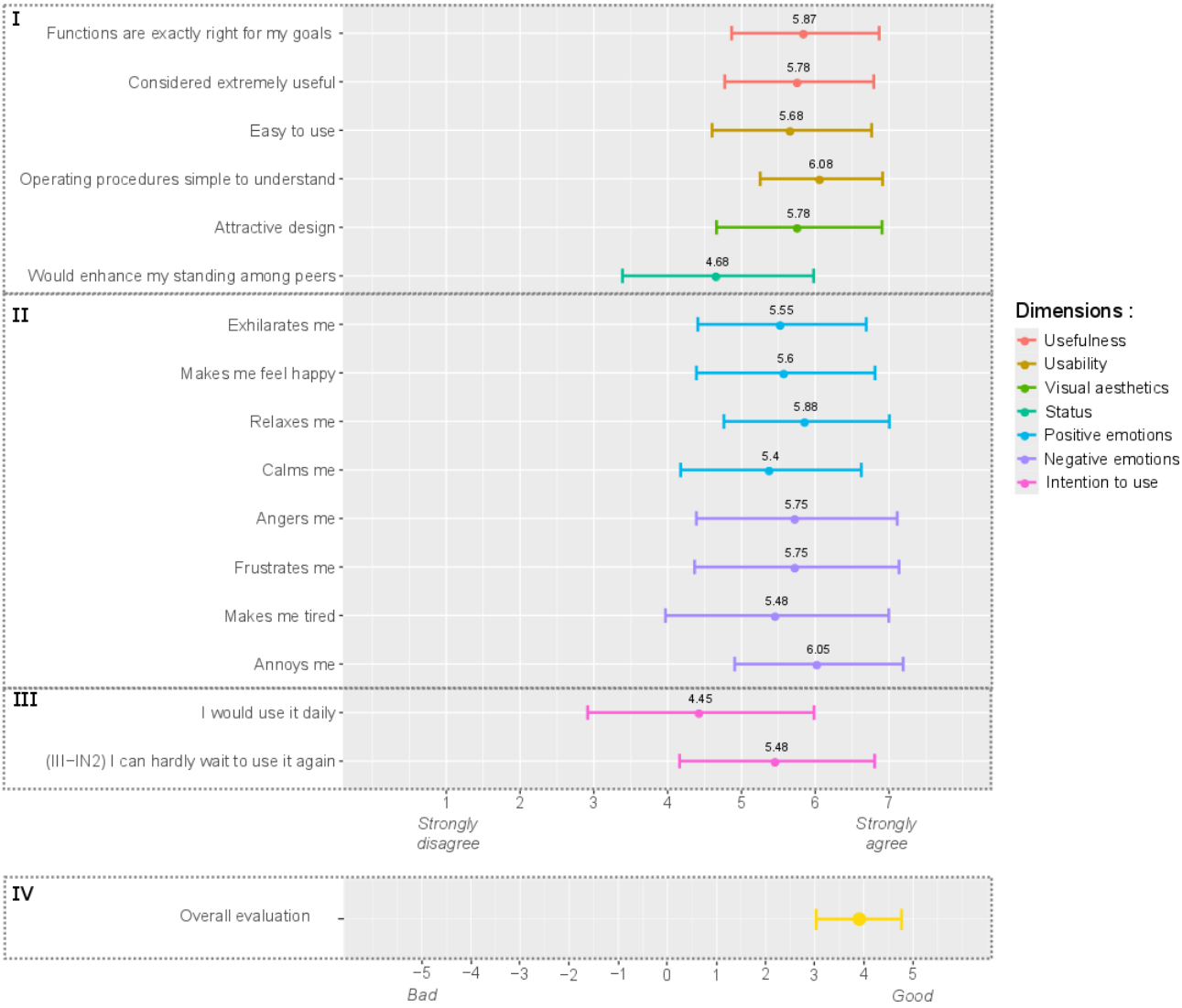
Figures 4 and 5 provide the UX results for the groups of patients with and without MNCD. Among all UX items, no statistical difference was observed between the groups of patients with and without MNCD, with the exception of one item (1/29), the meCUE module I—"The product would enhance my standing among peers" (P=.04).

Figure 2. Mean (SD) scores to the AttrakDiff items. ATT: attractiveness; HQI: hedonic quality identification; HQS: hedonic quality stimulation; PQ: pragmatic quality.



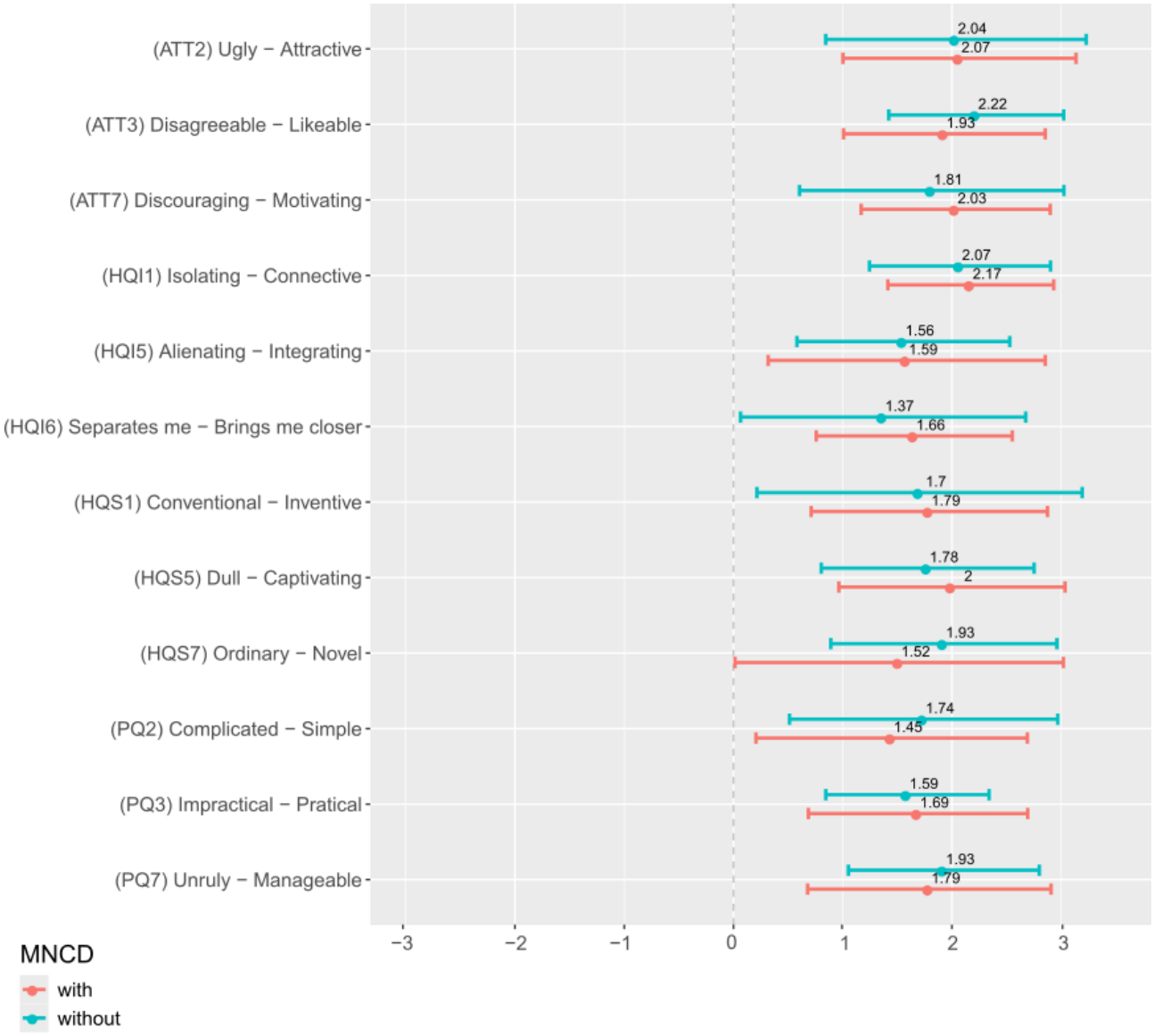
XSL•FO
RenderX

Figure 3. Mean (and SD) scores to the meCUE (modular evaluation of the components of user experience) items: (I) module I: product perceptions; (II) module II: emotions; (III) module III: consequences; (IV) module IV: global assessment.



XSL•FO
RenderX

Figure 4. Mean (SD) scores to the AttrakDiff items for the 2 groups of patients with and without major neurocognitive disorder (MNCD). ATT: attractiveness; HQI: hedonic quality identification; HQS: hedonic quality stimulation; PQ:pragmatic quality.



XSL•FO
RenderX

Figure 5. Mean (SD) scores to the meCUE (modular evaluation of the components of user experience) items for the 2 groups of patients with and without major neurocognitive disorder (MNCD). * stands for statistical difference (*P*=.04) between the groups, tested through Wilcoxon unpaired test.



Regarding the levels of engagement, 10 (17%) out of 60 participants were passive, 23 (38%) were active, 21 (35%) were constructive, and 6 (10%) were interactive. None of the participants was rated as disengaged.

Internal consistencies of the UX dimensions assessed are presented in Table 1.

Emotions and perceptions modules showed good and acceptable internal consistency, respectively. ATT, PQ, and consequences modules showed questionable to poor consistency. HQI and HQS showed unacceptable consistency.

XSL•FO
RenderX

Table . Internal consistency of the assessed user experience dimensions.

		Cronbach α (95% CI)
AttrakDiff		
	Pragmatic quality	0.579 (0.256 - 0.738)
	Hedonic quality identification	0.411 (0.097 - 0.612)
	Hedonic quality stimulation	0.422 (0.011 - 0.664)
	Attractiveness	0.608 (0.397 - 0.752)
meCUE[a]		
	Module I: Perceptions	0.701 (0.541 - 0.804)
	Module II: Emotions	0.844 (0.754 - 0.899)
	Module III: Consequences	0.655 (0.471 - 0.779)

[a]meCUE: modular evaluation of the components of user experience.

Discussion

Principal Findings

This study aimed to assess the UX of geriatric inpatients regarding a semi-immersive musical game designed to stimulate cognitive functions. The findings showed that the game was largely evaluated positively by the targeted population. The patients were mostly actively engaged in the experimentation showing great interest in the game. Patients with and without MNCD appreciated the game similarly.

Comparison to Prior Work

The instrumental (pragmatic) qualities of the game were well perceived. The patients considered the system useful and usable (meCUE, module I) and simple, practical, and manageable (AttrakDiff, PQ). These results were beyond what could have been expected assuming that the use of VR could be limited due to factors associated with old age [9]. None of the participants complained about discomfort wearing the glasses, nor criticized some aspects of the game such as fluidity, realism, vision difficulties, or cybersickness. In fact, cybersickness and presence were not quantitatively assessed like in studies testing VR in older adults [9,13,32], considering that the game was semi-immersive and the patient saw the therapist guiding the whole session.

It could have been argued that the diversity of support (music, living room visuals, objects, and events) could bring confusion to the patients, especially those with neurocognitive impairment, but this study showed the contrary. Patients acknowledged the variety (“I liked the diversity” [patient 23]). Although most participants indicated they would not play the game daily (meCUE item close to the neutral value), their overall intention to use it was largely positive. The strength of the game was to propose multiple angles of attack to ensure immersion and eventually trigger reminiscence, unlike any other traditional test and therapy [6].

Noninstrumental qualities were also appreciated with good rates on visual aesthetics especially ”meCUE, module I”. While both hedonic qualities were highly recognized, HQIs were particularly highlighted. Notably the connective aspect of the game (AttrakDiff item “HQI1: isolating-connective”) received the highest score. Based on Lallemand et al’s guidelines [16], scores higher than 2 on the AttrakDiff scale can be considered particularly interesting for the product. This was again not expected at this level. This finding demonstrated the success of the game regarding its potential to open the dialogue between the patient and the therapist. A participant commented that the game “is full of life, brings back memories, is cheerful, brings comfort, relaxation and conviviality” [patient 40]. The “status” dimension (item “the product would enhance my standing among peers”) of the meCUE was very close to the neutral value (with the SD extending beyond) and was one of the lowest scores attributed to the game. This could reflect the item being too complex or unsuitable for the population and the product in question.

As reflected by some patients’ comments—“it moved me” [patient 53] or “it pleased me and did me good” [patient 56], and the scores to the “meCUE, module II”, emotions were also highly positive. The current version of the components of the UX model by Thüring and Mahlke [33] claims that emotions, in addition to perceived product qualities, determine the overall judgment of a product, thus the consequences of use [24]. Our results showed that the overall evaluation (meCUE, module IV) and ATT were closely aligned with the emotions appreciation achieving high scores, and 100% of the participants rated the game positively (minimum rate=2 on a scale of −5 to 5).

From an interviewed psychologist’s point of view, one advantage of the game is to promote the patient’s persistent cognitive capacities without putting him or her in a difficult position. This aspect was also evidenced by patients’ comments (“it makes me realize that I still have memory capacities” [patients 28 and 33]. Game sessions were thus seen as moments of pleasure. This was supported by the results of the AttrakDiff in the ATT dimension, notably with the items “ATT3: disagreeable-likeable” and “ATT7: discouraging-motivating.”

The game was natively designed to suit every patient in geriatric wards. For the purpose of this study, only those capable of answering the UX questionnaires were included. It is worth mentioning that they were all able to play at least with 2 decades of the game, and they perceived it very positively. Their overall judgment (meCUE, module IV) was rated between 3 and 5 (on a scale of −5 to 5). The objective of the game session was not

XSL•FO
RenderX

to test the patients' memory by asking them to name songs or singers but rather to evoke memories and stimulate conversation. This was particularly true for patients with severe neurocognitive impairments (MMSE≤18) [34]. A comparison between patients with lower and higher levels of neurocognitive impairment revealed similar game perceptions. This finding aligns with a larger-scale study involving 313 patients with varying degrees of MNCD, which were enrolled in "task-based experiment"-type UX studies. The conclusions indicated that although engagement may be less active and participation rates lower, patients with severe MNCD were still able to complete the experiments as long as the sessions were properly guided [35].

Strength and Limitations

Study limitations must be acknowledged. First, nonstandardized sets of UX items were used. The AttrakDiff and meCUE were not used in their complete versions to avoid difficulties with patients presenting neurocognitive disorders. This is why an analysis of internal consistency through Cronbach α tests was carried out. This analysis showed that some dimensions were not consistent (HQI and HQS notably). This is not likely to be attributed to the insufficient number of participants, since the number of participants was higher than usually found in the literature (between 5 and 57 participants) [14]. The absence of using complete questionnaires likely limited the ability to accurately measure certain nuances of the UX and prevented an analysis by dimensions, leading instead to an item-level analysis. Additionally, the ICAPD scale was an innovative method designed to assess engagement. However, since its validity and reliability have not yet been confirmed, this could also be a potential drawback. The ICAPD provided a preliminary estimate of engagement levels, but further validation is needed. In total, the results obtained with the UX and engagement scale may not be comparable with those of other studies using validated methods.

Second, the heterogeneity of the cohort regarding their level of neurocognitive functions should be acknowledged and may have contributed to the low internal consistency of the questionnaires. Detailed results for each item are thus provided in Multimedia Appendix 2, to provide a better insight into the heterogeneity of the results. Third, this study reported verbatim comments heard during the experimental sessions but no proper qualitative study, such as semidirective interviews, was conducted. The investigator did his best to accurately transcribe the verbal comments in order to enrich the present quantitative data. Fourth, cybersickness was not assessed via structured validated scores such as the "virtual reality sickness questionnaire" [36] as recently suggested by Bruno et al [32] for all VR or AR studies. However, it was assessed through spontaneous questions to the participant, and no symptom related to cybersickness was reported.

Future Directions

Further studies are now needed to assess the impact of the game on clinical aspects. Music therapy or music-based interventions have shown some benefit in people with MNCD [37]. Small short-term benefits on cognitive functions associated with reminiscence therapy were found in a recent Cochrane meta-analysis [10]. The changes in cognitive performance, notably in memory capacity, should thus be assessed when using "A Life in Songs." It is also possible that not only cognitive function but also quality of life and communication, as well as mood, functioning in daily activities, agitation, and relationship quality, may be improved with the serious musical game [10]. This should not be limited to older populations.

Conclusions

Because feedback from users is essential in the process of developing new tools, the UX of geriatric inpatients was evaluated regarding the semi-immersive musical game that is aimed to be used for the assessment, rehabilitation, and prevention of neurocognitive disorders in older adults. The 60 patients who tested the musical game in VR reported highly positive UX. Instrumental and noninstrumental qualities as well as emotions were positively perceived, leading to excellent scores of attractiveness and global assessment. Further studies are needed to examine long-term benefits on cognitive functions, mood, and quality of life.

Acknowledgments

A very special thank is extended to Edith Gire and Johanne Daycard, who significantly facilitated the inclusion process; without their support, the experiments would have been much more difficult. There was no compensation for this contribution. The authors are grateful to the whole medical team of the medical and rehabilitation care unit of the University Hospital of Angers, France, for having greatly encouraged the project. The authors wish to thank Evan Delaunay and Marc Travers from Imagin-VR for their support and competence in transforming an idea into (virtual) reality. Spelling and syntax checking of this manuscript was performed with the assistance of ChatGPT (version GPT-4, OpenAI). This study was supported by Angers Loire Métropole and SATT Ouest Valorisation through the Maturation de Projets Innovants en Anjou grant in 2022. Angers University Hospitals also supported this work by providing the expertise of the living lab Allegro (an internal hospital structure that connects innovative companies with users in the hospital setting) for methodological support.

Data Availability

The datasets generated during this study are available from the corresponding author on reasonable request.

XSL•FO
RenderX

Authors' Contributions

Conceptualization was performed by LS and FN. Methodology was carried out by FN. Investigation was conducted by LS. Formal analysis was performed by LC. Data curation was handled by LS and LC. Visualization was conducted by LC. Writing—original draft preparation was done by LC. Writing—review and editing was done by LS, LC, FN, SN, HVL, and CA. Funding acquisition was managed by FN, LS, and CA. Resources were handled by FN, SN, HVL, and CA. Project administration was performed by LS, HVL, SN, and FN. Supervision was performed by CA.

Conflicts of Interest

The first author, LS, was the inventor of the game but has no direct financial interest related to the publication of this study. The developers of the virtual reality solution (Imagin-VR) were not involved in the data collection, analysis, or writing of this paper.

Multimedia Appendix 1
The internally developed ICAPD (interactive, constructive, active, passive, and disengaged) scale.
[DOCX File, 21 KB - games_v13i1e57030_app1.docx]

Multimedia Appendix 2
Detail of the scores obtained for each item of the AttrakDiff and the meCUE (modular evaluation of the components of user experience).
[XLSX File, 18 KB - games_v13i1e57030_app2.xlsx]

References

1. Arai H, Satake S, Kozaki K. Cognitive frailty in geriatrics. Clin Geriatr Med 2018 Nov;34(4):667-675. [doi: 10.1016/j.cger.2018.06.011] [Medline: 30336994]
2. Scott TL. Silver tsunamis, gray ceilings, and 'the problem of older adults': commentary on "Combatting ageism through virtual embodiment? Using explicit and implicit measures" by Ayalon et al. Int Psychogeriatr 2023;35(3):127-130. [doi: 10.1017/S1041610222001004]
3. Petersen RC, Smith GE, Waring SC, Ivnik RJ, Kokmen E, Tangelos EG. Aging, memory, and mild cognitive impairment. Int Psychogeriatr 1997;9(S1):65-69. [doi: 10.1017/s1041610297004717] [Medline: 9447429]
4. Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, Text Revision (DSM-5-TR): American Psychiatric Association; 2022.
5. Harada CN, Natelson Love MC, Triebel KL. Normal cognitive aging. Clin Geriatr Med 2013 Nov;29(4):737-752. [doi: 10.1016/j.cger.2013.07.002] [Medline: 24094294]
6. Abichou K, La Corte V, Piolino P. Does virtual reality have a future for the study of episodic memory in aging? Geriatr Psychol Neuropsychiatr Vieil 2017 Mar 1;15(1):65-74. [doi: 10.1684/pnv.2016.0648] [Medline: 28266343]
7. Plancher G, Tirard A, Gyselinck V, Nicolas S, Piolino P. Using virtual reality to characterize episodic memory profiles in amnestic mild cognitive impairment and Alzheimer's disease: influence of active and passive encoding. Neuropsychologia 2012 Apr;50(5):592-602. [doi: 10.1016/j.neuropsychologia.2011.12.013] [Medline: 22261400]
8. McDermott O, Charlesworth G, Hogervorst E, et al. Psychosocial interventions for people with dementia: a synthesis of systematic reviews. Aging Ment Health 2019 Apr;23(4):393-403. [doi: 10.1080/13607863.2017.1423031] [Medline: 29338323]
9. Benoit M, Guerchouche R, Petit PD, et al. Is it possible to use highly realistic virtual reality in the elderly? A feasibility study with image-based rendering. Neuropsychiatr Dis Treat 2015;11:557-563. [doi: 10.2147/NDT.S73179] [Medline: 25834437]
10. Woods B, O'Philbin L, Farrell E, Spector A, Orrell M. Reminiscence therapy for dementia. Issues Ment Health Nurs 2020 Mar 3;41(3):265-266. [doi: 10.1080/01612840.2019.1654572]
11. Yeung AWK, Tosevska A, Klager E, et al. Virtual and augmented reality applications in medicine: analysis of the scientific literature. J Med Internet Res 2021;23(2):e25499. [doi: 10.2196/25499]
12. Maranesi E, Casoni E, Baldoni R, et al. The effect of non-immersive virtual reality exergames versus traditional physiotherapy in Parkinson's disease older patients: preliminary results from a randomized-controlled trial. Int J Environ Res Public Health 2022 Nov 10;19(22):14818. [doi: 10.3390/ijerph192214818] [Medline: 36429537]
13. Lecavalier NC, Ouellet É, Boller B, Belleville S. Use of immersive virtual reality to assess episodic memory: a validation study in older adults. Neuropsychol Rehabil 2020 Apr;30(3):462-480. [doi: 10.1080/09602011.2018.1477684] [Medline: 29807474]
14. D'Cunha NM, Nguyen D, Naumovski N, et al. A mini-review of virtual reality-based interventions to promote well-being for people living with dementia and mild cognitive impairment. Gerontology 2019;65(4):430-440. [doi: 10.1159/000500040] [Medline: 31108489]

XSL•FO
RenderX

15. Bevilacqua R, Maranesi E, Riccardi GR, et al. Non-immersive virtual reality for rehabilitation of the older people: a systematic review into efficacy and effectiveness. J Clin Med 2019 Nov 5;8(11):1-14. [doi: 10.3390/jcm8111882] [Medline: 31694337]
16. Lallemand C, Gronier G. Méthodes de Design UX: Eyrolles; 2018.
17. Dhouib A, Trabelsi A, Kolski C, Neji M. A classification and comparison of usability evaluation methods for interactive adaptive systems. Presented at: 2016 9th International Conference on Human System Interactions (HSI); Jul 6-8, 2016; Portsmouth, United Kingdom. [doi: 10.1109/HSI.2016.7529639]
18. Noublanche F, Jaglin-Grimonprez C, Sacco G, et al. The development of gerontechnology for hospitalized frail elderly people: the ALLEGRO hospital-based geriatric living lab. Maturitas 2019 Jul;125(March):17-19. [doi: 10.1016/j.maturitas.2019.04.002] [Medline: 31133211]
19. Toulouse E, Lafont B, Granier S, Mcgurk G, Bazin JE. French legal approach to patient consent in clinical research. Anaesth Crit Care Pain Med 2020 Dec;39(6):883-885. [doi: 10.1016/j.accpm.2020.10.012] [Medline: 33130015]
20. Hassenzahl M, Burmester M, Koller F. AttrakDiff: ein fragebogen zur messung wahrgenommener hedonischer und pragmatischer qualität [Chapter in German]. In: Szwillus G, Ziegler J, editors. Mensch & Computer 2003: Vieweg+Teubner Verlag; 2003:187-196. [doi: 10.1007/978-3-322-80058-9_19]
21. Minge M, Riedel L. meCUE: ein modularer fragebogen zur erfassung des nutzungserlebens [meCUE: a modular questionnaire for capturing the user experience] [Chapter in German]. In: Mensch & Computer 2013 - Tagungsband: Oldenbourg Wissenschaftsverlag; 2013:89-98. [doi: 10.1524/9783486781229.89]
22. Lallemand C, Koenig V, Gronier G, Martin R. A French version of the AttrakDiff scale: translation and validation study of a user experience assessment tool. Rev Eur Psychol Appl 2015;65(5):239-252. [doi: 10.1016/j.erap.2015.08.002]
23. Lallemand C, Koenig V. How could an intranet be like a friend to me? Why standardized UX scales don't always fit. Presented at: ECCE 2017: European Conference on Cognitive Ergonomics 2017; Sep 19-22, 2017; Umeå, Sweden. [doi: 10.1145/3121283.3121288]
24. Minge M, Thüring M, Wagner I. Developing and validating an English version of the meCUE questionnaire for measuring user experience. Proc Hum Factors Ergon Soc Annu Meet 2016 Sep;60(1):2063-2067. [doi: 10.1177/1541931213601468]
25. Chi MTH, Wylie R. The ICAP framework: linking cognitive engagement to active learning outcomes. Educ Psychol 2014 Oct 2;49(4):219-243. [doi: 10.1080/00461520.2014.965823]
26. Chi MTH, Adams J, Bogusch EB, et al. Translating the ICAP theory of cognitive engagement into practice. Cogn Sci 2018 Jun 28;42(6):1777-1832. [doi: 10.1111/cogs.12626] [Medline: 29954048]
27. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975 Nov;12(3):189-198. [doi: 10.1016/0022-3956(75)90026-6] [Medline: 1202204]
28. Chun CT, Seward K, Patterson A, Melton A, MacDonald-Wicks L. Evaluation of available cognitive tools used to measure mild cognitive decline: a scoping review. Nutrients 2021 Nov 8;13(11):1-27. [doi: 10.3390/nu13113974] [Medline: 34836228]
29. R: A Language and Environment for Statistical Computing: R Foundation for Statistical Computing; 2021.
30. Nunnally J. Psychometric Theory: McGraw-Hill; 1994.
31. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126. [doi: 10.2196/jmir.1923] [Medline: 22209829]
32. Bruno RR, Wolff G, Wernly B, et al. Virtual and augmented reality in critical care medicine: the patient's, clinician's, and researcher's perspective. Crit Care 2022 Oct 25;26(1):1-13. [doi: 10.1186/s13054-022-04202-x] [Medline: 36284350]
33. Thüring M, Mahlke S. Usability, aesthetics and emotions in human–technology interaction. Int J Psychol 2007 Aug;42(4):253-264. [doi: 10.1080/00207590701396674]
34. Mungas D. In-office mental status testing: a practical guide. Geriatrics 1991 Jul;46(7):54-58. [Medline: 2060803]
35. Noublanche F, Ben-Sadoun G, Carcreff L, Gautier J, Allain P, Annweiler C. Interest of a hospital-based geriatric living lab among inpatients with neurocognitive disorders: the ALLEGRIA cross-sectional study. Dement Geriatr Cogn Disord 2024;53(4):1-10. [doi: 10.1159/000538144] [Medline: 38679008]
36. Kim HK, Park J, Choi Y, Choe M. Virtual reality sickness questionnaire (VRSQ): motion sickness measurement index in a virtual reality environment. Appl Ergon 2018 May;69:66-73. [doi: 10.1016/j.apergo.2017.12.016] [Medline: 29477332]
37. Vink A, Hanser S. Music-based therapeutic interventions for people with dementia: a mini-review. Medicines (Basel) 2018 Oct 8;5(4):109. [doi: 10.3390/medicines5040109] [Medline: 30297605]

Abbreviations

AR: augmented reality
ATT: attractiveness
CONSORT-eHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
HQI: hedonic quality identification
HQS: hedonic quality stimulation

XSL•FO
RenderX

ICAP: interactive, constructive, active, and passive
ICAPD: interactive, constructive, active, passive, and disengaged
meCUE: modular evaluation of the components of user experience
MMSE: Mini-Mental State Examination
MNCD: major neurocognitive disorder
PQ: pragmatic quality
UX: user experience
VR: virtual reality

Edited by T Leung; submitted 02.02.24; peer-reviewed by E Maranesi, Y Asada; revised version received 03.10.24; accepted 31.10.24; published 06.01.25.

Please cite as:
Samson L, Carcreff L, Noublanche F, Noublanche S, Vermersch-Leiber H, Annweiler C
User Experience of a Semi-Immersive Musical Serious Game to Stimulate Cognitive Functions in Hospitalized Older Patients: Questionnaire Study
JMIR Serious Games 2025;13:e57030
URL: https://games.jmir.org/2025/1/e57030
doi:10.2196/57030

© Laurent Samson, Lena Carcreff, Frédéric Noublanche, Sophie Noublanche, Hélène Vermersch-Leiber, Cédric Annweiler. Originally published in JMIR Serious Games (https://games.jmir.org), 6.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Effectiveness of a Virtual Reality Serious Video Game (The Secret Trail of Moon) for Emotional Regulation in Children With Attention-Deficit/Hyperactivity Disorder: Randomized Clinical Trial

Marina Martin-Moratinos[1,2], MSc; Marcos Bella-Fernández[1,3,4,5], PhD; María Rodrigo-Yanguas[2], PhD; Carlos González-Tardón[6], PhD; Chao Li[1,2], MD; Ping Wang[1,2,7], MD, PhD; Ana Royuela[8,9], PhD; Pilar Lopez-Garcia[2,10], MD, PhD; Hilario Blasco-Fontecilla[10,11,12], MD, PhD

[1]Department of Psychiatry, Health Research Institute of the Puerta de Hierro Majadahonda-Segovia de Arana University Hospital (IDIPHISA), Puerta de Hierro Majadahonda University Hospital, Madrid, Spain

[2]Faculty of Medicine, Autonoma University of Madrid, Madrid, Spain

[3]Faculty of Psychology, Autonoma University of Madrid, Madrid, Spain

[4]Department of Psychology, Pontifical University of Comillas, Madrid, Spain

[5]Faculty of Health, International Business University, Madrid, Spain

[6]Department of Video Games, Design and Technology University, Madrid, Spain

[7]Sichuan Provincial Center for Mental Health, University of Electronic Science and Technology of China, Chengdu, China

[8]Clinical Biostatistics Unit, Puerta de Hierro Majadahonda University Hospital, Madrid, Spain

[9]Consortium for Biomedical Research Network in Epidemiology and Public Health, Madrid, Spain

[10]Spain Biomedical Research Networking Center for Mental Health Network, Madrid, Spain

[11]Health Sciences and School of Doctoral Programs, International University of La Rioja, Logroño, Spain

[12]Emooti, Institute for Mental Health and Wellness, Madrid, Spain

Corresponding Author:
Hilario Blasco-Fontecilla, MD, PhD
Health Sciences and School of Doctoral Programs
International University of La Rioja
Avda de la Paz, 137
Logroño, 26006
Spain
Phone: 34 910149575
Email: hilariomanuel.blasco@unir.net

Abstract

Background: Difficulties in emotional regulation are often observed in children and adolescents with attention-deficit/hyperactivity disorder (ADHD). Innovative complementary treatments, such as video games and virtual reality, have become increasingly appealing to patients. *The Secret Trail of Moon* (MOON) is a serious video game developed by a multidisciplinary team featuring cognitive training exercises. In this second randomized clinical trial, we evaluated the impact of a 20-session treatment with MOON on emotional regulation, as measured by the Strengths and Difficulties Questionnaire.

Objective: We hypothesize that patients with ADHD using MOON will show improvements in (1) emotional regulation, (2) core ADHD symptoms, (3) cognitive functioning, and (4) academic performance, compared to a control group; additionally, we anticipate that (5) changing the platform (from face-to-face using virtual reality to the web) will not affect emotional regulation scores; and (6) the video game will not cause any clinically significant side effects.

Methods: This was a prospective, unicentric, randomized, unblinded, pre- and postintervention study with block-randomized sequence masking. Participants included individuals aged between 7 and 18 years who had a clinical diagnosis of ADHD and were receiving pharmacological treatment. They were randomized into 2 groups using an electronic case report form: the MOON group, receiving standard pharmacological treatment plus personalized cognitive training via a serious video game, and the control group, receiving standard pharmacological treatment. We provided both the groups with psychoeducational support on ADHD. Analysis was conducted using the Student 2-tailed *t* test and 2-factor ANOVA. An independent monitor supervised the study.

XSL•FO
RenderX

Results: A total of 76 patients with ADHD participated in the trial, with an equal randomization (MOON: n=38, 50% and control: n=38, 50%) and a total dropout rate of 7. The primary hypothesis, a 3- or 4-point reduction in the global Strengths and Difficulties Questionnaire score, was not met. However, significant improvements were observed in material organization (*P*=.03), working memory (*P*=.04), and inhibition (*P*=.05), particularly among patients more engaged with the MOON treatment.

Conclusions: Serious video games, when integrated into a multimodal treatment plan, can enhance outcomes for symptoms associated with ADHD.

Trial Registration: ClinicalTrials.gov NCT06006871; https://clinicaltrials.gov/study/NCT06006871

International Registered Report Identifier (IRRID): RR2-10.2196/53191

(JMIR Serious Games 2025;13:e59124) doi:10.2196/59124

KEYWORDS

attention-deficit/hyperactivity disorder; ADHD; emotional regulation; serious video games; virtual reality; cognitive training; music; chess

Introduction

Attention-deficit/hyperactivity disorder (ADHD) is characterized by a persistent pattern of inattention, hyperactivity, and impulsivity, affecting approximately 5% of children and adolescents worldwide [1,2]. These symptoms significantly impair daily functioning across social, academic, cognitive, behavioral, and family domains. In addition to the core symptoms, emotional problems and poor self-regulation are increasingly seen as central features of ADHD [3-5]. Emotional regulation allows modifying the emotional state by promoting adaptive behaviors to achieve a goal [6]. Children and adolescents with ADHD often struggle with emotional regulation, exhibiting low frustration tolerance, irritability, and emotional lability. A recent meta-analysis highlighted difficulties associated with ADHD in emotional lability (*d*=0.95), emotional regulation (*d*=0.80), empathy (*d*=0.68), and emotion recognition or understanding (*d*=0.64), with differences associated with gender, age, and the existence of executive dysfunction or behavioral problems [7]. Internalizing and externalizing problems are common among individuals with ADHD [8]. In fact, comorbidity is highly prevalent in ADHD with internalizing disorders such as anxiety or depression as well as externalizing disorders such as conduct disorder and oppositional defiant disorder [9]. The camouflaging of symptoms [10] often leads to underdiagnosis and treatment delays, resulting in misdiagnosis, higher rates of accidents, use of public services, increased likelihood of substance use, delinquency, and even suicide attempts [11-13].

Multimodal treatment for ADHD includes pharmacological treatment, psychological treatment, and psychoeducation for parents and teachers. While multimodal treatment is the most effective for ADHD, it can sometimes be insufficient, requiring complementary treatments [14]. Pharmacological treatment is the treatment of choice in children and adolescents with severe ADHD [15]. Stimulant treatment reduced behaviors linked to emotional instability, such as substance use, suicide rate, or criminal behavior [5,13]. Behavioral therapy, particularly focused on problem-solving and social skills training, is the recommended form of psychotherapy for ADHD [16,17]. Nevertheless, some areas, such as executive function, emotional regulation, and core ADHD symptoms, may show limited improvement through multimodal treatment alone. Motivation, especially intrinsic motivation, and responsiveness to insufficient reward plays a critical role in ADHD [18]. Individuals with ADHD often seek higher levels of immediate reward, which can lead to neglect tasks, dropped jobs, or incomplete treatments. Complementary treatments have emerged that are more attractive to patients, such as the use of games or video games or new technologies such as virtual reality (VR) [19,20]. Children with ADHD who engaged in working memory training incorporating game elements demonstrated longer training durations, completed more sequences with fewer errors, and achieved higher scores compared to those who underwent standard cognitive training [21]. When motivated, children with ADHD perform comparably to typically developing children on most cognitive tasks and may even outperform them on engaging computer-based tasks, despite the need for sustained attention [18].

In this prospective, unicentric, unblinded, pre- and postintervention randomized clinical trial (NCT06006871), we investigated the efficacy of *The Secret Trail of Moon* (MOON) for improving emotional regulation in people with ADHD aged between 7 and 18 years. MOON is a serious video game designed for health purposes, developed by a multidisciplinary team; the environment is composed of a forest with the intention of immersing the player through the sensation of presence that VR facilitates [22,23]. VR favors a lower perception of the real physical environment [22] and, therefore, a lower distraction by stimulation, a key aspect for a treatment for individuals with ADHD [15,24]. The video game experience can be immersive and exciting. A very low level of arousal can lead to boredom and reduced attention to the game. Conversely, if the player experiences too much arousal, it can also divert attention [22]. Music can help modulate these emotional states besides facilitating flow and immersion while playing video games [25]. Particularly in ADHD, music can be beneficial in reducing environmental distraction, especially in monotonous tasks [25,26]. Elements of music therapy have been both passively and interactively incorporated into this new version of the game [25,27]. Finally, an additional rhythm-based game mechanic (active music therapy) could not be added in the clinical trial due to its nonapproval by the ethics committee. The video game integrates bright but calm colors in a cartoon style, friendly characters that accompany the patient in the process, and varied cognitive training games based on neuropsychological tests: (1)

XSL•FO
RenderX

Smasher (sustained attention and inhibitory control); (2) Teka Teki (planning); and (3) Kuburi (visuospatial capacity), Chess (reasoning), and Enigma (working memory). The games increase in difficulty level according to patient performance [28]. A previous clinical trial (NCT04355065) showed promising results in emotional intelligence, emotional regulation, and performance in the school context in both self-reports and parent reports [29]. Contrary to our expectations, no significant differences were observed in the improvement of executive functions, as measured by the Behavior Rating Inventory Executive Function, version 2 (BRIEF-2), between the video game group and the control group. Compared to the previous clinical trial, this time the study design is simplified into 2 branches: experimental (play MOON in the hospital and at home during 20 sessions) versus control (do not play MOON and telephone follow-up to parents). Both groups have stable pharmacological treatment and psychoeducation to parents. We anticipated a 3- to 4-point reduction in the global Strengths and Difficulties Questionnaire (SDQ) score between the baseline assessment (day 0) and the final assessment (day 90), reflecting an improvement in emotional regulation. As secondary measures, we propose 6 additional hypotheses:

- Hypothesis 1—ADHD patients using the MOON video game improve emotional regulation compared to the control group
- Hypothesis 2—patients with ADHD using the MOON video game improve core ADHD symptoms compared with the control group
- Hypothesis 3—patients with ADHD using MOON improve their cognitive functioning compared to the control group
- Hypothesis 4–patients with ADHD using MOON improve in academic performance with respect to the control group
- Hypothesis 5—a change in platform (hospital-based, face-to-face VR vs home-based, web-based computer version) will not result in differences in emotional regulation outcomes
- Hypothesis 6—there is no clinically meaningful side effects associated with the video game

Methods

Study Design

Prospective, unicentric, randomized, unblinded, pre- and postintervention study with a masked randomization sequence by blocks (NCT06006871) [30]. Participants were randomized by electronic case report form (eCRF) in 2 groups: group 1 (MOON), which received standard pharmacological treatment combined with personalized cognitive training through a serious video game designed for patients with ADHD, along with psychoeducational support for parents; and group 2 (control), which received standard pharmacological treatment and psychoeducational support for parents, without the video game intervention. The study followed a parallel assignment model (MOON vs control) with a 1:1 allocation ratio.

No major changes were made to the methodology before the commencement of the clinical trial. In this randomized clinical trial, the study design and reporting adhered to the CONSORT 2010 guidelines, specifically following the CONSORT checklist for reporting information in randomized trials (Multimedia Appendix 1). The research questions were formulated using the Population, Intervention, Comparison, Outcomes and Study framework. The study protocol is published [30] and registered on ClinicalTrials.gov (NCT06006871) with International Registered Report Identifier (DERR1-10.2196/53191).

Ethical Considerations

This study received approval from the Research Ethics Committee of the Puerta de Hierro University Hospital on December 14, 2022 (PI 106/22). Authorization was subsequently granted by the Spanish Agency of Medicines and Health Products on February 14, 2023 (1061/22/EC-R). The study was monitored by an independent monitor. All participants signed the informed consent form. Data were anonymized through the assignment of a clinical trial-specific code. The data were treated confidentially in accordance with the Organic Law 3/2018 of December 5 on the protection of personal data and guarantee of digital rights. This study complied with the standards of good clinical practice and the declaration of Helsinki. Participants were not compensated financially.

Study Population

Children and adolescents aged between 7 and 18 years (mean age 12.68, SD 2.75 years) with a clinical diagnosis of ADHD according to the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition*, were enrolled in the study. All participants had a clinical diagnosis of ADHD in any presentation and were receiving stable pharmacological treatment for ADHD. Patients were clinically stable, with a Clinical Global Impression (CGI) score between 3 and 6 before entering the clinical trial. Comorbidity was not an exclusion criterion in the study, except for those patients with suicide risk, lack of ability to follow verbal instructions, or motor difficulties that make it difficult to play a video game. Exclusion criteria also included participation in similar video games studies or the intention to initiate psychotherapeutic treatment during the 3-month study period. Medication was not altered during the study, unless adjustments were necessary for clinical reasons. A dropout rate of 15% was expected.

The randomized clinical trial had a total duration of 7 months, concluding on December 15, 2023. Recruitment took place between May and October 2023. The first evaluation was performed on May 9, 2023. The last evaluation was performed on October 31, 2023.

The intention-to-treat (ITT) population was defined as randomized participants who completed at least 10 sessions. The per-protocol (PP) population was defined as participants who completed 100% of the training. Only 32% (12/38) participants in the MOON group completed all 20 treatment sessions (PP), so we have not performed the analyses with this small sample size. In the ADHD population, adherence difficulties are common in long-term treatments; therefore, we planned to conduct the analyses on the premise of the ITT analysis. An additional analysis was also included to evaluate those participants who engaged with the intervention by completing 80% treatment (16 sessions in the MOON group or engagement for 8 weeks in the control group). Recognizing that

XSL•FO
RenderX

the number of sessions could be a relevant covariate, we performed an analysis of covariance (ANCOVA), which was not initially included in the protocol. Only half of the parents (38/76, 50%; Multimedia Appendix 2) provided information on academic grades. Therefore this primary outcome (hypothesis 4) was excluded due to insufficient data.

Procedure

All participants were recruited from the child and adolescent psychiatry outpatient clinics of Hospital Universitario Puerta de Hierro Majadahonda. The MOON group had 12 face-to-face visits, whereas the control group had only 2 presential visits: one at the preevaluation visit (D0) and another at the postevaluation visit (D90). The total duration of the research was 3 months for each participant. During the preinclusion phase, the principal investigator (HB-F) informed eligible patients about the procedure and research protocol. A total of 155 participants were assessed for enrollment. The study coordinator (MM-M) provided a detailed explanation of the research procedure and summoned the prerecruited participants. In total, 7.7% (12/155) of the patients did not meet the inclusion criteria (main reasons were clinical CGI, no medication, or age); 43.2% (67/155) of the patients declined to participate (the most common reason being lack of availability).

During the inclusion visit (D0), the research procedure was thoroughly explained, the informed consent form was signed, and the preevaluation was conducted (refer to the Study Outcomes section). Afterward, participants were randomly assigned to groups. Group randomization (MOON vs control) was performed using block randomization through the electronic eCRF REDCap (Research Electronic Data Capture; Vanderbilt University) with a 1:1 allocation ratio.

The intervention training (D1-D90) varied according to the assigned group. Parents in the control group received weekly telephone monitoring. For the MOON group, 20 sessions with the video game were scheduled: 10 sessions were conducted in the hospital with the researchers, and 10 sessions took place on the web at home, with participants using their computers under the researchers' supervision (twice a week, adjusted according to the availability of the participants). Patients aged >12 years played MOON with VR, while those aged <12 years performed the video game sessions on a computer, following PlayStation guidelines. MM-M and MB-F assisted participants during their initial use of the video game, providing instructions, encouraging progress, and helping them manage any frustration caused by mistakes made while playing MOON. They also monitored for signs of motion sickness with VR, ensured proper use of the software, and addressed any potential frustration resulting from bugs in the video game. If a patient reported dizziness, even mild, they were instructed to remove the device and were given the option to decide whether or not to continue using it. All participants followed the same order pattern of the MOON games during 20 minutes of gameplay (refer to the study by Martin-Moratinos et al [30] for details).

In week 5 (D45), all participants' parents receive a follow-up questionnaire with the main variable (SDQ). For the MOON group, both parents and children were trained on how to use the video game at home, using a flash drive. Parents and patients were additionally given an instruction manual. Monitoring was provided by telephone where the researchers checked the correct playability at home. The web-based sessions conducted were monitored by MM-M through the PlayFab data server (Microsoft Corporation). All participants had their personal username and password connected with PlayFab data server.

In the final visit (D90), a postevaluation was conducted using the same questionnaires, and participants were provided with feedback on the most relevant changes observed.

Materials

The MOON video game was designed with the Unity (version 2020.3.7; Unity Technologies) software tool by a multidisciplinary team. Autodesk 3ds Max, Autodesk Maya, Adobe Photoshop, and PreSonus Studio 4 programs were used for the modeling of characters and landscape elements. In relation to the narrative and identity features [31,32], the player is immersed in a forest with archaic chess statues of an ancient civilization [29,31]. The raccoon Movi and the fox One will help the player to solve the puzzles (game mechanics). MOON is mainly composed of a menu screen and 5 different game mechanics focused on cognitive training.

To provide manipulation and control features [32], each player has their own user profile with saved progress. Each test is set in a closed environment with distinct esthetics and game mechanics. The manipulation of the environment and interactivity are carefully controlled to avoid distractions or dizziness in the VR setting [28]. For example, in Kuburi, blocks can be rotated and moved within the space, while in Chess, the player can shift positions to view the board from different perspectives. The mobility within environments is restricted based on the individual's cognitive capacity. In cases where mobility could cause distraction—such as in tests of sustained attention or working memory, which are more challenging for individuals with ADHD—exploration is limited.

In all game mechanics, players begin at a base level (tutorial, level 1) to ensure they understand the task. The difficulty increases in a personalized manner based on performance, ensuring the game is neither too easy (and thus boring) nor too difficult (and thus frustrating). Emotional regulation maturity, specifically the ability to reappraise, can be trained through controlled exposure to negative emotional stimuli, helping players overcome levels and obstacles within the game [31]. Regarding reward and punishment features [32], the MOON video game uses a star system based on player performance (0 stars=poor performance, must repeat the level; 1 star=acceptable; 2 stars=good; and 3 stars=excellent). Player progress in each game mechanic is communicated through level advancements, performance parameters, and the number of stars earned [30].

In terms of music, each game mechanic features different tracks [25]. The music was specifically designed to align with cognitive abilities. For example, sustained attention, which requires more effort for individuals with ADHD [24], benefits from dynamic soundtracks. To prevent monotony and boredom, more epic, fast-paced music was composed for Smasher to maintain motivation through its rhythm. Feedback is not only visual but also auditory, with distinct sounds depending on user

XSL•FO
RenderX

interactions. Different sounds were created to indicate level selection and hits or misses, as well as to signal conditions of victory (successfully completing a level) or defeat (repeating the level after earning 0 stars due to poor performance) [25]. In this clinical trial, the same version of the video game was proposed in 2 platforms: VR and computer (Table 1) [30].

Table 1. Hardware and software specifications categorized by The Secret Trail of Moon (MOON) gameplay on virtual reality (VR) or computer.

MOON	Patients' target	Hardware	Software
VR	• For children aged >12 years (following PlayStation guidelines) • Only the first 10 sessions with on-site supervision at the hospital by researchers	• PlayStation 4 Dev Kit and Test Kit • PlayStation VR headset • PlayStation camera • Computer screen • Controller: DualShock 4 • Headphones with 3D stereo sound and noise cancelation	• Gameplay: the player's camera is the VR viewer
Computer	• For patients aged between 7 and 11 years • All participants played MOON on their home computers after session 10	• Any computer with operating system 8/10/11, DirectX 11 support (32 or 64 bits), 1.9-GH processor (Intel core i5 or AMD equivalent), 8 GB RAM, NVIDIA GTX 660 or AMD Radeon HD 7950 with at least 3-GB storage space, and internet connection • Controller: mouse and keyboard	• Gameplay: the player's camera depends on the mouse pointer

A PlayStation 4 Dev Kit and Test Kit were used for VR version. The Dev Kit and Test Kit are development consoles equipped with debug firmware that allow the execution of unsigned builds; access to memory and performance analysis tools; real-time error trace logging; as well as software compatibility, optimization, and development testing in a controlled environment. The PlayStation VR for PlayStation 4 is a VR headset featuring a 5.7-inch OLED display with a resolution of 1920×1080 (960×1080 per eye), a field of view of approximately 100°, and a refresh rate of 90/120 Hz. A PlayStation camera was used for motion tracking through an accelerometer, gyroscope, and 9 LEDs. The participants had to be placed at a distance of 1.5 to 2 m with a not very intense light for the correct calibration of the PlayStation camera with the LED lights. The researchers monitored what each participant experienced in VR by viewing it on a computer screen.

The PlayStation 4 VR goggles weigh approximately 600 g, so they could be uncomfortable for some participants. Regarding the added weight of sound equipment, patients were given a choice between basic stereo headphones and over-the-ear headphones. Both had 3D stereo sound and noise cancelation.

A Stealth 15M A11SDK (MSI) laptop was used for participants aged between 7 and 11 years to play MOON in the hospital. The laptop specifications were as follows: Microsoft Windows 11 Home ×64 bits; 11th Gen Intel Core i7-1185G7, 3.00 GHz, 16 GB RAM, 8 GB storage memory; GeForce GTX 1660 Ti.

Study Outcomes

Primary Outcome

The SDQ is a brief 25-item questionnaire that examines difficulties related to emotions, behavior, and social relationships. SDQ includes 5 subscales: emotional symptoms (somatic, worries, unhappy, clingy, and fears), behavioral problems (tempers, not obedient, fights, lies, and steals), hyperactivity (restless, fidgety, distractible, not reflective, and not persistent), peer problems (solitary, not good friend, not popular, bullied, and best with adults), and prosocial behavior (considerate, shares, caring, kind to kids, and helps) [33,34]. The authors propose an alternative three-factor model: (1) prosocial behavior; (2) internalizing difficulties (composed of the subscales of "emotional symptoms" and "peer problems"; cutoff point=7) as an indicator of the presence of excessive worry, somatization with headache, depressive mood, nervousness in unfamiliar situations, and tendency to startle easily; and (3) externalizing difficulties (composed of the subscales "hyperactivity" and "behavior problems"; cutoff point=4), to assess the tendency to lose control, fighting with other children, deceiving, and stealing. The questionnaire has shown adequate psychometric properties in the grouping of these 3 factors and good sensitivity and specificity for detecting children with ADHD [35,36].

The face-to-face web-based change will be measured by comparing the 3 SDQ measures. Subsequently, 3 parent SDQ measures were assessed: initial assessment (D0), midterm assessment (D45), and final assessment (D90) to evaluate hypothesis 5 (switching from face-to-face to web-based methods do not lead to differences in emotional regulation).

Secondary Outcomes

The core ADHD symptomatology was measured using 3 subjective scales for parents: the Swanson, Nolan, and Pelham Rating Scale (SNAP-IV); the Conners Abbreviated Symptom Questionnaire (CPRS); and CGI. SNAP-IV [37] is an 18-item questionnaire that assess attention deficit (cutoff point is 1.78) and hyperactive impulsive (cutoff point is 1.44) with a Likert scale ranging from 0 to 4. The CPRS [38] also assesses ADHD symptoms with cutoff points of 16 for boys and 12 for girls. It is a 10-item questionnaire using a Likert scale of 0 to 3 (0=not true at all or never, 1=just a little true or occasionally, 2=pretty much true or often, and 3=very much true or very often). The CGI-40 [39], adapted for parents, allows them to rate their child's general condition using a Likert scale ranging from 1 to 10.

XSL•FO
RenderX

Executive function refers to a set of top-down processes that allow us to be flexible and to direct behavior toward a goal. Executive function was measured with a parent questionnaire and 3 objective tests for the participants. BRIEF-2 [40] is a 63-item questionnaire with 3 answer options (never, sometimes, and frequently) that provides 4 indices: emotional regulation, cognitive regulation, behavioral regulation, and global index of executive function. For objective testing of patients, we assessed them using the Conners' Continuous Performance Test Third Edition (CPT-3); Corsi cubes; and the Comprehensive Trail-Making Test, Second Edition (CTMT-2). CPT-3 [41,42] is a computerized, standardized, and validated application test for screening ADHD in clinical practice. CPT-3 provides results on hits, errors of omission, commission error, hit mean reaction time, and variability. Corsi cubes are used to measure visuospatial working memory. CTMT-2 is used to measure cognitive flexibility, with 3 indexes: inhibitory control, task switching, and total index.

Additional school information was collected due to its potential impact on emotional regulation and executive functions. Information about academic performance will be obtained via the patients' grades.

Safety Outcomes

During the development of the video game, risk control measures were applied, especially those related to the possible adverse consequences of VR. We followed recommendations from Sony Interactive Entertainment (Project Morpheus PlayStation Virtual_000010933011) regarding best practices for VR interfaces, including interaction, navigation, user interface, game environments, player comfort, and safety. In addition, risk prevention measures, particularly those related to VR, were implemented (Textbox 1). Before conducting the clinical trial, programming adjustments were made. The game was tested for correct playability using a quality assurance tester. Development was an iterative process using a user-centered model in which game bugs were corrected based on the usability study and the previously conducted clinical trial [28,29].

Textbox 1. Risk prevention and control measures in The Secret Trail of Moon clinical trial.

Programming

- Control latency and frame rate, maintaining a minimum of 90 frames per second. Ensure synchronization between player movements and virtual reality (VR) visuals
- Minimize motion blur
- Maintain a field of view between 90 and 110°
- Avoid artificial or conflicting movements that may cause motion sickness by incorporating static or low-motion scenes where possible
- Reduce loading times and ensure smooth transitions to prevent player discomfort
- Limit intense visual effects that may cause fatigue and provide regular breaks
- Manage proximity and positioning of the user interface
- Control the height of the viewfinder or horizon line

Optimal playability and adjustment

- Maintain a player distance of 1.5 m from the PlayStation Camera, with the camera positioned slightly above the eye level to capture head and body movements
- Ensure moderate lighting for better motion detection
- Address visual accommodation: calibrate eye settings and clean lenses regularly
- Assess user comfort, including helmet weight, sound volume, and potential issues such as blurred vision

Predictable misuse of software and hardware in children and adolescents with attention-deficit/hyperactivity disorder

- Reduce sudden camera movements with aid of the researcher
- Ensure a clear play space of approximately 2 m around the player, taking special care to manage cables
- Monitor for excessive hyperkinesia or hypokinesia, assessing the potential stress of using new technologies
- Remove VR gear gradually to prevent dizziness
- Evaluate the risk of video game addiction using the Game Addiction Scale for Adolescents and Udvalg für Kliniske Undersogelser (UKU) tests
- Assess fatigue and motivation levels related to gameplay using the UKU test and usability measures
- Monitor drowsiness and arousal levels with the UKU and Sleep Disturbance Scale for Children tests
- Cease use immediately if dizziness occurs. If dizziness arises in ≥2 sessions, consider restricting gameplay to the computer only

Safety outcomes focused on measuring the risk of video game addiction and associated sleep difficulties. Motion sickness may occur with VR, especially in the early adaptation sessions. The group that played MOON with VR was additionally given a

XSL•FO
RenderX

measure to control for this possible effect. According to the protocol, the sponsor must notify the Spanish Agency of Medicines and Health Products within 15 calendar days of receiving the Serious Adverse Event report, in the case of any serious adverse events.

The Game Addiction Scale for Adolescents (GASA) is a 7-item questionnaire to assess video game addiction [43,44]. The version of GASA used was the 7-item Spanish version, which measures salience, tolerance, mood modification, relapse, withdrawal, conflict, and problems. A higher score indicates more problematic gaming behavior. Video game addiction is calculated by obtaining a score of ≥3 on at least 4 items or a total score with a cutoff point of 21.

The Sleep Disturbance Scale for Children is a 26-item questionnaire whose subscales measure disorders of initiating and maintaining sleep, sleep breathing disorders, disorders of arousal nightmares, sleep-wake transition disorders, disorders of excessive somnolence, and sleep hyperhydrosis [45]. A higher score indicates more sleep disturbance. Sleep problems are calculated by obtaining a total score with a cutoff point of 39.

Only for MOON group participants aged >12 years and mainly due to the possibility of motion sickness with VR, the questionnaire Udvalg für Kliniske Undersogelser (UKU) was applied after each session. UKU evaluate psychic, neurological, autonomic, and other side effects [46]. The most relevant side effects evaluated with UKU were those associated with ADHD symptoms (concentration difficulties, failing memory, depression, tension, emotional difference, and hyperkinesia), those associated with side effects after using VR (accommodation disturbances, nausea, dizziness, and headache), those associated with the use of video games (physical and psychic dependence), and those associated with sleep (fatigue, sleepiness or sedation, reduced or increased duration of sleep, and increased dream activity).

Statistical Analyses

Sample Size Calculation

To calculate the sample size, we referenced the means of the main measure (SDQ) in similar studies. To account for most possibilities with a significance level of P=.05 and a statistical power of 80%, while considering a dropout rate of 15%, the author AR calculated that a total of 152 participants with an ADHD diagnosis were needed. We performed the sample size calculation to contrast a mean decrease of 3 points (symptomatologic improvement) in the experimental group with a SD of 4.

Randomization and Masking: Sequence Generation and Allocation

A randomized block sequence of 76 times “number 1 block” (experimental) and 76 times “number 2 block” (control) was generated in 4 blocks of 38 numbers (19 times for number 1 and 19 times for number 2) using the program R (R Foundation for Statistical Computing) by MB-F. The sequence was unknown to the recruiters and was enclosed in the eCRF. The randomization was done through REDCap (eCRF) after the initial evaluation was completed.

Statistical Methods

A 2-way ANOVA were performed, with time (pre and post; pre-mid-term-post in the SDQ questionnaire) and group (experimental vs control) as factors and each measure as dependent variable. The difference between the groups was considered significant when P<.05. In addition, clinical trends of improvement have been described due to noncompliance with the proposed sample size. These analyses were performed with SPSS (IBM Corp) and R software.

Analyses were performed under the ITT assumption following the protocol; analyses were not performed on PP (only 12 participants completed 20 sessions); an additional analysis has been added regarding the MOON subgroup that completed 80% of the treatment (16 sessions) and, therefore, had a higher involvement in the treatment. As this variable was relevant to the results, an ANCOVA was performed using the number of sessions as a covariate, which had a mean of 18.22 (SD 3.12; range 0-20) sessions. The adjusted means of the groups were calculated based on this mean value.

The ANOVA for hypothesis 4 could not be conducted due to insufficient data. The mean and SD for each academic grade are provided in Multimedia Appendix 2.

Results

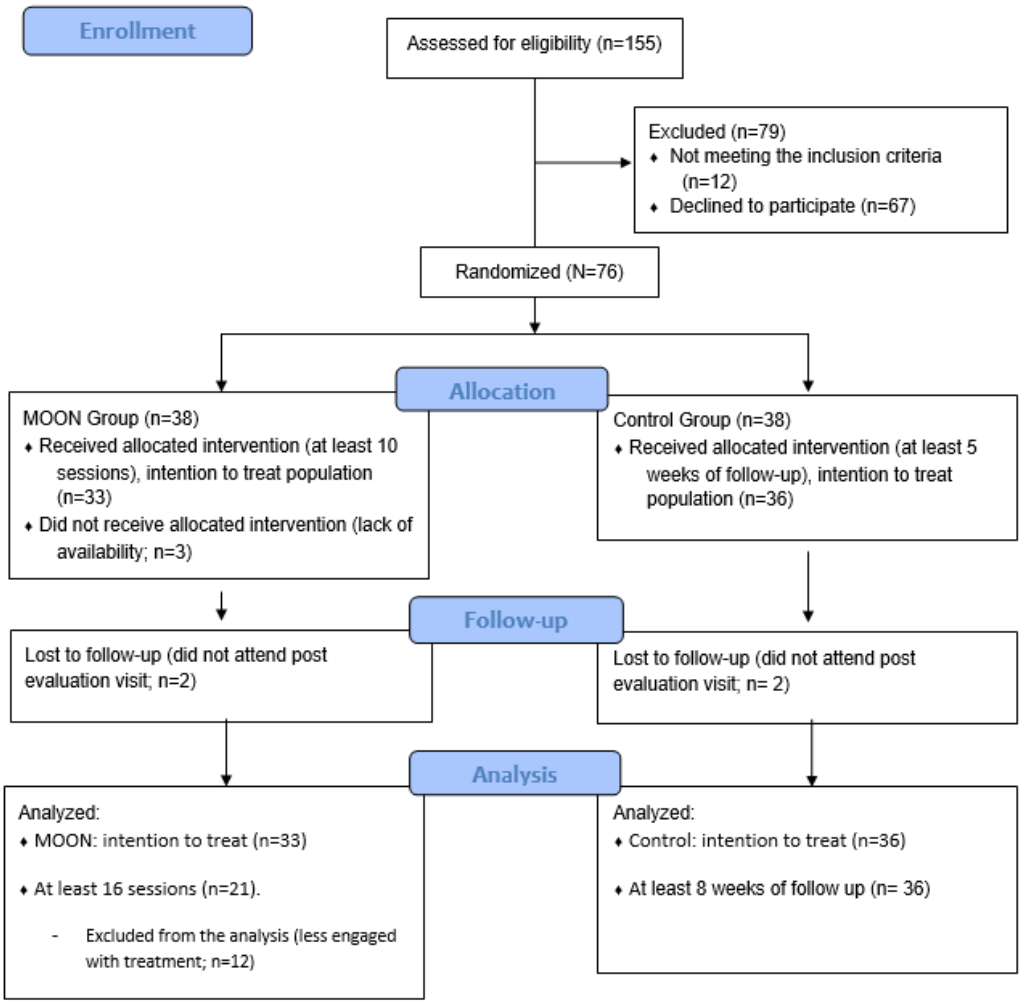
Recruitment: Patient Inclusion and Randomization

The recruitment period was between May 9, 2023, and October 31, 2023. The enrollment target was not completed (n=152) [30]. The last post evaluation was performed on December 13, 2023. The clinical trial ended on December 15, 2023. A total of 76 patients with ADHD participated in the clinical trial and signed a reported consent form. No patients were excluded. The 76 participants were randomized 1:1 (MOON: n=38, 50% and control: n=38, 50%).

The total dropout rate was 9% (7/76) of the participants (n=5, 71% MOON and n=2, 29% control) and did not exceed expectations (12/76, 15%). Lack of availability was the reason mentioned by the dropout cases for not participating. In total, 3% (2/38) of the patients requested to switch to the control group after preassessment and randomization to MOON due to insufficient time to commit to the study, which was not allowed. Another MOON participant had to drop out of treatment due to lack of availability after 2 sessions. In total, 5% (4/76) of the participants (n=2, 50% MOON and n=2, 50% control) were not assessed at postevaluation and were lost to follow-up (Figure 1). The ITT population was analyzed for the primary and secondary outcomes. The ITT population included 87% (33/38) of the patients assigned to the MOON group and 95% (36/38) of the patients assigned to the control group.

XSL•FO
RenderX

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flow diagram for the number of participants in the clinical trial. ITT: intention-to-treat; MOON: The Secret Trail of Moon.



Baseline Characteristics

Table 2 shows the sociodemographic variables for all participants who completed a preevaluation (N=76). The mean age of the groups were 12.79 (SD 2.86) years for MOON and 12.58 (SD 2.67) years for control. In total, 80% (61/76) of the sample was composed of male participants.

Regarding multimodal treatment, 30% (23/76) of the participants were receiving >1 medication per day. The remaining 70% (53/76) were taking only their ADHD pharmacological treatment. As much as possible, ADHD patients were instructed not to change their medication during the study, except for clinical reasons (eg, side effects or inadequate medication adjustment at the time of examination). For ethical reasons, the needs of the patients were prioritized, and there were 11 changes: 8 dose increases of the usual medication and 3 dose decreases of the usual medication. Of the abovementioned changes, 3 occurred in the MOON group (1 decrease and 2 dose increases) and 8 occurred in the control group (2 decreases and 6 dose increases). Currently, 41% (30/73) of the participants are receiving psychological treatment, and 30% (22/73) had received it before the study. In addition, all participants received weekly psychoeducation and counseling on ADHD.

School, social, and recreational data have been included in Table 3. Regarding difficulties with peers, 19% (14/76) of the participants experienced bullying and 31% (23/76) had difficulty making friends, as reported by parents. Additional clinical information was collected. Regarding medical history, 45% (34/76) of the mothers had difficulties during pregnancy. Regarding the developmental alterations of the participants, 26% (20/76) had delayed speech development, 15% (11/76) had slow gait development, and 24% (18/76) had delayed sphincter control. Of the participants, 35% (26/76) needed previous speech therapy support, 9% (7/76) needed psychomotor therapy, and 11% (8/76) needed early stimulation.

The results are listed in the primary and secondary outcome sections, ordered according to the hypotheses. Analysis was performed with the ITT population (69/76, 91%) and patients who completed 80% of treatment (57/76, 75%).

XSL•FO
RenderX

Table 2. Baseline demographic variables.

Baseline variables	Total (N=76)	MOON[a] (n=38)	Control[b] (n=38)	Chi-square (*df*)	*t* test (*df*)	*P* value
Sex (female), n (%)	15 (20)	7 (18)	8 (21)	0.0 (1)	N/A[c]	.77
Age (y), mean (SD)	12.68 (2.75)	12.79 (2.86)	12.58 (2.67)	N/A	0.3 (74)	.74
Nationality of birth, n (%)				0.3 (2)	N/A	.84
European (Spain)	71 (93)	35 (92)	36 (95)			
Hispanic America	2 (3)	1 (3)	1 (3)			
Asian	3 (4)	2 (5)	1 (3)			
Adopted, n (%)	4 (5)	2 (5)	2 (5)	__[d]	—	—
Family background, n (%)						
Lives with both parents	63 (83)	34 (90)	29 (76)	2.3 (1)	N/A	.12
Lives with mother	12 (16)	4 (11)	8 (21)	1.5 (1)	N/A	.20
Family socioeconomic status (€ per month; €1=US $1.4), n (%)				0.4 (3)	N/A	.93
500-1500	5 (7)	2 (5)	3 (8)			
1500-2000	7 (9)	4 (11)	3 (8)			
2000-2500	13 (17)	6 (16)	7 (18)			
>2500	51 (67)	26 (68)	25 (66)			

[a]MOON: The Secret Trail of Moon (the experimental group with the serious video game).

[b]Control: control group without the serious video game.

[c]N/A: not applicable.

[c]Not available.

XSL•FO
RenderX

Table 3. Clinical characteristics of each group.

Clinical variables	Total (N=76), n (%)	MOON[a] (n=38), n (%)	Control (n=38), n (%)	Chi-square (*df*)	*P* value
Medical history					
ADHD[b] subtypes					
Combined	48 (66)	23 (62)	25 (69)	0.4 (1)	.51
Primarily inattentive	25 (34)	14 (37)	11 (31)	__[c]	—
Medication					
Elvanse	53 (55)	29 (62)	24 (49)	8.5 (10)	.58
Equasym	8 (9)	3 (7)	5 (11)	—	—
Intuniv	7 (7)	3 (6)	4 (8)	—	—
Medikinet	5 (5)	2 (4)	3 (6)	—	—
Risperdal	6 (6)	2 (4)	4 (8)	—	—
Concerta	3 (3)	1 (2)	2 (4)	—	—
Other: atenza, atomoxetina, rubifen retard, and rubicrono	4 (4)	2 (4)	2 (4)	—	—
Invega and zyprexa	2 (2)	1 (2)	1 (2)	—	—
Fluoxetina and sertralina	8 (9)	4 (9)	4 (8)	—	—
Psychological therapy				1.2 (2)	.52
Current	30 (41)	13 (34)	9 (25)		
Previous	22 (29)	13 (34)	17 (47)		
School background					
Repeating a school year	21 (8)	11 (29)	10 (27)	0.3 (1)	.85
Special education	8 (11)	3 (8)	5 (14)	0.6 (1)	.43
Curricular adaptation	35 (47)	21 (55)	14 (37)	2.2 (1)	.13
Private teacher	24 (32)	10 (26)	14 (37)	1.1 (1)	.28
High abilities program	6 (8)	5 (13)	1 (3)	2.7 (1)	.09
Peer problems					
Bullying	14 (19)	4 (11)	10 (26)	3.3 (1)	.06
Tendency to isolate	14 (19)	7 (18)	7 (19)	0.0 (1)	.95
Difficulty in making friends	23 (31)	10 (26)	13 (35)	0.6 (1)	.40
Behavioral problems	13 (17)	5 (13)	8 (22)	0.9 (1)	.33
Recreation					
Sports	60 (79)	30 (83)	30 (88)	—	—
Play a musical instrument	11 (14)	5 (13)	4 (11)	—	—
Regular video game use (h)	53 (74)	28 (76)	25 (71)	0.1 (2)	.68
0.25-1	21 (44)	10 (39)	11 (50)	0.9 (1)	.62
1.5-2	14 (29)	9 (35)	5 (23)	—	—
2.5-5	13 (27)	7 (18)	6 (27)	—	—
Regular mobile use (h)	50 (67)	—	—	—	—
0.25-1	10 (24)	6 (27)	4 (20)	0.8 (2)	.66
1.5-2	16 (38)	7 (32)	9 (45)	—	—
2.5-6	16 (38)	9 (41)	7 (35)	—	—

[a]MOON: The Secret Trail of Moon.

XSL•FO
RenderX

[b]ADHD: attention-deficit/hyperactivity disorder.

[c]Not available.

Primary Outcome: Emotional Regulation

The main hypothesis of a 3- or 4-point drop in the primary end point of global SDQ was not achieved. We found no statistically significant differences in the primary variable (SDQ) in our pretest and posttest comparison ($F_{1,64}$=0.51; P=.47; η^2=0.00; 1–β=0.11). The preassessment overall SDQ scores for the experimental group (mean 27.48, SD 5.64; 95% CI 25.45-29,51) and the control group (mean 26.61, SD 6.03; 95% CI 24.57-28.63) were higher than the post scores of the experimental group (mean 26.09, SD 5.74; 95% CI 24.06-28.12) and the post scores of the control group (mean 25.94, SD 5.93; 95% CI 23.90-27.97). Significant differences were found in the prosocial scale in favor of the control group ($F_{1,66}$=3.73; P=.05; η^2=0.05; 1–β=0.47). These results are shown in Table 4 and Multimedia Appendix 3.

Table 4. Primary outcome: pre-post comparison of the Strengths and Difficulties Questionnaire (SDQ) scores.

	Emotional symptoms	Behavior problems	Hyperactivity	Peer problems	Prosocial scale	Internalizing	Externalizing	SDQ global
Intention-to-treat (at least 10 sessions; n=69)								
MOON[a] (n=33), mean (SD)								
Pre	9.18 (2.54)	6.39 (1.56)	8.03 (2.00)	3.88 (2.53)	12.18 (2.37)	13.06 (4.39)	14.42 (2.68)	27.48 (5.64)
Post	8.85 (2.19)	6.27 (1.66)	7.27 (2.28)	3.70 (2.55)	11.78 (2.56)	12.55 (4.10)	13.55 (3.15)	26.09 (5.74)
Control (n=36), mean (SD)								
Pre	9.00 (2.44)	6.19 (1.93)	7.83 (2.39)	3.91 (2.35)	12.78 (1.88)	12.86 (4.04)	14.09 (3.52)	26.79 (6.03)
Post	8.89 (2.23)	6.08 (1.85)	7.33 (2.49)	4.00 (2.60)	13.08 (1.76)	12.74 (3.99)	13.42 (3.82)	26.09 (5.91)
F test (*df*)	0.50 (1)	0 (1)	0.40 (1)	0.70 (1)	3.73 (1)	0.96 (1)	0.21 (1)	0.51 (1)
P value	.48	.97	.52	.40	.05	.33	.64	.47
η^2	0.00	0.00	0.00	0.01	0.05	0.01	0.00	0.00
16 sessions (n=57)								
MOON (n=21), mean (SD)								
Pre	9.67 (2.45)	6.57 (1.56)	7.90 (1.84)	4.24 (2.54)	11.33 (2.24)	13.90 (4.19)	14.48 (2.54)	28.38 (5.54)
Post	9.00 (2.12)	6.33 (1.56)	7.38 (2.24)	3.76 (2.36)	11.14 (2.15)	12.76 (3.83)	13.71 (3.28)	26.48 (5.76)
Control (n=36), mean (SD)								
Pre	9.00 (2.44)	6.19 (1.93)	7.83 (2.39)	3.91 (2.35)	12.78 (1.88)	12.86 (4.04)	14.09 (3.52)	26.79 (6.03)
Post	8.89 (2.23)	6.08 (1.85)	7.33 (2.49)	4.00 (2.60)	13.08 (1.76)	12.74 (3.99)	13.42 (3.82)	26.09 (5.91)
F test (*df*)	1.95 (1)	0.10 (1)	0.02 (1)	2.28 (1)	1.85 (1)	3.35 (1)	0.05 (1)	1.08 (1)
P value	.16	.75	.87	.13	.17	.07	.81	.30
η^2	0.03	0.00	0.00	0.04	0.03	0.05	0.00	0.02

[a]MOON: *The Secret Trail of Moon*.

In the most engaged subgroup (16 sessions), clinical trends of improvement on the SDQ of 2 points were observed but did not approach significance. This improvement trend is particularly notable in internalizing difficulties ($F_{1,53}$=3.35; P=.07; η^2=0.05; 1–β=0.43), peer problems ($F_{1,54}$=2.28; P=.13; η^2=0.04; 1–β=0.31) and emotional symptoms ($F_{1,54}$=1.95; P=.16; η^2=0.03; 1–β=0.27; Table 4; Multimedia Appendix 3). At the SDQ follow-up, parents in the MOON group reported that their children felt "a little better," while parents in the control group reported their children felt "about the same."

All participants completed the 10 face-to-face sessions. However, the number of sessions decreased when there was no direct face-to-face supervision by the investigators in the MOON web-based intervention. The results of the main variable (SDQ) of the 3 measures (pre-mid-term-post comparison) are shown in Table 5 and Multimedia Appendix 4. The change of modality (face-to-face and on the web) could have influenced adherence to treatment. The MOON group had an average of 16 sessions, with 3 participants having the minimum number of sessions (10) and 12 patients having the maximum number (20; Multimedia Appendix 5). Most participants in the MOON group (26/33, 79%) played the video game in VR, as they were aged >12 years, following PlayStation's recommendations. A comparative statistical analysis was conducted within the MOON group, comparing participants who played the face-to-face VR version (26/33, 79%) with those who used the computer version for the entire treatment (7/33, 21%). We found

XSL•FO
RenderX

that the VR subgroup improved most on emotional symptoms ($F_{1, 31}$=7.79; P=.004; η^2=0.20; 1–β=0.77), internalizing ($F_{1, 31}$=4.18; P=.04; η^2=0.11; 1–β=0.50) and, to a lesser extent, global SDQ ($F_{1, 31}$=2.99; P=.09; η^2=0.08; 1–β=0.38).

The ANCOVA for the main hypothesis—which posited that the MOON group would exhibit improved emotional regulation, reflected by a 3- to 4-point decrease in global SDQ scores compared to the control group—was conducted with the number of sessions included as a covariate. The results approached significance ($F_{1, 63}$=3.25; P=.07; η^2=0.04; 1–β=0.42; Figure 2; Table 6). Significant improvements were found in the MOON group in the variables of emotional symptoms ($F_{1, 65}$=4.47; P=.03; η^2=0.06; 1–β=0.54) and internalizing problems ($F_{1, 64}$=6.57; P=.01; η^2=0.09; 1–β=0.71).

Table 5. Primary outcome: pre-mid-term-post comparison of the Strengths and Difficulties Questionnaire (SDQ) scores.

	Emotional symptoms	Behavior problems	Hyperactivity	Peer problems	Prosocial scale	Internalizing	Externalizing	SDQ global
Intention-to-treat (at least 10 sessions; n=69)								
MOON[a] (n=33),mean (SD)								
Pre	9.18 (2.54)	6.39 (1.56)	8.03 (2)	3.88 (2.53)	12.18 (2.37)	13.06 (4.39)	14.42 (2.68)	27.48 (5.64)
Midterm	8.91 (2.28)	6.09 (1.52)	7.58 (2.18)	3.88 (2.80)	12.06 (2.46)	12.79 (4.38)	13.67 (2.88)	26.45 (5.71)
Post	8.85 (2.19)	6.27 (1.66)	7.27 (2.28)	3.70 (2.55)	11.78 (2.56)	12.55 (4.10)	13.55 (3.15)	26.09 (5.74)
Control (n=36), m ean (SD)								
Pre	9 (2.44)	6.19 (1.93)	7.83 (2.39)	3.91 (2.35)	12.78 (1.88)	12.86 (4.04)	14.09 (3.52)	26.79 (6.03)
Midterm	8.69 (2.30)	6.25 (2.08)	6.97 (2.55)	3.97 (2.43)	12.72 (1.95)	12.67 (3.89)	13.22 (3.72)	25.89 (5.80)
Post	8.89 (2.23)	6.08 (1.85)	7.33 (2.49)	4 (2.60)	13.08 (1.76)	12.74 (3.99)	13.42 (3.82)	26.09 (5.91)
F test (*df*)	0.44 (2)	0.71 (2)	0.94 (2)	0.72 (2)	2.33 (2)	0.86 (2)	0.13 (2)	0.39 (2)
P value	.64	.49	.39	.48	.10	.42	.87	.67
η^2	0	0.01	0.01	0.01	0.03	0.01	0	0
16 sessions (n=57)								
MOON (n=21), mean (SD)								
Pre	9.67 (2.45)	6.57 (1.56)	7.90 (1.84)	4.24 (2.54)	11.33 (2.24)	13.90 (4.19)	14.48 (2.54)	28.38 (5.54)
Midterm	9.29 (2.10)	6.10 (1.37)	7.57 (1.96)	4.05 (2.90)	11.76 (2.11)	13.33 (4.16)	13.62 (2.72)	26.95 (5.81)
Post	9 (2.12)	6.33 (1.56)	7.38 (2.24)	3.76 (2.36)	11.14 (2.15)	12.76 (3.83)	13.71 (3.28)	26.48 (5.76)
Control (n=36), mean (SD)								
Pre	9 (2.44)	6.19 (1.93)	7.83 (2.39)	3.91 (2.35)	12.78 (1.88)	12.86 (4.04)	14.09 (3.52)	26.79 (6.03)
Midterm	8.69 (2.30)	6.25 (2.08)	6.97 (2.55)	3.97 (2.43)	12.73 (1.95)	12.67 (3.89)	13.22 (3.72)	25.89 (5.80)
Post	8.89 (2.23)	6.08 (1.85)	7.33 (2.49)	4 (2.60)	13.08 (1.76)	12.74 (3.99)	13.42 (3.82)	26.09 (5.91)
F test (*df*)	1.30 (2)	1.01 (2)	0.45 (2)	1.51 (2)	3.15 (2)	2.41 (2)	0.03 (2)	0.65 (2)
P value	.27	.36	.63	.22	*.04*	.09	.96	.51
η^2	0.02	0.01	0	0.02	0.05	0.04	0	0.01

[a]MOON: *The Secret Trail of Moon*.

XSL•FO
RenderX

Figure 2. Results of analysis of covariance of the main hypothesis (decrease in The Secret Trail of Moon [MOON] group of 3 or 4 points in global Strengths and Difficulties Questionnaire [SDQ] compared to the control group). The number of sessions was used as a covariate.

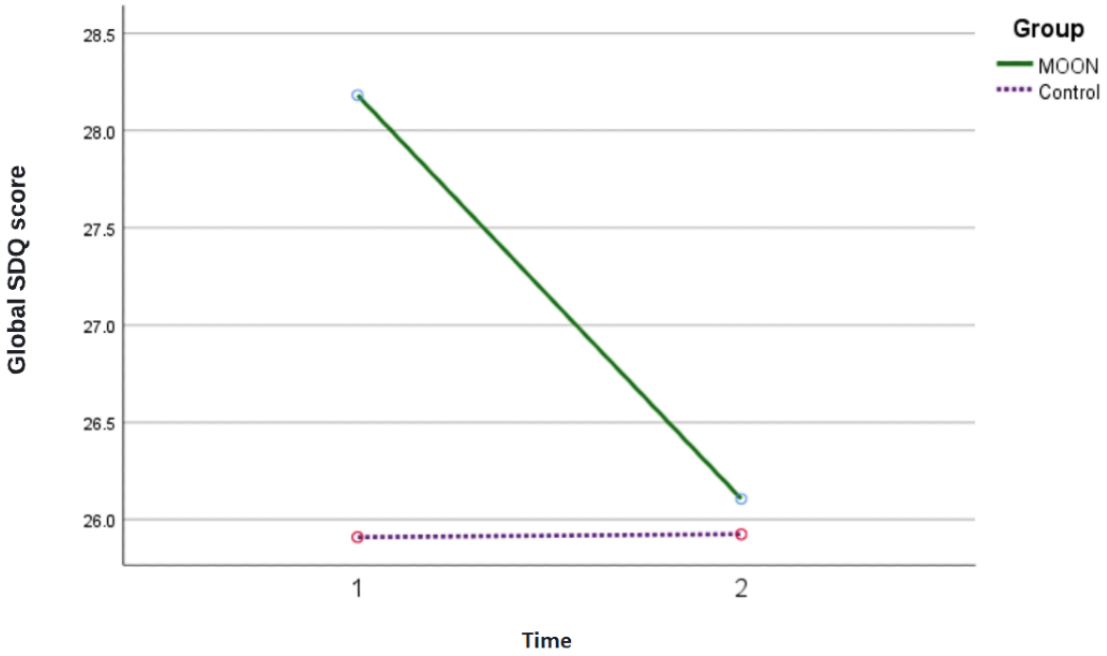


Table 6. Analysis of covariance of the Strengths and Difficulties Questionnaire (SDQ) using the number of sessions as a covariate (N=69).

SDQ	Intention-to treat (at least 10 sessions)							
	MOON[a] (n=33), mean (SD)		Control (n=36), mean (SD)		*F* test (*df*)	*P* value	η^2	1–β
	Pre	Post	Pre	Post				
Emotional symptoms	9.18 (2.54)	8.85 (2.19)	8.91 (2.43)	8.89 (2.23)	4.47 (1)	.03	0.06	0.54
Behavior problems	6.39 (1.56)	6.27 (1.66)	6.19 (1.93)	6.08 (1.85)	0.32 (1)	.56	0	0.08
Hyperactivity	8.03 (2.00)	7.27 (2.28)	7.83 (2.39)	7.40 (2.49)	0.43 (1)	.51	0	0.10
Peer problems	3.88 (2.53)	3.70 (2.55)	3.91 (2.35)	4.00 (2.60)	2.74 (1)	.10	0.04	0.37
Prosocial scale	12.09 (2.36)	11.78 (2.56)	12.78 (1.88)	13.08 (1.76)	1.19 (1)	.27	0.01	0.19
Internalizing	13.06 (4.39)	12.55 (4.10)	12.68 (3.96)	12.74 (3.99)	6.57 (1)	*.01*	0.09	0.71
Externalizing	14.42 (2.68)	13.55 (3.15)	14.09 (3.52)	13.51 (3.83)	0.57 (1)	.44	0	0.11
SDQ global	28.18 (1.10)	26.10 (1.12)	25.90 (1.10)	25.94 (1.12)	3.25 (1)	.07	0.04	0.42

[a]MOON: *The Secret Trail of Moon*.

Secondary Outcomes

Core ADHD Symptoms

No significant differences (*P*>.05) were found in the main ADHD symptoms in SNAP-IV and CPRS.

Clinical trends of improvement were seen in the MOON group versus control in the subscales of hyperactivity ($F_{1, 66}$=3.06; *P*=.08; η^2=0.04) and inattention ($F_{1, 66}$=2.41; *P*=.12; η^2=0.03) and in the total scale of SNAP-IV ($F_{1, 67}$=2.63; *P*=.11; η^2=0.03) and CPRS ($F_{1, 65}$=2.06; *P*=.15; η^2=0.03). In the participants who more engaged in the MOON treatment (at least 16 sessions), clinical trends of improvement were also observed in hyperactivity ($F_{1, 54}$=2.51; *P*=.11; η^2=0.04) and CPRS total scale ($F_{1, 53}$=2.05; *P*=.15; η^2=0.03; Table 7; Multimedia Appendix 6).

XSL•FO
RenderX

Table 7. Secondary outcomes related to hypotheses 2 and 3 (patients with attention-deficit/hyperactivity disorder [ADHD] using The Secret Trail of Moon [MOON] improve in core ADHD symptoms [hypothesis 2] and cognitive functioning compared with the control group; N=69).

	Intention-to-treat (at least 10 sessions; n=69)							16 sessions (n=57)						
	MOON (n=33), mean (SD)		Control (n=36), mean (SD)		*F* test (*df*)	*P* value	ηp^2	MOON (n=21), mean (SD)		Control (n=36), mean (SD)		*F* test (*df*)	*P* value	ηp^2
	Pre	Post	Pre	Post				Pre	Post	Pre	Post			
Core ADHD														
SNAP-IV[a]														
Inattention	28.73 (5.51)	25.91 (6.44)	26.67 (6.59)	25.78 (6.06)	2.41 (1)	.12	0.03	28.05 (6.16)	26.33 (6.32)	26.67 (6.59)	25.78 (6.06)	0.37 (1)	.54	0.00
Hyperactivity	23.12 (5.93)	20.18 (5.37)	23.42 (7.81)	21.91 (7.50)	3.06 (1)	.08	0.04	24.43 (5.10)	21.33 (4.99)	23.42 (7.81)	21.91 (7.50)	2.51 (1)	.11	0.04
Total	51.85 (8.96)	45.30 (10.97)	50.08 (12.69)	47.08 (11.88)	2.63 (1)	.11	0.03	52.48 (8.86)	47.67 (9.11)	50.08 (12.69)	47.08 (11.88)	0.53 (1)	.46	0.01
CPRS[b]														
ADHD symptoms	26.67 (5.17)	23.97 (6.02)	25.23 (6.17)	23.86 (6.60)	2.06 (1)	.15	0.03	27.95 (4.45)	25.00 (5.74)	25.23 (6.17)	23.86 (6.60)	2.05 (1)	.15	0.03
CGI[c]														
General condition rate	6.9 (11.28)	7.1 (12.65)	6.6 (15.59)	6.7 (12.55)	0.13 (1)	.71	0.00	7.2 (9.82)	7.0 (10.94)	6.6 (15.59)	6.7 (12.55)	0.25 (1)	.61	0.00
Cognitive functioning														
BRIEF-2[d]														
Inhibition	66.00 (12.64)	63.09 (12.60)	63.92 (12.19)	63.67 (12.86)	1.79 (1)	.18	0.02	68.48 (12.17)	63.81 (12.47)	63.92 (12.19)	63.67 (12.86)	3.89 (1)	.05	0.06
Self-supervision	66.58 (12.20)	63.76 (12.52)	63.19 (10.07)	61.14 (14.18)	0.17 (1)	.68	0.00	68.76 (11.36)	65.43 (12.08)	63.19 (10.07)	61.14 (14.18)	0.17 (1)	.68	0.00
Flexibility	69.00 (16.96)	67.82 (16.18)	66.81 (15.00)	65.86 (17.41)	0 (1)	.93	0.00	72.67 (16.53)	70.90 (15.46)	66.81 (15.00)	65.86 (17.41)	0.05 (1)	.80	0.00
Emotional control	63.45 (12.07)	59.76 (11.65)	62.25 (10.91)	61.03 (11.08)	1.25 (1)	.26	0.01	66.24 (10.09)	61.86 (10.72)	62.25 (10.91)	61.03 (11.08)	1.51 (1)	.22	0.02
Initiative	65.85 (11.71)	64.18 (12.38)	59.39 (11.76)	59.86 (13.68)	0.90 (1)	.34	0.01	65.62 (12.91)	65.62 (13.17)	59.39 (11.76)	59.86 (13.68)	0.03 (1)	.85	0.00
Working memory	71.27 (10.15)	66.36 (12.40)	67.22 (11.52)	65.94 (12.90)	3.49 (1)	.06	0.05	71.48 (11.69)	64.33 (16.17)	67.22 (11.52)	65.94 (12.90)	4.08 (1)	.04	0.06
Planning and organization	67.88 (9.83)	66.52 (11.29)	62.61 (11.21)	62.39 (11.70)	0.49 (1)	.48	0.00	68.62 (10.70)	67.14 (11.72)	62.61 (11.21)	62.39 (11.70)	0.56 (1)	.45	0.01
Task supervision	65.36 (11.92)	62.82 (13.00)	62.47 (10.12)	61.03 (12.50)	0.40 (1)	.52	0.00	64.57 (13.73)	61.95 (13.82)	62.47 (10.12)	61.03 (12.50)	0.44 (1)	.65	0.00
Material organization	73.76 (13.01)	71.03 (12.71)	65.44 (13.96)	64.36 (16.19)	0.61 (1)	.43	0.00	75.95 (13.23)	70.00 (12.16)	65.44 (13.96)	64.36 (16.19)	4.75 (1)	.03	0.08
Behavioral regulation Index	67.73 (12.38)	64.70 (12.95)	64.92 (12.25)	63.94 (12.41)	1.06 (1)	.30	0.01	70.29 (11.77)	65.95 (12.72)	64.92 (12.25)	63.94 (12.41)	2.29 (1)	.13	0.04
Emotional regulation Index	68.55 (15.28)	65.30 (13.97)	66.39 (12.27)	64.97 (13.78)	0.64 (1)	.42	0.01	72.14 (14.01)	68.10 (12.19)	66.39 (12.27)	64.97 (13.78)	0.96 (1)	.33	0.01
Cognitive regulation index	72.64 (11.31)	69.36 (12.98)	66.28 (11.93)	65.56 (13.90)	2.52 (1)	.11	0.03	73.10 (13.09)	69.33 (13.92)	66.28 (11.93)	65.56 (13.90)	3.42 (1)	.07	0.05
Global executive function	75.88 (16.29)	70.39 (13.13)	71.44 (17.28)	70.44 (19.40)	3.56 (1)	.06	0.03	76.05 (11.83)	71.38 (12.87)	71.44 (17.28)	70.44 (19.40)	3.58 (1)	.06	0.06

XSL•FO
RenderX

	Intention-to-treat (at least 10 sessions; n=69)							16 sessions (n=57)						
	MOON (n=33), mean (SD)		Control (n=36), mean (SD)		*F* test (*df*)	*P* value	ηp^2	MOON (n=21), mean (SD)		Control (n=36), mean (SD)		*F* test (*df*)	*P* value	ηp^2
	Pre	Post	Pre	Post				Pre	Post	Pre	Post			
CPT-3[e]														
Detectability	45.97 (11.71)	46.88 (13.11)	47.67 (8.70)	45.25 (9.35)	3.29 (1)	.07	0.04	46.81 (11.38)	49.57 (13.60)	47.67 (8.70)	45.25 (9.35)	5.92 (1)	.01	0.09
Omissions	50.67 (13.62)	52.27 (15.08)	49.00 (8.07)	48.97 (11.69)	0.43 (1)	.51	0.00	52.95 (15.99)	56.10 (17.68)	49.00 (8.07)	48.97 (11.69)	1.08 (1)	.30	0.01
Commissions	44.73 (9.78)	45.33 (10.29)	46.25 (8.65)	43.97 (7.77)	2.92 (1)	.09	0.04	43.81 (8.75)	46.00 (10.20)	46.25 (8.65)	43.97 (7.77)	5 (1)	.02	0.08
Perseverations	51.55 (11.57)	54.15 (15.76)	50.64 (10.97)	50.11 (8.71)	1.07 (1)	.30	0.01	52.67 (13.39)	58.10 (18.55)	50.64 (10.97)	50.11 (8.71)	2.66 (1)	.10	0.04
HRT[f]	57.76 (12.82)	56.91 (11.14)	56.22 (8.58)	58.19 (10.95)	2.61 (1)	.11	0.03	59.81 (13.67)	58.76 (11.47)	56.22 (8.58)	58.19 (10.95)	1.98 (1)	.16	0.03
HRTsd[g]	51.21 (14.38)	53.52 (15.64)	51.19 (9.21)	51.56 (11.98)	0.63 (1)	.43	0.00	53.90 (16.18)	58.81 (16.28)	51.19 (9.21)	51.56 (11.98)	2.49 (1)	.12	0.04
Variability	50.69 (12.80)	49.52 (9.87)	51.53 (10.19)	50.69 (8.77)	0.40 (1)	.52	0.00	50.90 (13.73)	52.53 (11.00)	51.53 (10.19)	50.69 (8.77)	0 (1)	.93	0.00
Block change	48.28 (8.69)	50.73 (9.84)	50.42 (10.09)	52.08 (12.43)	0.03 (1)	.85	0.00	48.45 (10.25)	51.52 (12.01)	50.42 (10.09)	52.08 (12.43)	0.08 (1)	.76	0.00
Inter stimulus change	52.67 (12.77)	57.30 (13.44)	51.97 (10.31)	54.14 (12.01)	1.25 (1)	.26	0.01	55.67 (14.06)	61.81 (13.83)	51.97 (10.31)	54.14 (12.01)	2.20 (1)	.14	0.03
Corsi														
N span (items recalled)	5.66 (1.15)	5.79 (1.21)	5.17 (0.98)	5.03 (1.20)	0.38 (1)	.53	0.00	5.55 (1.35)	5.67 (1.35)	5.17 (0.98)	5.03 (1.20)	0.27 (1)	.60	.00
CTMT-2[h]														
Inhibitory control	44.24 (14.41)	52.33 (15.42)	46.06 (13.77)	53.72 (13.62)	0.15 (1)	.69	0.00	41.43 (15.66)	50.95 (17.40)	46.06 (13.77)	53.72 (13.62)	0.74 (1)	.39	.01
Set shifting	43.55 (16.72)	52.24 (15.33)	45.60 (12.79)	49.97 (12.09)	4.06 (1)	.04	0.05	41.24 (17.19)	50.00 (16.15)	45.60 (12.79)	49.97 (12.09)	3.05 (1)	.08	.05
Total	43.82 (14.86)	52.76 (15.59)	45.74 (12.37)	52.39 (11.82)	1.95 (1)	.16	0.02	41.19 (15.58)	51.00 (17.33)	45.74 (12.37)	52.39 (11.82)	2.45 (1)	.12	.04

[a]SNAP-IV: Swanson, Nolan, and Pelham Rating Scale.

[b]CPRS: Conners Abbreviated Symptom Questionnaire.

[c]CGI: Clinical Global Impression.

[d]BRIEF: Behavior Rating Inventory Executive Function, version 2.

[e]CPT-3: Conners' Continuous Performance Test Third Edition.

[f]HRT: Hit Reaction Time.

[g]HRTsd: Hit Reaction Time SD.

[h]CTMT-2: Comprehensive Trail-Making Test, Second Edition.

Cognitive Functioning

In the BRIEF-2 questionnaire evaluated by parents, statistically significant differences were observed in the MOON group that was more involved (at least 16 sessions) in material organization ($F_{1, 55}$=4.75; *P*=.03; η^2=0.08), working memory ($F_{1, 55}$=4.08; *P*=.05; η^2=0.06), and inhibition ($F_{1, 55}$=3.89; *P*=.05; η^2=0.06). Some statistically significant differences were also found in the cognitive tests performed by the participants. In the CPT-3 test, improvements in detectability were found ($F_{1, 55}$=5.92; *P*=.01; η^2=0.09) in MOON group; however, significant differences in commissions ($F_{1, 55}$=5.00; *P*=.02; η^2=0.08) were found in favor of the control group; in the CTMT-2 test, an improvement in the MOON group was found in set shifting ($F_{1, 66}$=4.06; *P*=.04; η^2=0.05; Table 7; Multimedia Appendices 6 and 7).

Clinical trends of improvement with MOON versus control (ITT=69) were observed in working memory ($F_{1, 67}$=3.49; *P*=.06; η^2=0.05), inhibition ($F_{1, 67}$=1.79; *P*=.18; η^2=0.02),

XSL•FO
RenderX

cognitive regulation index ($F_{1, 67}$=2.52; P=.11; η^2=0.03), and global executive function ($F_{1, 6}$=3.56; P=.06; η^2=0.03).

Side Effects

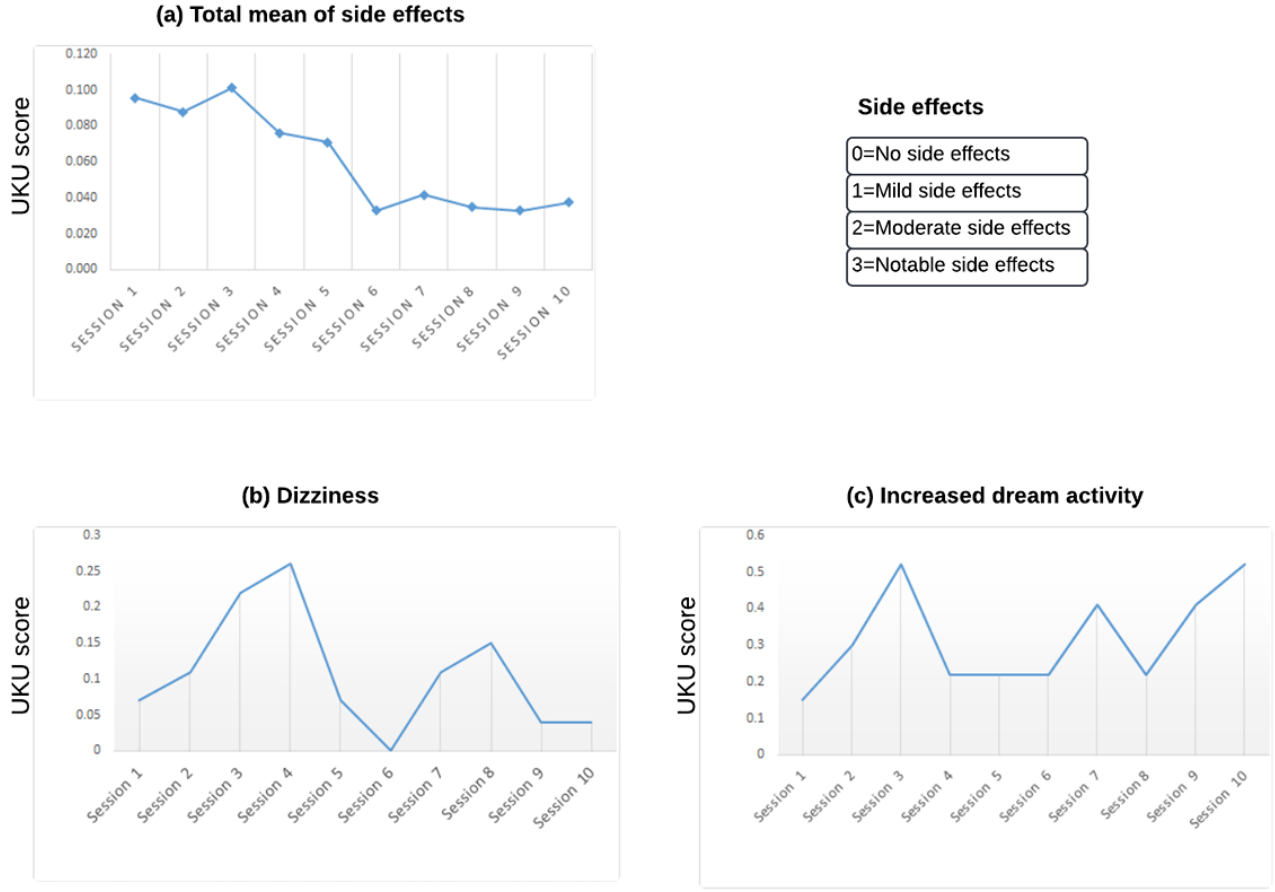
There were no interruptions due to adverse events. None of the participants had side effects related to the VR treatment lasting >1 hour after use. In total, 2 patients reported discomfort from the weight of the VR headset, so the researchers offered the possibility of using the computer game.

We found no significant differences in video game addiction between the groups assessed with the GASA questionnaire.

No mild, moderate, or notable side effects were found (Multimedia Appendix 8). Patients who described a causal relationship with the video game through the UKU questionnaire (possible or probable) reported dizziness and increased dream activity (Figure 3; Multimedia Appendices 9 and 10). The mean score for dizziness across the 10 sessions was 0.10 (SD 0.08); the mean score for increased dream activity was 0.31 (SD 0.13; 0=no side effects and 1=mild side effects that do not interfere with patient performance) and was decreasing across sessions; however, some participants (15/26, 58%) reported increased dream activity across sessions.

In relation to sleep problems measured with the Sleep Disturbance Scale for Children questionnaire, no significant differences were found in the number of hours slept (mean in both groups=8-9 h) or in the time taken to fall asleep (mean 15-30 min). Nonetheless, an improvement in favor of the MOON group was found in the disorders of excessive somnolence scale ($F_{1, 65}$=4.55; P=.03; η^2=0.06; Table 8).

Figure 3. (A) Total mean of the side effects per session, specifically for the most relevant side effects: (B) dizziness (mean 0.10) and (C) increased dream activity (mean 0.31). UKU: Udvalg für Kliniske Undersogelser.



XSL•FO
RenderX

Table 8. Secondary outcomes related to hypothesis 6: there are no clinically meaningful side effects associated with the video game (N=69).

Safety outcomes	Intention-to-treat (at least 10 sessions)					
	MOON[a] (n=33), mean (SD)		Control (n=36), mean (SD)		*F* test (*df*)	*P* value
	Pre	Post	Pre	Post		
Game addiction (GASA[b])	17.42 (5.75)	15.70 (5.48)	17.06 (5.25)	15.89 (5.24)	0.32 (1)	.57
Sleep disorders (SDSC[c])						
Sleeping hours	2.37 (0.94)	2.39 (0.65)	2.33 (0.92)	2.50 (0.56)	0.57 (1)	.44
Time to sleep	2.52 (1.30)	2.15 (0.79)	2.38 (1.24)	2.05 (0.71)	0.00 (1)	.99
DIMS[d]	15.53 (4.70)	15.72 (4.60)	15.24 (5.97)	14.74 (5.62)	0.04 (1)	.83
SBD[e]	4.13 (1.99)	3.87 (1.47)	4.22 (1.80)	4.36 (1.93)	1.70 (1)	.19
DA[f]	4.38 (1.86)	4.15 (1.58)	3.97 (1.40)	3.86 (1.24)	0.13 (1)	.71
SWTD[g]	11.88 (4.15)	11.56 (3.78)	11.37 (4.56)	11.64 (5.05)	0.18 (1)	.67
DOES[h]	12.09 (4.05)	11.09 (3.97)	9.51 (3.28)	10.00 (4.27)	4.55 (1)	*.03*
SHY[i]	3.47 (1.91)	3.42 (1.78)	4.19 (2.60)	4.11 (2.43)	0.01 (1)	.90
Total	55.40 (10.63)	55.00 (10.57)	55.43 (14.59)	53.28 (14.79)	1.04 (1)	.31

[a]MOON: *The Secret Trail of Moon*.

[b]GASA: Game Addiction Scale for Adolescents.

[c]SDSC: Sleep Disturbance Scale for Children.

[d]DIMS: disorders of initiating and maintaining sleep.

[e]SBD: sleep breathing disorders.

[f]DA: disorders of arousal nightmares.

[g]SWTD: sleep-wake transition disorders.

[h]DOES: disorders of excessive somnolence.

[i]SHY: sleep hyperhidrosis.

Adherence and Acceptability

Additional information was collected using an unvalidated scale (previously applied in our usability study [47] to obtain the patients' opinion of the video game). Of the 33 participants who played MOON, 21 (64%) played video games on a habitual basis and 20 (61%) had tried VR before. In total, 82% (27/33) of the participants liked the experience; 79% (26/33) of the participants would recommend it to other people with ADHD.

Additional information was collected on fatigue, boredom, or dizziness (in the case of VR). In total, 24% (8/33) of the participants experienced fatigue (not surprising in cognitive training), 13% (4/33) found it boring, and 7% (2/33) were dizzy on ≥1 occasions. The overall experience was considered positive, as reported by the parents. Some parents mentioned perceiving improvement in thinking before acting, responsibility, autonomy, and organization; others did not perceive changes. Some parents considered that we should have held more weekly sessions.

Most of the participants and their parents preferred face-to-face sessions to web-based sessions. Bugs (video game errors) were considered especially frustrating in the web-based sessions by both children and parents, especially in the interface or in games such as chess. The levels were considered well balanced, except for Kuburi (too demanding and frustrating). Overall, the video game music was liked by the participants; some patients said it helped them to concentrate.

Discussion

Principal Findings

In this prospective, single-center, randomized, unblinded pre-post evaluation study, our primary hypothesis of a 3-point decrease in the global SDQ score for emotional regulation after 3 months of MOON cognitive training in clinically stable, medicated children and adolescents with ADHD was not confirmed. However, we observed improvement trends in specific areas among patients who were more engaged with the MOON treatment. After conducting an ANCOVA with the number of sessions as a covariate, we found values approaching significance for the main hypothesis, with a 2-point drop in the global SDQ. This suggests that motivation may have played a key role in the improvement of emotional regulation in the participants. Given that motivation is a crucial factor in individuals with ADHD, it likely influenced these outcomes [21,24,31].

Considering the number of sessions as a covariate, significant differences in emotional symptoms and internalizing problems were found in favor of the experimental group. No behavioral improvement trends were observed. While externalizing difficulties are related to temperamental negative affect and

XSL•FO
RenderX

anger dysregulation (bottom-up processes), internalizing difficulties suggest an overregulation of negative emotion through maladaptive cognitive strategies such as blaming and ruminating (top-down processes, more closely linked to executive functions) [48-50]. Participants with externalizing problems may have found the video game lacking in external rewards (eg, prizes), unlike those participants with greater intrinsic motivation (eg, self-improvement and sense of progress). This clinical trial had significantly more sessions than the previous one (20 vs 12) [29], increasing the need for novelty, change, and reinforcement. However, as participants noted in their feedback, the game did not seem to meet these reward-based needs effectively.

Our results also showed improvements in material organization, working memory, and inhibition in the 16-session MOON group. The improvement in working memory is in accordance with other investigations whose improvement in working memory occurred especially in those individuals who had a higher level of motivation and voluntary commitment to training [21,51] as well as the organization [52]. The ITT group also had an improvement close to significance. However, in the Corsi cubes (objective task of working memory), no significant differences were found; the MOON group had higher memory span than the control group, but this difference was very small. These results are consistent with other randomized controlled trials that showed significant improvements in ADHD symptoms and working memory but no improvements in other cognitive functions tested, such as planning or self-monitoring [53]. Regarding cognitive flexibility, significant differences were found in favor of the MOON group in set shifting on the CTMT-2 test, but it was not significant on the parent-reported flexibility subscale on the BRIEF-2 questionnaire.

Findings may point at a more powerful experience when using the VR experience. At least theoretically, VR may be more effective in tricking the brain than regular video games [28-30,47] suggesting a heightened sense of presence in the virtual environment and reduced perception of the real world, leading to fewer distractions [22]. In addition our findings may also just be a matter of age and brain maturity, perhaps related to the age of the VR users (aged >12 years years). Another putative explanation is that those who played on the computer experienced a more bugged and, consequently, more frustrating version of the game.

Furthermore, the use of MOON was well tolerated, as no clinically significant side effects were reported. There were no adverse events to report to the Spanish Agency of Medicines and Health Products. The VR intervention reported a mean dizziness score of 0.31 on the UKU scale, where 0 indicates no side effects and 1 indicates mild side effects, which did not interfere with patient performance. No significant differences were found in the GASA scale on video game addiction. In ADHD, this variable is especially important as there is a higher risk of addiction [54]. Serious video games versus commercial games allow controlling design variables about addiction [28].

Here, we found significant differences related to sleep in the subscale disorders of excessive somnolence in favor of the MOON group. As for the MOON group that played through VR, we also found that some of them remembered more of their dreams after the sessions, as assessed with the UKU scale. This phenomenon (although neither significant nor interfering) was consistent with the previous clinical trial [29]. While we do not consider this a concerning side effect, it is worth exploring the potential impact of VR on sleep.

This clinical trial has important limitations, particularly the smaller-than-planned sample size (n=152) that affected its statistical power [30]. Trends of improvement were observed in various areas, but a larger sample may have yielded statistically significant results. In addition, this investigation, such as the previous one, was not blinded [29]. The effect sizes are small, and we must interpret these results cautiously. The sessions were decreasing with the intervention with MOON on the web, and this suggests that the game was insufficiently developed, as mentioned by the users themselves. The reward system in the video game was only partially developed, which may have impacted motivation and caused the game to become repetitive. In addition, usability study would have been necessary before the clinical trial (as in the previous study [47] to refine the gameplay and correct bugs). Another limitation of the study was that hypothesis 4 could not be tested, and the corresponding analyses could not be performed, as academic grades were not collected consistently (only half of the parents provided this information).

Nonetheless, our results provide additional, albeit partial, support for incorporating serious video games into the multimodal treatment approach for ADHD. After conducting our second clinical trial, we considered that video games do not have to be for everyone. Heterogeneity in ADHD is substantial. While some patients found improvements related to cognitive training with MOON, others did not. Other training methods exist. To optimize the treatment of patients with ADHD, it is crucial to focus on individual interests and their strengths and difficulties to personalize the treatment to what the person needs. Video games can be another addition to complement multimodal treatment along with other interventions such as music training, physical exercise, regular interaction with nature, or learning a new language accompanied by a healthy peer support network and appropriate parenting. For example, to improve externalizing symptomatology, behavioral psychological therapy intervention strategies may be more effective, especially early in the preschool years [17]. Combining training with pharmacological treatment seems to enhance the benefits more than pharmacological treatment alone [23].

Conclusions

Serious video games combined with multimodal treatment can improve symptoms associated with ADHD. In this study, significant differences in inhibition, working memory, and material organization were observed in participants who were more engaged in the MOON treatment.

XSL•FO
RenderX

Acknowledgments

The authors thank in advance all the staff involved in the research and the Gammera Nest company, especially Miguel Ruiz, the main programmer and composer of the game music, and the trainees who made the development of the video game possible. The authors extend special thanks to all children, adolescents, and parents for their future participation in this study. This research is funded by the Comunidad de Madrid (Spain) through the Industrial Doctorate (IND2020/BMD-1755) and Grants Public-Private Partnership Projects 2022, Comunidad de Madrid (The Secret Trail of Moon; CPP2022-009788); Ministry of Science, Innovation and Universities, European Regional Development Fund. This research was previously funded by the National Project (RTI2018-101857-B-I00 for HB-F), the Foundation for Health Innovation and Prospective in Spain (HB-F), the Center for Technological Development and Innovation (MM-M), and the Carlos III Health Institute (predoctoral contracts for training in health research for MRY and HB-F). The Shibuya and Health Start awards also provided funding for this project. The clinical trial is funded by the Community of Madrid (2020 Industrial Doctorates IND2020/BMD-17544).

Authors' Contributions

MM-M, MB-F, and HB-F carried out the research. HB-F was the principal investigator. MM-M and HB-F recruited the patients, and they performed the data collection. MM-M monitored the Calendly appointments and *The Secret Trail of Moon* home training data collection via PlayFab and designed the electronic case report form (REDCap [Research Electronic Data Capture]). MB-F entered the data into the program. MM-M, MR-Y, CG-T, and HB-F designed the video game. CL and PW provided research design ideas and statistical input. MB-F, AR, MM-M, and HB-F performed the statistical analyses. PL-G and HB-F supervised the study.

Conflicts of Interest

HB-F has received lecture fees from Takeda Pharmaceuticals and the laboratories BIAL, Rubio, and Rovi. He has also been granted 3 prizes for developing serious video games to treat attention-deficit/hyperactivity disorder: the Shibuya Prize by Takeda, first prize from the College of Psychologists of Madrid, and a prize for the best innovative health initiative within the Health Start campaign. He is the principal investigator of predoctoral contracts for training in health research (IFI16/00039); the coprincipal investigator of a Ministry of Economy, Trade, and Enterprise research grant (RTI2018-101857-B-I00); and the principal investigator of a Sincronia research project funded by the Start-up Bitsphi. Moreover, he is the recipient of (1) a Foundation for Innovation and Foresight in Health in Spain grant and (2) a Puerta de Hierro Segovia de Arana Institute of Health Research intensification grant and is involved in 2 clinical trials (Mensia Koala, Newrofeed study; ESKETSUI2002). He is also a cofounder of Haglaia Solutions and an employee and member of the advisory board of ITA Salud Mental (Korian).

Multimedia Appendix 1
CONSORT (Consolidated Standards of Reporting Trials) 2010 checklist of information to include when reporting a randomized trial.
[PDF File (Adobe PDF File), 69 KB - games_v13i1e59124_app1.pdf]

Multimedia Appendix 2
Secondary outcomes related to hypotheses 4: Patients with attention-deficit/hyperactivity disorder using The Secret Trail of Moon improve in academic performance compared with the control group.
[DOCX File , 17 KB - games_v13i1e59124_app2.docx]

Multimedia Appendix 3
Estimated marginal means pre-post measured with the Strengths and Difficulties Questionnaire (SDQ) for group 1: The Secret Trail of Moon (green solid line) and group 2: control (purple dashed line). The drop in score indicates symptom improvement. The prosocial scale of the SDQ is an exception, where improvement is indicated by an increase in the score.
[DOCX File , 219 KB - games_v13i1e59124_app3.docx]

Multimedia Appendix 4
Estimated marginal means measured with the Strengths and Difficulties Questionnaire (SDQ) at 3 time points (pre, mid, and post) for group 1: The Secret Trail of Moon (green solid line) and group 2: control (purple dashed line). The drop in score indicates symptom improvement. The prosocial scale of the SDQ is an exception, where improvement is indicated by an increase in the score.
[DOCX File , 101 KB - games_v13i1e59124_app4.docx]

Multimedia Appendix 5
Number of web-based sessions completed by The Secret Trail of Moon group.

XSL•FO
RenderX

[PNG File , 11 KB - games_v13i1e59124_app5.png]

Multimedia Appendix 6
Estimated marginal means of pre-post attention-deficit/hyperactivity disorder symptoms measured with the Swanson, Nolan, and Pelham Rating Scale and Conners Abbreviated Symptom Questionnaire for group 1: The Secret Trail of Moon (green solid line) and group 2: control (purple dashed line). The drop in both scores indicates symptom improvement.
[DOCX File , 110 KB - games_v13i1e59124_app6.docx]

Multimedia Appendix 7
Estimated marginal means of pre-post cognitive functioning measured with T scores of the Behavior Rating Inventory Executive Function, version 2, questionnaire for parents. Pre-post comparisons for the 2 groups: The Secret Trail of Moon (green solid line) and group 2: control (purple dashed line). Decreasing scores indicate symptom improvement.
[DOCX File , 300 KB - games_v13i1e59124_app7.docx]

Multimedia Appendix 8
Estimated marginal means of pre-post cognitive functioning measured with T scores of neuropsychological tests (Corsi cubes; and Comprehensive Trail-Making Test, Second Edition [CTMT-2]) for patients with attention-deficit/hyperactivity disorder. Pre-post comparisons for the 2 groups: The Secret Trail of Moon (green solid line) and group 2: control (purple dashed line). Decreasing scores indicate symptom improvement, except in Corsi cubes and CTMT-2 whose improvement is a rise in score.
[DOCX File , 306 KB - games_v13i1e59124_app8.docx]

Multimedia Appendix 9
Total means of relevant symptoms and side effects assessed with the Udvalg für Kliniske Undersogelser questionnaire (all 1).
[PNG File , 161 KB - games_v13i1e59124_app9.png]

Multimedia Appendix 10
List of side effects measured with the Udvalg für Kliniske Undersogelser test during 10 face-to-face sessions.
[DOCX File , 32 KB - games_v13i1e59124_app10.docx]

References

1. American Psychological Association (APA). URL: https://www.apa.org [accessed 2021-09-08]
2. Polanczyk GV, Willcutt EG, Salum GA, Kieling C, Rohde LA. ADHD prevalence estimates across three decades: an updated systematic review and meta-regression analysis. Int J Epidemiol 2014 Apr;43(2):434-442 [FREE Full text] [doi: 10.1093/ije/dyt261] [Medline: 24464188]
3. Christiansen H, Hirsch O, Albrecht B, Chavanon ML. Attention-Deficit/Hyperactivity Disorder (ADHD) and emotion regulation over the life span. Curr Psychiatry Rep 2019 Mar 02;21(3):17. [doi: 10.1007/s11920-019-1003-6] [Medline: 30826879]
4. Groves NB, Wells EL, Soto EF, Marsh CL, Jaisle EM, Harvey TK, et al. Executive functioning and emotion regulation in children with and without ADHD. Res Child Adolesc Psychopathol 2022 Jun 11;50(6):721-735 [FREE Full text] [doi: 10.1007/s10802-021-00883-0] [Medline: 34762251]
5. Petrovic P, Castellanos FX. Top-down dysregulation-from ADHD to emotional instability. Front Behav Neurosci 2016 May 23;10:70 [FREE Full text] [doi: 10.3389/fnbeh.2016.00070] [Medline: 27242456]
6. Thompson RA. Emotion regulation: a theme in search of definition. Monogr Soc Res Child Dev 1994;59(2-3):25-52. [Medline: 7984164]
7. Graziano PA, Garcia A. Attention-deficit hyperactivity disorder and children's emotion dysregulation: a meta-analysis. Clin Psychol Rev 2016 Jun;46:106-123. [doi: 10.1016/j.cpr.2016.04.011] [Medline: 27180913]
8. Deotto A, Eastwood JD, Toplak ME. Temperament profiles associated with internalizing symptoms and externalizing behavior in adolescents with ADHD. Child Psychiatry Hum Dev 2022 Feb 05;53(1):109-123. [doi: 10.1007/s10578-020-01116-z] [Medline: 33398690]
9. Mak AD, Lee S, Sampson NA, Albor Y, Alonso J, Auerbach RP, et al. ADHD comorbidity structure and impairment: results of the WHO World Mental Health Surveys International College Student Project (WMH-ICS). J Atten Disord 2022 Jun 10;26(8):1078-1096 [FREE Full text] [doi: 10.1177/10870547211057275] [Medline: 34753324]
10. van der Putten WJ, Mol AJ, Groenman AP, Radhoe TA, Torenvliet C, van Rentergem JA, et al. Is camouflaging unique for autism? A comparison of camouflaging between adults with autism and ADHD. Autism Res 2024 Apr 07;17(4):812-823. [doi: 10.1002/aur.3099] [Medline: 38323512]

XSL•FO
RenderX

11. Dalsgaard S, Østergaard SD, Leckman JF, Mortensen PB, Pedersen MG. Mortality in children, adolescents, and adults with attention deficit hyperactivity disorder: a nationwide cohort study. Lancet 2015 May 30;385(9983):2190-2196. [doi: 10.1016/S0140-6736(14)61684-6] [Medline: 25726514]
12. Holden SE, Jenkins-Jones S, Poole CD, Morgan CL, Coghill D, Currie CJ. The prevalence and incidence, resource use and financial costs of treating people with attention deficit/hyperactivity disorder (ADHD) in the United Kingdom (1998 to 2010). Child Adolesc Psychiatry Ment Health 2013 Oct 11;7(1):34 [FREE Full text] [doi: 10.1186/1753-2000-7-34] [Medline: 24119376]
13. Rice KM, Peltzman T, Gottlieb D, Shiner B, Watts BV. Stimulant medication and suicide mortality in attention-deficit hyperactivity disorder. BJPsych Open 2024 Jan 22;10(1):e33 [FREE Full text] [doi: 10.1192/bjo.2023.643] [Medline: 38251676]
14. Castellanos FX, Sonuga-Barke EJ, Milham MP, Tannock R. Characterizing cognition in ADHD: beyond executive dysfunction. Trends Cogn Sci 2006 Mar;10(3):117-123. [doi: 10.1016/j.tics.2006.01.011] [Medline: 16460990]
15. Faltinsen E, Zwi M, Castells X, Gluud C, Simonsen E, Storebø OJ. Updated 2018 NICE guideline on pharmacological treatments for people with ADHD: a critical look. BMJ Evid Based Med 2019 Jun;24(3):99-102. [doi: 10.1136/bmjebm-2018-111110] [Medline: 30530806]
16. Dalrymple RA, McKenna Maxwell L, Russell S, Duthie J. NICE guideline review: attention deficit hyperactivity disorder: diagnosis and management (NG87). Arch Dis Child Educ Pract Ed 2020 Oct 27;105(5):289-293. [doi: 10.1136/archdischild-2019-316928] [Medline: 31776172]
17. Sonuga-Barke EJ, Brandeis D, Cortese S, Daley D, Ferrin M, Holtmann M, European ADHD Guidelines Group. Nonpharmacological interventions for ADHD: systematic review and meta-analyses of randomized controlled trials of dietary and psychological treatments. Am J Psychiatry 2013 Mar;170(3):275-289. [doi: 10.1176/appi.ajp.2012.12070991] [Medline: 23360949]
18. Gallen CL, Schachtner JN, Anguera-Singla R, Anguera JA, Gazzaley A. Influence of game features on attention in adults. Front Psychol 2023 May 9;14:1123306 [FREE Full text] [doi: 10.3389/fpsyg.2023.1123306] [Medline: 37228349]
19. Peñuelas-Calvo I, Jiang-Lin LK, Girela-Serrano B, Delgado-Gomez D, Navarro-Jimenez R, Baca-Garcia E, et al. Video games for the assessment and treatment of attention-deficit/hyperactivity disorder: a systematic review. Eur Child Adolesc Psychiatry 2022 Jan 18;31(1):5-20. [doi: 10.1007/s00787-020-01557-w] [Medline: 32424511]
20. Romero-Ayuso D, Toledano-González A, Rodríguez-Martínez MD, Arroyo-Castillo P, Triviño-Juárez JM, González P, et al. Effectiveness of virtual reality-based interventions for children and adolescents with ADHD: a systematic review and meta-analysis. Children (Basel) 2021 Jan 21;8(2):70 [FREE Full text] [doi: 10.3390/children8020070] [Medline: 33494272]
21. Prins PJ, Dovis S, Ponsioen A, ten Brink E, van der Oord S. Does computerized working memory training with game elements enhance motivation and training efficacy in children with ADHD? Cyberpsychol Behav Soc Netw 2011 Mar;14(3):115-122. [doi: 10.1089/cyber.2009.0206] [Medline: 20649448]
22. Roettl J, Terlutter R. The same video game in 2D, 3D or virtual reality - how does technology impact game evaluation and brand placements? PLoS One 2018 Jul 20;13(7):e0200724 [FREE Full text] [doi: 10.1371/journal.pone.0200724] [Medline: 30028839]
23. Rabipour S, Raz A. Training the brain: fact and fad in cognitive and behavioral remediation. Brain Cogn 2012 Jul;79(2):159-179. [doi: 10.1016/j.bandc.2012.02.006] [Medline: 22463872]
24. Skalski S, Pochwatko G, Balas R. Impact of motivation on selected aspects of attention in children with ADHD. Child Psychiatry Hum Dev 2021 Aug 20;52(4):586-595 [FREE Full text] [doi: 10.1007/s10578-020-01042-0] [Medline: 32816140]
25. Martin-Moratinos M, Bella-Fernández M, Blasco-Fontecilla H. Effects of music on attention-deficit/hyperactivity disorder (ADHD) and potential application in serious video games: systematic review. J Med Internet Res 2023 May 12;25:e37742 [FREE Full text] [doi: 10.2196/37742] [Medline: 37171837]
26. Mayer-Benarous H, Benarous X, Vonthron F, Cohen D. Music therapy for children with autistic spectrum disorder and/or other neurodevelopmental disorders: a systematic review. Front Psychiatry 2021 Apr 9;12:643234 [FREE Full text] [doi: 10.3389/fpsyt.2021.643234] [Medline: 33897497]
27. Bégel V, Di Loreto I, Seilles A, Dalla Bella S. Music games: potential application and considerations for rhythmic training. Front Hum Neurosci 2017 May 29;11:273 [FREE Full text] [doi: 10.3389/fnhum.2017.00273] [Medline: 28611610]
28. Sújar A, Martín-Moratinos M, Rodrigo-Yanguas M, Bella-Fernández M, González-Tardón C, Delgado-Gómez D, et al. Developing serious video games to treat attention deficit hyperactivity disorder: tutorial guide. JMIR Serious Games 2022 Aug 01;10(3):e33884 [FREE Full text] [doi: 10.2196/33884] [Medline: 35916694]
29. Rodrigo-Yanguas M, Martín-Moratinos M, González-Tardón C, Sanchez-Sanchez F, Royuela A, Bella-Fernández M, et al. Effectiveness of a personalized, chess-based training serious video game in the treatment of adolescents and young adults with attention-deficit/hyperactivity disorder: randomized controlled trial. JMIR Serious Games 2023 Apr 24;11:e39874 [FREE Full text] [doi: 10.2196/39874] [Medline: 37093628]
30. Martin-Moratinos M, Bella-Fernández M, Rodrigo-Yanguas M, González-Tardón C, Sújar A, Li C, et al. Effectiveness of a serious video game (MOON) for attention deficit hyperactivity disorder: protocol for a randomized clinical trial. JMIR Res Protoc 2024 Feb 23;13:e53191 [FREE Full text] [doi: 10.2196/53191] [Medline: 38393773]

XSL•FO
RenderX

31. Villani D, Carissoli C, Triberti S, Marchetti A, Gilli G, Riva G. Videogames for emotion regulation: a systematic review. Games Health J 2018 Apr;7(2):85-99 [FREE Full text] [doi: 10.1089/g4h.2017.0108] [Medline: 29424555]
32. King D, Delfabbro P, Griffiths M. Video game structural characteristics: a new psychological taxonomy. Int J Ment Health Addiction 2009 Apr 7;8(1):90-106. [doi: 10.1007/S11469-009-9206-4]
33. Goodman R. The strengths and difficulties questionnaire: a research note. J Child Psychol Psychiatry 1997 Jul;38(5):581-586. [doi: 10.1111/j.1469-7610.1997.tb01545.x] [Medline: 9255702]
34. Goodman A, Goodman R. Strengths and difficulties questionnaire as a dimensional measure of child mental health. J Am Acad Child Adolesc Psychiatry 2009 Apr;48(4):400-403. [doi: 10.1097/chi.0b013e3181985068]
35. Goodman A, Lamping DL, Ploubidis GB. When to use broader internalising and externalising subscales instead of the hypothesised five subscales on the Strengths and Difficulties Questionnaire (SDQ): data from British parents, teachers and children. J Abnorm Child Psychol 2010 Nov 10;38(8):1179-1191. [doi: 10.1007/s10802-010-9434-x] [Medline: 20623175]
36. Karlsson P, Larm P, Svensson J, Raninen J. The factor structure of the strength and difficulties questionnaire in a national sample of Swedish adolescents: comparing 3 and 5-factor models. PLoS One 2022 Mar 14;17(3):e0265481 [FREE Full text] [doi: 10.1371/journal.pone.0265481] [Medline: 35286366]
37. Swanson JM, Schuck S, Porter MM, Carlson C, Hartman CA, Sergeant JA, et al. Categorical and dimensional definitions and evaluations of symptoms of ADHD: history of the SNAP and the SWAN rating scales. Int J Educ Psychol Assess 2012 Apr;10(1):51-70 [FREE Full text] [Medline: 26504617]
38. Conners CK, Sitarenios G, Parker JD, Epstein JN. The revised Conners' Parent Rating Scale (CPRS-R): factor structure, reliability, and criterion validity. J Abnorm Child Psychol 1998;26(4):257-268. [doi: 10.1037/T81067-000]
39. Busner J, Targum SD. The clinical global impressions scale: applying a research tool in clinical practice. Psychiatry (Edgmont) 2007 Jul;4(7):28-37 [FREE Full text] [Medline: 20526405]
40. Gioia GA, Isquith PK, Guy SC, Kenworthy L. BRIEF®2: Behavior rating inventory of executive function®, second edition. PAR. 2017. URL: https://www.parinc.com/products/pkey/24 [accessed 2023-09-11]
41. Conners CK. Conners' continuous performance test 3rd edition. Pearson. URL: https://tinyurl.com/256en5cm [accessed 2024-04-29]
42. Corsi PM. Human memory and the medial temporal region of the brain. McGill University. 1972. URL: https://escholarship.mcgill.ca/concern/theses/05741s554 [accessed 2023-03-16]
43. Lloret Irles D, Morell Gomis R, Marzo Campos JC, Tirado González S. [Spanish validation of Game Addiction Scale for Adolescents (GASA)]. Aten Primaria 2018 Jun;50(6):350-358 [FREE Full text] [doi: 10.1016/j.aprim.2017.03.015] [Medline: 28939247]
44. Lemmens JS, Valkenburg PM, Peter J. Development and validation of a game addiction scale for adolescents. Media Psychol 2009 Mar 05;12(1):77-95. [doi: 10.1080/15213260802669458]
45. Bruni O, Ottaviano S, Guidetti V, Romoli M, Innocenzi M, Cortesi F, et al. The Sleep Disturbance Scale for Children (SDSC). Construction and validation of an instrument to evaluate sleep disturbances in childhood and adolescence. J Sleep Res 1996 Dec;5(4):251-261. [doi: 10.1111/j.1365-2869.1996.00251.x] [Medline: 9065877]
46. Lingjaerde O, Ahlfors UG, Bech P, Dencker S, Elgen K. The UKU side effect rating scale. A new comprehensive rating scale for psychotropic drugs and a cross-sectional study of side effects in neuroleptic-treated patients. Acta Psychiatr Scand Suppl 1987 Aug 23;334(s334):1-100. [doi: 10.1111/j.1600-0447.1987.tb10566.x] [Medline: 2887090]
47. Rodrigo-Yanguas M, Martin-Moratinos M, Menendez-Garcia A, Gonzalez-Tardon C, Royuela A, Blasco-Fontecilla H. A virtual reality game (The Secret Trail of Moon) for treating attention-deficit/hyperactivity disorder: development and usability study. JMIR Serious Games 2021 Sep 01;9(3):e26824 [FREE Full text] [doi: 10.2196/26824] [Medline: 34468332]
48. Leaberry KD, Rosen PJ, Fogleman ND, Walerius DM, Slaughter KE. Physiological emotion regulation in children with ADHD with and without comorbid internalizing disorders: a preliminary study. J Psychopathol Behav Assess 2018 Jan 18;40(3):452-464. [doi: 10.1007/S10862-018-9644-Z]
49. Graziano PA, McNamara JP, Geffken GR, Reid AM. Differentiating co-occurring behavior problems in children with ADHD: patterns of emotional reactivity and executive functioning. J Atten Disord 2013 Apr 29;17(3):249-260. [doi: 10.1177/1087054711428741] [Medline: 22210802]
50. Faraone SV, Rostain AL, Blader J, Busch B, Childress AC, Connor DF, et al. Practitioner review: emotional dysregulation in attention-deficit/hyperactivity disorder - implications for clinical recognition and intervention. J Child Psychol Psychiatry 2019 Feb 06;60(2):133-150. [doi: 10.1111/jcpp.12899] [Medline: 29624671]
51. Choi E, Shin SH, Ryu JK, Jung KI, Kim SY, Park MH. Commercial video games and cognitive functions: video game genres and modulating factors of cognitive enhancement. Behav Brain Funct 2020 Feb 03;16(1):2 [FREE Full text] [doi: 10.1186/s12993-020-0165-z] [Medline: 32014027]
52. Bul KC, Franken IH, Van der Oord S, Kato PM, Danckaerts M, Vreeke LJ, et al. Development and user satisfaction of "Plan-It Commander," a serious game for children with ADHD. Games Health J 2015 Dec;4(6):502-512. [doi: 10.1089/g4h.2015.0021] [Medline: 26325247]
53. Shephard E, Zuccolo PF, Idrees I, Godoy PB, Salomone E, Ferrante C, et al. Systematic review and meta-analysis: the science of early-life precursors and interventions for attention-deficit/hyperactivity disorder. J Am Acad Child Adolesc Psychiatry 2022 Feb;61(2):187-226. [doi: 10.1016/j.jaac.2021.03.016] [Medline: 33864938]

XSL•FO
RenderX

54. Rojas-Jara C, Polanco-Carrasco R, Navarro-Castillo R, Faúndez-Castillo F, Chamorro-Gallardo M. “Game (not) Over”: a systematic review of video game disorder in adolescents. Rev Colomb Psicol 2022 Jul 18;31(2):45-64. [doi: 10.15446/rcp.v31n2.90741]

Abbreviations

ADHD: attention-deficit/hyperactivity disorder
ANCOVA: analysis of covariance
BRIEF-2: Behavior Rating Inventory Executive Function, version 2
CGI: Clinical Global Impression
CPRS: Conners Abbreviated Symptom Questionnaire
CPT-3: Conners’ Continuous Performance Test Third Edition
eCRF: electronic case report form
GASA: Game Addiction Scale for Adolescents
ITT: intention-to-treat
MOON: The Secret Trail of Moon
PP: per-protocol
REDCap: Research Electronic Data Capture
SDQ: Strengths and Difficulties Questionnaire
SNAP-IV: Swanson, Nolan, and Pelham Rating Scale
UKU: Udvalg für Kliniske Undersogelser
VR: virtual reality

Edited by T Leung, A Bustillo; submitted 02.04.24; peer-reviewed by D Baron, H Bhurawala, J Ferrer Costa; comments to author 16.08.24; revised version received 09.10.24; accepted 25.10.24; published 08.01.25.

Please cite as:
Martin-Moratinos M, Bella-Fernández M, Rodrigo-Yanguas M, González-Tardón C, Li C, Wang P, Royuela A, Lopez-Garcia P, Blasco-Fontecilla H
Effectiveness of a Virtual Reality Serious Video Game (The Secret Trail of Moon) for Emotional Regulation in Children With Attention-Deficit/Hyperactivity Disorder: Randomized Clinical Trial
JMIR Serious Games 2025;13:e59124
URL: https://games.jmir.org/2025/1/e59124
doi:10.2196/59124
PMID:

©Marina Martin-Moratinos, Marcos Bella-Fernández, María Rodrigo-Yanguas, Carlos González-Tardón, Chao Li, Ping Wang, Ana Royuela, Pilar Lopez-Garcia, Hilario Blasco-Fontecilla. Originally published in JMIR Serious Games (https://games.jmir.org), 08.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Gamified Web-Delivered Attentional Bias Modification Training for Adults With Chronic Pain: Randomized, Double-Blind, Placebo-Controlled Trial

Julie F Vermeir[1], PhD; Melanie J White[1], PhD; Daniel Johnson[2], PhD; Geert Crombez[3], PhD; Dimitri M L Van Ryckeghem[3,4,5], PhD

[1]School of Psychology and Counselling, Faculty of Health, Queensland University of Technology (QUT), Brisbane, Australia

[2]School of Computer Science, Faculty of Science, Queensland University of Technology (QUT), Brisbane, Australia

[3]Department of Experimental Clinical and Health Psychology, Ghent University, Ghent, Belgium

[4]Department of Clinical Psychological Science, Maastricht University, Maastricht, Netherlands

[5]Department of Behavioural and Cognitive Sciences, University of Luxembourg, Esch-sur-Alzette, Luxembourg

Corresponding Author:
Julie F Vermeir, PhD
School of Psychology and Counselling
Faculty of Health
Queensland University of Technology (QUT)
170 Victoria Park Road.
Brisbane, 4059
Australia
Phone: 61 731384714
Email: julie.vermeir@hdr.qut.edu.au

Abstract

Background: Attentional bias to pain-related information has been implicated in pain chronicity. To date, research investigating attentional bias modification training (ABMT) procedures in people with chronic pain has found variable success, perhaps because training paradigms are typically repetitive and monotonous, which could negatively affect engagement and adherence. Increasing engagement through the gamification (ie, the use of game elements) of ABMT may provide the opportunity to overcome some of these barriers. However, ABMT studies applied to the chronic pain field have not yet incorporated gamification elements.

Objective: This study aimed to investigate the effects of a gamified web-delivered ABMT intervention in a sample of adults with chronic pain via a randomized, double-blind, placebo-controlled trial.

Methods: A final sample of 129 adults with chronic musculoskeletal pain, recruited from clinical (hospital outpatient waiting list) and nonclinical (wider community) settings, were included in this randomized, double-blind, placebo-controlled, 3-arm trial. Participants were randomly assigned to complete 6 web-based sessions of nongamified standard ABMT (n=43), gamified ABMT (n=41), or a control condition (nongamified sham ABMT; n=45) over a period of 3 weeks. Active ABMT conditions trained attention away from pain-related words. The gamified task included a combination of 5 game elements. Participant outcomes were assessed before training, during training, immediately after training, and at 1-month follow-up. Primary outcomes included self-reported and behavioral engagement, pain intensity, and pain interference. Secondary outcomes included anxiety, depression, cognitive biases, and perceived improvement.

Results: Results of the linear mixed model analyses suggest that across all conditions, there was an overall small to medium decline in self-reported task-related engagement between sessions 1 and 2 (P<.001; Cohen d=0.257; 95% CI 0.13-0.39), sessions 1 and 3 (P<.001; Cohen d=0.368; 95% CI 0.23-0.50), sessions 1 and 4 (P<.001; Cohen d=0.473; 95% CI 0.34-0.61), sessions 1 and 5 (P<.001; Cohen d=0.488; 95% CI 0.35-0.63), and sessions 1 and 6 (P<.001; Cohen d=0.596; 95% CI 0.46-0.73). There was also an overall small decrease in depressive symptoms from baseline to posttraining assessment (P=.007; Cohen d=0.180; 95% CI 0.05-0.31) and in pain intensity (P=.008; Cohen d=0.180; 95% CI 0.05-0.31) and pain interference (P<.001; Cohen d=0.237; 95% CI 0.10-0.37) from baseline to follow-up assessment. However, no differential effects were observed over time between the 3 conditions on measures of engagement, pain intensity, pain interference, attentional bias, anxiety, depression, interpretation bias, or perceived improvement (all P values>.05).

XSL•FO
RenderX

Conclusions: These findings suggest that gamification, in this context, was not effective at enhancing engagement, and they do not support the widespread clinical use of web-delivered ABMT in treating individuals with chronic musculoskeletal pain. The implications of these findings are discussed, and future directions for research are suggested.

Trial Registration: Australian New Zealand Clinical Trials Registry (ANZCTR) ACTRN12620000803998; https://anzctr.org.au/ACTRN12620000803998.aspx

International Registered Report Identifier (IRRID): RR2-10.2196/32359

(JMIR Serious Games 2025;13:e50635) doi:10.2196/50635

KEYWORDS

chronic pain; cognition; attentional bias; gamification; motivation; randomized controlled trial; web-based intervention; pain management; digital intervention; digital health

Introduction

Background

Cognitive theories of pain posit that biases in attention contribute to the development and maintenance of chronic pain problems [1-3]. For example, the fear-avoidance model of chronic pain [4] assigns a causal role to attentional bias, such that individuals who have an attentional bias are more likely to have higher levels of pain and pain-related disability and, subsequently, be at a greater risk of developing chronic pain. Several reviews and meta-analyses have found that individuals with chronic pain exhibit an attentional bias toward pain-related words or pictures [5-7], and these biases have been positively associated with pain intensity, pain-related disability, and emotional distress [8,9]. This has led research to explore whether these attentional biases can be directly altered using a computer-based attentional bias modification training (ABMT) program and whether this modification results in concomitant changes in pain intensity and health outcomes associated with pain [10]. ABMT protocols typically use a modified version of the dot-probe task [11] to alter the bias by training individuals to disengage from pain-related cues and redirect attention to the competing non–pain-related (neutral) cues. It is suggested that repeated ABMT trials create a strategy for individuals to disengage from pain-related information and to facilitate attentional engagement toward other non–pain-related information. If this strategy is transferred to everyday life, it is expected to result in a reduction of pain, pain interference, and disability [10,12].

ABMT has been proposed as a promising tool for chronic pain based on its successful use for various conditions such as anxiety [13] and depression [14]. So far, however, results of pain ABMT interventions in pain (refer to the study by Van Ryckeghem et al [15] for an overview) and chronic pain samples [10,12,16-20] have been mixed, with most studies reporting at least some short- to medium-term therapeutic benefits of ABMT for clinically relevant pain outcomes [10,12,16,18,19].

A total of 2 factors that may contribute to these mixed findings are boredom and low motivation. ABMT procedures require participants to complete numerous trials over multiple sessions across several weeks [10,12,16-20] and typically use a basic layout. There is evidence that participants view the dot-probe task as monotonous, repetitive, and boring [21-23]. Consequently, this may lead to (temporal) low task engagement, low motivation to complete the sessions, and high dropout rates, which in turn may compromise the effectiveness of the intervention. Increasing engagement through the gamification of ABMT may provide the opportunity to overcome some of these barriers. Gamification refers to the use of digital game elements (eg, points) in nonentertainment settings [24]. Several reviews on gamified cognitive training tasks have reported that adding game-like elements to repetitive tasks improves motivation and engagement [25,26]. This is further supported by a recent study investigating gamification of interpretation bias modification for anxiety, which found that gamification could increase engagement and enjoyment [27]. However, it is still difficult to draw any firm conclusions regarding the effectiveness of gamified cognitive bias modification interventions [27,28], and more rigorously designed and theory-driven research is needed, particularly in the field of chronic pain.

Aims and Hypotheses

The aim of this study was to investigate the effects of a gamified web-delivered ABMT intervention using an empirically supported set of pain-related word stimuli on behavioral and self-reported engagement, pain intensity, pain interference, anxiety, depression, cognitive biases, and perceived improvement in a sample of adults with chronic musculoskeletal pain. We hypothesized that (1) the gamified ABMT condition would be more engaging than the nongamified (ie, standard ABMT and control) conditions; (2) both the standard and gamified ABMT conditions, compared to the control condition, would be more effective in improving outcomes of interest over time; and (3) these improvements would be greater in the gamified ABMT condition compared to the standard ABMT condition [29].

Methods

The trial protocol has been published elsewhere [29]. This study is conducted and described according to the CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) checklist (Multimedia Appendix 1) [30].

Ethical Considerations

The study was approved by the Human Research Ethics Committees of the Royal Brisbane and Women's Hospital (HREC/2020/QRBW/61743) and Queensland University of Technology (2000000395) and prospectively registered on the

XSL•FO
RenderX

Australian New Zealand Clinical Trials Registry (ACTRN12620000803998). All participants provided informed consent before their inclusion in the study, with the option to withdraw at any time without any consequences. Only approved study team members had access to participant data. The data were deidentified before analysis to safeguard participants' privacy. No incentives were offered to the participants.

Study Design and Setting

This study was a randomized, double-blind, placebo-controlled, 3-arm, parallel-group trial examining the efficacy of a gamified web-delivered ABMT for chronic pain. Participants were involved in the study for approximately 2 months, which included a 3-week intervention period, followed by a 1-month follow-up period. All training sessions were conducted via the internet at the participants' time and place of convenience using their own computers, and all outcome assessments (ie, at baseline, during training, immediately after training, and at 1-month posttraining) were self-assessed via web-based questionnaires and computerized tasks. Participants were randomly allocated to 1 of the 3 training conditions: nongamified standard ABMT, gamified ABMT, or nongamified sham ABMT (control). The control condition comprised a dot-probe paradigm without training direction (ie, the probe was located in the position of the pain-related vs non–pain-related words with equal probability), whereas the standard and gamified ABMT conditions aimed to train attention away from pain-related words (ie, the probe was located in the non–pain-related word location most of the time). Those in the control condition were offered the opportunity to do the standard ABMT at the conclusion of the study. Concomitant care (eg, rehabilitation program and pain medications) was permitted during the trial and was monitored through a pain treatment question that probed participants' pain treatments and frequency since the commencement of the study or previous assessment. Electronic informed consent (ie, e-consent) was obtained for all study participants.

Participants

Participants were recruited from a large Australian public hospital outpatient waitlist for pain management (clinical setting) and from the wider community (nonclinical setting). Individuals on the hospital outpatient waiting list were invited to participate through personalized mail correspondence, whereas individuals from the wider community were recruited through university electronic mailing lists, social media, and community channels (eg, Facebook advertising, Pain Australia, Chronic Pain Australia, and word of mouth). The inclusion and exclusion criteria for participants are listed in Textbox 1.

Textbox 1. Eligibility criteria.

Inclusion criteria

- Aged ≥18 years
- Experiencing chronic musculoskeletal pain, that is, pain in bones, joints, muscles, or related soft tissues (eg, rheumatoid arthritis pain, nonspecific back pain, or fibromyalgia pain)
- Meeting the criteria for chronic pain, that is, self-reported pain that lasts or recurs for >3 months [31]
- Having normal or corrected-to-normal (eg, glasses or contact lenses) vision

Exclusion criteria

- Not being a native English speaker or fluent in reading and writing English (as participants' reaction time to English words was used as an index of attentional bias to semantically related pain memory networks)
- Not having access to a desktop or laptop computer connected to reliable internet (as the trial was conducted on the web)
- Not being able or willing to provide informed consent to participate

As this was the first study to assess the effects of gamification techniques in a pain ABMT intervention, no prior effect size was available to guide sample size estimation during the study design phase. Therefore, a minimum sample size of 30 per training condition was planned on the basis that this exceeded the sample size determined by several similar pain ABMT and gamified training studies [12,20,32]. Considering attrition rates of previous trials in chronic pain treatment [33] and given the 1-month follow-up assessment, a dropout rate of approximately 30% was expected for this trial. Therefore, a total target sample size of 120 participants (40 participants per condition) was sought. We monitored attrition throughout the study, and recruitment ended when approximately 120 participants had completed at least 1 training session (ie, minimum threshold for exposure to ABMT).

Randomization, Allocation Concealment, and Blinding

After the baseline assessment, eligible consenting adults were randomly allocated to 1 of the 3 training conditions: standard ABMT, gamified ABMT, or control. Participant randomization was performed by an independent researcher with no involvement in the study using a computerized random number generator, Sealed Envelope (Sealed Envelope Ltd). A block randomization technique was used, allowing 6 participants at a time to be randomized in equal proportions to the 3 training arms. Participants and researchers were both blinded to the training condition to which participants were assigned. Participants were not provided information about the 3 training conditions but were informed that they might receive the intervention or complete a similar task (the control condition). This approach kept participants blinded to their allocation, as it would have been easy for them to recognize if they were in

XSL•FO
RenderX

the active gamified condition (particularly because the control condition did not include game elements). The outcome data were blinded, as assessments and training occurred on the web in the absence of the investigators. The first author (JFV) could access the training data to monitor the data collection process and was responsible to respond to participants who had questions or technical issues. However, it is unlikely that this caused problems of bias allocation or assessment because of the web-based nature of the study.

Study Program

Task Stimuli

Word stimuli were taken from a large pool of pain-related and non–pain-related linguistic stimuli, previously created and evaluated for use in chronic pain samples [34]. Specifically, for this study, we selected sensory and affective pain words that were rated as most related to chronic musculoskeletal pain and were categorized the fastest as pain-related by adults with self-reported chronic pain. As shown in Table 1, three sets of word stimuli were used: (1) 8 non–pain-related word pairs related to the categories of natural and man-made resources for the practice trials, (2) 8 pain-related and non–pain-related word pairs for the training trials, and (3) 8 pain-related and non–pain-related word pairs for the pretraining and posttraining assessment of attentional bias trials. Each pain-related word was matched with a non–pain-related word for length and frequency of word use in the English language, according to SubtlexUS [35]. Word stimuli in each set were not replicated in any other set, and each word stimulus was presented in a black 28-point uppercase Courier New font on a white background.

Table 1. Word pairs used in the practice, training, and assessment trials.

Practice set		Training set		Assessment set	
Non–pain-related	Non–pain-related	Pain-related	Non–pain-related	Pain-related	Non–pain-related
Cork	Fork	Aching (S[a])	Sponge	Agonizing (A[b])	Octagonal
Cotton	Carpet	Burning (S)	Streets	Cramping (S)	Cupboard
Goat	Lamp	Debilitating (A)	Strawberries	Cutting (S)	Crystal
Iron	Huts	Pain (S)	Bird	Excruciating (A)	Entertaining
Log	Pot	Stabbing (S)	Stocking	Hurting (S)	Trailer
Plant	Table	Suffering (A)	Newspaper	Sharp (S)	Plate
Rice	Bowl	Throbbing (S)	Resorting	Spasm (S)	Tiles
Waterfalls	Microwaves	Unbearable (A)	Metabolite	Tortured (A)	Currents

[a]S: sensory pain-related word.

[b]A: affective pain-related word.

Experimental Tasks

Overview

Tasks were programmed and presented using Inquisit 6.4 (Millisecond Software) on participants' internet-connected computers. To account for different screen sizes and ensure consistency in the display of word stimuli across participants, a calibration process was performed at the start of each session, asking participants to place a credit card on the screen and adjust the length of a horizontal line until it matched the width of the credit card.

The 3 training tasks were delivered using modified versions of the dot-probe task [11]. Each task began with a centered fixation cross for 500 milliseconds, after which a randomly selected stimulus pair of words (ie, a pain-related and a non–pain-related word) was presented horizontally on the screen for 500 milliseconds, with one word located at the top of the screen and the other at the bottom. Next, the paired words disappeared, and a probe (*p* or *q*) replaced the location of 1 of the stimuli. Participants were instructed to determine whether a *p* or a *q* had appeared and respond as quickly and as accurately as possible by pressing either the *P* or *Q* key on the keyboard, with the right and left index finger, respectively. The probe disappeared after 2500 milliseconds or sooner upon response. The intertrial interval was 500 milliseconds.

In addition, digit trials, that is, trials during which a random digit number between 1 and 9 replaced the fixation cross for a duration of 150 milliseconds, were included to ensure that participants' attention was directed to the center of the screen [20,36]. The intertrial interval was 1000 milliseconds after digit trials so that participants could reposition their index fingers on the *P* and *Q* keys. In the context of this study, trials in which the probe appeared in the location previously occupied by the pain-related word were considered congruent trials, and trials in which the probe appeared in the opposite location to that previously occupied by the pain-related word were considered incongruent trials.

Nongamified Standard ABMT

Each session started with a practice block of 17 trials (16 non–pain-related stimulus pairs and 1 digit trial) where participants received feedback after every correct (ie, *Correct!*) and erroneous (ie, *Incorrect!*) response. The training phase comprised 4 training blocks, each comprising 68 experimental trials (8 congruent trials, 56 incongruent trials, and 4 digit trials), totaling 272 experimental trials. The probe replaced

XSL•FO
RenderX

non–pain-related cues in 87.5% (224/256) of trials and pain cues in 12.5% (32/256) of trials, thereby directing attention away from pain-related words. This distribution was selected to reduce the obviousness of the probe contingency [37], and participants were not made aware of it. Word pairs were randomly presented in each of the 4 possible combinations (probe up and target down, probe down and target up, probe up and target up, and probe down and target down). Stimuli were presented in a randomized order across trials and participants, and trials were intermixed and randomly presented in 4 blocks, with a rest offered between each block of trials.

Nongamified Sham (Control) ABMT

The control and standard ABMT conditions were identical in all respects, except that in the control condition, the probe appeared with equal frequency in the position of the pain-related and non–pain-related words.

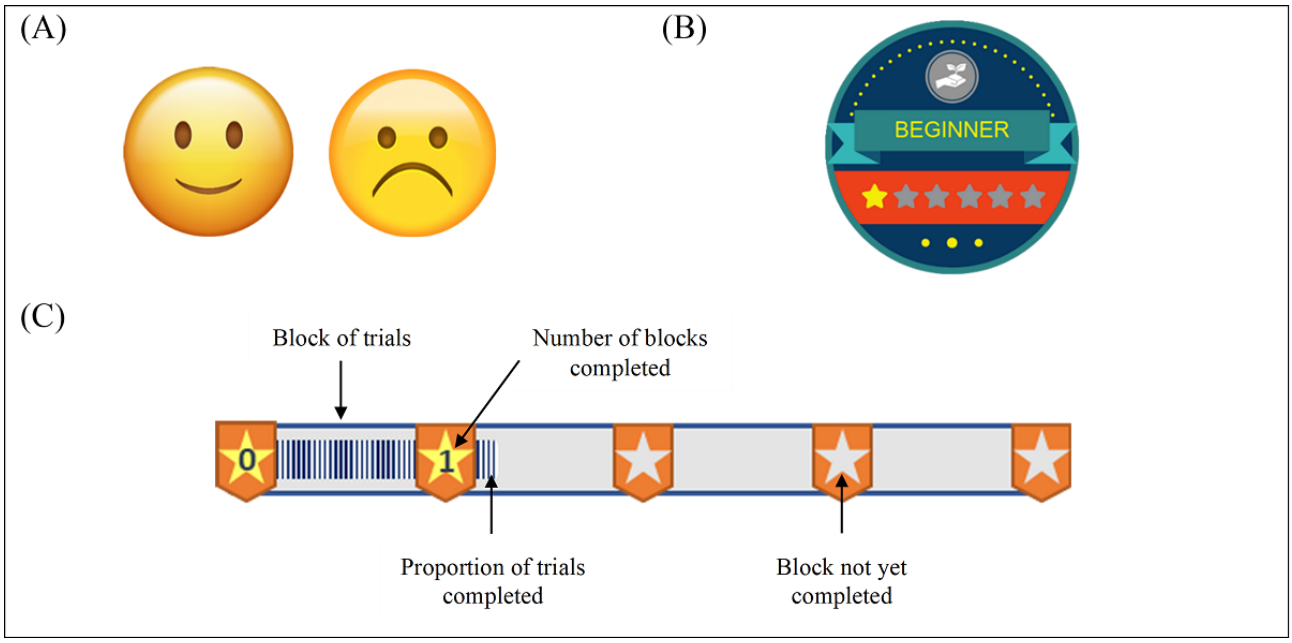
Gamified ABMT

Gamified ABMT was based on the standard ABMT but with the addition of game elements. Details regarding the development of the gamified task can be found in the published study protocol [29]. In brief, the development of gamified ABMT followed the Medical Research Council framework for complex interventions [38], incorporating theory, evidence from reviews, and expert input. The selection of game elements was guided by concepts of self-determination theory [39,40] and self-regulation [41] and informed by a recent qualitative and quantitative review assessing the effectiveness of gamification applied to cognitive training tasks [25]. According to self-determination theory [39,40], competence, relatedness, and autonomy are the 3 basic psychological needs that determine intrinsic motivation, sustained engagement, and psychological well-being. Self-regulation techniques such as goal setting and self-monitoring can also motivate users to engage and sustain in activities [42-44].

Specifically, a combination of 5 game elements was incorporated in the gamified task to facilitate participants' motivation and engagement in the ABMT procedure. These features were specifically chosen and implemented in a way designed to minimize cognitive disruption (ie, aiming to avoid interfering with the key cognitive mechanisms involved in the procedure). First, at the beginning of each training session, a clear gamified performance goal was set for the task: to earn as many points as possible and receive badges along the way (game element: clear gamified goal). Goals that are specific and reasonably challenging are the most effective at increasing motivation and task performance [45] and are likely to increase the satisfaction of the need for competence [46]. Second, during the practice phase, immediate gamified feedback was given (game element: feedback loops). For each correct trial, the word *Correct!* and a smiling emoticon appeared on the screen, whereas the word *Incorrect!* and a frowning emoticon occurred in every incorrect practice trial (Figure 1). This type of feedback has been shown to facilitate self-monitoring [43,47] and feelings of competence [46]. Third, during the training phase, a constantly visible progress bar at the top of the screen indicated the proportion of trials remaining in each block, and a written indicator reflected the number of blocks completed (game element: task-related progress; Figure 1). Such gamification features have been shown to facilitate self-tracking and motivate participants toward the attainment of goals [42,43] and fulfill their desire for competence [48]. Fourth, between blocks of trials, participants received feedback about their performance in the form of points (game element: rewards), calculated for each block of trials (1 point is earned for each correct trial; maximum of 68 points could be earned per block). To ensure the flow of training was uninterrupted, feedback was provided after each block of trials rather than after each trial. In addition, an auditory and visual reward in the form of a firework was incorporated into the task (game element: sound effect with reward). All the participants in the gamified ABMT condition experienced the fireworks after the first block of trials to ensure that everyone was exposed to the same type of game elements. However, for subsequent blocks of trials, only those who obtained at least 60% accuracy experienced the fireworks. Audio-visual rewards have been shown to emphasize feelings of competence [49], and the criterion of 60% involves an element of uncertainty that could further increase motivation. Finally, at the end of each training session, the participants were rewarded with a badge (game element: rewards). There were 6 different badges, and each badge had a number on it corresponding to the number of sessions completed (Figure 1). Collectible points and badges have been shown to facilitate goal setting [42,43] and satisfy the needs for competence, autonomy, and relatedness [46].

XSL•FO
RenderX

Figure 1. Sample of game elements used in the gamified task: (A) smiling (left) and frowning (right) emoticons received during practice trials for correct and incorrect responses, respectively; (B) badge earned at the end of the first training session; and (C) progress bar representing the proportion of trials completed in each block and a written indicator reflecting the number of blocks completed.



Outcomes and Measures

Data were collected via the internet using the web-based system Qualtrics (Qualtrics International Inc) for survey responses and Inquisit 6.4 (Millisecond Software) for task data and self-reported engagement responses. Participants' outcomes were assessed before beginning training (baseline), during training (for self-reported task-related engagement), immediately after completion of training (posttraining), and 1 month after the last training session (follow-up).

Baseline Information

At baseline, demographic information (eg, age); health data (eg, general health status); pain experience details (eg, duration of primary pain condition); and data on type of computer, keyboard, and screen size used were collected. To provide further information about the participants' chronic pain severity, the Graded Chronic Pain Scale (GCPS) [50] was administered. The GCPS is a 7-item self-report instrument that assesses pain intensity and pain-related disability in the past 6 months. All items, except for the number of days disabled, are scored on an 11-point Likert scale, with responses ranging from 0 to 10. Subscale scores for the pain intensity and pain-related dimensions are combined to calculate a chronic pain grade that enables individuals to be classified into pain severity grades ranging from grade 0 (*no pain problem*) to grade 4 (*high disability-high intensity*). The reliability and validity of the GCPS have been demonstrated in previous studies [50-52].

Primary Outcome Measures

Engagement

A total of 2 self-report measures were used to assess participants' experiences of engagement. Task-related engagement was measured after each training session with a single-item question: How engaging was this session? This item was rated on an 11-point Likert scale, ranging from 0 (*not at all*) to 10 (*very much*). Higher scores represent greater engagement. Task-related interest and enjoyment was assessed with the Intrinsic Motivation Inventory (IMI) Interest and Enjoyment subscale [53-55], which comprises 7 items. Each item is scored on a 7-point Likert scale that ranges from 1 (*not at all*) to 7 (*very true*), with higher scores representing higher levels of interest and enjoyment. The reliability and validity of this subscale have been demonstrated in previous studies [56,57].

A total of 2 behavioral measures were used to assess participants' engagement: nonuse intervention attrition [58], defined as the number of participants who discontinued using the intervention at each training session, and completion rates, defined as the number of training sessions (out of 6) that each participant completed during the intervention period.

Pain Intensity

The Patient-Reported Outcomes Measurement Information System (PROMIS) Pain Intensity Short Form 3a (version 1.0; 3 items) [59] was used to measure pain intensity. The first 2 items assess pain intensity over the last 7 days (average and worst pain), and the last item assesses pain intensity *right now*, each scored using a 5-point Likert scale, with responses ranging from 1 (*had no pain* or *no pain*) to 5 (*very severe*). As recommended by PROMIS, the Pain Intensity scale (version 1.0) was rescored into T-scores by using the free web-based HealthMeasures Scoring Service [59]. Higher T-scores represent worse pain. This measure has shown to be valid for assessing pain intensity in various settings [60].

Pain Interference

The PROMIS Pain Interference Short Form 8a (version 1.0; 8 items) [59] was used to assess the impact of pain on daily life over the last 7 days using a 5-point Likert scale, with responses ranging from 1 (*not at all*) to 5 (*very much*). As recommended by PROMIS, the Pain Interference scale (version 1.0) scale was

XSL•FO
RenderX

rescored into T-scores using the free web-based HealthMeasures Scoring Service [59]. Higher T-scores represent greater pain interference. This measure has been assessed and validated in both general and clinical populations [60,61].

Secondary Outcome Measures

Measure of Attentional Bias for Pain

The dot-probe paradigm [62] was used to measure pain-related attentional biases. This task is similar to the nongamified sham (control) ABMT task, except for the total number of trials. The assessment phase included a practice block of 17 trials (16 non–pain-related stimulus pairs and 1 digit trial) and 2 assessment blocks of 68 trials (32 congruent trials, 32 incongruent trials, and 4 digit trials), totaling 136 assessment trials. Stimuli were presented in a randomized order across trials and participants, and trials were intermixed and randomly presented in 2 blocks, with a rest offered between the blocks.

Anxiety and Depression

A total of 2 PROMIS measures comprising PROMIS Anxiety 8a (version 1.0; 8 items) and PROMIS Depression 8b (version 1.0; 8 items) [59] were used to assess negative affect over the last 7 days using a 5-point Likert scale, with responses ranging from 1 (*never*) to 5 (*always*). The scales were scored using the free web-based HealthMeasures Scoring Service [59]. Higher T-scores represent greater symptoms of anxiety or depression. These PROMIS measures have demonstrated excellent psychometric properties in both population-based [63] and clinical samples [64,65].

Perceived Improvement

The Patient Global Impression of Change (PGIC) scale [66] was used to assess participants' perception of overall pain-related improvement following training. It is composed of a single item rated on a 7-point Likert scale, ranging from 1 (*very much improved*) to 7 (*very much worse*). For descriptive purposes, participants were classified into 3 categories according to the PGIC score: disease deterioration (1-3 points), stable disease (4 points), or disease improvement (5-7 points) since the start of the program [67]. The PGIC is widely used in chronic pain research [67,68].

Measure of Interpretation Bias for Pain

An adapted version of the computerized interpretation bias task [69] was used to measure pain-related interpretation biases, which contains 16 incomplete vignettes that describe 8 ambiguous situations relating to bodily threat or pain and 8 ambiguous situations relating to social evaluations. Vignettes were adapted to reflect events that may occur in the workplace, home, or during an adult's everyday life. Participants are asked to rate how likely each ending came to their mind on a scale of 1 (*does not come to mind*) to 5 (*definitely comes to mind*) and to select the interpretation (word) that first came to their mind. Next, participants are presented with the same scenarios again; however, this time, they are asked to rate the likelihood that each resolution would actually happen in that situation on a scale of 1 (*not likely*) to 5 (*very likely*) and to select the word that they believe is most likely to end the sentence. All items and interpretations were presented in a fixed random order.

Similar to previous reports [69-71], this study only used the rating data for interpretation belief (ie, belief that the interpretation is likely to be true); however, all other data are available upon request from the authors. An interpretation bias score for each domain was calculated by subtracting the mean ratings of benign endings from the mean ratings of negative endings, with higher scores indicating a higher tendency to believe that negative interpretations are likely to be true. Previous studies using this task have found evidence of interpretation biases for pain in individuals with chronic pain, particularly for interpretation belief [70].

Exploratory and Other Measures

Exploratory Measures

As detailed in the preregistered protocol, further measures were collected for exploratory purposes, all of which are reported in more detail in Multimedia Appendix 2 [20,29,72-77]. These measures included the Attentional Control Scale (ACS) [72], the Behavioral Inhibition System and Behavioral Activation System scales [73], and the Pain Catastrophizing Scale [74].

Pain Treatment Information

The posttraining and 1-month follow-up assessments included a question that probed participants' use of pain treatments and frequency of health care use since the commencement of the study or previous assessment.

Manipulation Check

The posttraining assessment included a manipulation check question asking participants whether they believed they had received the intervention or sham training (ie, no intervention).

Validity Check

Instructional questions (eg, "Please select 5=Always") were included in the baseline, posttraining, and 1-month follow-up assessments to identify careless responding patterns [78]. Participants were excluded if they answered all instructional questions incorrectly.

Procedure

Interested participants first provided informed consent before being taken to the screening questions, and then to the baseline assessment (approximately 35 min), consisting of questions relating to demographic characteristics; general health status; current mental health status; pain experience information; as well as the PROMIS Pain Intensity 3a, PROMIS Pain Interference 8a, PROMIS Anxiety 8a, PROMIS Depression 8b, ACS, Behavioral Inhibition System and Behavioral Activation System scales, and Pain Catastrophizing Scale. At the end of the assessment, participants selected their preferred days for training (Monday and Thursday or Tuesday and Friday) and provided an email address so that the research team could send links to the training sessions. Participants who completed the baseline assessment were randomized into 1 of 3 conditions (standard ABMT, gamified ABMT, or control) and invited by email to start their first training session. This email included a web link to the appropriate version of the training as well as instructions on how to download and install the software used to run the program.

XSL•FO
RenderX

Participants performed the training sessions on the web at their time and place of convenience, twice a week on a separate pair of days (Monday and Thursday or Tuesday and Friday) for 3 consecutive weeks, totaling 6 training sessions. This dosage was based on previous pain ABMT literature [10,12,16], which has shown positive training effects for dosage ranging between 4 and 8 sessions. After each training session, participants were asked to rate their engagement with the task. The first and final sessions took approximately 30 minutes, as they included cognitive assessment measures (ie, the dot-probe and interpretation bias assessment tasks were administered at the beginning, before training at session 1, and at the end, after the last training block in session 6), whereas sessions 2 to 5 took approximately 15 minutes to complete. Participants were asked to complete the sessions within 24 hours of receiving a web link on a computer with a proper keyboard and to create a quiet and private environment free from distractions for at least 30 minutes. Each session started with the same instructions, similar to those used in previous research (eg, [20]). A combination of SMS text message and email message reminders were sent to those who did not complete the scheduled session within 24 hours of receiving the web link. Participants were allowed to skip training sessions but were excluded from the analyses if they did not successfully complete at least 1 session (ie, minimum threshold for exposure to ABMT). They could contact the first author (JFV) by email or phone if they had any questions or technical problems.

Immediately after completion of the final training session, participants were automatically invited to the posttraining assessment (approximately 15 min), consisting of questions about pain treatments, a manipulation check, and the PROMIS Pain Intensity 3a, PROMIS Pain Interference 8a, PROMIS Anxiety 8a, PROMIS Depression 8b, ACS, IMI (interest and enjoyment), and PGIC. Participants were encouraged to complete the posttraining assessment regardless of whether they completed all training sessions. Finally, 1 month after the end of the last training session, all participants were sent a link to complete the follow-up assessment (approximately 10 min), which included the same questions as that of the posttraining assessment, except for the manipulation check, the ACS, and IMI (interest and enjoyment). Participants who failed to complete the follow-up assessment received up to 2 emails or SMS text message reminders.

Data Preparation and Data Analysis

Statistical analyses were performed using SPSS (version 27.0; IBM Corp). Analyses included all randomized participants who successfully completed at least 1 training session. To manage missing data, linear mixed model analyses were used (where appropriate), as it allows the inclusion of all available data. Missing PROMIS measures data were handled according to the recommendations in the scoring manual, using the HealthMeasures Scoring Service [59]. Concomitant care received during the study was categorized into pain treatment received (*yes* or *no*) and use of medication (*yes* or *no*). Significance for all statistical tests were set at $P<.05$ (2-tailed). Effect sizes were presented by the test's most appropriate effect size [79]. For the linear mixed models, a standardized Cohen *d* was calculated from the estimated marginal means tables [80]. No analyses were performed until recruitment and data collection were completed.

Participant characteristics were analyzed using descriptive statistics, and chi-square test, Fisher-Freeman-Halton exact test, 1-way ANOVA, and nonparametric Kruskal-Wallis test were used for group comparisons. To prepare reaction time (RT) data for analyses, practice trials, digit trials, incorrect trials, and outliers (ie, responses <200 or >1000 ms) were excluded from the calculation of mean RTs [20,36]. In addition, the data of 2 participants at pretraining assessment and 1 participant at posttraining assessment were excluded from the attentional bias analyses because they committed errors on >30% of the trials, showing suboptimal dot-probe task performance. Furthermore, the data of another 2 participants at pretraining assessment had to be discarded due to having a very high percentage of outliers (>96.8%). The mean percentage of dot-probe errors made by participants was 3.2%, and the mean percentage of outliers was 5.6%. An attentional bias index was calculated using the following formula: (tupl–tlpl)+(tlpu–tupu)/2, where t=target (pain) stimulus, p=probe, u=upper location, and l=lower location. Positive scores indicated an attentional bias toward pain-related stimuli, whereas negative scores reflected an attentional bias toward non–pain-related stimuli (or away from pain-related stimuli). To determine whether pain-related attentional biases were present at baseline, 1-sample 2-tailed *t* tests (vs 0) were performed on the pretraining attentional bias index scores for each condition separately.

A series of linear mixed model analyses were conducted to examine changes over time in symptoms (ie, pain intensity, pain interference, anxiety, and depression), self-reported task-related engagement, cognitive biases (ie, attentional bias and interpretation bias), and perceived improvement in the different training conditions. The categories baseline, control, married or in a relationship, employed, and session 1 were used as reference categories for time, training condition, marital status, work status, and session, respectively. A backward modeling approach was used to build the most parsimonious model to test the hypotheses [81], using the Akaike information criterion (AIC) and Bayesian information criterion (BIC) to identify the most appropriate model. Specifically, a 3-step model-building procedure was used to identify the best-fit model. First (model 1), we constructed a model that included the main effects and covariates. Second (model 2), we removed from the model covariates that were nonsignificant to see if it improved the fit of the model. Third (model 3), we added the interaction terms to the previously best-fit model (model 1 or model 2) and compared the 3 models on their fit to the data using the AIC and BIC. All models incorporated a random intercept for participants and used the maximum likelihood estimator. Finally, sensitivity analyses were conducted on all primary analyses without controlling for the covariates. Comparisons of models for each primary and secondary outcome variable are available in Multimedia Appendix 3.

To analyze the impact of gamification on interest and enjoyment, a 1-way analysis of covariance in a general linear model was performed (no data were missing). Regarding behavioral engagement, Kaplan-Meier survival curves [82] were calculated to assess the time at which attrition occurred in each training

XSL•FO
RenderX

condition and compared statistically using a log-rank test. The number of sessions completed was the time variable, and the event variable was specified as the moment of ceasing participation. Participants were classified as noncompleters if they did not complete all 6 training sessions. In addition, a 1-way analysis of covariance in a general linear model was performed to determine whether there were differences in the mean number of sessions completed between the training conditions. Pearson correlations assessed the relationship between changes in attentional bias magnitude from pretraining to posttraining assessment and changes in scores on symptom measures (ie, pain intensity, pain interference, anxiety, and depression). Change in attentional bias was calculated by subtracting the attentional bias score in the pretraining session from the attentional bias score in the posttraining session.

Finally, exploratory analyses were conducted to address additional questions. These included the role of engagement metrics (ie, number of training sessions completed) and individual differences (ie, attentional control, pain-related worrying, personality characteristics, and recruitment setting) in the impact of training conditions on pain intensity and pain interference. Models for each exploratory analysis are available in Multimedia Appendix 2.

Results

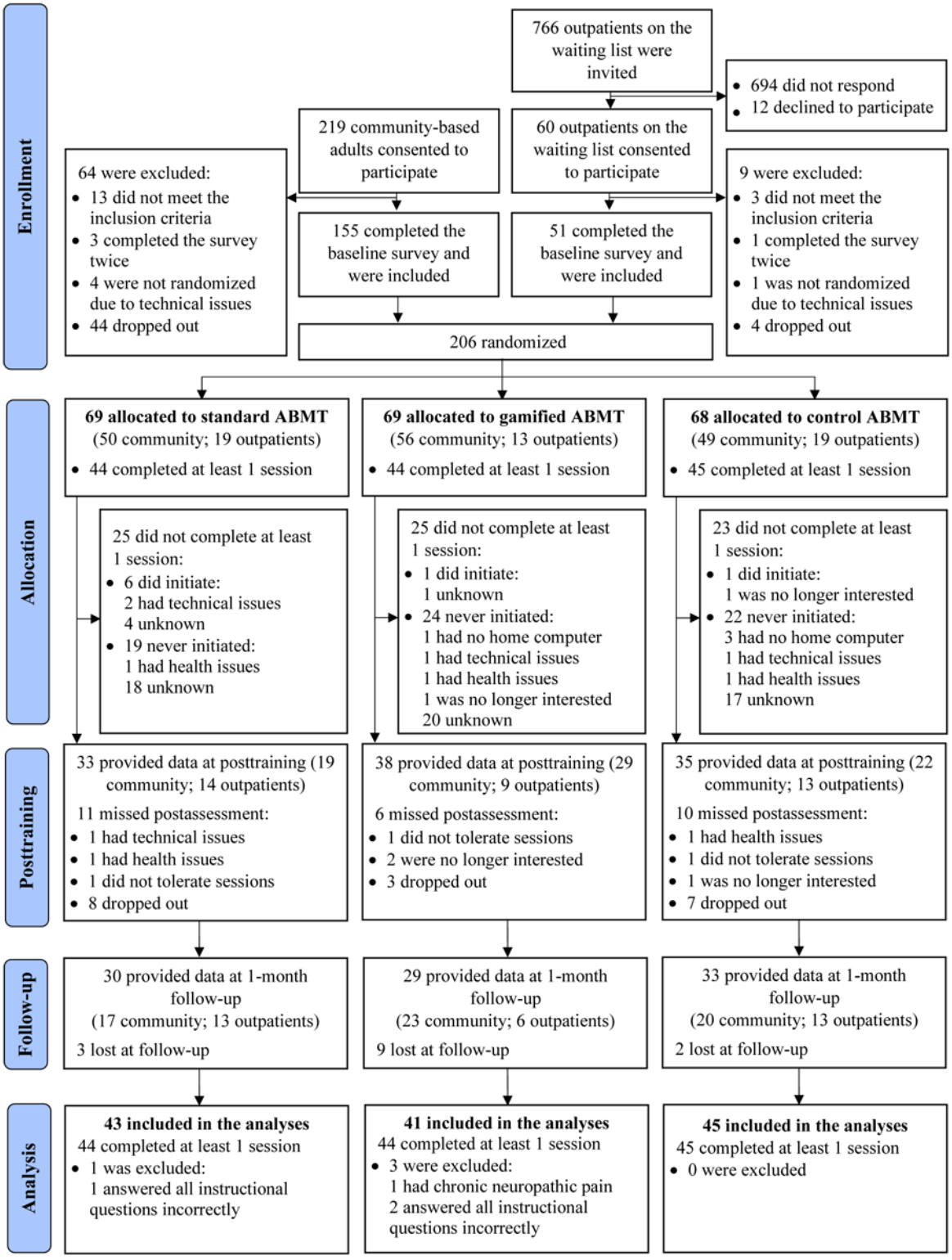
Participant Flow

The first participant for this study was enrolled in August 2021, and the final follow-up assessment was completed in June 2022.

Figure 2 shows the CONSORT (Consolidated Standards of Reporting Trials) flow diagram for the study. Of the 766 outpatients on the waiting list that were invited to participate, 60 (7.8% response rate) consented to participate, and of those 60, 51 (85%) were included and randomized to conditions. Of the 219 community-based adults that consented to participate, 155 (70.8%) were included and randomized to conditions. In total, 206 participants were randomized into the standard ABMT (n=69, 33.5%), gamified ABMT (n=69, 33.5%), or control condition (n=68, 33%). Of these 206 participants, 106 (51.5%) completed the posttraining survey (n=33, 16% standard ABMT; n=38, 18.5% gamified ABMT; and n=35, 17% control) and 92 (44.7%) completed the 1-month follow-up survey (n=30, 14.6% standard ABMT; n=29, 14.1% gamified ABMT; and n=33, 16% control). Of the 206 participants who were randomized, 73 (35.4%) were excluded from all analyses due to not successfully completing at least 1 training session, which was the minimum threshold for exposure to ABMT (n=25, 12.1% standard ABMT; n=25, 12.1% gamified ABMT; and n=23, 11.2% control), 3 (1.5%) were excluded for answering all the instructional questions incorrectly (n=1, 0.5% standard ABMT and n=2, 1% gamified ABMT), and 1 (0.5%) was excluded for having only chronic neuropathic pain (gamified ABMT). The final sample size included in the analysis was 129 adults with chronic musculoskeletal pain (n=43, 33.3% standard ABMT; n=41, 31.8% gamified ABMT; and n=45, 34.9% control).

XSL•FO
RenderX

Figure 2. CONSORT (Consolidated Standards of Reporting Trials) diagram of flow of participants through the study. ABMT: attentional bias modification training.



Sample Characteristics and Baseline Group Differences

Descriptive statistics for baseline key characteristics and outcome measures are presented in Tables 2 and 3, respectively. The mean age of the 129 participants was 49.49 (SD 12.50) years. The participants were primarily female, born in Australia, tertiary educated (ie, university, college, or posthigh school qualifications), married or in a relationship, employed, and right-handed. Most participants (59/129, 45.7%) reported that their general health was *fair*, but that they had mental health problems (eg, anxiety). Nearly all participants (124/129, 96.1%) in the sample had received a diagnosis for their pain. Chronic lower back pain was the most common pain problem. The mean duration of pain was 13.56 (SD 10.58) years. On the GCPS, participants' mean current pain was 5.81 (SD 1.94), and calculation of the pain grade showed that most participants

XSL•FO
RenderX

(69/129, 53.5%) were classified in grade 4 (*high disability-high intensity*). Most participants had consulted a physician for their pain in the past 4 weeks, were taking regular medication (ie, nonprescription and prescription medicines) to alleviate their pain, and were receiving pain treatment (eg, physical interventions).

Table 2. Baseline key characteristics of the sample by training condition (N=129).

Variable	Full sample	By training condition			*P* value
		Standard ABMT[a] (n=43)	Gamified ABMT (n=41)	Control (n=45)	
Key demographics					
Age (y), mean (SD)	49.49 (12.50)	48.09 (12.51)	47.78 (11.88)	52.38 (12.79)	.16
Gender (female), n (%)	104 (80.6)	35 (81.4)	36 (87.8)	33 (73.3)	.44
Country of birth (Australia), n (%)	105 (81.4)	34 (79.1)	32 (78)	39 (86.7)	.53
Education level (tertiary), n (%)	101 (78.3)	37 (86)	37 (90.2)	27 (60)	.003[b]
Relationship status (married or in a relationship), n (%)	75 (58.1)	28 (65.1)	16 (39)	31 (68.9)	.01[b]
Employment status (employed), n (%)	55 (42.6)	18 (41.9)	27 (65.9)	10 (22.2)	<.001[b]
Handedness (right-handed), n (%)	114 (88.4)	39 (90.7)	37 (90.2)	38 (84.4)	.88
General health status, n (%)					.45
Very good	3 (2.3)	1 (2.3)	1 (2.4)	1 (2.2)	
Good	38 (29.5)	10 (23.3)	15 (36.6)	13 (28.9)	
Fair	59 (45.7)	18 (41.9)	18 (43.9)	23 (51.1)	
Bad	23 (17.8)	11 (25.5)	7 (17.1)	5 (11.1)	
Very bad	6 (4.7)	3 (7)	0 (0)	3 (6.7)	
Mental health condition (yes), n (%)	71 (55)	26 (60.5)	20 (48.8)	25 (55.6)	.56
Diagnosis (yes), n (%)	124 (96.1)	41 (95.3)	38 (92.7)	45 (100)	.16
Type of pain (chronic lower back pain), n (%)	92 (71.3)	34 (79.1)	25 (61)	33 (73.3)	.19
Primary pain site (lower back), n (%)	61 (47.3)	22 (51.2)	16 (39)	23 (51.1)	.78
Duration of chronic pain[c] (y), mean (SD)	13.56 (10.58)	14.02 (9.87)	12.91 (10.48)	13.70 (11.50)	.41
GCPS[d] pain intensity, mean (SD)	5.81 (1.94)	6.14 (1.77)	5.37 (2.13)	5.89 (1.87)	.18
GCPS scales, n (%)					.60
Grade 0: no pain problem	0 (0)	0 (0)	0 (0)	0 (0)	
Grade 1: low disability-low intensity	8 (6.2)	4 (9.3)	2 (4.9)	2 (4.4)	
Grade 2: low disability-high intensity	20 (15.5)	4 (9.3)	10 (24.4)	6 (13.3)	
Grade 3: high disability-low intensity	32 (24.8)	11 (25.6)	10 (24.4)	11 (24.4)	
Grade 4: high disability-high intensity	69 (53.5)	24 (55.8)	19 (46.3)	26 (57.8)	
Medical visit in the past 4 week (yes), n (%)	95 (73.6)	33 (76.7)	28 (68.3)	34 (75.6)	.64
Use of medication (yes), n (%)	126 (97.7)	41 (95.3)	41 (100)	44 (97.8)	.65
Pain treatment (yes), n (%)	114 (88.4)	38 (88.4)	38 (92.7)	38 (84.4)	.55
Computer type (laptop), n (%)	79 (61.2)	28 (65.1)	23 (56.1)	28 (62.2)	.65
Computer screen (≤50.8 cm), n (%)	64 (49.6)	20 (46.5)	18 (43.9)	26 (57.8)	.16

[a]ABMT: attentional bias modification training.

[b]Statistical significance: $P<.05$, 2-tailed.

[c]One missing value in the gamified ABMT condition.

[d]GCPS: Graded Chronic Pain Scale.

XSL•FO
RenderX

Table 3. Summary statistics on outcome measures by training condition and assessment time points.

Variable	Training condition						*P* value[a]
	Standard ABMT[b] (n=43)		Gamified ABMT (n=41)		Control (n=45)		
	Values, n (%)	Values, mean (SD)	Values, n (%)	Values, mean (SD)	Values, n (%)	Values, mean (SD)	
Primary outcomes							
Task-related engagement							
Session 1	37 (86)	5.81 (3.10)	38 (93)	6.03 (2.95)	43 (96)	6.14 (2.74)	—[c]
Session 2	34 (79)	4.79 (3.06)	38 (93)	5.79 (2.92)	38 (84)	5.95 (3.11)	—
Session 3	29 (67)	4.83 (3.34)	35 (85)	5.49 (2.96)	36 (80)	5.47 (3.05)	—
Session 4	28 (65)	5.14 (3.34)	30 (73)	5.00 (2.91)	34 (76)	5.59 (2.89)	—
Session 5	27 (63)	4.96 (3.35)	28 (68)	4.46 (2.85)	30 (67)	5.90 (2.77)	—
Session 6	26 (61)	4.81 (3.18)	31 (76)	4.23 (3.01)	31 (69)	5.32 (3.05)	—
Number of sessions completed	43 (100)	4.21 (2.01)	41 (100)	4.88 (1.62)	45 (100)	4.71 (1.80)	—
IMI[d]	32 (74)	3.45 (1.50)	35 (85)	2.89 (1.59)	35 (78)	3.14 (1.21)	—
PROMIS[e] pain intensity							
Baseline	43 (100)	66.06 (7.23)	41 (100)	65.51 (7.00)	45 (100)	65.86 (6.40)	.94
Posttraining	32 (74)	67.23 (8.59)	35 (85)	63.25 (7.28)	35 (78)	65.34 (6.66)	—
Follow-up	29 (67)	65.33 (8.64)	28 (68)	61.31 (6.48)	33 (73)	64.04 (8.93)	—
PROMIS pain interference							
Baseline	43 (100)	66.00 (7.53)	41 (100)	63.73 (6.23)	45 (100)	66.88 (6.42)	.09
Posttraining	32 (74)	64.77 (7.40)	35 (85)	62.85 (7.52)	35 (78)	66.01 (6.72)	—
Follow-up	29 (67)	64.05 (7.64)	28 (68)	62.70 (6.28)	33 (73)	62.85 (8.50)	—
Secondary outcomes							
Attentional bias index							
Baseline	40 (93)	–0.58 (18.61)	41 (100)	–3.72 (24.41)	44 (98)	1.30 (20.23)	.55
Posttraining	26 (61)	2.12 (15.08)	30 (73)	3.98 (16.90)	31 (69)	–1.65 (20.52)	—
PROMIS anxiety							
Baseline	43 (100)	59.97 (11.39)	41 (100)	57.86 (9.05)	45 (100)	61.24 (8.03)	.26
Posttraining	32 (74)	61.00 (12.11)	35 (85)	58.78 (9.47)	35 (78)	59.15 (8.33)	—
Follow-up	29 (67)	59.55 (10.13)	28 (68)	56.80 (9.71)	33 (73)	59.16 (8.49)	—
PROMIS depression							
Baseline	43 (100)	61.28 (11.05)	41 (100)	56.97 (10.09)	45 (100)	61.92 (8.07)	.04[g]
Posttraining	32 (74)	61.28 (11.17)	35 (85)	55.68 (10.40)	35 (78)	58.64 (8.53)	—
Follow-up	29 (67)	61.20 (9.39)	28 (68)	55.20 (10.19)	33 (73)	60.05 (10.58)	—
PGIC[f]							
Posttraining							
Illness improvement	6 (14)	—	6 (15)	—	4 (9)	—	—
Stable illness	20 (47)	—	23 (56)	—	29 (64)	—	—
Illness deterioration	6 (14)	—	6 (15)	—	2 (4)	—	—
Follow-up							
Disease improvement	7 (16)	—	6 (15)	—	11 (24)	—	—

XSL•FO
RenderX

Variable	Training condition						*P* value[a]
	Standard ABMT[b] (n=43)		Gamified ABMT (n=41)		Control (n=45)		
	Values, n (%)	Values, mean (SD)	Values, n (%)	Values, mean (SD)	Values, n (%)	Values, mean (SD)	
Stable disease	17 (40)	—	18 (44)	—	16 (36)	—	—
Disease deterioration	5 (12)	—	4 (10)	—	6 (13)	—	—
Interpretation bias							
Baseline							
Belief: health and pain	43 (100)	0.41 (1.12)	41 (100)	–0.12 (0.93)	45 (100)	0.04 (0.94)	.07
Belief: social	43 (100)	–0.66 (1.44)	41 (100)	–1.30 (1.15)	45 (100)	–0.94 (1.07)	.06
Posttraining							
Belief: health and pain	27 (63)	0.51 (0.96)	30 (73)	–0.16 (0.82)	31 (69)	–0.08 (0.77)	—
Belief: social	27 (63)	–0.66 (1.29)	30 (73)	–0.98 (1.11)	31 (69)	–0.92 (0.86)	—

[a]*P* values calculated for baseline differences only.

[b]ABMT: attentional bias modification training.

[c]Not applicable as differences between groups are examined with linear mixed model analyses and are reported in Tables 4 and 5.

[d]IMI: Intrinsic Motivation Inventory, subscales interest and enjoyment.

[e]PROMIS: Patient-Reported Outcomes Measurement Information System.

[f]PGIC: Patient Global Impression of Change.

[g]Statistical significance *P*<.05, 2-tailed

A series of analyses were conducted to determine whether there were any baseline differences between the ABMT conditions on key characteristics (Table 2) and outcome measures (Table 3). The training conditions did differ significantly at baseline on marital status, educational level, employment status, and depressive symptomology. Specifically, the participants in the gamified ABMT condition reported significantly lower scores for depression ($F_{2,126}$=3.19; *P*=.04) and were significantly less likely to be married or in a relationship (χ^2_4=12.7; *P*=.01). Furthermore, the participants in the control condition were significantly less likely to have a tertiary education (Fisher-Freeman-Halton exact test=13.43; *P*=.003) and to be working (Fisher-Freeman-Halton exact test=23.01; *P*<.001). The baseline covariates were accounted for in the analyses as per the statistical plan.

Primary Outcomes

Engagement

Table 4 displays the results of the linear mixed model for self-reported task-related engagement. There was a significant main effect of session between sessions 1 and 2 (*P*<.001), sessions 1 and 3 (*P*<.001), sessions 1 and 4 (*P*<.001), sessions 1 and 5 (*P*<.001), and sessions 1 and 6 (*P*<.001), with small to medium effect sizes, reflecting an overall decrease in task-related engagement ratings after session 1 over the 6 sessions (session 1: mean 6.00, SD 2.90; session 2: mean 5.54, SD 3.04; session 3: mean 5.29, SD 3.09; session 4: mean 5.26, SD 3.02; session 5: mean 5.13, SD 3.02; and session 6: mean 4.78, SD 3.07). However, there was no significant main effect of training condition (*P* values>.05; AIC=2384.19; BIC=2449.96). The model that included the interaction effects (Table S1 in Multimedia Appendix 3) did not fit the data better than the previous model, and all interactions were nonsignificant for session×condition (*P* values>.05; AIC=2396.60; BIC=2506.23). The 1-way analysis of covariance revealed that there was no significant difference in the mean interest and enjoyment (IMI) scores between training conditions ($F_{2,91}$=0.94; *P*=.40; η^2_p=0.020).

On average, participants completed 4.60 (SD 1.83) out of 6 training sessions, which did not differ significantly between conditions ($F_{2,118}$=1.24; *P*=.29; η^2_p=0.021). The attrition pattern of the 3-week intervention period is presented in the Kaplan-Meier plot in Figure 3. Nonuse intervention attrition occurred for 58% (25/43), 44% (18/41), and 44% (20/45) of the standard ABMT, gamified ABMT, and control conditions, respectively, which is comparable to other web-based health interventions [83-86]. Dropout occurred throughout the course of the intervention, with no significant differences between conditions in the time of attrition (log-rank $\chi_2{}^2$=2.8; *P*=.25). For those who provided unsolicited feedback (25/63, 40%), the reasons for nonuse intervention attrition were technical issues (8/25, 32%), sick or in hospital (5/25, 20%), unable to tolerate the training sessions (4/25, 16%), lack of interest (4/25, 16%), too busy or away (2/25, 8%), and no access to a computer due to the COVID-19 pandemic lockdown or floods (2/25, 8%).

XSL•FO
RenderX

Table 4. Linear mixed models for primary outcomes.

Variable	B (SE; 95% CI)	*P* value	Cohen *d* (95% CI)
Engagement			
Intercept	5.72 (0.59; 4.55 to 6.90)	<.001[a]	—[b]
Session 2	–0.69 (0.18; –1.03 to –0.34)	<.001[a]	0.257 (0.13 to 0.39)
Session 3	–0.99 (0.18; –1.35 to –0.63)	<.001[a]	0.368 (0.23 to 0.50)
Session 4	–1.28 (0.19; –1.65 to –0.91)	<.001[a]	0.473 (0.34 to 0.61)
Session 5	–1.33 (0.19; –1.71 to –0.96)	<.001[a]	0.488 (0.35 to 0.63)
Session 6	–1.63 (0.19; –2.00 to –1.25)	<.001[a]	0.596 (0.46 to 0.73)
Gamified ABMT[c]	0.55 (0.67; –0.77 to 1.87)	.41	–0.040 (–0.14 to 0.06)
Standard ABMT	–0.20 (0.60; –1.38 to 0.98)	.74	0.017 (–0.08 to 0.12)
Single (covariate)	–1.54 (0.56; –2.64 to –0.43)	.007[a]	—
Divorced or separated (covariate)	0.22 (0.75; –1.26 to 1.69)	.77	—
Student (covariate)	1.09 (0.87; –0.64 to 2.81)	.22	—
Retired (covariate)	1.71 (0.67; 0.37 to 3.04)	.01[a]	—
Not employed (covariate)	0.67 (0.63; –0.58 to 1.92)	.29	—
Pain intensity			
Intercept	51.17 (3.50; 44.24 to 58.10)	<.001[a]	—
Posttraining	–0.36 (0.68; –1.70 to 0.99)	.60	0.034 (–0.10 to 0.16)
Follow-up	–1.90 (0.71; –3.31 to –0.50)	.008[a]	0.180 (0.05 to 0.31)
Gamified ABMT	–0.48 (1.31; –3.06 to 2.11)	.72	0.025 (–0.11 to 0.16)
Standard ABMT	1.01 (1.27; –1.50 to 3.52)	.43	–0.054 (–0.19 to 0.08)
Depression (covariate)	0.24 (0.06; 0.13 to 0.35)	<.001[a]	—
Pain interference			
Intercept	43.53 (2.89; 37.83 to 49.24)	<.001[a]	—
Posttraining	–0.88 (0.58; –2.02 to 0.26)	.13	0.100 (–0.03 to 0.23)
Follow-up	–2.12 (0.61; –3.31 to –0.93)	<.001[a]	0.237 (0.10 to 0.37)
Gamified ABMT	1.63 (1.19; –0.72 to 3.99)	.17	–0.093 (–0.23 to 0.04)
Standard ABMT	0.27 (1.06; –1.84 to 2.37)	.80	–0.017 (–0.15 to 0.12)
Single (covariate)	–3.32 (1.02; –5.34 to –1.30)	.001[a]	—
Divorced or separated (covariate)	–1.11 (1.32; –3.72 to 1.51)	.41	—
Student (covariate)	0.77 (1.56; –2.32 to 3.86)	.62	—
Retired (covariate)	4.40 (1.21; 2.01 to 6.78)	<.001[a]	—
Not employed (covariate)	3.94 (1.16; 1.65 to 6.23)	<.001[a]	—
Depression (covariate)	0.34 (0.05; 0.25 to 0.43)	<.001[a]	—

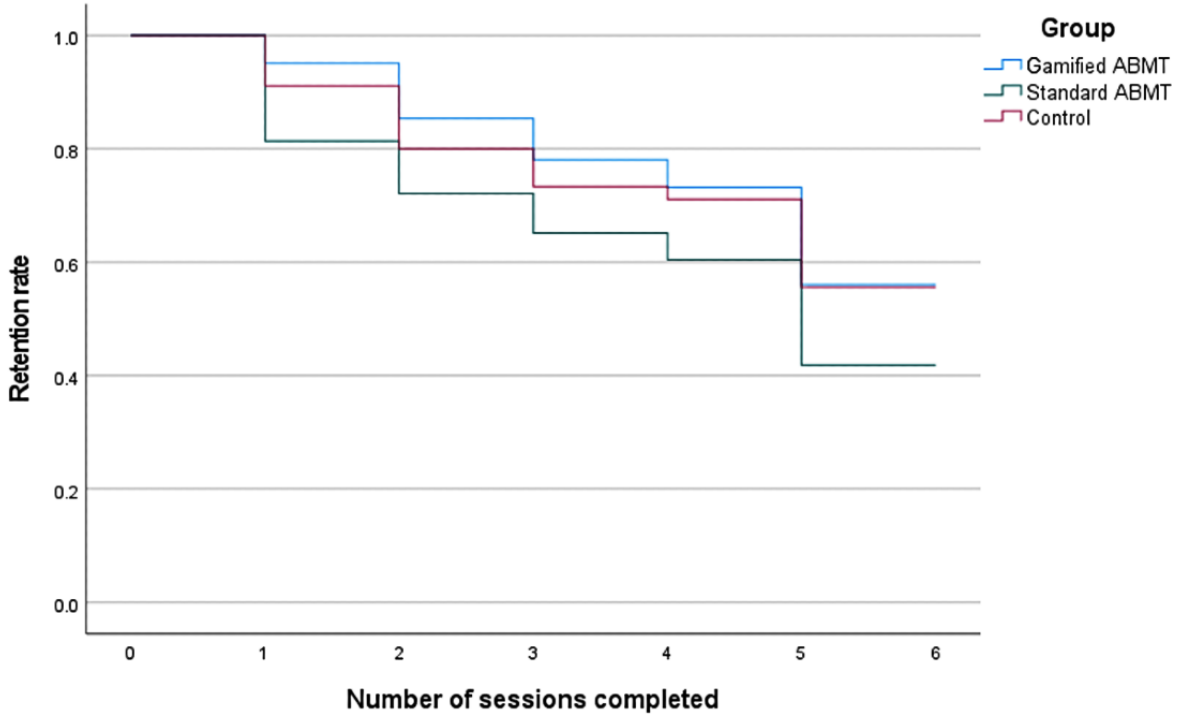
[a]Statistical significance: $P<.05$, 2-tailed.

[b]Not applicable.

[c]ABMT: attentional bias modification training.

XSL•FO
RenderX

Figure 3. Kaplan-Meier survival curves for dropout by training condition over the 3-week intervention period. ABMT: attentional bias modification training.



Pain Intensity

Table 4 displays the results of the linear mixed model for PROMIS Pain Intensity. There was a significant main effect of time from baseline to follow-up assessment (P=.008), with small effect size, indicating an overall small decrease in pain intensity scores over time across all conditions (baseline: mean 65.81, SD 6.83; follow-up: mean 63.61, SD 8.22). However, there was no significant main effect of time from baseline to posttraining assessment (P=.60) or of training condition (P values>.05; AIC=2113.76; BIC=2143.93). The model that included the interaction effects (Table S2 in Multimedia Appendix 3) did not fit the data better than the previous model, and all interactions were nonsignificant for time×condition (P values>.05; AIC=2115.34; BIC=2160.60).

Pain Interference

Table 4 displays the results of the linear mixed model for PROMIS Pain Interference scores. There was a significant main effect of time from baseline to follow-up assessment (P<001), with small effect size, indicating an overall small decrease in pain interference scores over time across all conditions (baseline: mean 65.58, SD 6.83; follow-up: mean 63.19, SD 7.53). However, there was no significant main effect of time from baseline to posttraining assessment (P=.13) or of training condition (P values>.05; AIC=2006.30; BIC=2055.33). The model that included the interaction effects (Table S3 in Multimedia Appendix 3) did not fit the data better than the previous model. The model, however, showed a significant interaction for follow-up×gamified (P=.03; AIC=2007.61; BIC=2071.72). All the other interactions were nonsignificant (P values>.05).

Secondary Outcomes

Measure of Attentional Bias for Pain

The 1-sample t tests against 0 indicated that there were no significant attentional biases toward pain-related stimuli observed at baseline in the standard ABMT (mean –0.58, SD 18.61; t_{39}=–0.20; P=.84; Cohen d=–0.031; 95% CI –0.34 to 0.28), gamified ABMT (mean –3.72, SD 24.41; t_{40}=–0.98; P=.34; Cohen d=–0.152; 95% CI –0.46 to 0.16), or control (mean 1.30, SD 20.23; t_{43}=0.43; P=.67; Cohen d=0.064; 95% CI –0.23 to 0.36) conditions. Table 5 displays the results of the linear mixed model for attentional bias index. There was no significant main effect of time on attentional bias index scores from baseline to posttraining assessment (P=.39) or of training condition (P values>.05; AIC=1878.14; BIC=1898.28). The model that included the interaction effects (Table S4 in Multimedia Appendix 3) did not fit the data better than the previous model, and all interactions were nonsignificant for time×condition (P values>.05; AIC=1879.54; BIC=1906.39). Correlations between changes in attentional bias magnitude and changes on pain intensity, pain interference, anxiety, and depression measures from baseline to posttraining assessment were all nonsignificant (P values>.05; Table S5 in Multimedia Appendix 3).

XSL•FO
RenderX

Table 5. Linear mixed models for secondary outcomes.

Variable	B (SE; 95% CI)	*P* value	Cohen *d* (95% CI)
Attentional bias index			
Intercept	–0.91 (2.55; –5.93 to 4.12)	.72	—[a]
Posttraining	2.39 (2.76; –3.05 to 7.82)	.39	–0.059 (–0.19 to 0.08)
Gamified ABMT[b]	–0.57 (3.27; –7.01 to 5.88)	.86	0.014 (–0.15 to 0.18)
Standard ABMT	0.45 (3.33; –6.12 to 7.02)	.89	–0.011 (–0.18 to 0.15)
Anxiety			
Intercept	16.40 (3.45; 9.58 to 23.22)	<.001[c]	—
Posttraining	–0.43 (0.65; –1.71 to 0.86)	.51	0.043 (–0.09 to 0.17)
Follow-up	–1.16 (0.68; –2.51 to 0.18)	.09	0.116 (–0.02 to 0.25)
Gamified ABMT	1.46 (1.29; –1.09 to 4.00)	.26	–0.077 (–0.21 to 0.06)
Standard ABMT	–0.06 (1.25; –2.54 to 2.42)	.96	0.003 (–0.13 to 0.14)
Depression (covariate)	0.71 (0.05; 0.61 to 0.82)	<.001[c]	—
Depression			
Intercept	54.65 (1.94; 50.81 to 58.49)	<.001[c]	—
Posttraining	–1.82 (0.66; –3.12 to –0.51)	.007[c]	0.180 (0.05 to 0.31)
Follow-up	–1.17 (0.70; –2.54 to 0.20)	.09	0.114 (–0.02 to 0.25)
Gamified ABMT	–0.90 (2.10; –5.06 to 3.26)	.67	0.029 (–0.11 to 0.16)
Standard ABMT	2.07 (1.90; –1.68 to 5.82)	.28	–0.074 (–0.21 to 0.06)
Primary (covariate)	–8.46 (8.39; –25.08 to 8.16)	.32	—
Secondary (covariate)	5.04 (1.92; 1.24 to 8.85)	.01[c]	—
Single (covariate)	3.94 (1.73; 0.51 to 7.37)	.03[c]	—
Divorced or separated (covariate)	–0.78 (2.31; –5.36 to 3.79)	.74	—
Student (covariate)	5.89 (2.69; 0.57 to 11.22)	.03[c]	—
Retired (covariate)	3.14 (2.10; –1.02 to 7.31)	.14	—
Not employed (covariate)	7.09 (1.95; 3.22 to 10.95)	<.001[c]	—
Perceived improvement			
Intercept	3.76 (0.18; 3.40 to 4.12)	<.001[c]	—
Follow-up	–0.03 (0.09; –0.22 to 0.15)	.73	0.025 (–0.12 to 0.17)
Gamified ABMT	0.20 (0.19; –0.17 to 0.57)	.28	–0.095 (–0.27 to 0.08)
Standard ABMT	0.11 (0.18; –0.24 to 0.46)	.53	–0.056 (–0.23 to 0.12)
Primary (covariate)	–1.94 (0.69; –3.30 to –0.58)	.006[c]	—
Secondary (covariate)	0.11 (0.17; –0.24 to 0.45)	.54	—
Student (covariate)	–0.04 (0.24; –0.52 to 0.45)	.88	—
Retired (covariate)	0.20 (0.20; –0.19 to 0.58)	.32	—
Not employed (covariate)	0.42 (0.18; 0.07 to 0.77)	.02[c]	—
Interpretation bias: pain and health belief			
Intercept	–0.96 (0.49; –1.93 to 0.02)	.06	—
Posttraining	–0.04 (0.09; –0.21 to 0.14)	.68	0.028 (–0.11 to 0.16)
Gamified ABMT	–0.08 (0.19; –0.45 to 0.29)	.67	0.036 (–0.13 to 0.20)

XSL•FO
RenderX

Variable	B (SE; 95% CI)	*P* value	Cohen *d* (95% CI)
Standard ABMT	0.42 (0.18; 0.07 to 0.78)	.02[c]	–0.195 (–0.36 to –0.31)
Depression (covariate)	0.02 (0.01; 0.00 to 0.03)	.04[c]	—
Interpretation bias: social belief			
Intercept	–2.62 (0.64; –3.88 to –1.36)	<.001[c]	—
Posttraining	0.05 (0.09; –0.12 to 0.23)	.56	–0.040 (–0.18 to 0.09)
Gamified ABMT	–0.27 (0.24; –0.74 to 0.20)	.26	0.093 (–0.07 to 0.26)
Standard ABMT	0.31 (0.23; –0.15 to 0.77)	.18	–0.112 (–0.28 to 0.05)
Depression (covariate)	0.03 (0.01; 0.01 to 0.05)	.007[c]	—

[a] Not applicable

[b]ABMT: attentional bias modification training.

[c]Statistical significance: *P*<.05, 2-tailed.

Anxiety and Depression

Table 5 displays the results of the linear mixed model for PROMIS Anxiety and PROMIS Depression. There was no significant main effect of time on anxiety scores from baseline to posttraining assessment (*P*=.51) or from baseline to follow-up assessment (*P*=.09). Moreover, there was no significant main effect of training condition (*P* values>.05; AIC=2091.03; BIC=2121.20). The model that included the interaction effects (Table S6 in Multimedia Appendix 3) did not fit the data better than the previous model, and all interactions were nonsignificant for time×condition (*P* values>.05; AIC=2095.35; BIC=2140.60).

Regarding depressive symptomatology, there was a significant main effect of time on depression scores from baseline (mean 60.13, SD 9.95) to posttraining assessment (mean 58.45, SD 10.21; *P*=.007), with small effect size, indicating an overall small decrease in depressive symptoms immediately after the intervention across all training conditions. However, there was no significant main effect of time from baseline to follow-up assessment (*P*=.09) or of training condition (*P* values>.05; AIC=2196.33; BIC=2249.13). The model that included the interaction effects (Table S7 in Multimedia Appendix 3) did not fit the data better than the previous model, and all interactions were nonsignificant for time×condition (*P* values>.05; AIC=2201.58; BIC=2269.47).

Perceived Improvement

Table 3 presents the statistics of the PGIC after training and at 1-month follow-up. At posttraining assessment, 14% (6/43), 15% (6/41), and 9% (4/45) of participants in the standard ABMT, gamified ABMT, and control conditions, respectively, reported that their symptoms improved. At the 1-month follow-up assessment, 16% (7/43), 15% (6/41), and 24% (11/45) of participants in the standard ABMT, gamified ABMT, and control conditions, respectively, reported that their symptoms improved. As shown in Table 5, there was no significant main effect of time on perceived overall pain-related improvement scores from posttraining to follow-up assessment (*P*=.73) or of training condition (*P* values>.05; AIC=468.12; BIC=503.95). The model that included the interaction effects (Table S8 in Multimedia Appendix 3) did not fit the data better than the previous model, and all interactions were nonsignificant for time×condition (*P* values>.05; AIC=471.87; BIC=514.22).

Measure of Interpretation Bias for Pain

Table 5 displays the results of the linear mixed model for interpretation biases for pain. There was no significant main effect of time on interpretation bias scores for pain and bodily threat situations from baseline to posttraining assessment (*P*=.68). However, there was a significant main effect of training condition (*P*=.02), with small effect size, such that participants in the standard ABMT condition (mean 0.45, SD 1.06) endorsed more negative interpretations for scenarios related to pain and bodily threat compared to those in the control condition (mean –0.01, SD 0.87). There was no significant difference in interpretation bias scores for pain and bodily threat between the gamified ABMT and control conditions (*P*=.67; AIC=558.34; BIC=582.00). The model that included the interaction effects (Table S9 in Multimedia Appendix 3) did not fit the data better than the previous model, and all interactions were nonsignificant for time×condition (*P* values>.05; AIC=561.97; BIC=592.39).

Regarding social interpretation, there was no significant main effect of time on interpretation bias scores for social situations from baseline to posttraining assessment (*P*=.56) or of training condition (*P* values>.05; AIC=621.84; BIC=645.50). The model that included the interaction effects (Table S10 in Multimedia Appendix 3) did not fit the data better than the previous model, and all interactions were nonsignificant for time×condition (*P* values>.05; AIC=625.11; BIC=655.53).

Additional and Exploratory Analyses

Exploratory Analyses

Multimedia Appendix 2 displays the results of the additional exploratory analyses. Overall, none of these variables were found to moderate the effects of ABMT on pain outcomes, except for Behavioral Activation System on pain interference (refer to Multimedia Appendix 2 for further details).

Manipulation Check

At the end of the posttraining assessment, participants were asked which treatment they believed they had received (ie, intervention or sham training). Overall, 41.2% (42/102) of the

XSL•FO
RenderX

participants who responded to the question identified their allocated treatment condition correctly. The proportion of participants who thought they were in the treatment condition was low, with 22% (7/32) for the standard ABMT, 20% (7/35) for the gamified ABMT, and 20% (7/35) for the control condition, a difference that was not statistically significant (χ^2_2=0.1; *P*=.99).

Sensitivity Analyses

Sensitivity analyses were conducted on all primary analyses without controlling for marital status, educational level, employment status, and baseline depressive symptomology as covariates. The pattern of results remained unchanged for pain intensity, pain interference, and engagement (Tables S11-S14 in Multimedia Appendix 3).

Adverse Events and Negative Experiences

Similar to previous studies (eg, [20]), no serious adverse events were reported. However, 5 participants (3/41, 7% gamified ABMT; 1/43, 2% standard ABMT; and 1/45, 2% control) discontinued their participation due to training-related nonserious adverse events. Specifically, despite being offered a rest between each block of trials, 4 participants reported that the physical action required to complete the tasks (ie, repetitive movements and sitting for prolonged periods) triggered pain in their hand, wrist, back, or neck, and 1 participant reported that the task caused dizziness. Of these 5 participants, 2 (2/129, 1.6%) specified that their difficulty in performing the tasks was aggravated by the COVID-19 pandemic lockdowns, which prevented them from receiving regular treatment for their pain. Another participant in the standard ABMT condition discontinued participation due to fear that the thin black writing on a white background might be a migraine trigger and suggested offering the option of a dark mode (ie, white or gray text presented against a dark or back background) and large text (ie, bolding and increasing the font size).

Discussion

Principal Findings

This study investigated the effects of a gamified web-delivered pain ABMT intervention in adults with chronic musculoskeletal pain. The findings can be readily summarized. First, there was no evidence that gamified ABMT was effective in enhancing engagement over and above the standard ABMT and sham training control conditions. Second, there were also no differential effects over time between the active ABMT and control conditions on measures of pain intensity, pain interference, depression, anxiety, or perceived improvement. Third, depressive symptoms, pain intensity, and pain interference reduced over time across all conditions. Finally, there was no indication for the presence of pretraining attentional biases in any of the conditions and no evidence that cognitive biases changed after training for the active ABMT conditions compared to the control condition.

Comparison With Previous Work

This study extends previous pain ABMT research [10,12,16-20] by examining the effects of gamification of a web-delivered pain ABMT intervention in adults with chronic musculoskeletal pain. Unexpectedly, there was no indication that participants in the gamified version of the ABMT task rated higher on self-reported task-related engagement, interest, and enjoyment or that they completed more training sessions or had less dropout compared to those in the nongamified versions. These results are inconsistent with the findings of several reviews [25,26] and recent studies [27,87-89], which found evidence for the potential of gamification to enhance motivation, engagement, and enjoyment in cognitive tasks. In fact, self-reported task-related engagement ratings were only moderate even at the first session (mean 6.00, SD 2.90; rating: 0-11), and all conditions on average reported relatively low levels of interest and enjoyment (IMI measure) immediately after the training (mean 3.15, SD 1.44; rating: 1-7). There was also an overall decline in self-reported task-related engagement over the 6 sessions. This is in line with the findings found in the study by Boendermaker et al [90], who found in their gamified ABMT for alcohol study that motivation to train decreased over time across both gamified and nongamified conditions. Thus, it seems that adding minimal game elements alone is not enough to counteract the monotony of the dot-probe task, which is of ongoing concern [21-23].

The finding that there was no superiority of the gamified ABMT condition over the other conditions was also unexpected but is in accordance with the results of a recent meta-analysis, which found no effect of gamification on cognitive or clinical outcomes [25]. Together, these findings do not imply that gamification cannot have positive effects on engagement and other outcomes, but rather highlight the complexity of designing effective gamified interventions. The development of the gamified ABMT task followed the Medical Research Council framework for complex interventions [38], using theory (ie, self-determination theory and self-regulation), reviewing evidence, and expert involvement. Despite this, it is possible that the type and combination of game elements we selected were not optimal. To date, it is still unclear as to which and how many game elements should be combined to optimize motivation [91,92]. There might also have been a ceiling effect (ie, participants completed, on average, 4.60, SD 1.83 out of 6 training sessions) where all participants were already, at baseline, highly motivated to complete the treatment course, making it difficult for the gamification to have improved further on this already high motivation to engage with the task. In accordance with the preregistered protocol, we excluded participants who did not complete at least 1 training session. As shown in Figure 2, a small minority of these excluded participants (8/73, 11%) had initiated a session, and it is possible that they did not find it engaging, leading to their decision to drop out before completing a session. As such, engagement could be lower than reported in this study if they had been included. Another explanation is that gamification was unsuccessful simply because the ABMT procedure lacks intervention credibility, which was inferred from participants' guess about their treatment allocation. This view was further supported by 2 participants who explicitly stated that they did not understand how clicking on *P*s and *Q*s could help with their pain.

XSL•FO
RenderX

In contrast to the hypothesis that ABMT would yield therapeutic benefit, there were also no differential effects over time between the active ABMT and control conditions on measures of pain intensity, pain interference, anxiety, depression, or perceived improvement. The results add to the mixed literature on the effectiveness of pain ABMT [10,12,16-20]. There was, however, an overall small decrease in depressive symptoms from baseline to posttraining assessment and in pain intensity and pain interference from baseline to follow-up assessment across all conditions. Similar results have been reported in the study by Carleton et al [17], who found a reduction in the experience of pain in both the ABMT and control groups. It is unclear why these reductions were seen in the control condition. It is possible that sham training (ie, the control condition) is not inert and has an active training component (ie, training at 50/50 contingency), which could influence attentional processes [93] or increase attentional control [94]. However, there was no evidence for change in attentional biases or attentional control following training in any of the conditions. Another possibility is that sham training may in fact be training cognitive flexibility [17]. Results could alternatively reflect natural fluctuation of symptoms over time, known as regression to the mean [95].

Treatment expectancies may have accounted, in part, for the observed improvements [96]; although, intervention credibility was low. It is also possible that treatment effects were observed due to improved standard of care or management. The impacts of the COVID-19 pandemic on our results are also unknown. Research has found that COVID-19 pandemic lockdowns affected individuals with chronic pain disproportionately, with increased pain levels and greater effects on mood and physical activity [97]. Although the exact number is unknown, it is certain that some participants started participation during or shortly after a lockdown period, when their levels of pain, pain interference, and depression may have been elevated.

One explanation for the nonsignificant effects is that pretraining attentional biases were not observed in the active ABMT conditions, which makes it difficult to alter them. However, the absence of baseline biases does not mean that ABMT interventions cannot be effective, as there have been reports of samples without an initial bias where ABMT interventions still led to pain-related improvements [12]. A potential explanation for the lack of baseline biases is that attentional bias in chronic pain may not be as strong as attentional bias in anxiety [98], despite our efforts to use stimuli related to pain. Alternatively, attentional biases toward pain may not be a stable trait-like characteristics of individuals with chronic pain, as often presumed, and may vary considerably within and between participants depending on their personal goals and contexts [15]. This would make it difficult to capture them comprehensively or validly with the dot-probe task, which only provides a snapshot of attention at the time when the probe appears. This might explain why the evidence for attentional biases toward pain-related words or pictures has been mixed [5,6,99].

In this study, there was also no evidence that biases changed after training for the active ABMT conditions compared to the control condition. The finding is, although contrary to our hypotheses, consistent with other pain ABMT studies [12,17,20]. Moreover, there was no relationship between changes in attentional bias magnitude from pretraining to posttraining assessment and changes in scores on measures of pain intensity, pain interference, and mood. Another explanation for the null findings is that the effectiveness of ABMT was attenuated by the relatively uncontrolled setting, with the potential presence of multiple distractions (eg, noise) that can impact attentional resources [100] and also influence the RT-based attentional bias index. It is also possible that ABMT does not target attentional bias or other underlying mechanisms (eg, interpretation bias), as suggested by some researchers [101,102]. Our data indicate no evidence of changes in interpretation biases or attentional control following training between the conditions. The findings are consistent with the study by Todd et al [102], which reported no significant effects of ABMT on interpretation bias in healthy undergraduate students under conditions of low threat and with the study by Heathcote et al [20], which reported no training effects on attentional control in adolescents with chronic pain. It should be noted, however, that Todd et al [102] used a more indirect measure of interpretation bias (ie, Incidental Learning Task [103]), whereas the task we used was constrained by a forced-choice response format and did not include control scenarios, which raises questions about whether other processes (eg, priming and negative expectancy bias) may have influenced the scores on this measure.

Limitations and Future Work

This study has several limitations. First, our sample was restricted to individuals with chronic musculoskeletal pain, thereby limiting the generalizability of findings to other types of pain conditions. Second, despite a rigorous randomization procedure, the conditions differed on several baseline characteristics. Although these were controlled for in our analyses, replications with matched samples are warranted as there might still be some residual confounding. Third, the use of a web-based delivery mode provided limited control over the environment in which ABMT was completed. Given there are suggestions in the anxiety ABMT literature that the efficacy of ABMT may be more difficult to achieve with remote training delivery compared to training in the laboratory [104], future research may want to replicate our findings in a more controlled setting. Fourth, although we opted for the dot-probe assessment paradigm as it is frequently used in web-based attentional bias research [17], we did not assess its reliability, which has been called into question [105]. Future studies may wish to use alternative measures of attentional bias such as eye-tracking technology. Fifth, the selection of a limited number of stimuli (8 stimulus pairs) may have affected the results. Despite our stimulus material consisting of word stimuli that have a strong association to pain, it is possible that a habituation effect to the words reduced their meaningfulness or that it contributed to the decrease in engagement over time. Future research may consider the use of more diverse stimulus types (eg, pictures as well). Finally, we chose to not include a gamified sham (control) condition to maximize power, which limited our ability to determine the impact of cognitive load on the processes underlying ABMT. Future research may consider using a "neutral" or no-contingency control condition such as sham-n training to ascertain whether attentional bias is affected or if cognitive flexibility is trained [93,106]. Future studies would

XSL•FO
RenderX

also benefit from using a more specific gamification framework (eg, [107]) that involves an interdisciplinary team of individuals with chronic pain, cognitive experts, and gamification designers to guide the co-design of their interventions.

Conclusions

In summary, the results of this first study investigating the effects of a gamified web-delivered pain ABMT intervention suggest that it has no additional effects over and above the standard ABMT and sham training control conditions. Adding minimal game elements alone does not seem to be enough to counteract the monotony of the dot-probe task. There is also no robust evidence that ABMT is of additional clinical value for individuals with chronic pain, at least not in its current form. A change of attentional bias, which is the presumed mechanism of action, was also not observed. While theoretically justified, the findings of this study do not provide support for the widespread clinical use of gamified and nongamified pain ABMT paradigms delivered over the internet.

Acknowledgments

The authors would like to thank the Royal Brisbane and Women's Hospital, particularly Associate Professor Paul Gray and Dr Andrew Claus, for their assistance with the recruitment of participants. The authors would like to acknowledge Dr Zoe Dettrick for her guidance with the statistical analyses. JFV received an Australian Government Research Training Program Scholarship and a Queensland University of Technology Excellence Top-Up Scholarship. DMLVR is supported by the Luxembourg National Research Fund (FNR) Core Junior programme (Painflex 12671141).

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

All authors made contributions to the conception and design of the study, which was undertaken as part of a PhD for JFV under the supervision of the other authors. DMLVR developed the experimental tasks for this study in collaboration with JFV. JFV conducted the study and analyzed all study data. JFV drafted the manuscript, and the other authors provided critical feedback and revisions. All authors read and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) checklist.
[PDF File (Adobe PDF File), 1050 KB - games_v13i1e50635_app1.pdf]

Multimedia Appendix 2
Additional and exploratory analyses.
[PDF File (Adobe PDF File), 1004 KB - games_v13i1e50635_app2.pdf]

Multimedia Appendix 3
Primary and secondary outcomes.
[PDF File (Adobe PDF File), 875 KB - games_v13i1e50635_app3.pdf]

References

1. Crombez G, Eccleston C, Van Damme S, Vlaeyen J, Karoly P. Fear-avoidance model of chronic pain: the next generation. Clin J Pain 2012 Jul;28(6):475-483. [doi: 10.1097/AJP.0b013e3182385392] [Medline: 22673479]
2. Todd J, Sharpe L, Johnson A, Nicholson Perry K, Colagiuri B, Dear BF. Towards a new model of attentional biases in the development, maintenance, and management of pain. Pain 2015 Sep;156(9):1589-1600 [FREE Full text] [doi: 10.1097/j.pain.0000000000000214] [Medline: 26291997]
3. Eccleston C, Crombez G. Worry and chronic pain: a misdirected problem solving model. Pain 2007 Dec 05;132(3):233-236. [doi: 10.1016/j.pain.2007.09.014] [Medline: 17961924]
4. Vlaeyen JW, Linton SJ. Fear-avoidance and its consequences in chronic musculoskeletal pain: a state of the art. Pain 2000 Apr;85(3):317-332 [FREE Full text] [doi: 10.1016/S0304-3959(99)00242-0] [Medline: 10781906]
5. Crombez G, Van Ryckeghem DM, Eccleston C, Van Damme S. Attentional bias to pain-related information: a meta-analysis. Pain 2013 Apr;154(4):497-510 [FREE Full text] [doi: 10.1016/j.pain.2012.11.013] [Medline: 23333054]

XSL•FO
RenderX

6. Todd J, van Ryckeghem DM, Sharpe L, Crombez G. Attentional bias to pain-related information: a meta-analysis of dot-probe studies. Health Psychol Rev 2018 Dec 19;12(4):419-436. [doi: 10.1080/17437199.2018.1521729] [Medline: 30205757]
7. Schoth DE, Nunes VD, Liossi C. Attentional bias towards pain-related information in chronic pain; a meta-analysis of visual-probe investigations. Clin Psychol Rev 2012 Feb;32(1):13-25. [doi: 10.1016/j.cpr.2011.09.004] [Medline: 22100743]
8. Shigetoh H. Hypervigilance to pain affects activities of daily living: an examination using the Japanese version of the pain vigilance awareness questionnaire. J Phys Ther Sci 2017 Dec;29(12):2094-2096 [FREE Full text] [doi: 10.1589/jpts.29.2094] [Medline: 29643581]
9. Rogers AH, Farris SG. A meta-analysis of the associations of elements of the fear-avoidance model of chronic pain with negative affect, depression, anxiety, pain-related disability and pain intensity. Eur J Pain 2022 Sep 07;26(8):1611-1635 [FREE Full text] [doi: 10.1002/ejp.1994] [Medline: 35727200]
10. Carleton RN, Richter AA, Asmundson GJ. Attention modification in persons with fibromyalgia: a double blind, randomized clinical trial. Cogn Behav Ther 2011 Dec;40(4):279-290. [doi: 10.1080/16506073.2011.616218] [Medline: 22060250]
11. MacLeod C, Rutherford E, Campbell L, Ebsworthy G, Holker L. Selective attention and emotional vulnerability: assessing the causal basis of their association through the experimental manipulation of attentional bias. J Abnorm Psychol 2002 Feb;111(1):107-123. [Medline: 11866165]
12. Sharpe L, Ianiello M, Dear BF, Perry KN, Refshauge K, Nicholas MK. Is there a potential role for attention bias modification in pain patients? Results of 2 randomised, controlled trials. Pain 2012 Mar;153(3):722-731. [doi: 10.1016/j.pain.2011.12.014] [Medline: 22281100]
13. Hakamata Y, Lissek S, Bar-Haim Y, Britton JC, Fox NA, Leibenluft E, et al. Attention bias modification treatment: a meta-analysis toward the establishment of novel treatment for anxiety. Biol Psychiatry 2010 Dec 01;68(11):982-990 [FREE Full text] [doi: 10.1016/j.biopsych.2010.07.021] [Medline: 20887977]
14. Yang W, Ding Z, Dai T, Peng F, Zhang JX. Attention Bias Modification training in individuals with depressive symptoms: a randomized controlled trial. J Behav Ther Exp Psychiatry 2015 Dec;49(Pt A):101-111 [FREE Full text] [doi: 10.1016/j.jbtep.2014.08.005] [Medline: 25245928]
15. Van Ryckeghem DM, Noel M, Sharpe L, Pincus T, Van Damme S. Cognitive biases in pain: an integrated functional-contextual framework. Pain 2019 Jul 29;160(7):1489-1493. [doi: 10.1097/j.pain.0000000000001508] [Medline: 30913168]
16. Schoth DE, Georgallis T, Liossi C. Attentional bias modification in people with chronic pain: a proof of concept study. Cogn Behav Ther 2013 Jun 04;42(3):233-243. [doi: 10.1080/16506073.2013.777105] [Medline: 23731349]
17. Carleton RN, Asmundson GJ, Korol SL, LeBouthillier DM, Hozempa K, Katz JD, et al. Evaluating the efficacy of an attention modification program for patients with fibromyalgia: a randomized controlled trial. Pain 2020 Mar 4;161(3):584-594 [FREE Full text] [doi: 10.1097/j.pain.0000000000001746] [Medline: 31693540]
18. Hasegawa T, Nishi K, Nakashima A, Moriuchi T, Iso N, Koseki H, et al. Effects of attentional bias modification on chronic low back pain in older outpatients: a randomized crossover trial (pilot study). Medicine (Baltimore) 2021 Nov 12;100(45):e27738 [FREE Full text] [doi: 10.1097/MD.0000000000027738] [Medline: 34766584]
19. Shiasy Y, Shakiba S, Taremian F, Akhavan Hejazi SM, Abasi A. The effectiveness of attention bias modification with and without trans cranial direct current stimulation in chronic low back pain. Iran J Psychiatry 2020 Apr 17;15(2):112-125. [doi: 10.18502/ijps.v15i2.2683]
20. Heathcote LC, Jacobs K, Van Ryckeghem DM, Fisher E, Eccleston C, Fox E, et al. Attention bias modification training for adolescents with chronic pain: a randomized placebo-controlled trial. Pain 2018 Feb;159(2):239-251. [doi: 10.1097/j.pain.0000000000001084] [Medline: 28968342]
21. Beard C, Weisberg RB, Primack J. Socially anxious primary care patients' attitudes toward cognitive bias modification (CBM): a qualitative study. Behav Cogn Psychother 2012 Oct;40(5):618-633 [FREE Full text] [doi: 10.1017/S1352465811000671] [Medline: 22127022]
22. Kuckertz JM, Schofield CA, Clerkin EM, Primack J, Boettcher H, Weisberg RB, et al. Attentional bias modification for social anxiety disorder: what do patients think and why does it matter? Behav Cogn Psychother 2019 Jan;47(1):16-38 [FREE Full text] [doi: 10.1017/S1352465818000231] [Medline: 29729676]
23. Wen S, Larsen H, Boffo M, Grasman RP, Pronk T, van Wijngaarden JB, et al. Combining web-based attentional bias modification and approach bias modification as a self-help smoking intervention for adult smokers seeking online help: double-blind randomized controlled trial. JMIR Ment Health 2020 May 08;7(5):e16342 [FREE Full text] [doi: 10.2196/16342] [Medline: 32383682]
24. Deterding S, Dixon D, Khaled R, Nacke L. From game design elements to gamefulness: defining "gamification". In: Proceedings of the 15th International Academic MindTrek Conference on Envisioning Future Media. 2011 Presented at: MindTrek '11; September 28-30, 2011; Tampere, Finland p. 9-15 URL: https://dl.acm.org/doi/10.1145/2181037.2181040 [doi: 10.1145/2181037.2181040]
25. Vermeir JF, White MJ, Johnson D, Crombez G, Van Ryckeghem DM. The effects of gamification on computerized cognitive training: systematic review and meta-analysis. JMIR Serious Games 2020 Aug 10;8(3):e18644 [FREE Full text] [doi: 10.2196/18644] [Medline: 32773374]

XSL•FO
RenderX

26. Lumsden J, Edwards EA, Lawrence NS, Coyle D, Munafò MR. Gamification of cognitive assessment and cognitive training: a systematic review of applications and efficacy. JMIR Serious Games 2016 Jul 15;4(2):e11 [FREE Full text] [doi: 10.2196/games.5888] [Medline: 27421244]
27. Salemink E, de Jong SR, Notebaert L, MacLeod C, Van Bockstaele B. Gamification of cognitive bias modification for interpretations in anxiety increases training engagement and enjoyment. J Behav Ther Exp Psychiatry 2022 Sep;76:101727 [FREE Full text] [doi: 10.1016/j.jbtep.2022.101727] [Medline: 35217211]
28. Zhang M, Ying J, Song G, Fung DS, Smith H. Gamified cognitive bias modification interventions for psychiatric disorders: review. JMIR Ment Health 2018 Oct 25;5(4):e11640 [FREE Full text] [doi: 10.2196/11640] [Medline: 30361194]
29. Vermeir JF, White MJ, Johnson D, Crombez G, Van Ryckeghem DM. Gamified web-delivered attentional bias modification training for adults with chronic pain: protocol for a randomized, double-blind, placebo-controlled trial. JMIR Res Protoc 2022 Jan 27;11(1):e32359 [FREE Full text] [doi: 10.2196/32359] [Medline: 35084344]
30. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
31. Treede RD, Rief W, Barke A, Aziz Q, Bennett MI, Benoliel R, et al. Chronic pain as a symptom or a disease: the IASP classification of chronic pain for the International Classification of Diseases (ICD-11). Pain 2019 Jan;160(1):19-27. [doi: 10.1097/j.pain.0000000000001384] [Medline: 30586067]
32. Boendermaker WJ, Boffo M, Wiers RW. Exploring elements of fun to motivate youth to do cognitive bias modification. Games Health J 2015 Dec;4(6):434-443. [doi: 10.1089/g4h.2015.0053] [Medline: 26421349]
33. McBeth J, Prescott G, Scotland G, Lovell K, Keeley P, Hannaford P, et al. Cognitive behavior therapy, exercise, or both for treating chronic widespread pain. Arch Intern Med 2012 Jan 09;172(1):48-57. [doi: 10.1001/archinternmed.2011.555] [Medline: 22082706]
34. Vermeir JF, White MJ, Johnson D, Crombez G, Van Ryckeghem DM. Development and evaluation of linguistic stimuli for pain research. J Pain 2023 Oct;24(10):1843-1858 [FREE Full text] [doi: 10.1016/j.jpain.2023.05.011] [Medline: 37268166]
35. Brysbaert M, New B. Moving beyond Kučera and Francis: a critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. Behav Res Methods 2009 Nov;41(4):977-990. [doi: 10.3758/brm.41.4.977]
36. Van Ryckeghem DM, Crombez G, Van Hulle L, Van Damme S. Attentional bias towards pain-related information diminishes the efficacy of distraction. Pain 2012 Dec;153(12):2345-2351. [doi: 10.1016/j.pain.2012.07.032] [Medline: 22989933]
37. Kemps E, Tiggemann M, Cibich M, Cabala A. Cognitive bias modification for energy drink cues. PLoS One 2019;14(12):e0226387 [FREE Full text] [doi: 10.1371/journal.pone.0226387] [Medline: 31841542]
38. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, Medical Research Council Guidance. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ 2008 Sep 29;337:a1655 [FREE Full text] [doi: 10.1136/bmj.a1655] [Medline: 18824488]
39. Deci EL, Ryan RM. Intrinsic Motivation and Self-Determination in Human Behavior. New York, NY: Springer; 1985.
40. Ryan RM, Deci EL. Intrinsic and extrinsic motivations: classic definitions and new directions. Contemp Educ Psychol 2000 Jan;25(1):54-67. [doi: 10.1006/ceps.1999.1020] [Medline: 10620381]
41. Bandura A. Social cognitive theory of self-regulation. Organ Behav Hum Decis Process 1991 Dec;50(2):248-287. [doi: 10.1016/0749-5978(91)90022-L]
42. Tondello GF, Premsukh H, Nacke L. A theory of gamification principles through goal-setting theory. In: Proceedings of the 51st Hawaii International Conference on System Sciences. 2018 Presented at: HICSS '18; January 3-6, 2018; Waikoloa. HI p. 1118-1127 URL: https://tinyurl.com/5h3kvce9 [doi: 10.24251/hicss.2018.140]
43. Hassan L, Xi N, Gurkan B, Koivisto J, Hamari J. Gameful self-regulation: a study on how gamified self-tracking features evoke gameful experiences. In: Proceedings of the 53rd Hawaii International Conference on System Sciences. 2020 Presented at: HICSS '20; January 7-10, 2020; Maui, HI p. 1103-1112 URL: https://tinyurl.com/47uyu8ff [doi: 10.24251/hicss.2020.138]
44. Lister C, West JH, Cannon B, Sax T, Brodegard D. Just a fad? Gamification in health and fitness apps. JMIR Serious Games 2014 Aug 04;2(2):e9 [FREE Full text] [doi: 10.2196/games.3413] [Medline: 25654660]
45. Locke EA, Latham GP. Building a practically useful theory of goal setting and task motivation. A 35-year odyssey. Am Psychol 2002 Sep;57(9):705-717. [doi: 10.1037//0003-066x.57.9.705] [Medline: 12237980]
46. Xi N, Hamari J. Does gamification satisfy needs? A study on the relationship between gamification features and intrinsic need satisfaction. Int J Inf Manage 2019 Jun;46:210-221. [doi: 10.1016/j.ijinfomgt.2018.12.002]
47. Weiser P, Bucher D, Cellina F, De Luca V. A taxonomy of motivational affordances for meaningful gamified and persuasive technologies. In: Proceedings of the 29th International Conference on Informatics for Environmental Protection & 3rd International Conference on ICT for Sustainability. 2015 Presented at: EnviroInfo/ICT4S '15; September 7-9, 2015; Copenhagen, Denmark p. 271-280 URL: https://www.atlantis-press.com/proceedings/ict4s-env-15/25836175 [doi: 10.2991/ict4s-env-15.2015.31]

XSL•FO
RenderX

48. Sailer M, Hense JU, Mayr SK, Mandl H. How gamification motivates: an experimental study of the effects of specific game design elements on psychological need satisfaction. Comput Hum Behav 2017 Apr;69:371-380. [doi: 10.1016/j.chb.2016.12.033]
49. Phillips C, Johnson D, Klarkowski M, White MJ, Hides L. The impact of rewards and trait reward responsiveness on player motivation. In: Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play. 2018 Presented at: CHI PLAY '18; October 28-31, 2018; Melbourne, Australia p. 393-404 URL: https://doi.org/10.1145/3242671.3242713 [doi: 10.1145/3242671.3242713]
50. Von Korff M, Ormel J, Keefe FJ, Dworkin SF. Grading the severity of chronic pain. Pain 1992 Aug;50(2):133-149. [doi: 10.1016/0304-3959(92)90154-4] [Medline: 1408309]
51. Smith BH, Penny KI, Purves AM, Munro C, Wilson B, Grimshaw J, et al. The Chronic Pain Grade questionnaire: validation and reliability in postal research. Pain 1997 Jun;71(2):141-147. [doi: 10.1016/s0304-3959(97)03347-2] [Medline: 9211475]
52. Dunn KM, Jordan K, Croft PR. Does questionnaire structure influence response in postal surveys? J Clin Epidemiol 2003 Jan;56(1):10-16. [doi: 10.1016/s0895-4356(02)00567-x] [Medline: 12589865]
53. Plant RW, Ryan RM. Intrinsic motivation and the effects of self‐consciousness, self‐awareness, and ego‐involvement: an investigation of internally controlling styles. J Pers 2006 Apr 28;53(3):435-449. [doi: 10.1111/J.1467-6494.1985.tb00375.X]
54. Ryan RM. Control and information in the intrapersonal sphere: an extension of cognitive evaluation theory. J Pers Soc Psychol 1982 Sep;43(3):450-461. [doi: 10.1037/0022-3514.43.3.450]
55. Ryan RM, Mims V, Koestner R. Relation of reward contingency and interpersonal context to intrinsic motivation: a review and test using cognitive evaluation theory. J Pers Soc Psychol 1983 Oct;45(4):736-750. [doi: 10.1037/0022-3514.45.4.736]
56. McAuley E, Duncan T, Tammen VV. Psychometric properties of the Intrinsic Motivation Inventory in a competitive sport setting: a confirmatory factor analysis. Res Q Exerc Sport 1989 Mar;60(1):48-58. [doi: 10.1080/02701367.1989.10607413] [Medline: 2489825]
57. Tsigilis N, Theodosiou A. Temporal stability of the intrinsic motivation inventory. Percept Mot Skills 2003 Aug;97(1):271-280. [doi: 10.2466/pms.2003.97.1.271] [Medline: 14604050]
58. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
59. Patient-reported outcomes measurement information system (PROMIS). Health Measures. URL: http://www.healthmeasures.net/explore-measurement-systems/promis [accessed 2024-12-21]
60. Bartlett SJ, Orbai AM, Duncan T, DeLeon E, Ruffing V, Clegg-Smith K, et al. Reliability and validity of selected PROMIS measures in people with rheumatoid arthritis. PLoS One 2015 Sep 17;10(9):e0138543 [FREE Full text] [doi: 10.1371/journal.pone.0138543] [Medline: 26379233]
61. Amtmann D, Cook KF, Jensen MP, Chen WH, Choi S, Revicki D, et al. Development of a PROMIS item bank to measure pain interference. Pain 2010 Jul;150(1):173-182 [FREE Full text] [doi: 10.1016/j.pain.2010.04.025] [Medline: 20554116]
62. MacLeod C, Mathews A, Tata P. Attentional bias in emotional disorders. J Abnorm Psychol 1986 Feb;95(1):15-20. [doi: 10.1037//0021-843x.95.1.15] [Medline: 3700842]
63. Batterham PJ, Sunderland M, Carragher N, Calear AL. Psychometric properties of 7- and 30-day versions of the PROMIS emotional distress item banks in an Australian adult sample. Assessment 2019 Mar 04;26(2):249-259. [doi: 10.1177/1073191116685809] [Medline: 28052687]
64. Schalet BD, Pilkonis PA, Yu L, Dodds N, Johnston KL, Yount S, et al. Clinical validity of PROMIS depression, anxiety, and anger across diverse clinical samples. J Clin Epidemiol 2016 May;73:119-127 [FREE Full text] [doi: 10.1016/j.jclinepi.2015.08.036] [Medline: 26931289]
65. Amtmann D, Kim J, Chung H, Bamer AM, Askew RL, Wu S, et al. Comparing CESD-10, PHQ-9, and PROMIS depression instruments in individuals with multiple sclerosis. Rehabil Psychol 2014 May;59(2):220-229 [FREE Full text] [doi: 10.1037/a0035919] [Medline: 24661030]
66. Guy W. Clinical Global Impressions: ECDEU assessment manual for psychopharmacology. Revised edition (DHEW Publ. No. ADM 76-338). National Institute of Mental Health. 1976. URL: https://www.scirp.org/reference/ReferencesPapers?ReferenceID=1153932 [accessed 2024-04-29]
67. Rampakakis E, Ste-Marie PA, Sampalis JS, Karellis A, Shir Y, Fitzcharles M. Real-life assessment of the validity of patient global impression of change in fibromyalgia. RMD Open 2015 Sep 14;1(1):e000146 [FREE Full text] [doi: 10.1136/rmdopen-2015-000146] [Medline: 26535150]
68. Scott W, McCracken LM. Patients' impression of change following treatment for chronic pain: global, specific, a single dimension, or many? J Pain 2015 Jun;16(6):518-526 [FREE Full text] [doi: 10.1016/j.jpain.2015.02.007] [Medline: 25746196]
69. Heathcote LC, Koopmans M, Eccleston C, Fox E, Jacobs K, Wilkinson N, et al. Negative interpretation bias and the experience of pain in adolescents. J Pain 2016 Sep;17(9):972-981 [FREE Full text] [doi: 10.1016/j.jpain.2016.05.009] [Medline: 27263991]

XSL•FO
RenderX

70. Heathcote LC, Jacobs K, Eccleston C, Fox E, Lau JY. Biased interpretations of ambiguous bodily threat information in adolescents with chronic pain. Pain 2017 Mar;158(3):471-478 [FREE Full text] [doi: 10.1097/j.pain.0000000000000781] [Medline: 28067692]
71. Lau JY, Badaoui M, Meehan AJ, Heathcote LC, Barker ED, Rimes KA. Assessing the content specificity of interpretation biases in community adolescents with persistent and interfering pain. Pain 2020 Feb 17;161(2):319-327. [doi: 10.1097/j.pain.0000000000001723] [Medline: 31634340]
72. Derryberry D, Reed MA. Anxiety-related attentional biases and their regulation by attentional control. J Abnorm Psychol 2002 May;111(2):225-236. [doi: 10.1037//0021-843x.111.2.225] [Medline: 12003445]
73. Carver CS, White TL. Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: the BIS/BAS scales. J Pers Soc Psychol 1994 Aug;67(2):319-333. [doi: 10.1037/0022-3514.67.2.319]
74. Sullivan MJ, Bishop SR, Pivik J. The pain catastrophizing scale: development and validation. Psychol Assess 1995 Dec;7(4):524-532. [doi: 10.1037/1040-3590.7.4.524]
75. Meyer B, Johnson SL, Carver CS. Exploring behavioral activation and inhibition sensitivities among college students at risk for bipolar spectrum symptomatology. J Psychopathol Behav Assess 1999 Dec 01;21(4):275-292 [FREE Full text] [doi: 10.1023/A:1022119414440] [Medline: 21765591]
76. Crombez G, De Paepe AL, Veirman E, Eccleston C, Verleysen G, Van Ryckeghem DM. Let's talk about pain catastrophizing measures: an item content analysis. PeerJ 2020;8:e8643 [FREE Full text] [doi: 10.7717/peerj.8643] [Medline: 32181053]
77. Van Damme S, Crombez G, Bijttebier P, Goubert L, Van Houdenhove B. A confirmatory factor analysis of the Pain Catastrophizing Scale: invariant factor structure across clinical and non-clinical populations. Pain 2002 Apr;96(3):319-324. [doi: 10.1016/S0304-3959(01)00463-8] [Medline: 11973004]
78. Oppenheimer DM, Meyvis T, Davidenko N. Instructional manipulation checks: detecting satisficing to increase statistical power. J Exp Soc Psychol 2009 Jul;45(4):867-872. [doi: 10.1016/j.jesp.2009.03.009]
79. Lakens D. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. Front Psychol 2013 Nov 26;4:863 [FREE Full text] [doi: 10.3389/fpsyg.2013.00863] [Medline: 24324449]
80. Dunlap WP, Cortina JM, Vaslow JB, Burke MJ. Meta-analysis of experiments with matched groups or repeated measures designs. Psychol Methods 1996 Jun;1(2):170-177. [doi: 10.1037/1082-989X.1.2.170]
81. Barr DJ, Levy R, Scheepers C, Tily HJ. Random effects structure for confirmatory hypothesis testing: keep it maximal. J Mem Lang 2013 Apr;68(3):1 [FREE Full text] [doi: 10.1016/j.jml.2012.11.001] [Medline: 24403724]
82. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. J Am Stat Assoc 1958 Jun;53(282):457-481. [doi: 10.1080/01621459.1958.10501452]
83. Meyerowitz-Katz G, Ravi S, Arnolda L, Feng X, Maberly G, Astell-Burt T. Rates of attrition and dropout in app-based interventions for chronic disease: systematic review and meta-analysis. J Med Internet Res 2020 Sep 29;22(9):e20283 [FREE Full text] [doi: 10.2196/20283] [Medline: 32990635]
84. Edney S, Ryan JC, Olds T, Monroe C, Fraysse F, Vandelanotte C, et al. User engagement and attrition in an app-based physical activity intervention: secondary analysis of a randomized controlled trial. J Med Internet Res 2019 Nov 27;21(11):e14645 [FREE Full text] [doi: 10.2196/14645] [Medline: 31774402]
85. Vandelanotte C, Duncan MJ, Kolt GS, Caperchione CM, Savage TN, Van Itallie A, et al. More real-world trials are needed to establish if web-based physical activity interventions are effective. Br J Sports Med 2019 Dec 03;53(24):1553-1554. [doi: 10.1136/bjsports-2018-099437] [Medline: 29970409]
86. Guertler D, Vandelanotte C, Kirwan M, Duncan MJ. Engagement and nonusage attrition with a free physical activity promotion program: the case of 10,000 steps Australia. J Med Internet Res 2015 Jul 15;17(7):e176 [FREE Full text] [doi: 10.2196/jmir.4339] [Medline: 26180040]
87. Vervaeke J, Hoorelbeke K, Baeken C, Van Looy J, Koster EH. Transfer and motivation after cognitive control training for remitted depression in healthy sample. J Cogn Enhanc 2019 Apr 29;4(1):49-61. [doi: 10.1007/S41465-019-00135-6]
88. Gabele M, Weicker J, Wagner S, Thoms A, Hußlein S, Hansen C. Effects and ways of tailored gamification in software-based training in cognitive rehabilitation. In: Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization. 2021 Presented at: UMAP '21; June 21-25, 2021; Utrecht, Netherlands p. 158-168 URL: https://dl.acm.org/doi/10.1145/3450613.3456828 [doi: 10.1145/3450613.3456828]
89. Shaban A, Pearson E. Evaluation of user experience and cognitive load of a gamified cognitive training application for children with learning disabilities. In: Proceedings of the 17th International Web for All Conference. 2020 Presented at: W4A '20; April 20-21, 2020; Taipei, Taiwan p. 1-10 URL: https://dl.acm.org/doi/10.1145/3371300.3383341 [doi: 10.1145/3371300.3383341]
90. Boendermaker WJ, Sanchez Maceiras S, Boffo M, Wiers RW. Attentional bias modification with serious game elements: evaluating the shots game. JMIR Serious Games 2016 Dec 06;4(2):e20 [FREE Full text] [doi: 10.2196/games.6464] [Medline: 27923780]
91. Mazarakis A, Bräuer P. Gamification is working, but which one exactly? Results from an experiment with four game design elements. Int J Hum Comput Interact 2022 Apr 06;39(3):612-627. [doi: 10.1080/10447318.2022.2041909]

XSL•FO
RenderX

92. Groening C, Binnewies C. The more, the merrier? - How adding and removing game design elements impact motivation and performance in a gamification environment. Int J Hum Comput Interact 2021 Jan 27;37(12):1130-1150. [doi: 10.1080/10447318.2020.1870828]
93. Tiggemann M, Kemps E. Is sham training still training? An alternative control group for attentional bias modification. Front Psychol 2020 Nov 4;11:583518 [FREE Full text] [doi: 10.3389/fpsyg.2020.583518] [Medline: 33250822]
94. Heeren A, Mogoaşe C, McNally RJ, Schmitz A, Philippot P. Does attention bias modification improve attentional control? A double-blind randomized experiment with individuals with social anxiety disorder. J Anxiety Disord 2015 Jan;29:35-42 [FREE Full text] [doi: 10.1016/j.janxdis.2014.10.007] [Medline: 25465885]
95. Kamerman PR, Vollert J. Greater baseline pain inclusion criteria in clinical trials increase regression to the mean effect: a modelling study. Pain 2022 Jun 01;163(6):e748-e758. [doi: 10.1097/j.pain.0000000000002468] [Medline: 34510140]
96. Smeets RJ, Beelen S, Goossens ME, Schouten EG, Knottnerus JA, Vlaeyen JW. Treatment expectancy and credibility are associated with the outcome of both physical and cognitive-behavioral treatment in chronic low back pain. Clin J Pain 2008 May;24(4):305-315. [doi: 10.1097/AJP.0b013e318164aa75] [Medline: 18427229]
97. Fallon N, Brown C, Twiddy H, Brian E, Frank B, Nurmikko T, et al. Adverse effects of COVID-19-related lockdown on pain, physical activity and psychological well-being in people with chronic pain. Br J Pain 2021 Aug 21;15(3):357-368 [FREE Full text] [doi: 10.1177/2049463720973703] [Medline: 34377461]
98. Bar-Haim Y, Lamy D, Pergamin L, Bakermans-Kranenburg MJ, van IJzendoorn MH. Threat-related attentional bias in anxious and nonanxious individuals: a meta-analytic study. Psychol Bull 2007 Jan;133(1):1-24. [doi: 10.1037/0033-2909.133.1.1] [Medline: 17201568]
99. Chan FH, Suen H, Jackson T, Vlaeyen JW, Barry TJ. Pain-related attentional processes: a systematic review of eye-tracking research. Clin Psychol Rev 2020 Aug;80:101884. [doi: 10.1016/j.cpr.2020.101884] [Medline: 32585493]
100. MacLeod C, Clarke PJ. The attentional bias modification approach to anxiety intervention. Clin Psychol Sci 2015 Jan 06;3(1):58-78. [doi: 10.1177/2167702614560749]
101. Heeren A, De Raedt R, Koster EH, Philippot P. The (neuro)cognitive mechanisms behind attention bias modification in anxiety: proposals based on theoretical accounts of attentional bias. Front Hum Neurosci 2013;7:119 [FREE Full text] [doi: 10.3389/fnhum.2013.00119] [Medline: 23576969]
102. Todd J, Sharpe L, Colagiuri B. Attentional bias modification and pain: the role of sensory and affective stimuli. Behav Res Ther 2016 Aug;83:53-61. [doi: 10.1016/j.brat.2016.06.002] [Medline: 27327608]
103. Khatibi A, Sharpe L, Jafari H, Gholami S, Dehghani M. Interpretation biases in chronic pain patients: an incidental learning task. Eur J Pain 2015 Sep 18;19(8):1139-1147. [doi: 10.1002/ejp.637] [Medline: 25523038]
104. Heeren A, Mogoa e C, Philippot P, McNally RJ. Attention bias modification for social anxiety: a systematic review and meta-analysis. Clin Psychol Rev 2015 Aug;40:76-90. [doi: 10.1016/j.cpr.2015.06.001] [Medline: 26080314]
105. Waechter S, Nelson AL, Wright C, Hyatt A, Oakman J. Measuring attentional bias to threat: reliability of dot probe and eye movement indices. Cogn Ther Res 2013 Oct 17;38(3):313-333. [doi: 10.1007/S10608-013-9588-2]
106. Kemps E, Tiggemann M. Attentional bias modification for chocolate: sham-n training as a new control group. PLoS One 2021 Nov 19;16(11):e0260294 [FREE Full text] [doi: 10.1371/journal.pone.0260294] [Medline: 34797898]
107. Khaleghi A, Aghaei Z, Mahdavi MA. A gamification framework for cognitive assessment and cognitive training: qualitative study. JMIR Serious Games 2021 May 18;9(2):e21900 [FREE Full text] [doi: 10.2196/21900] [Medline: 33819164]

Abbreviations

ABMT: attentional bias modification training
ACS: Attentional Control Scale
AIC: Akaike information criterion
BIC: Bayesian information criterion
CONSORT: Consolidated Standards of Reporting Trials
CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
GCPS: Graded Chronic Pain Scale
IMI: Intrinsic Motivation Inventory
PGIC: Patient Global Impression of Change
PROMIS: Patient-Reported Outcomes Measurement Information System
RT: reaction time

XSL•FO
RenderX

Edited by T Leung, A Coristine; submitted 09.07.23; peer-reviewed by A Bucher, J Curcic; comments to author 04.02.24; revised version received 13.05.24; accepted 18.09.24; published 16.01.25.

Please cite as:
Vermeir JF, White MJ, Johnson D, Crombez G, Van Ryckeghem DML
Gamified Web-Delivered Attentional Bias Modification Training for Adults With Chronic Pain: Randomized, Double-Blind, Placebo-Controlled Trial
JMIR Serious Games 2025;13:e50635
URL: https://games.jmir.org/2025/1/e50635
doi:10.2196/50635
PMID:

©Julie F Vermeir, Melanie J White, Daniel Johnson, Geert Crombez, Dimitri M L Van Ryckeghem. Originally published in JMIR Serious Games (https://games.jmir.org), 16.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Enhancing Immersion in Virtual Reality–Based Advanced Life Support Training: Randomized Controlled Trial

Dilek Kitapcioglu[1,2], MD, Prof Dr Med, PhD; Mehmet Emin Aksoy[1], MD, Prof Dr Med, MSc, PhD; Arun Ekin Ozkan[3], MSci; Tuba Usseli[4], MS; Dilan Cabuk Colak[5], PhD; Tugrul Torun[5], MS

[1]Center of Advanced Simulation and Education, Acibadem Mehmet Ali Aydinlar University, Istanbul, Turkey

[2]Department of Medical Education, Medical Faculty, Acibadem Mehmet Ali Aydinlar University, Istanbul, Turkey

[3]Institute of Biomedical Engineering, Boğaziçi University, Istanbul, Turkey

[4]Department of Anesthesia, Vocational School of Health Services, Acibadem Mehmet Ali Aydinlar University, Istanbul, Turkey

[5]Department of Psychology, Faculty of Humanities and Social Sciences, Acibadem Mehmet Ali Aydinlar University, Istanbul, Turkey

Corresponding Author:
Mehmet Emin Aksoy, MD, Prof Dr Med, MSc, PhD
Center of Advanced Simulation and Education
Acibadem Mehmet Ali Aydinlar University
Kayisdagi cad No 32 Atasehir
Istanbul, 34752
Turkey
Phone: 90 05052685158
Email: emin.aksoy@acibadem.edu.tr

Abstract

Background: Serious game–based training modules are pivotal for simulation-based health care training. With advancements in artificial intelligence (AI) and natural language processing, voice command interfaces offer an intuitive alternative to traditional virtual reality (VR) controllers in VR applications.

Objective: This study aims to compare AI-supported voice command interfaces and traditional VR controllers in terms of user performance, exam scores, presence, and confidence in advanced cardiac life support (ACLS) training.

Methods: A total of 62 volunteer students from Acibadem Mehmet Ali Aydinlar University Vocational School for Anesthesiology, aged 20-22 years, participated in the study. All the participants completed a pretest consisting of 10 multiple-choice questions about ACLS. Following the pretest, participants were randomly divided into 2 groups: the voice command group (n=31) and the VR controller group (n=31). The voice command group members completed the VR-based ACLS serious game in training mode twice, using an AI-supported voice command as the game interface. The VR controller group members also completed the VR-based ACLS serious game in training mode twice, but they used VR controllers as the game interface. The participants completed a survey to assess their level of presence and confidence during gameplay. Following the survey, participants completed the exam module of the VR-based serious gaming module. At the final stage of the study, participants completed a posttest, which had the same content as the pretest. VR-based exam scores of the voice command and VR controller groups were compared using a 2-tailed, independent-samples *t* test, and linear regression analysis was conducted to examine the effect of presence and confidence rating.

Results: Both groups showed an improvement in performance from pretest to posttest, with no significant difference in the magnitude of improvement between the 2 groups (P=.83). When comparing presence ratings, there was no significant difference between the voice command group (mean 5.18, SD 0.83) and VR controller group (mean 5.42, SD 0.75; P=.25). However, when comparing VR-based exam scores, the VR controller group (mean 80.47, SD 13.12) significantly outperformed the voice command group (mean 66.70, SD 21.65; P=.005), despite both groups having similar time allocations for the exam (voice command group: mean 18.59, SD 5.28 minutes and VR controller group: mean 17.3, SD 4.83 minutes). Confidence levels were similar between the groups (voice command group: mean 3.79, SD 0.77 and VR controller group: mean 3.60, SD 0.72), but the voice command group displayed a significant overconfidence bias (voice command group: mean 0.09, SD 0.24 and VR controller group: mean –0.09, SD 0.18; P=.002).

Conclusions: VR-based ACLS training demonstrated effectiveness; however, the use of voice commands did not result in improved performance. Further research should explore ways to optimize AI's role in education through VR.

XSL•FO
RenderX

Trial Registration: ClinicalTrials.gov NCT06458452; https://clinicaltrials.gov/ct2/show/NCT06458452

(JMIR Serious Games 2025;13:e68272) doi:10.2196/68272

KEYWORDS

artificial intelligence; voice recognition; serious gaming; immersion; virtual reality

Introduction

Serious game–based training modules have become one of the training modalities for simulation-based health care training and have an important role in the training of health care professionals [1-5]. Training using serious gaming modules can be conducted through various platforms such as personal computers, tablets, virtual reality (VR), or mixed reality systems like augmented reality or augmented virtuality. The portability and increased affordability of VR systems in recent years have contributed to significant growth in VR-based learning. Providing immersive and interactive experiences of VR technology has revolutionized various fields including gaming, education, and training of health care professionals [1,6-8].

Enhanced student engagement makes VR and mixed reality modules particularly favored for health care training, such as advanced cardiac life support (ACLS) training, which aims to train health care professionals in the management of cardiopulmonary arrest. More than a million health care professionals participate in ACLS or advanced life support (ALS) courses globally each year [9]. Since these courses must be repeated at specific intervals, as mandated by local regulations or institutional requirements, software-based training modules and VR-based serious games will allow trainees to refresh their knowledge and skills anytime and anywhere [10].

VR-based training modules have used hand-held controllers for navigation and interaction with the virtual environment so far. In parallel to the advancements in artificial intelligence (AI) and natural language processing (NLP) potential for voice command interfaces, they offer an alternative method of control in VR applications by providing more intuitive and accessible means of interaction for users [11-15]. Voice commands, powered by NLP algorithms, enable users to navigate and manipulate virtual environments using natural language and have the potential to reduce the learning curve associated with VR controller inputs. Voice interfaces have also the potential to offer accessibility benefits for users with physical disabilities, for whom conventional controllers may be difficult to use [14].

Despite the potential advantages, using voice command systems in VR gaming has also its limitations like speech recognition accuracy, latency, the system's ability to understand context-specific commands, problems caused by environmental noise, and the user's accent or speech patterns [16,17].

There has been limited research comparing the user performance and user satisfaction of AI-supported voice command interfaces against traditional VR controllers [11,13,15]. This study aims to compare the 2 control methods using various metrics like task completion time, participants' exam scores, level of confidence, and level of presence. As the level of presence gives insight into how deeply the learner is immersed in the game, while confidence ratings reveal how the learner perceives their own performance and decision-making, the level of presence and confidence ratings during gameplay were also compared [18-20]. The hypothesis of this study is that AI-supported voice commands have the potential to offer a superior user experience and user performance in VR gaming by providing a natural form of interaction. The research questions of the study are the following.

- How do participants' pretest and posttest scores differ when using AI-supported voice commands versus traditional VR controllers?
- Do AI-supported voice command interfaces provide a superior user experience compared to traditional VR controllers in terms of performance using VR-based exam scores?
- How do confidence ratings vary between users of AI-supported voice command interfaces and traditional VR controllers during gameplay?
- How does the perceived level of presence differ between users of AI-supported voice commands and traditional VR controllers in VR gaming?

Methods

Recruitment

This study included 67 volunteer students from Acibadem Mehmet Ali Aydinlar University Vocational School for Anesthesiology. The participants, aged between 20-22 years, were in their fourth semester (spring semester 2023-2024) of the anesthesiology program. Five students declined to participate in this study. The remaining participants were randomly assigned to 2 groups: 31 students using VR controllers for game interaction (VR controller group) and 31 students using AI-supported voice recognition for game interaction (voice command group). The exclusion criteria for the study included prior ACLS training, a history of VR-induced motion sickness, and medical conditions such as vertigo attacks or using medications causing vertigo-like symptoms. Three participants from the voice command group were excluded for not meeting the inclusion criteria. The randomization was based on the participants' university ID numbers, with odd numbers assigned to one group and even numbers to the other. The clinical trial registration identifier of the study is ClinicalTrials.gov NCT06458452.

Ethical Considerations

This study was conducted in compliance with ethical guidelines for human subject research. The research protocol was reviewed and approved by the Scientific Ethical Committee of Acibadem Mehmet Ali Aydinlar University (ATADEK 2024-6). For the primary data collection, informed consent was obtained from all participants in accordance with the guidelines set forth by

XSL•FO
RenderX

the scientific ethical committee of Acibadem Mehmet Ali Aydinlar University. The informed consent obtained for the primary study explicitly allowed for the deidentified data to be used in subsequent research, including secondary analyses, without requiring additional consent. This study was conducted with strict adherence to privacy and confidentiality standards to protect the data of all participants. The study data were deidentified prior to analysis, ensuring that no personally identifiable information was accessible during the research process. Any potential identifiers were removed or coded to maintain confidentiality throughout the study. Participation in the study was entirely voluntary, and no compensation was provided to the participants. All images included in the study and supplementary materials do not feature individual participant images or any identifying clues that could reveal the identity of the users.

Study Flow

All the participants completed a pretest consisting of 10 multiple-choice questions about ACLS, which served to assess the participants' initial knowledge level and whether the content was compatible with the latest version of ACLS (Multimedia Appendix 1) [21].

Following the pretest, participants were randomly divided into 2 groups: the voice command group (n=31) and the VR controller group (n=31). The voice command group members completed the VR-based ACLS serious game in training mode twice, using an AI-supported voice command as the game interface. The VR controller group members also completed the VR-based ACLS serious game in training mode twice, but they used VR controllers as the game interface. Although having volunteered for the study, 3 participants from the voice command group did not attend the study.

The participants from both groups completed the Turkish version of the Presence Questionnaire (PQ), a 7-point Likert scale, to estimate the level of presence during gameplay [22]. The participants rated their experience on a scale from 1=not at all to 7=completely by selecting the option that best reflects their experience. The PQ was first developed by Witmer and Singer [23] to subjectively measure the level of presence in 3D virtual environments. The scale was revised in a subsequent study conducted in 2005, which identified its factor structure as consisting of 4 factors [24]. The adaptation of the PQ to Turkish was conducted by Gokoglu and Cakiroglu [22] in 2019. The factor analysis of the Turkish version of the PQ found a 5-factor structure—involvement, sensory fidelity, adaptation or immersion, interaction, and interface quality—which was confirmed through a confirmatory factor analysis. The Turkish version of the questionnaire used in this study, along with its translated English version, are provided in Multimedia Appendices 2 and 3.

In the involvement factor of the scale, there are items for evaluating a situation obtained because of focusing the individual's attention and mental potential consistently or meaningfully on relevant stimuli, activities, or events. For the sensory fidelity factor, items related to perceiving the VR scenario visually, auditorily, and haptically were included. The adaptation or immersion factor includes items related to being included in and interacting with a continuous flow of experiences and stimuli with a sense of being surrounded, while the interface quality factor includes items to evaluate the effect of visual and control interfaces in the VR experience. Finally, in the interaction factor, items related to interaction between individuals and the virtual environment were included. The adapted scale was found to be valid and reliable with a Cronbach α=.84 and applicable with its 29 items under 5 factors [22]. Participants' presence score was calculated as the mean of their responses to all items in the scale.

Following the survey, participants were asked to indicate how they believed they would perform in the upcoming exam on a scale of 1 to 5, where 1=very bad and 5=very good, to assess confidence in future performance. After confidence assessments, participants completed the exam module of the VR-based serious gaming module, which evaluates the technical and nontechnical of the participants. The voice command group used a voice interface, while the VR controller group used VR controllers to complete the exam. At the final stage of the study, participants completed a posttest, which had the same content as the pretest.

Serious Gaming Module Used for the Study

The software used in this study is a VR-based ALS serious game that has been developed in accordance with the ACLS guidelines European Resuscitation Council and American Heart Association [21,25]. The development team worked closely with clinicians to ensure compliance with these guidelines as well as crisis resource management criteria [26-28].

This serious game works with a learning management system and keeps user credentials as well as game results in a shared database with the help of a Learning Record Store [29-31]. The interactions of the players with the virtual world are tracked using a 3D visualization engine and experience application programming interface calls are created accordingly [30,32]. These calls are sent to the Learning Record Store servers with the help of a unity extension library that uses HTTP protocol [33]. This library also includes security measures and authentication methods included with HTTP protocols required for the safekeeping of user data.

The VR-based ALS serious game includes an AI-driven voice recognition feature that uses NLP algorithms of a service called Wit.ai [34]. Wit.ai is an open-source service that has existing Unity plug-ins that facilitate the implementation process. Through this service, the sentences of the user are sent to wit.ai servers, processed into game commands according to predefined keywords, and returned to determine the action taken by the user.

The serious gaming module offers 2 language options: English and Turkish. Participants were given the choice between the 2 options. As all volunteers preferred Turkish during gameplay, both the VR controller group and voice command groups selected Turkish as the game language. Since the voice command group also chose Turkish, the AI-driven voice recognition was based on the Turkish model. There are 2 stages in this serious game: beginner training and VR-based exam. These stages can be played by either voice command or using the controller of the VR headset. The simulated environment

XSL•FO
RenderX

consists of a virtual hospital room and equipment to immerse the user as much as possible in this experience.

The beginner training stage is directed toward users who are willing to familiarize themselves with the ACLS algorithm and the VR environment. In this stage, the user is completely guided from the beginning to the end using visual and audio cues. Which objects to interact with or which voice commands to say are given to the users? Examples of guidance can be seen in Figure 1 and Figure 2 respectively. There are also no timing constraints in this mode which removes the stress element and allows an inexperienced user to focus on learning the algorithm [8].

Figure 1. Screen capture of the laser pointer–controlled mode of the advanced cardiac life support serious game. CPR: cardiopulmonary resuscitation.



Figure 2. Screen capture of the voice-control mode of the advanced cardiac life support serious game.



XSL•FO
RenderX

The exam stage is designed for users who are confident in their abilities after spending some time in the basic training stage. Each action of the user is scored according to its accuracy, order, and timing. Each action taken by the user is graded based on its timing, order, and accuracy. The final score of the VR-based exam, out of 100 points, is derived from 2 categories: 70% (70/100) from ACLS assessments and 30% (30/100) from crisis resource management performance. The scores of the exam mode were stored and could be accessed afterward.

Data Analysis

Data from the comparison of pretest and posttest scores, VR-based exam scores, confidence ratings, presence ratings, and time spent in training mode were evaluated to compare the 2 groups. Additionally, the relationship between VR-based exam scores and time spent in training was assessed. Given the presence of multiple variables, a variety of statistical tests were used for data analyses based on the nature of the addressed question. To compare the performance change from pretest to posttest across the voice command and VR controller groups, a 2×2 mixed ANOVA was used. VR-based exam scores of voice command and VR controller groups were compared using 2-tailed, Welch independent-sample *t* test as the homogeneity of variances assumption was not met. Confidence ratings of the voice command and VR controller groups were compared using a Shapiro-Wilk test as the data were not normally distributed. Confidence bias was assessed using a 2-tailed, 1-sample *t* test. Presence scores of the groups and time spent in training mode across groups were compared using a 2-tailed, Student independent-sample *t* test.

Prior to analysis, the dataset was screened for outliers across all variables. No outliers were detected; thus, no data points were removed. The central tendency of the data was described using the mean. Effect sizes were reported as Cohen *d* (0.2=small, 0.5=medium, and 0.8=large) for *t* tests and η^2P (0.01=small, 0.06=medium, and 0.14=large) for ANOVA.

Each reported analyses were deemed appropriate based on data distribution and variance homogeneity. Data distribution was checked using the Shapiro-Wilk test of normality and homogeneity of variances was assessed by performing the Levene test of equality of variances. Details regarding the selection of the statistical analyses used can be found in Multimedia Appendix 4.

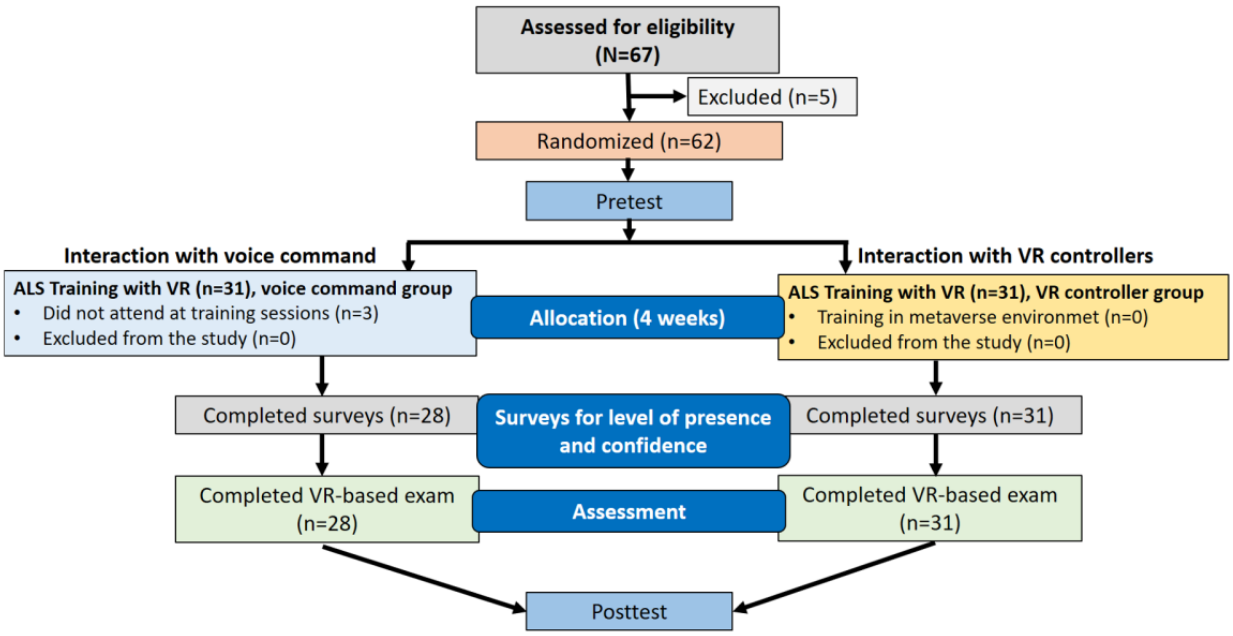
Data was analyzed using JASP (Version 0.19.3, JASP Team), an open-source software for statistical analysis. The figures were created using JASP.

Results

Overview

The following analysis was conducted with 59 participants, as shown in Figure 3.

Figure 3. CONSORT (Consolidated Standards of Reporting Trials) flow diagram of the study. ALS: advanced life support; VR: virtual reality.



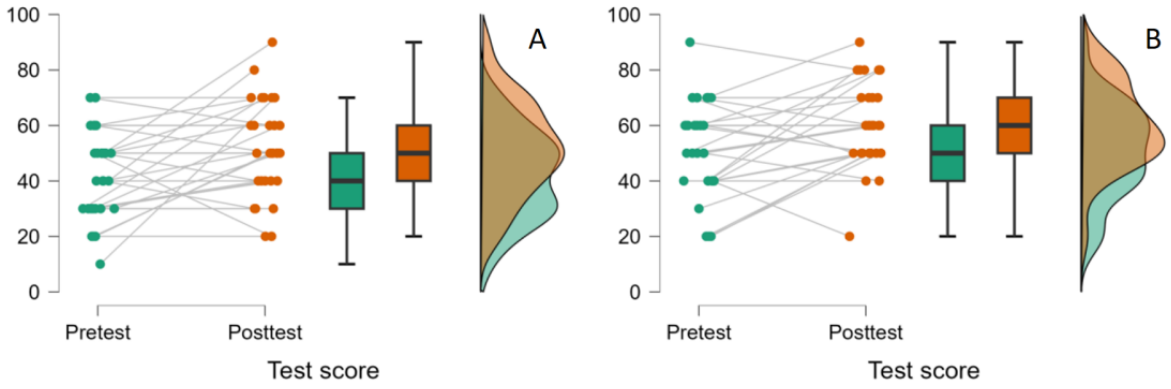
Change of Scores from Pretest to Posttest

To reveal whether performance change from the pretest to the posttest is different between groups, a 2 (group: voice command group vs VR controller group) × 2 (test: pretest vs posttest) mixed ANOVA where the first factor is the between-subjects variable and the second is the within-subjects variable was conducted. Levene test showed that the assumption of homogeneity of variances across groups was met for both pretest (P=.85) and posttest scores (P=.54). Furthermore, a Shapiro-Wilk test for normality showed that pretest scores of the voice command group (P=.14) and VR controller group (P=.15) were normally distributed. Similarly, posttest scores of both groups were also normally distributed ($P_{voice\ command}$

XSL•FO
RenderX

$_{group}$=.09 and $P_{VR\ controller\ group}$ =.43). Considering that the within-subjects variable has only 2 levels, a test for sphericity assumption was not required. Hence, the analysis was conducted without any correction as all assumptions were met. The results showed that the voice command group performed better than the VR controller group in general, ($F_{1,57}$=6.43; P=.01; n^2p=0.10). Moreover, performance was higher in the posttest than in the pretest for both groups ($F_{1,57}$=15.69; P<.001; n^2p=0.22). This increase in performance from pretest to posttest, however, was not different for voice command and VR controller groups, as the interaction between group and test was not significant ($F_{1,57}$= 0.05; P=.83; n^2p=8.18×10^{-4}). These findings indicate that while the voice command group had a higher performance in general, both groups showed a performance increment from pretest to posttest, the increase in performance was not higher in one group than the other (Figure 4).

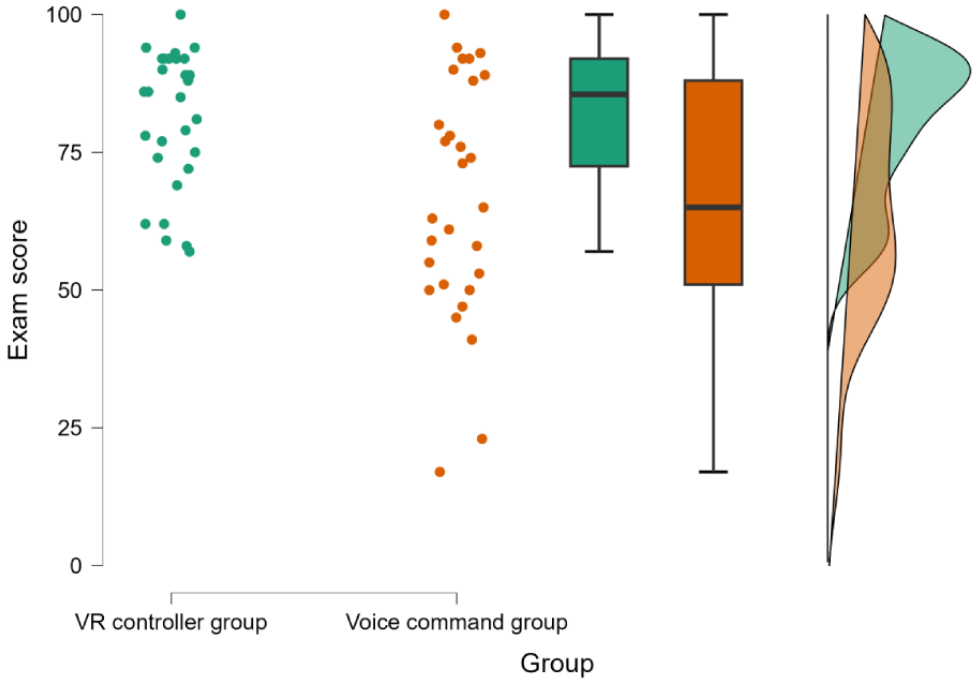
Figure 4. Pretest and posttest scores for (A) VR controller group and (B) voice command group, respectively. Dots indicate each participant's test score. The boxes indicate IQRs; the line in the middle of the boxes indicates the median. Lines between dots connect each participant's pretest and posttest scores. VR: virtual reality.



Comparison of VR-Based Exam Scores

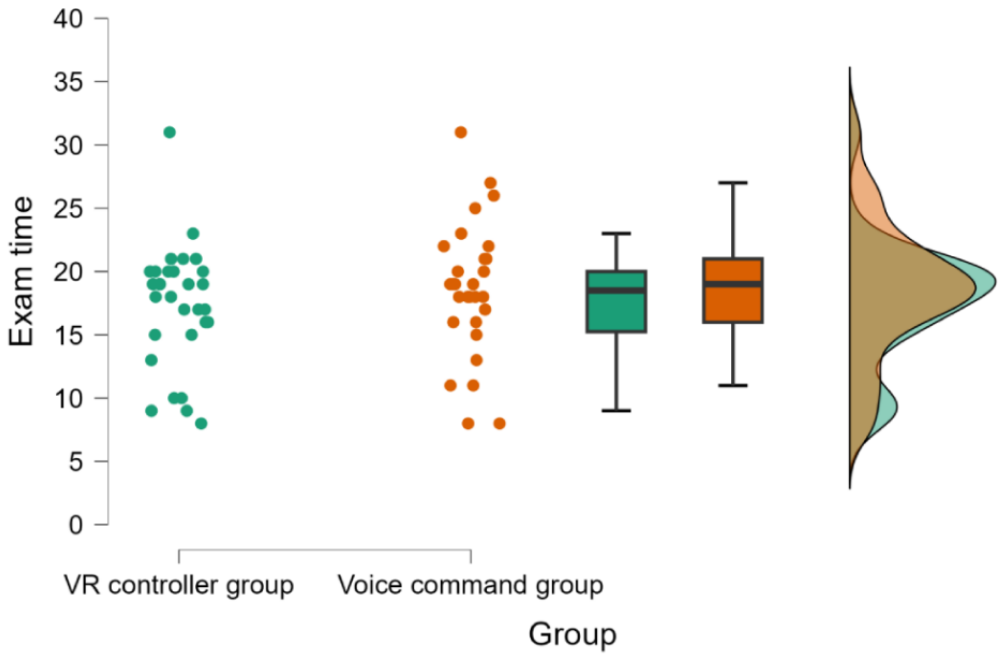
Crucial to the purpose of the study, an independent-sample, 2-tailed Welch t test showed that the VR controller group's performance in the VR-based exam (mean 80.47, SD 13.12) was higher than the voice command group's performance (mean 66.70, SD 21.65; $t_{45.83}$=2.94; P=.005; d=0.77, 95% CI 0.23-1.30; Figure 5). Moreover, both groups spent a similar amount of time in the exam (t_{57}=–0.95; P=.35; d=–0.25, 95% CI –0.76 to 0.27; voice command group: mean 18.59, SD 5.28 and VR controller group: mean 17.33, SD 4.83; Figure 6).

Figure 5. Scores of VR-based exam mode for VR controller and voice command groups. Dots indicate each participant's exam score. The boxes indicate IQR; the line in the middle indicates the median. VR: virtual reality.



XSL•FO
RenderX

Figure 6. Time spent on the advanced training stage for VR controller and voice command groups in minutes. Dots indicate each participant's time. The boxes indicate IQRs; the line in the middle indicates the median. VR: virtual reality.



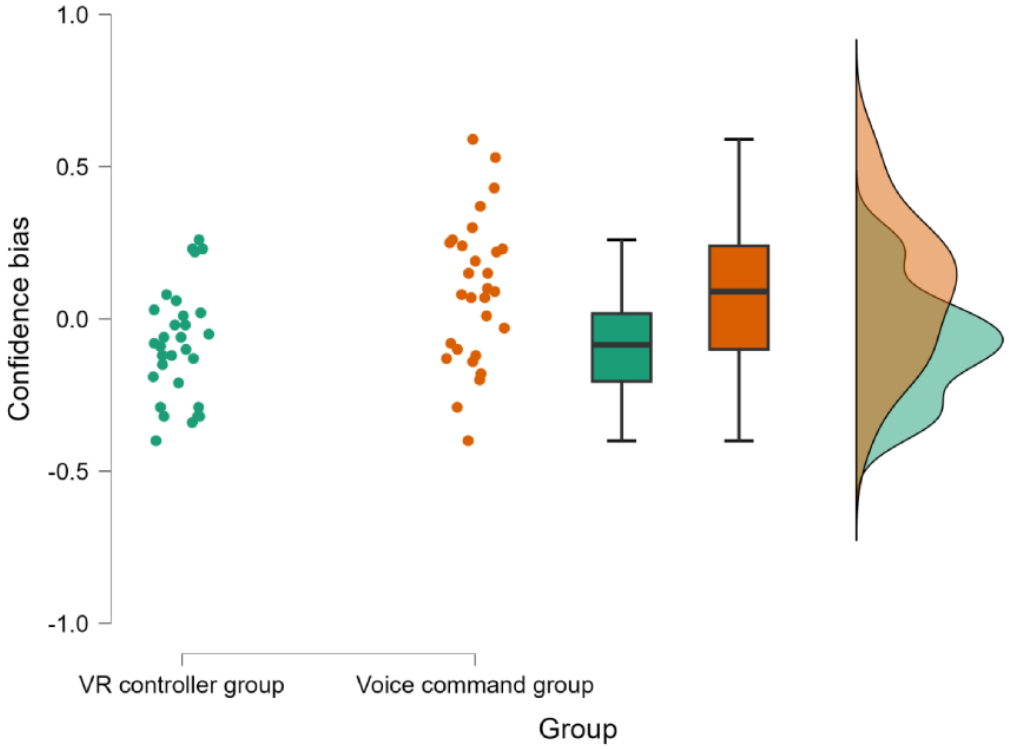
Confidence Ratings

Results revealed that there was no significant difference between groups' confidence ratings (W=392.00; *P*=.49, r_b=.15, 95% CI –0.38 to 0.20; voice command group: mean 3.79, SD 0.77; VR controller group: mean 3.60, SD 0.72).

In general, there is a positive relationship between performance and confidence levels, such that higher performance is accompanied by higher confidence levels [18,35]. Therefore, given the lower performance of the voice command group, it can be expected that their confidence ratings would also be lower. The absence of this expected difference points to a confidence bias in groups voice command group, that is, a mismatch between confidence and performance [36]. To see if this is the case, confidence bias for each participant was calculated by scaling their confidence ratings and advanced training performance over 1 and subtracting the scaled performance from the scaled confidence rating, a commonly used method for confidence bias quantification [37]. A 2-tailed, independent-sample *t* test showed a significant difference, with the voice command group's performance and confidence difference (mean 0.09, SD 0.24) being bigger than the VR controller group's performance and confidence difference (mean –0.09, SD 0.18; t_{57}=–3.21; *P*=.002; *d*=–0.83, 95 CI –1.37 to –0.30; Figure 7). To clarify the possible source of the group differences in confidence bias, a 2-tailed, 1-sample *t* test was conducted for each group. The results showed that the mean confidence bias of the VR controller group was significantly lower than 0, indicating an underconfidence bias (t_{29}=–2.56; *P*=.02; *d*=–0.47), and the mean confidence bias of the voice command group was significantly higher than 0, indicating an overconfidence bias (t_{28}=2.08; *P*=.047; *d*=0.39).

XSL•FO
RenderX

Figure 7. Confidence bias for VR controller and voice command groups. Dots indicate each participant's confidence bias. The boxes indicate IQRs; the line in the middle indicates the median. VR: virtual reality.



Comparison of Presence Ratings

The 2-tailed, independent-sample *t* test showed that the presence ratings did not differ across groups (t_{57}=1.16; *P*=.25; *d*=0.30, 95% CI –0.21 to 0.81; voice command group: mean 5.18, SD 0.83; VR controller group: mean 5.42, SD 0.75).

Discussion

Principal Findings

This study explored the potential of AI-driven voice control in VR-based ACLS training. The findings revealed that both groups improved from pretest to posttest and the VR controller group achieved higher scores in the VR-based exam compared to the voice command group. This suggests that traditional controllers may currently provide a more reliable interaction method in VR environments. Analyses confirmed that the data followed a normal distribution and satisfied the assumption of equal variances, with the exceptions being exam scores and confidence ratings, for which appropriate corrections were applied to address the assumption violations. These corrections, along with the reported effect sizes demonstrating medium to large effects across the majority of significant findings, underscored the robustness of the results and the adequacy of the sample size.

Despite the potential advantages, using voice command systems in VR gaming has also its limitations in speech recognition accuracy, latency, the system's ability to understand context-specific commands, problems caused by environmental noise, and the user's accent or speech patterns. Voice command systems in VR gaming present both promising advantages and notable limitations [17,38-41]. While these systems offer an intuitive, hands-free interaction method, their effectiveness can be hindered by factors like speech recognition accuracy, latency issues, context-specific command understanding, environmental noise, user's accent, and speech patterns. Errors in recognizing words or phrases can disrupt gameplay and may lead to reduced immersion. Delays in system response to voice commands can significantly impact the flow of VR experiences, particularly in fast-paced gaming scenarios where real-time interaction is crucial. Research shows that even minor delays can break the sense of presence and reduce overall engagement. The ability of a system to interpret context-specific commands remains a challenge. For instance, players might use commands with implied meanings or colloquial expressions that the system struggles to process, affecting functionality and the seamlessness of gameplay. Research highlights that background noise significantly impacts voice recognition performance. In noisy environments, players might need to repeat commands, which can be both disruptive and exhausting, diminishing the immersive qualities of VR gaming. Studies also show that accents, dialects, and individual speech patterns affect recognition accuracy. Systems trained on limited datasets may struggle with diverse user inputs, leading to unequal performance across different user demographics [17,38-41].

These literature data underscore the potential problems of AI-guided voice recognition in terms of accuracy and performance in different languages [42,43]. Since Turkish was selected as the language for AI-driven voice recognition in this study, there were similar difficulties due to the challenges of voice recognition, which caused participants in the voice command group to lose time and achieve lower scores compared to the VR controller group.

XSL•FO
RenderX

The level of presence did not significantly differ between the 2 groups, which contrasts with prior expectations that voice control would enhance immersion [18,19]. The level of presence may be hindered by problems in word recognition and delays in system response [17,38-41].

Confidence ratings were comparable across groups; however, an interesting finding was the presence of an overconfidence bias in the voice command group and an underconfidence bias in the VR controller group. This suggests that the mode of interaction can influence learners' self-perception of their performance. Aligning perceived competence with actual performance remains a critical goal for optimizing educational outcomes. While they are related, confidence bias and task performance are dissociable [36], making it important to examine both to better understand the learning process. Although both groups reported comparable confidence levels, the voice command group exhibited a significant overconfidence bias while the VR controller group exhibited a significant under-confidence bias. This discrepancy highlights the importance of aligning perceived competence with actual performance outcomes in educational settings [44] and indicates that the means of education can differentially affect this alignment.

Although current limitations in language recognition pose challenges, ongoing advancements in AI and NLP technologies may close these gaps and provide more effective solutions in the future. These limitations indicate that while voice command systems present innovative possibilities in VR gaming, developers need to address these issues to enhance the user experience and ensure accessibility for a wider audience.

Limitations

The first limitation of the study is the potential for selection bias, as participants were recruited from a single institution, which limits the generalizability of the results to other educational settings or demographics. The second limitation is the use of AI-driven voice recognition, which is prone to issues such as speech accuracy, environmental noise, and user-specific factors like accent, all of which could impact user experience and performance. A third limitation is the lack of longitudinal follow-up assessments to evaluate the long-term retention of skills, which is essential for assessing the effectiveness of VR-based training methods.

Conclusions

The results of this study highlight the effectiveness of VR-based training for ACLS. Although voice commands did not enhance the exam performance, level of presence, and confidence level of the participants in this study, voice commands showed a potential promise as an innovative interaction method. Addressing challenges such as language-specific recognition accuracy could significantly improve their application in VR-based educational modules.

Future research should focus on optimizing AI-driven interfaces and exploring their role in fostering learner immersion and confidence. Enhancing these aspects could lead to improved educational outcomes and better preparation for real-world scenarios in health care.

Acknowledgments

The authors express their gratitude to the biomedical support team at the Center of Advanced Simulation and Education of Acibadem Mehmet Ali Aydinlar University for their invaluable support during this study. The authors did not use generative AI to write any portion of the manuscript.

Data Availability

The datasets generated or analyzed during this study, as well as the serious gaming module, are available from the corresponding author upon request.

Authors' Contributions

DK was responsible for study conceptualization, methodology, and reviewing and editing the manuscript.
MEA was responsible for study conceptualization, methodology, and resources and drafting, reviewing, and correspondence. AEO was responsible for the software and methodology. TU was responsible for the recruitment of the participants and data collection. DCC was responsible for data analysis and survey preparation. TT was responsible for running the surveys.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Advanced cardiac life support (ACLS) pre- and posttest exams.
[PDF File (Adobe PDF File), 300 KB - games_v13i1e68272_app1.pdf]

Multimedia Appendix 2
The Turkish version of the questionnaire.

XSL•FO
RenderX

[PDF File (Adobe PDF File), 185 KB - games_v13i1e68272_app2.pdf]

Multimedia Appendix 3
Translated version of the questionnaire.
[PDF File (Adobe PDF File), 119 KB - games_v13i1e68272_app3.pdf]

Multimedia Appendix 4
Statistical data.
[PDF File (Adobe PDF File), 584 KB - games_v13i1e68272_app4.pdf]

Multimedia Appendix 5
CONSORT-EHEALTH checklist.
[PDF File (Adobe PDF File), 1226 KB - games_v13i1e68272_app5.pdf]

References

1. Tudor Car L, Kyaw BM, Teo A, Fox TE, Vimalesvaran S, Apfelbacher C, et al. Outcomes, measurement instruments, and their validity evidence in randomized controlled trials on virtual, augmented, and mixed reality in undergraduate medical education: systematic mapping review. JMIR Serious Games 2022;10(2):e29594 [FREE Full text] [doi: 10.2196/29594] [Medline: 35416789]
2. Reis G, Yilmaz M, Rambach J, Pagani A, Suarez-Ibarrola R, Miernik A, et al. Mixed reality applications in urology: requirements and future potential. Ann Med Surg 2021;66:102394 [FREE Full text] [doi: 10.1016/j.amsu.2021.102394] [Medline: 34040777]
3. Stepan K, Zeiger J, Hanchuk S, Del Signore A, Shrivastava R, Govindaraj S, et al. Immersive virtual reality as a teaching tool for neuroanatomy. Int Forum Allergy Rhinol 2017;7(10):1006-1013. [doi: 10.1002/alr.21986] [Medline: 28719062]
4. Aksoy E, Izzetoglu K, Onaral B, Kitapcioglu D, Sayali ME, Guven F. Assessing Correlation Between Virtual Reality Based Serious Gaming Performance and Cognitive Workload Changes via Functional Near Infrared Spectroscopy. Cham, Switzerland: Springer International Publishing; 2019.
5. Aksoy E. Comparing the effects on learning outcomes of tablet-based and virtual reality–based serious gaming modules for basic life support training: randomized trial. JMIR Serious Games 2019;7(2):e13442 [FREE Full text] [doi: 10.2196/13442] [Medline: 31042153]
6. Liberatore MJ, Wagner WP. Virtual, mixed, and augmented reality: a systematic review for immersive systems research. Virtual Reality 2021;25(3):773-799. [doi: 10.1007/s10055-020-00492-0]
7. Maas MJ, Hughes JM. Virtual, augmented and mixed reality in K-12 education: a review of the literature. Technol Pedagogy Educ 2020;29(2):231-249. [doi: 10.1080/1475939x.2020.1737210]
8. Aksoy ME, Özkan AE, Kitapcioglu D, Usseli T. Comparing the outcomes of virtual reality–based serious gaming and lecture-based training for advanced life support training: randomized controlled trial. JMIR Serious Games 2023;11:e46964 [FREE Full text] [doi: 10.2196/46964] [Medline: 37768719]
9. Lockey AY, Lin Y, Cheng A. Impact of adult advanced cardiac life support course participation on patient outcomes—a systematic review and meta-analysis. Resuscitation 2018;129:48-54. [doi: 10.1016/j.resuscitation.2018.05.034] [Medline: 29902494]
10. Moore N, Yoo S, Poronnik P, Brown M, Ahmadpour N. Exploring user needs in the development of a virtual reality–based advanced life support training platform: exploratory usability study. JMIR Serious Games 2020;8(3):e20797 [FREE Full text] [doi: 10.2196/20797] [Medline: 32763877]
11. Ishigaki SAK, Ismail AW, Abdul Halim NA, Suaib NM. Voice commands with virtual assistant in mixed reality telepresence. 2023 Presented at: ICAETA-2023: International Conference on Advanced Engineering, Technology and Applications; March 10-11, 2023; Istanbul, Turkey. [doi: 10.1007/978-3-031-50920-9_12]
12. Cha HS, Chang WD, Im CH. Deep-learning–based real-time silent speech recognition using facial electromyogram recorded around eyes for hands-free interfacing in a virtual reality environment. Virtual Reality 2022;26(3):1047-1057. [doi: 10.1007/s10055-021-00616-0]
13. Yang J, Chan M, Uribe-Quevedo; A, Kapralos B, Jaimes N, Dubrowski A. Prototyping virtual reality interactions in medical simulation employing speech recognition. 2020 Presented at: 2020 22nd Symposium on Virtual and Augmented Reality (SVR); Porto de Galinhas, Brazil; November 7-10, 2020. [doi: 10.1109/svr51698.2020.00059]
14. Masina F, Orso V, Pluchino P, Dainese G, Volpato S, Nelini C, et al. Investigating the accessibility of voice assistants with impaired users: mixed methods study. J Med Internet Res 2020;22(9):e18431 [FREE Full text] [doi: 10.2196/18431] [Medline: 32975525]
15. Zhu Z, Li S, You, J, Song Z. Application of speech recognition technology to virtual reality system. 2015 Presented at: 2015 International Industrial Informatics and Computer Engineering Conference (IIICEC 2015); January 10-11, 2015; Xi'an, China URL: https://www.atlantis-press.com/proceedings/iiicec-15/16911 [doi: 10.2991/iiicec-15.2015.25]

XSL•FO
RenderX

16. Meyer J, Dentel L, Meunier F. Speech recognition in natural background noise. PLoS ONE 2013;8(11):e79279. [doi: 10.1371/journal.pone.0079279]
17. Li D, Gao Y, Zhu C, Wang Q, Wang R. Improving speech recognition performance in noisy environments by enhancing lip reading accuracy. Sensors 2023;23(4):2053 [FREE Full text] [doi: 10.3390/s23042053] [Medline: 36850648]
18. McCurdy LY, Maniscalco B, Metcalfe J, Liu KY, de Lange FP, Lau H. Anatomical coupling between distinct metacognitive systems for memory and visual perception. J Neurosci 2013;33(5):1897-1906. [doi: 10.1523/jneurosci.1890-12.2013]
19. Osking H, Doucette JA. Enhancing emotional effectiveness of virtual-reality experiences with voice control interfaces. 2019 Presented at: 5th International Conference of the Immersive Learning Network (iLRN 2019); June 23–27, 2019; London, United Kingdom p. 199-209. [doi: 10.1007/978-3-030-23089-0_15]
20. de Bruin ABH, Dunlosky J, Cavalcanti RB. Monitoring and regulation of learning in medical education: the need for predictive cues. Med Educ 2017;51(6):575-584 [FREE Full text] [doi: 10.1111/medu.13267] [Medline: 28332224]
21. Panchal AR, Bartos JA, Cabañas JG, Donnino MW, Drennan IR, Hirsch KG, et al. Part 3: adult basic and advanced life support: 2020 American Heart Association guidelines for cardiopulmonary resuscitation and emergency cardiovascular care. Circulation 2020;142:S366-S468 [FREE Full text] [doi: 10.1161/CIR.0000000000000916] [Medline: 33081529]
22. Gokoglu S, Cakiroglu U. Measuring the sense of presence in virtual reality-based learning environments: adaptation of the presence scale to Turkish. Eğitim Teknolojisi Kuram ve Uygulama 2019;9(1):169-188. [doi: 10.17943/etku.441497]
23. Witmer BG, Singer MJ. Measuring presence in virtual environments: a presence questionnaire. Presence 1998;7(3):225-240. [doi: 10.1162/105474698565686]
24. Witmer BG, Jerome CJ, Singer MJ. The factor structure of the presence questionnaire. Presence 2005;14(3):298-312. [doi: 10.1162/105474605323384654]
25. Kane AD, Nolan JP. Changes to the European Resuscitation Council guidelines for adult resuscitation. BJA Educ 2022;22(7):265-272 [FREE Full text] [doi: 10.1016/j.bjae.2022.02.004] [Medline: 35754855]
26. Krage R, Zwaan L, Tjon Soei Len L, Kolenbrander MW, van Groeningen D, Loer SA, et al. Relationship between non-technical skills and technical performance during cardiopulmonary resuscitation: Does stress have an influence? Emerg Med J 2017;34(11):728-733 [FREE Full text] [doi: 10.1136/emermed-2016-205754] [Medline: 28844039]
27. Peltonen V, Peltonen L, Salanterä S, Hoppu S, Elomaa J, Pappila T, et al. An observational study of technical and non-technical skills in advanced life support in the clinical setting. Resuscitation 2020;153:162-168. [doi: 10.1016/j.resuscitation.2020.06.010] [Medline: 32561474]
28. Saravana-Bawan BB, Fulton C, Riley B, Katulka J, King S, Paton-Gay D, et al. Evaluating best methods for crisis resource management education: didactic teaching or noncontextual active learning. Simul Healthc 2019;14(6):366-371. [doi: 10.1097/SIH.0000000000000388] [Medline: 31490864]
29. Arancibia Muñoz ML, Halal Orfali C. The role of learning management systems (LMS) in educational environments: didactic and technical use. 2018 Presented at: 11th Annual International Conference of Education, Research and Innovation; November 12-14, 2018; Seville, Spain. [doi: 10.21125/iceri.2018.0976]
30. García Clemente, FJ, de la Torre L, Dormido S, Salzmann C, Gillet D. Collecting Experience Data From Remotely Hosted Learning Applications. Cham, Switzerland: Springer; 2018.
31. Learning Record Store (LRS). Rustici Software. URL: https://xapi.com/learning-record-store/ [accessed 2023-02-28]
32. Akbari F, Nasiri M, Rashidi N, Zonoori S, Amirmohseni L, Eslami J, et al. Comparison of the effects of virtual training by serious game and lecture on operating room novices' knowledge and performance about surgical instruments setup: a multi-center, two-arm study. BMC Med Educ 2022;22(1):268 [FREE Full text] [doi: 10.1186/s12909-022-03351-5] [Medline: 35410279]
33. Unity Technologies. URL: https://unity.com/ [accessed 2025-02-01]
34. wit.ai. URL: https://wit.ai/ [accessed 2025-02-01]
35. Rouault M, McWilliams A, Allen MG, Fleming SM. Human metacognition across domains: insights from individual differences and neuroimaging. Personal Neurosci 2018;1:e17 [FREE Full text] [doi: 10.1017/pen.2018.16] [Medline: 30411087]
36. Fleming SM, Lau HC. How to measure metacognition. Front Hum Neurosci 2014;8:443 [FREE Full text] [doi: 10.3389/fnhum.2014.00443] [Medline: 25076880]
37. Schraw G, Dunkle ME, Bendixen LD, Roedel TD. Does a general monitoring skill exist? J Educ Psychol 1995;87(3):433-444. [doi: 10.1037//0022-0663.87.3.433]
38. Basak S, Himanshi Agrawal H, Jena S, Gite S, Bachute M, Pradhan B, et al. Challenges and limitations in speech recognition technology: a critical review of speech signal processing algorithms, tools and systems. Comput Model Eng Sci 2023;135(2):1035-1089. [doi: 10.32604/cmes.2022.021755]
39. Onitilo AA, Shour AR, Puthoff DS, Tanimu Y, Joseph A, Sheehan MT. Evaluating the adoption of voice recognition technology for real-time dictation in a rural healthcare system: a retrospective analysis of dragon medical one. PLoS One 2023;18(3):e0272545 [FREE Full text] [doi: 10.1371/journal.pone.0272545] [Medline: 36952436]
40. Zargham N, Lamine Fetn M, Spillner L, Muender T, Malaka R. "I Know what you mean": Context-aware recognition to enhance speech-based games. 2024 Presented at: CHI '24: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems; May 11-16, 2024; Honolulu, HI. [doi: 10.1145/3613904.3642426]

XSL•FO
RenderX

41. Luo J. Research on speech recognition system based on virtual reality technology. 2023 Presented at: Proceedings of the 2023 4th International Conference on Education, Knowledge and Information Management (ICEKIM 2023); May 26-28, 2023; Nanjing, China.
42. Fang K, Wang J, Fang K. Interactive design with gesture and voice recognition in virtual teaching environments. IEEE Access 2024;12:4213-4224. [doi: 10.1109/access.2023.3348846]
43. Arslan RS, Barisci N. A detailed survey of Turkish automatic speech recognition. Turk J Elec Eng Comp Sci 2020;28(6):3253-3269. [doi: 10.3906/elk-2001-38]
44. Schraw G, Moshman D. Metacognitive theories. Educ Psychol Rev 1995;7(4):351-371. [doi: 10.1007/bf02212307]

Abbreviations

ACLS: advanced cardiac life support
AI: artificial intelligence
ALS: advanced life support
NLP: natural language processing
PQ: Presence Questionnaire
VR: virtual reality

Edited by A Coristine; submitted 01.11.24; peer-reviewed by O Tanriover, O Cakmakkaya, M Tanriverdi; comments to author 26.11.24; revised version received 27.12.24; accepted 24.01.25; published 14.02.25.

Please cite as:
Kitapcioglu D, Aksoy ME, Ozkan AE, Usseli T, Cabuk Colak D, Torun T
Enhancing Immersion in Virtual Reality–Based Advanced Life Support Training: Randomized Controlled Trial
JMIR Serious Games 2025;13:e68272
URL: https://games.jmir.org/2025/1/e68272
doi:10.2196/68272
PMID:

©Dilek Kitapcioglu, Mehmet Emin Aksoy, Arun Ekin Ozkan, Tuba Usseli, Dilan Cabuk Colak, Tugrul Torun. Originally published in JMIR Serious Games (https://games.jmir.org), 14.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Diaphragmatic Breathing Interfaces to Promote Relaxation for Mitigating Insomnia: Pilot Study

Yi-Jen Lai[1*], MDes; Hsiao-Yean Chiu[2*], PhD; Ko-Chiu Wu[1], PhD; Chun-Wei Chang[1], MDes

[1]Department of Interaction Design, National Taipei University of Technology, Taipei, Taiwan

[2]School of Nursing, Taipei Medical University, Taipei, Taiwan

[*]these authors contributed equally

Corresponding Author:
Ko-Chiu Wu, PhD
Department of Interaction Design
National Taipei University of Technology
Rm. 701-4, Design Building,
No.1 Sec.3 Zhongxiao E Rd, Da'an District
Taipei, 10608
Taiwan
Phone: 886 02 2771 2171 ext 4574
Fax: 886 02 8773 2913
Email: kochiuwu@mail.ntut.edu.tw

Abstract

Background: Brief behavioral treatment for insomnia is an effective short-term therapy focusing on stimulus control and sleep restriction to enhance sleep quality. As a crucial part of this therapy, diaphragmatic breathing is often recommended when patients fail to fall asleep within 30 minutes. With the rise of health apps and gamification, these tools are increasingly seen as effective ways to boost self-efficacy and user engagement; however, traditional games tend to increase attention, which can negatively impact sleep and contradicts the aim of sleep therapy. This study thus explored the potential for gamification techniques to promote relaxation without disrupting sleep processes.

Objective: The study developed 4 breathing guidance mechanisms, ranging from concrete to abstract: number countdown, zoom-in/out, up/down, and color gradients. The objective was to explore the relationship between game mechanics, cognitive load, relaxation effects, and attention as well as to understand how different designs impact users with varying levels of insomnia.

Methods: The study was conducted in 2 phases. The first phase involved a questionnaire on the 4 guidance mechanisms. In the second phase, 33 participants classified by insomnia severity completed a Sleep Self-Efficacy Scale. They then engaged in 5 minutes of diaphragmatic breathing using each of the 4 interfaces. Relaxation effects were measured using heart rate variability via a smartwatch, attention and relaxation levels via an electroencephalogram device, and respiratory rate via a smartphone. Participants also completed the Game Experience Questionnaire and NASA Task Load Index, followed by user interviews.

Results: The results indicated that competence, immersion, and challenge significantly influenced cognitive load. Specifically, competence and immersion reduced cognitive load, while challenge, negative affect, and positive affect were correlated with relaxation. Negative affect showed a positive correlation with the mean root mean square of successive differences, while positive affect exhibited a negative correlation with the mean root mean square of successive differences. Cognitive load was found to affect both relaxation and attention, with a negative correlation between mental demand and attention and a positive correlation between temporal demand and respiratory rate. Sleep self-efficacy was negatively correlated with temporal demand and negative affect and positively correlated with competence and immersion.

Conclusions: Interfaces offering moderate variability and neither overly abstract nor too concrete guidance are preferable. The up/down interface was most effective, showing the best overall relaxation effect. Conversely, the number countdown interface was stress-inducing, while the zoom-in/out interface had a significant impact on insomnia-related issues, making them less suitable for insomnia-related breathing exercises. Participants showed considerable variability in their response to the color gradient interface. These findings underscore the importance of carefully considering game design elements in relaxation training. It is essential that breathing guidance designs account for the impact of the game experience to effectively promote relaxation in users.

XSL•FO
RenderX

(JMIR Serious Games 2025;13:e67000) doi:10.2196/67000

KEYWORDS

brief behavioral treatment for insomnia; sleep self-efficacy; mobile health; mHealth; breathing training cognitive load; attention; gamification; diaphragmatic breathing; insomnia; sleep; games; relaxation; breathing; breathing guidance; questionnaire; mental; cognition

Introduction

Background

Insomnia is an increasingly severe issue, with approximately 30% of the population experiencing related symptoms [1]. It is characterized by difficulty in falling asleep, maintaining sleep, or waking up too early and being unable to return to sleep for at least 3 nights a week. This definition is based on the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* criteria. Acute insomnia lasts for 1 month, while chronic insomnia persists for over 3 months [2]. Cognitive behavioral therapy for insomnia (CBT-I) is the most common behavioral treatment [3]. CBT-I typically involves 4 to 8 sessions focusing on changing thoughts, emotions, and behaviors related to sleep [4]. Derived from CBT-I, the brief behavioral treatment for insomnia (BBTI) condenses the process into 4 weeks, focusing on sleep restriction and stimulus control [5]. Compared with CBT-I, BBTI requires less time and resources, making it accessible to a wider range of patients.

Many people attempt to compensate for poor sleep patterns by spending excessive time in bed, which often increases anxiety about sleep. Sleep restriction reduces the time spent in bed to approximately match the actual sleep time, while stimulus control helps to reestablish the connection between the bed and sleep, reducing anxiety and frustration [4]. Relaxation training is an important adjunct to insomnia treatment, particularly diaphragmatic breathing. Slow breathing can regulate the autonomic nervous system, increasing heart rate variability (HRV) and electroencephalogram (EEG) α power, leading to a state of calm and relaxation [6].

A variety of health apps focus on insomnia treatment, such as sleep diaries; however, there are few relaxation training apps specifically designed for insomniacs. Research on breathing relaxation has often focused on enhancing breathing techniques and analyzing different breathing guidance designs. One study examined 3 types of breathing guidance designs, including voice-only guidance and 2 combinations of voice and visuals. The results showed that wave visualization designs outperformed voice-only designs in improving both objective breathing performance and user experience [7]. Similarly, Faust-Christmann et al's [8] study confirmed that visual guidance effectively helps users achieve breathing goals and suggested further exploration of cognitive levels during training sessions as well as other psychophysiological data. A study on breathing training in a virtual reality environment revealed that users' progress was primarily concentrated in the first 2 sessions. Moreover, participants' subjective impression of progress was found to be more critical than objective improvements, as it significantly enhanced their motivation and willingness to participate [9].

The main challenge for health apps is ensuring long-term regular use. In this endeavor, understanding user needs is crucial [10]. Users initially engage with health apps to meet health needs, but over time, enjoyment becomes increasingly important [11]. Gamification, which involves adding interactive elements to nongame contexts [12], is often used in health apps to attract patients through engaging interactions, promote adherence, and facilitate positive behavioral change [13].

Albert Bandura proposed the theory of self-efficacy, which refers to the ability to perceive oneself as being able to achieve a goal. This theory is used to measure the degree or strength of an individual's belief in their ability to accomplish a task [14]. Feedback from games can enhance self-efficacy, thereby affecting user participation and persistence in activities [15]. Virtual reality biofeedback games can help to increase self-efficacy and reduce stress; there is evidence that higher self-efficacy is correlated with an increased ability to relax postgaming [16]. Enhanced self-efficacy also supports adherence to BBTI, while low self-efficacy is associated with poorer compliance with stimulus control therapy and sleep hygiene [17].

Sleep self-efficacy, an extension of general self-efficacy, measures an individual's perceived ability to influence their own sleep behavior to achieve healthy sleep. The Sleep Self-Efficacy Scale, which comprises 9 items, assesses this perception [18]. As insomnia severity, poor health, depression, and misconceptions about sleep increase, sleep self-efficacy decreases [19]. Both general self-efficacy and sleep self-efficacy are predictors of sleep health, with higher levels associated with better sleep quality, regularity, and scheduling [20].

Sweller [21] proposed the concept of cognitive load. It assumes that human cognitive resources are limited and that excessive cognitive load may affect the effectiveness of learning and task performance. Increased cognitive load before sleep can lead to arousal and longer sleep latency, affecting sleep quality [22]. Attention is also closely linked to sleep; the cognitive model of insomnia suggests that insomniacs often over-concern themselves with sleep, focusing attention on perceived sleep threats, such as bodily sensations or environmental factors, which can increase arousal and anxiety, leading to misconceptions that disrupt sleep [23]. Sleep should be an unconscious process [24]. Conscious attempts to fall asleep may disrupt sleep, and insomniacs are more likely than good sleepers to have difficulty falling asleep due to anxiety and environmental disruptions. Insomniacs are also more likely to think about not being able to fall asleep or things that have happened during the day before they go to bed. Using the NASA Task Load Index (NASA-TLX) scale to assess the association between serious gaming and cognitive load, Sevcenko [25] found that both task complexity and scene difficulty affected participants' cognitive load and impacted gaming performance.

XSL•FO
RenderX

With the proliferation of health apps and medical gamification, gamifying relaxation training has become an effective way to enhance training outcomes. Many apps target breathing training; however, few focus on patients with insomnia and presleep use, as traditional games often increase attention or excitement, which can interfere with sleep.

Objectives

This study aims to investigate how different breathing guidance mechanisms impact relaxation and whether these effects are related to the severity of insomnia and sleep self-efficacy. The selected variables were cognitive load, attention, gaming experience, and relaxation levels achieved through breathing guidance.

We hypothesize that concrete breathing guidance interfaces are more effective than abstract guidance in reducing respiration rates. Abstract breathing guidance interfaces are associated with lower cognitive load compared with concrete interfaces. In addition, we hypothesized that the positive dimensions of the Game Experience Questionnaire (GEQ; eg, competence, immersion, flow, and positive affect) are negatively associated with cognitive load, while the negative dimensions (eg, frustration, challenge, and negative affect) are positively associated with cognitive load.

Methods

Four Guidance Mechanisms

In this study, various breathing-related games were categorized into 4 guidance mechanisms: countdown, zoom-in/out, up/down movements, and color gradients. These methods range from concrete to abstract guidance and from small to large visual variations.

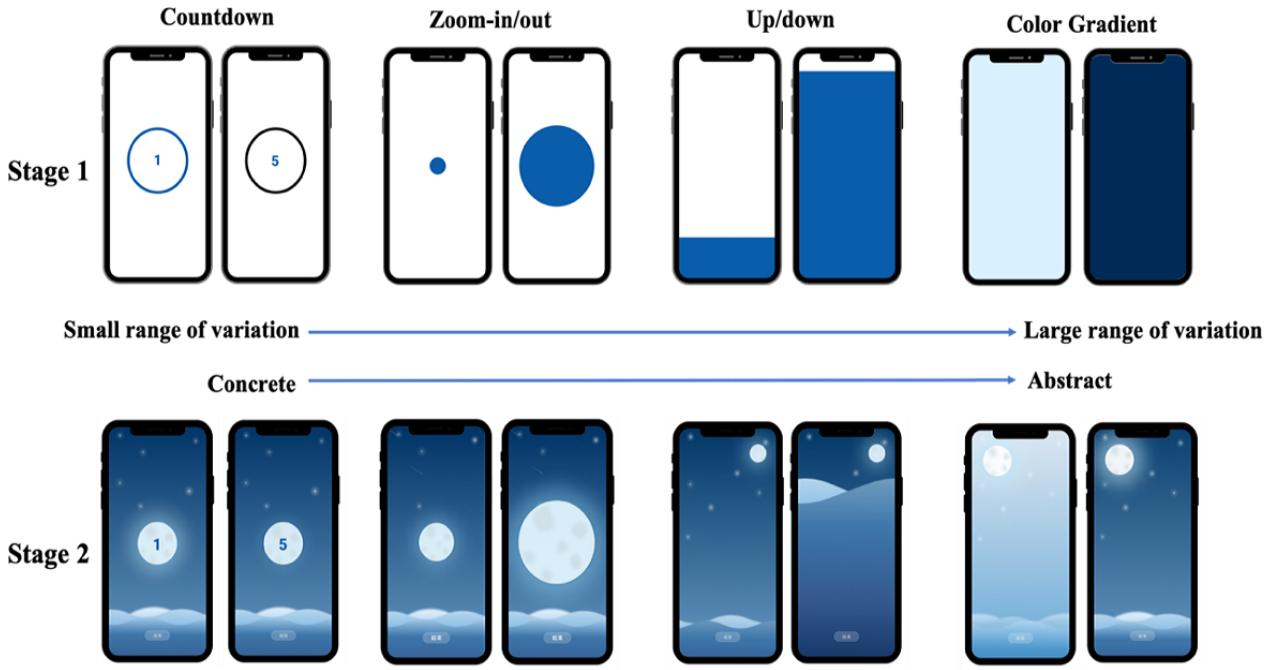
1. Countdown: Guidance is provided through a numerical countdown on the screen.
2. Zoom-in/out: Guidance is provided through the enlargement and reduction of a graphic, with enlargement indicating inhalation and reduction indicating exhalation.
3. Up/down movements: Guidance is provided through the vertical movement of the screen, with upward movement indicating inhalation and downward movement indicating exhalation.
4. Color gradients: Guidance is provided through changes in the color gradient on the screen.

Interface Design

Main Visuals

The interface for relaxation training through diaphragmatic breathing features nighttime ocean waves as the main visual (Figure 1). The color blue is associated with many positive emotions, including relaxation, calmness, happiness, and comfort [26], because blue is often linked to natural elements such as the ocean, water, and sky, which evoke feelings of relaxation and tranquility. It has also been found that blue is the most favored color among Asian populations [27]. In addition, images of natural landscapes produce higher levels of relaxation and are less stimulating than images of urban landscapes [28]. Among images of deserts, forests, snow, and water, researchers found images with water elements were the most relaxing [25].

Figure 1. Prototypes of interface design for 4 guidance mechanisms.



Music

The game uses white noise from ocean waves combined with relaxing music as background audio. White noise has been proven to improve sleep quality by reducing environmental noise and inducing relaxation [29]. Compared with natural white noise, natural soundscapes, and ambient music, listening to instrumental music or instrumental music mixed with natural soundscapes has been found to be the most relaxing [27]. The combination of natural white noise and music enhances the

XSL•FO
RenderX

sense of immersion and stimulates imagination. This allows participants to visualize the natural environment in their mind, creating a sense of being in the natural environment, while the diversity of the music provides a better listening experience.

Guidance Mechanism

Diaphragmatic breathing exercises are conducted in 5-minute sessions to ensure accurate HRV measurements [30]. The interface guides the user through 6 breaths per minute, as 6 breaths per minute achieve the highest HRV [31]. It also controls vagal activity, which helps insomniacs maintain a stable sleep cycle and shorten the latency to sleep through regular breathing, improving the quality of sleep [32].

Measurement

Overview

In the experiment, we used a Garmin smartwatch to measure HRV and paired it with a BrainLink dynamic EEG headset (60 Hz) to assess relaxation and attention levels during gameplay, relaxation and attention scores range from 0 to 100. The Google Fit mobile app was used to measure breathing rates before and after the game sessions. All these devices were noninvasive and collected physiological signals from users during the relaxation game, which were then used for subsequent statistical analysis.

Breathing Rate

The normal resting breathing rate for an adult is between 12 and 20 breaths per minute. However, breathing at a rate of 6 breaths per minute has been shown to achieve the highest HRV. Therefore, this study measured the changes in breathing rates before and after using the interface to determine which guiding method is most effective in reducing breathing rates.

Relaxation Level

HRV is an indicator of relaxation. The standard deviation of normal to normal intervals (SDNN) is used for assessing overall variability and heart health, whereas the root mean square of successive differences (RMSSD) is more influenced by the parasympathetic nervous system and is a suitable method for quantifying short-term HRV [33,34]. α power is closely related to psychological relaxation states. α power activity increases significantly when the individual is in a relaxed state. EEG data can also reflect specific psychological and emotional states [35-38]. Using a smartwatch to measure heart rate and HRV provides a direct reflection of physiological states, while EEG data reveal psychological relaxation activities and changes in attention during the relaxation process.

Questionnaire

Insomnia Severity Index

This scale assessed the severity of insomnia [39] in participants. Scores ranging from 0 to 7 indicate “no insomnia”; scores ranging from 8 to 14 indicate “mild insomnia”; and scores of 15 and above indicate “moderate to severe insomnia.”

NASA Task Load Index

This scale evaluated the cognitive load [40] experienced by users under different breathing guidance methods during the diaphragmatic breathing game. It consists of 6 items measuring mental, physical, and temporal demands as well as performance, effort, and frustration. Scores range from 0 to 100, with lower scores indicating better performance or lower subjective load in each dimension.

AttrakDiff Mini

AttrakDiff [41] is a tool used to assess user experience, analyzing the pragmatic quality (PQ), hedonic quality-stimulation (HQ-S), hedonic quality-identity (HQ-I), and attractiveness of interactive products. The full scale contains 28 items, while the AttrakDiff Mini reduces this to 10 items. This study used a 7-point scale (–3: negative, 3: positive) to measure the PQ, HQ-S, HQ-I, and attractiveness of the interfaces.

Sleep Self-Efficacy Scale

This scale [18] consisted of 9 items assessing participants’ confidence in their ability to sleep. The higher the total score, the more confident participants are in their ability to sleep.

GEQ

The GEQ [42] assessed players’ game experiences, containing 33 items covering 7 dimensions: competence, sensory and imaginative immersion, flow, tension, challenge, negative affect, and positive affect. This study used the GEQ to evaluate subjective game experiences across different game guidance methods and to analyze which aspects of the game influence relaxation.

Experimental Procedure

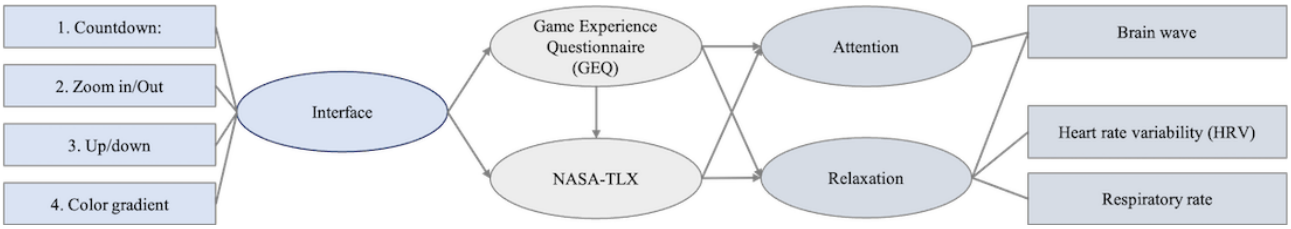
The anonymous questionnaire items covered age, gender, work status, past experience with relaxation apps, and the Insomnia Severity Index (ISI) to assess participants’ insomnia severity. Afterward, the participants watched videos of the 4 different breathing modalities following sequences: countdown, zoom-in/out, up/down movements, and color gradients, each of which was a 30-second video of each of the 4 interfaces, and then filled out the NASA-TLX workload scale and the AttrakDiff Mini scale to assess the cognitive load and experience of each modality. This process was repeated 4 times, once for each breathing guidance method. Based on the results and feedback from the first phase, the diaphragmatic breathing interfaces were refined and improved.

In the second phase, participants were recruited to perform 5-minute diaphragmatic breathing sessions. Participants included individuals with no insomnia, mild insomnia, and severe insomnia. They first completed a basic information form, which included demographic data, the ISI, and the Sleep Self-Efficacy Scale. Participants wore a Garmin smartwatch to collect HRV and a BrainLink Dynamic Brainwave Instrument to measure mental relaxation and attention as well as used a cell phone app, Google Fit, to measure respiratory rate before and after the experiment. Using a Latin square design to randomize the order, participants performed diaphragmatic breathing exercises with 4 different interfaces, each for 5 minutes. After each session, participants completed the NASA-TLX scale and the GEQ to evaluate the cognitive load and user experience of each interface. Upon completing all sessions, semistructured interviews were conducted to gather feedback on the design of the game

XSL•FO
RenderX

mechanisms and preferences as well as explore possible reasons for relaxation during the game. This experimental design allowed for a comprehensive evaluation of how participants with varying insomnia severity responded to the 4 breathing guidance methods, using both physiological signals and subjective assessments to identify the most effective relaxation training methods. Figure 2 shows the correlogram of the second stage of the experiment, which assessed the response of participants with different severities of insomnia to the 4 types of breathing guidance, and identified the most effective relaxation training method through physiological signals and subjective ratings.

Figure 2. Second phase: correlation between questionnaires and physiological signals. NASA-TLX: NASA Task Load Index.



Ethical Considerations

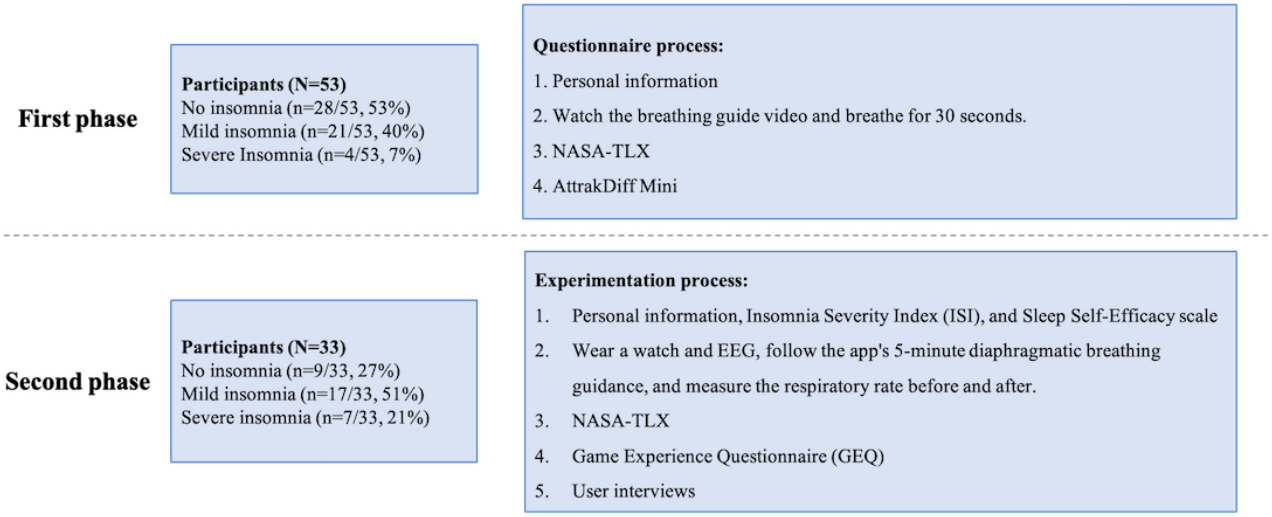
The study was approved by the TMU-Joint Institutional Review Board of Taipei Medical University and its affiliated hospital under the general review case number N20243070. Before participating in the second phase of the experiment, all participants provided written informed consent, confirming their full understanding of the study procedures and their agreement to participate. Additionally, participants were asked whether they had any symptoms listed in the exclusion criteria. All data collected were anonymized and deidentified to ensure the privacy and confidentiality of participants. No personally identifiable information was retained or shared during or after the study. Participants were provided with a gift voucher for a convenience store as a reward for their participation. Even if participants were unable to complete the experiment for any reason, the gift voucher was still given, and there was no requirement to return it. No images of individual participants or identifiable features are included in the paper or Multimedia Appendix 1.

Results

Overview

In the first phase of the breathing guidance prototype survey, conducted between January and February 2024, a total of 53 valid questionnaires were collected. The participants ranged in age from 18 to 65 years and included both genders. Among them, 28 (52.8%) participants had no insomnia, 21 (36.9%) participants had mild insomnia, and 4 (7.5%) participants had severe insomnia. In the second phase of the interface design experiment, conducted from May to July 2024, 33 participants were recruited, ranging in age from 20 to 65 years, including both genders, without excluding any participants. Among these participants, 9 (27.3%) participants had no insomnia, 17 (51.5%) participants had mild insomnia, and 7 (21.2%) participants had severe insomnia. The experimental steps and the number of participants are shown in Figure 3.

Figure 3. Number of participants and experimental procedure for 2-phase study. EEG: electroencephalogram; NASA-TLX: NASA Task Load Index.



Changes in Breathing Rate

As shown in Table 1, breathing rates decreased the least (1.13) after participants used the zoom-in/out interface. Breathing rates decreased respectively by 1.41 and 1.53 after participants used the countdown interface and the up/down interface. Breathing rates decreased the most (1.75) after participants used the color gradient interface. Paired sample *t* tests were used to analyze breathing rates before and after the intervention. There was a significant difference in breathing rates before and after participants used the up/down interface (P=.04) and the color

XSL•FO
RenderX

gradient interface (P=.02). This indicates that these 2 guidance methods can effectively regulate breathing frequency through relaxation training.

Table 1. Comparison of breathing rates before and after diaphragmatic breathing training.

Interface style	Mean (pretraining)	Mean (posttraining)	Mean difference (SD)	95% CI	*t* value	Significance (2-tailed)
Countdown	13.47	12.06	1.41 (4.063)	–0.059 to 2.871	1.958	0.059
Zoom-in/out	12.91	11.78	1.13 (4.549)	–0.515 to 2.765	1.399	0.172
Up/down	13.00	11.47	1.53 (3.992)	0.092 to 2.970	2.170	0.038
Color gradients	13.34	11.59	1.75 (3.976)	0.317 to 3.183	2.490	0.018

Interface Measurement

Figure 4 describes the average cognitive load, use experience, and gaming experience of the 4 pretest interventions. According to the NASA-TLX (1) results from the first phase of the experiment, the cognitive load induced by the countdown interface and the up/down interface was 39.86 and 35.14, respectively. The zoom-in/out interface induced the lowest cognitive load at an average of 29.72, while the color gradient interface induced the highest cognitive load at an average of 43.35. The AttrakDiff Mini results showed that overall satisfaction was the lowest for the countdown interface, particularly in HQ-I and HQ-S scores, which were negative. The zoom-in/out interface had the highest overall satisfaction, while the color gradient interface had the lowest PQ among the 4 interfaces.

Figure 4. Mean values of NASA-TLX (measuring cognitive load), Attrakdiff Mini (assessing user experience), and the Game Experience Questionnaire (evaluating gaming experience) across 4 interfaces. NASA-TLX: NASA Task Load Index.



In the second phase of the experiment, the NASA-TLX (2) results showed that the countdown interface induced the highest overall cognitive load, with high scores in psychological, physical, time, and frustration loads. The zoom-in/out interface induced the second-highest cognitive load, with the highest scores in performance and effort, indicating that this interface led to poorer performance and required more effort. The up/down interface induced the second lowest overall cognitive load, with the lowest scores in time pressure, effort, and frustration. The color gradient interface induced the lowest cognitive load, with the lowest scores in psychological, physical, and performance loads, indicating that it was less likely to cause psychological and physical strain.

Physiological Measurements

Table 2 illustrates the results of relaxation measurements from the second phase. The data included the average respiratory rate per minute after 5 minutes of diaphragmatic breathing, average heart rate, average SDNN, and average RMSSD measured during diaphragmatic breathing using a smartwatch as well as relaxation level, relaxation stability, attention, and attention stability measured using an EEG device. For the up-down interface, the respiratory rate and average heart rate were the lowest while the average RMSSD was the highest, indicating that the up-down interface provides better breathing guidance and relaxation effects. Additionally, the zoom-in/out interface

XSL•FO
RenderX

showed the highest psychological relaxation, the up-down interface had the highest relaxation stability and attention stability, and the color gradient interface had the highest attention.

Table 2. Physiological and psychological measurements of 4 interfaces.

Interface style and physiological signal	Minimum	Maximum	Mean (SD)
Countdown			
Breathing rate (bpm)	6	22	12.06 (3.98)
Mean heart rate (bpm)	62	107	81.58 (10.12)
Mean SDNN[a] (ms)	9.89	90.41	45.74 (16.09)
Mean RMSSD[b] (ms)	15.75	74.84	29.37 (12.08)
Relaxation	28	66	52.85 (7.95)
Relaxation stability	42	77	52.15 (6.15)
Attention	33	72	51.94 (9.09)
Attention stability	20	77	55.79 (12.46)
Zoom-in/out			
Breathing rate (bpm)	6	24	11.78 (4.76)
Mean heart rate (bpm)	66	110	81.03 (10.25)
Mean SDNN (ms)	29.77	102.18	50.43 (16.88)
Mean RMSSD (ms)	15.79	63.29	31.55 (11.28)
Relaxation	42	71	56.30 (8.60)
Relaxation stability	43	60	51.42 (4.02)
Attention	34	68	52.12 (7.47)
Attention stability	31	79	57.00 (10.70)
Up/down			
Breathing rate (bpm)	6	20	11.47 (4.12)
Mean heart rate (bpm)	64	108	80.36 (10.57)
Mean SDNN (ms)	29.70	86.17	50.22 (16.24)
Mean RMSSD (ms)	11.85	60.28	33.06 (12.21)
Relaxation	34	68	55.48 (7.76)
Relaxation stability	45	62	52.82 (4.33)
Attention	34	73	52.24 (9.57)
Attention stability	44	73	59.24 (8.39)
Color gradients			
Breathing rate (bpm)	6	24	11.59 (4.85)
Mean heart rate (bpm)	58	113	80.94 (11.13)
Mean SDNN (ms)	26.17	107.45	50.56 (16.67)
Mean RMSSD (ms)	12.29	77.00	32.47 (14.08)
Relaxation	43	69	55.70 (6.79)
Relaxation stability	41	67	50.88 (5.41)
Attention	37	73	52.33 (9.95)
Attention stability	26	77	57.85 (10.03)

[a]SDNN: standard deviation of normal to normal interval.

[b]RMSSD: root mean square of successive differences.

XSL•FO
RenderX

Impact of Gaming Experience on Cognitive Load

Tables 3-5 present the results of multivariate analysis of variance (MANOVA) analysis. Each participant completed the questionnaire 4 times, corresponding to 4 different interfaces. Therefore, we applied the Bonferroni correction, which adjusts the significance level for each individual test by dividing α (0.05) by 4, resulting in a new *P* value of .0125.

Table 3. Influence and correlations among gaming experience and cognitive load.

Source and dependent variable	*F* test (*df*)	*P* value	Correlation	Significance (2-tailed test)
Competence				
Mental demand	12.639 (1, 131)	.001[a]	–0.301	0.000[a]
Physical demand	2.836 (1, 131)	.10	–0.190	0.029
Temporal demand	4.660 (1, 131)	.03	–0.395	0.000[a]
Performance	9.902 (1, 131)	.002[a]	–0.547	0.000[a]
Effort	3.050 (1, 131)	.08	–0.367	0.000[a]
Frustration	10.168 (1, 131)	.002[a]	–0.561	0.000[a]
Immersion				
Mental demand	9.523 (1, 131)	.003[a]	–0.240	0.006[a]
Physical demand	5.606 (1, 131)	.02	–0.116	0.186
Temporal demand	0.429 (1, 131)	.51	–0.203	0.019
Performance	1.301 (1, 131)	.26	–0.259	0.003[a]
Effort	6.385 (1, 131)	.013	–0.278	0.001[a]
Frustration	1.592 (1, 131)	.21	–0.216	0.013
Challenge				
Mental demand	0.226 (1, 131)	.64	0.303	0.000[a]
Physical demand	3.592 (1, 131)	.06	0.410	0.000[a]
Temporal demand	3.241 (1, 131)	.07	0.422	0.000[a]
Performance	5.649 (1, 131)	.02	0.441	0.000[a]
Effort	17.471 (1, 131)	<.001[a]	0.586	0.000[a]
Frustration	8.012 (1, 131)	.005[a]	0.303	0.000[a]
Positive affect				
Mental demand	7.264 (1, 131)	.008[a]	–0.130	0.138
Physical demand	4.187 (1, 131)	.04	0.017	0.847
Temporal demand	1.578 (1, 131)	.21	–0.224	0.010[a]
Performance	1.893 (1, 131)	.17	–0.390	0.000[a]
Effort	0.677 (1, 131)	.41	–0.198	0.023
Frustration	0.131 (1, 131)	.72	–0.357	0.000[a]

[a]*P*<.0125.

XSL•FO
RenderX

Table 4. Impact of cognitive load on relaxation.

Source and dependent variable	*F* test (*df*)	*P* value	Correlation	Significance (2-tailed test)
Mental demand				
Breathing rate (bpm)	0.100 (1, 126)	.75	0.077	0.393
Mean heart rate (bpm)	0.315 (1, 126)	.58	–0.094	0.294
Mean SDNN[a] (ms)	0.384 (1, 126)	.54	0.185	0.038
Mean RMSSD[b] (ms)	0.034 (1, 126)	.85	0.101	0.260
Relaxation	2.599 (1, 126)	.11	0.000	0.996
Relaxation stability	0.050 (1, 126)	.82	0.027	0.763
Attention	13.395 (1, 126)	<.001[c]	–0.308	0.000[c]
Attention stability	0.343 (1, 126)	.56	0.146	0.101
Temporal demand				
Breathing rate (bpm)	7.900 (1, 126)	.006[c]	0.245	0.005[c]
Mean heart rate (bpm)	0.281 (1, 126)	.60	0.004	0.968
Mean SDNN (ms)	0.017 (1, 126)	.90	0.073	0.417
Mean RMSSD (ms)	0.082 (1, 126)	.78	0.017	0.850
Relaxation	3.491 (1, 126)	.06	0.159	0.074
Relaxation stability	0.429 (1, 126)	.51	–0.017	0.853
Attention	1.169 (1, 126)	.28	–0.072	0.423
Attention stability	0.598 (1, 126)	.44	0.116	0.194

[a]SDNN: standard deviation of normal to normal interval.

[b]RMSSD: root mean square of successive differences.

[c]*P*<.0125.

XSL•FO
RenderX

Table 5. Impact of gaming experience on relaxation.

Source and dependent variable	*F* test (*df*)	*P* value	Correlation	Significance (2-tailed test)
Flow				
Breathing rate (bpm)	0.015 (1, 126)	.90	–0.083	0.355
Mean heart rate (bpm)	7.926 (1, 126)	.006[a]	–0.081	0.367
Mean SDNN[b] (ms)	8.132 (1, 126)	.005[a]	0.063	0.481
Mean RMSSD[c] (ms)	11.171 (1, 126)	.001[a]	0.006	0.946
Relaxation	0.058 (1, 126)	.81	–0.021	0.814
Relaxation stability	0.000 (1, 126)	.99	0.096	0.280
Attention	7.420 (1, 126)	.007[a]	–0.076	0.393
Attention stability	0.001 (1, 126)	.97	0.026	0.769
Tension				
Breathing rate (bpm)	0.049 (1, 126)	.83	0.227[a]	0.010[a]
Mean heart rate (bpm)	6.938 (1, 126)	.01[a]	0.047	0.599
Mean SDNN (ms)	1.746 (1, 126)	.19	0.025	0.778
Mean RMSSD (ms)	2.423 (1, 126)	.12	0.087	0.332
Relaxation	0.591 (1, 126)	.44	0.150	0.093
Relaxation stability	1.936 (1, 126)	.17	–0.251[a]	0.004[a]
Attention	0.853 (1, 126)	.36	–0.022	0.809
Attention stability	0.347 (1, 126)	.56	0.010	0.914
Challenge				
Breathing rate (bpm)	4.301 (1, 126)	.04	0.296[a]	0.001[a]
Mean heart rate (bpm)	0.599 (1, 126)	.44	–0.046	0.604
Mean SDNN (ms)	0.133 (1, 126)	.72	0.098	0.274
Mean RMSSD (ms)	0.010 (1, 126)	.92	0.106	0.237
Relaxation	0.843 (1, 126)	.36	0.154	0.083
Relaxation stability	0.015 (1, 126)	.90	–0.182[a]	0.041
Attention	5.698 (1, 126)	.02	0.126	0.160
Attention stability	4.666 (1, 126)	.03	0.167	0.060
Negative affect				
Breathing rate (bpm)	0.061 (1, 126)	.81	0.177[a]	0.046
Mean heart rate (bpm)	6.880 (1, 126)	.01[a]	–0.133	0.137
Mean SDNN (ms)	1.704 (1, 126)	.19	0.106	0.237
Mean RMSSD (ms)	4.809 (1, 126)	.03	0.225[a]	0.011
Relaxation	0.060 (1, 126)	.81	0.103	0.247
Relaxation stability	1.510 (1, 126)	.22	–0.229[a]	0.010[a]
Attention	0.569 (1, 126)	.45	–0.062	0.489
Attention stability	3.928 (1, 126)	.05	–0.087	0.333

[a]*P*<.0125.

[b]SDNN: standard deviation of normal to normal interval.

[c]RMSSD: root mean square of successive differences.

XSL•FO
RenderX

We compared the 7 dimensions of game experience to assess their impact on cognitive load, followed by partial correlation analysis. The results indicate that competence, immersion, challenge, and positive affect significantly influenced overall cognitive load. The specific effects of each dimension are detailed below.

Competence significantly impacted mental load ($F_{1,131}$=12.639, P=.001) with a notable negative correlation (r=–0.301, P<.001). In addition, competence significantly influenced performance ($F_{1,131}$=9.902, P=.002) and frustration ($F_{1,131}$=10.168, P=.002), both with negative correlations (r=–0.542, P<.001) and (r=–0.554, P<.001), respectively.

Immersion affected mental load ($F_{1,131}$=9.523, P=.003) with a significant negative correlation (r=–0.240, P=.006).

Challenge significantly influenced effort ($F_{1,131}$=17.471, P<.001) and frustration ($F_{1,131}$=8.012, P=.005) both exhibiting significant positive correlations (r=0.441, P<.001), and (r=0.586, P<.001).

Positive affect influenced mental load ($F_{1,131}$=7.264, P=.008); however, the partial correlations were not significant, indicating a weaker direct impact under different interface conditions.

Impact of Cognitive Load on Relaxation

We compared the 6 dimensions of cognitive load on relaxation, with MANOVA and partial correlation analysis results presented in Table 4.

MANOVA results indicate that mental demand and temporal demand were the primary factors influencing relaxation. Mental demand significantly affected attention ($F_{1,126}$=13.395, P<.001), with a strong negative correlation (r=–0.308, P<.001); in other words, higher mental demand corresponds to lower attention. Temporal demand significantly impacted breathing rate ($F_{1,126}$=7.900, P=.006), with a significant positive correlation (r=0.245, P=.005); thus, higher temporal demand increases breathing rate.

Impact of Gaming Experience on Relaxation

Results from MANOVA and partial correlation analyses are shown in Table 5.

The study compared whether game experience influences relaxation. The statistical results indicate that flow, tension, and negative affect were the main gaming elements affecting overall relaxation. Flow significantly affected mean heart rate ($F_{1,126}$=7.926, P=.006), mean SDNN ($F_{1,126}$=8.132, P=.005), mean RMSSD ($F_{1,126}$=11.171, P=.001), and attention ($F_{1,126}$=7.420, P=.007). However, partial correlation analysis revealed no significant correlations between flow and these 4 variables, indicating a weak direct impact after controlling for different interface variables.

Tension significantly affected mean heart rate ($F_{1,126}$=6.938, P=.010); however, there was no significant correlation. Negative affect significantly affected mean heart rate ($F_{1,126}$=6.880, P=.01); however, there was no significant correlation.

Correlation Between Cognitive Load and Insomnia Severity

As shown in Figure 5, the correlation between cognitive load and insomnia severity was low. “Difficulty falling asleep” (ISI-1) was positively correlated with the physical demand of the zoom-in/out interface (r=0.364, P=.04) and negatively correlated with the mental demand of the up/down interface (r=–0.348, P=.047).

For the zoom-in/out interface, there was a positive correlation between performance and “sleep satisfaction” (ISI-4; r=0.399, P=.02).

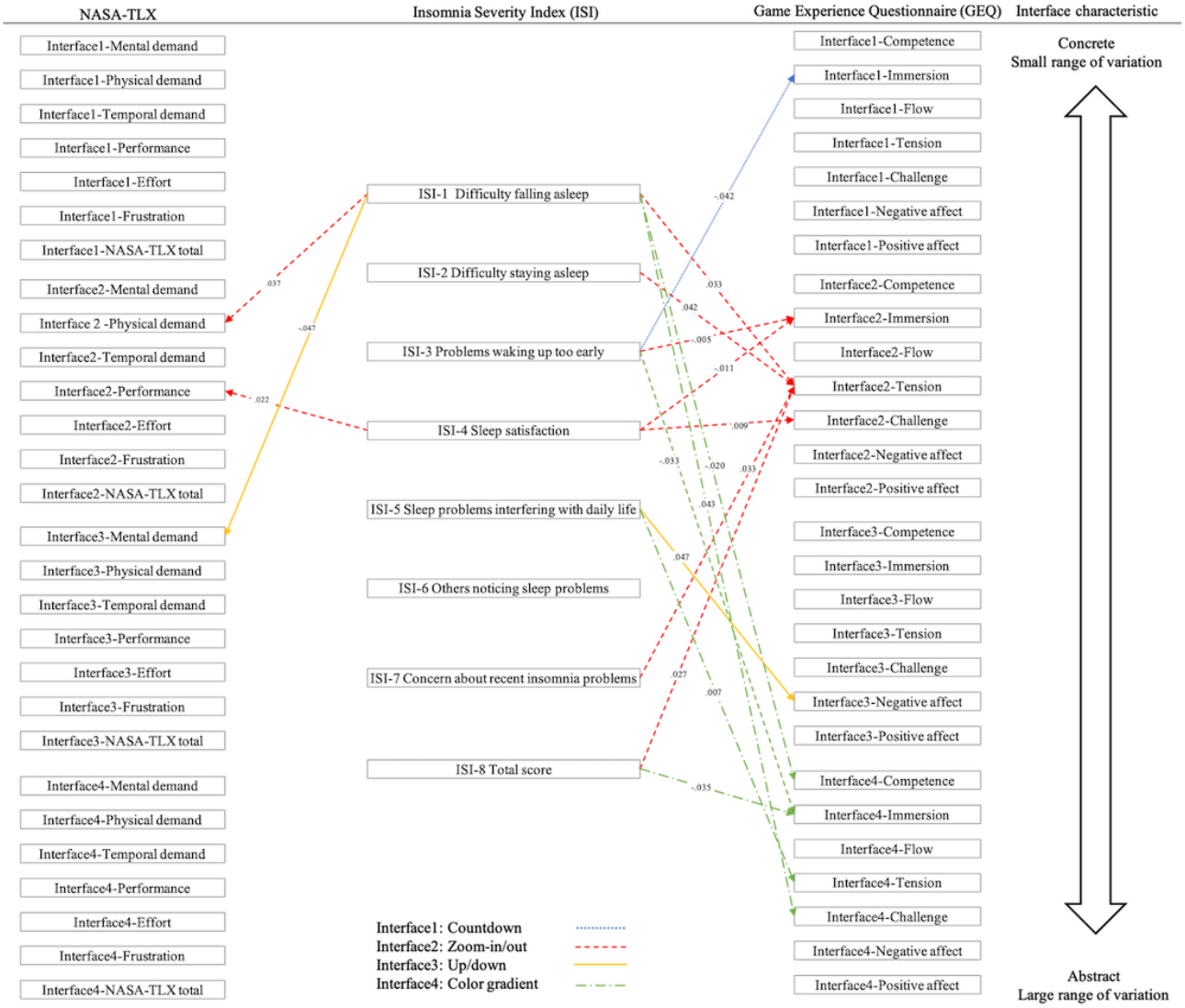
There was a negative correlation between participants who had problems waking up too early (ISI-3) and immersion when using the countdown (r=–0.356, P=.04), zoom-in/out (r=–0.481, P=.005), and color gradient (r=–0.373, P=.03) interfaces, suggesting that immersion is a design element that can significantly impact insomnia severity. The zoom-in/out interface showed a moderate positive correlation between tension and “difficulty falling asleep” (ISI-1; r=0.372, P=.03), “difficulty staying asleep” (ISI-2; r=0.357, P=.04), “concerns about recent insomnia problems” (ISI-7; r=0.372, P=.03), and the total ISI score (r=0.385, P=.03).

The negative affect triggered by the up/down interface was positively correlated with “sleep problems interfering with daily life” (ISI-5; r=0.348, P=.047). For the color gradient interface, there was a positive correlation between tension and “sleep problems interfering with daily life” (ISI-5; r=0.464, P=.007).

Overall, the gaming experiences of the zoom-in/out and color gradient interfaces show greater correlations with insomnia severity.

XSL•FO
RenderX

Figure 5. Correlations among insomnia severity, cognitive load, and gaming experience for 4 interfaces (countdown, zoom-in/out, up/down, and color gradients). NASA-TLX: NASA Task Load Index.



Sleep Self-Efficacy Correlation Analysis

As shown in Figure 6, the temporal demand of the countdown (r=–0.474, P=.005), up/down (r=–0.512, P=.002), and color gradient (r=–0.381, P=.03) interfaces were significantly negatively correlated with "wake after a poor night's sleep without feeling upset about it" (SSE-8 [sleep self-efficacy]). This suggests that temporal demand is a factor that easily affects sleep self-efficacy.

The cognitive load of the color gradient interface demonstrated several significant negative correlations with sleep self-efficacy (SSE-3), including performance (r=–0.442, P=.01), effort (r=–0.415, P=.02), and NASA-TLX total (r=–0.345, P=.049).

Overall, the cognitive load of the zoom-in/out interface was negatively correlated with many aspects of sleep self-efficacy. "Lie in bed with thoughts turned off" (SSE-3) showed significant negative correlations with performance (r=–0.364, P=.04) and effort (r=–0.356, P=.04). "Go back to sleep within 15 minutes of waking in the night" (SSE-6) demonstrated significant negative correlations with physical demand (r=–0.370, P=.03) and NASA-TLX total (r=–0.358, P=.04). "Wake after a poor night's sleep without feeling upset about it" (SSE-8) was significantly negatively correlated with mental demand (r=–0.402, P=.02), effort (r=–0.353, P=.04), and NASA-TLX total (r=–0.403, P=.02).

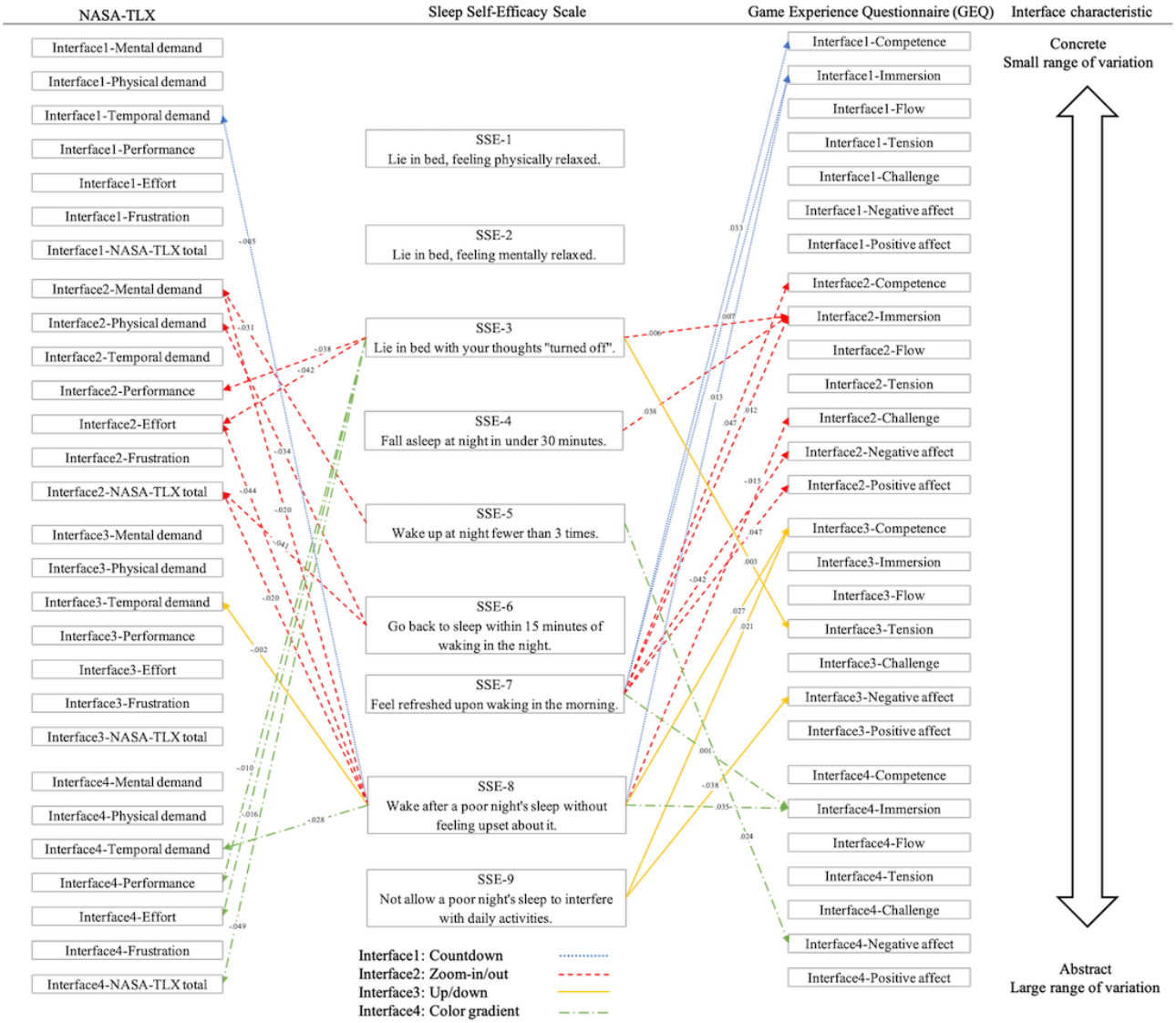
There were many correlations between the 7 dimensions of the GEQ and sleep self-efficacy, especially with the items "feel refreshed upon waking in the morning" (SSE-7) and "wake after a poor night's sleep without feeling upset about it" (SSE-8), indicating that there is a relationship between game design elements and mental states upon waking. Competence and immersion were the 2 game dimensions most related to sleep self-efficacy. "Feel refreshed upon waking in the morning" (SSE-7) was positively correlated with competence (r=0.372, P=.03) and immersion (r=0.458, P=.007) for the countdown interface, competence (r=0.348, P=.047) and immersion (r=0.434, P=.01) for the zoom-in/out interfaces, and immersion (r=0.567, P=.001) for the color gradient interface. "Waking after a poor night's sleep without feeling upset about it" (SSE-8) was positively correlated with immersion for the countdown interface (r=0.427, P=.01), competence (r=0.384, P=.03) for the up/down interface, and immersion (r=0.368, P=.04) for the

XSL•FO
RenderX

color gradient interface. This indicates that higher competence or immersion in these interfaces correlated with higher self-efficacy upon waking up.

Negative affect also has an important impact on sleep self-efficacy. There was a significant positive correlation between the negative affect triggered by the color gradient interface and "wake up at night fewer than 3 times" (SSE-5; r=0.393, P=.02), a negative correlation between negative affect triggered by the zoom-in/out interface and "feel refreshed upon waking in the morning" (SSE-7; r=–0.356, P=.04), and a negative correlation between negative affect triggered by the up/down interface and "wake after a poor night's sleep without feeling upset about it" (SSE-8; r=–0.363, P=.04). This indicates that negative affect can have both positive and negative impacts on sleep self-efficacy.

Figure 6. Correlations among sleep self-efficacy, cognitive load, and gaming experience for 4 interfaces (countdown, zoom-in/out, up/down, and color gradients). NASA-TLX: NASA Task Load Index; SSE: sleep self-efficacy.

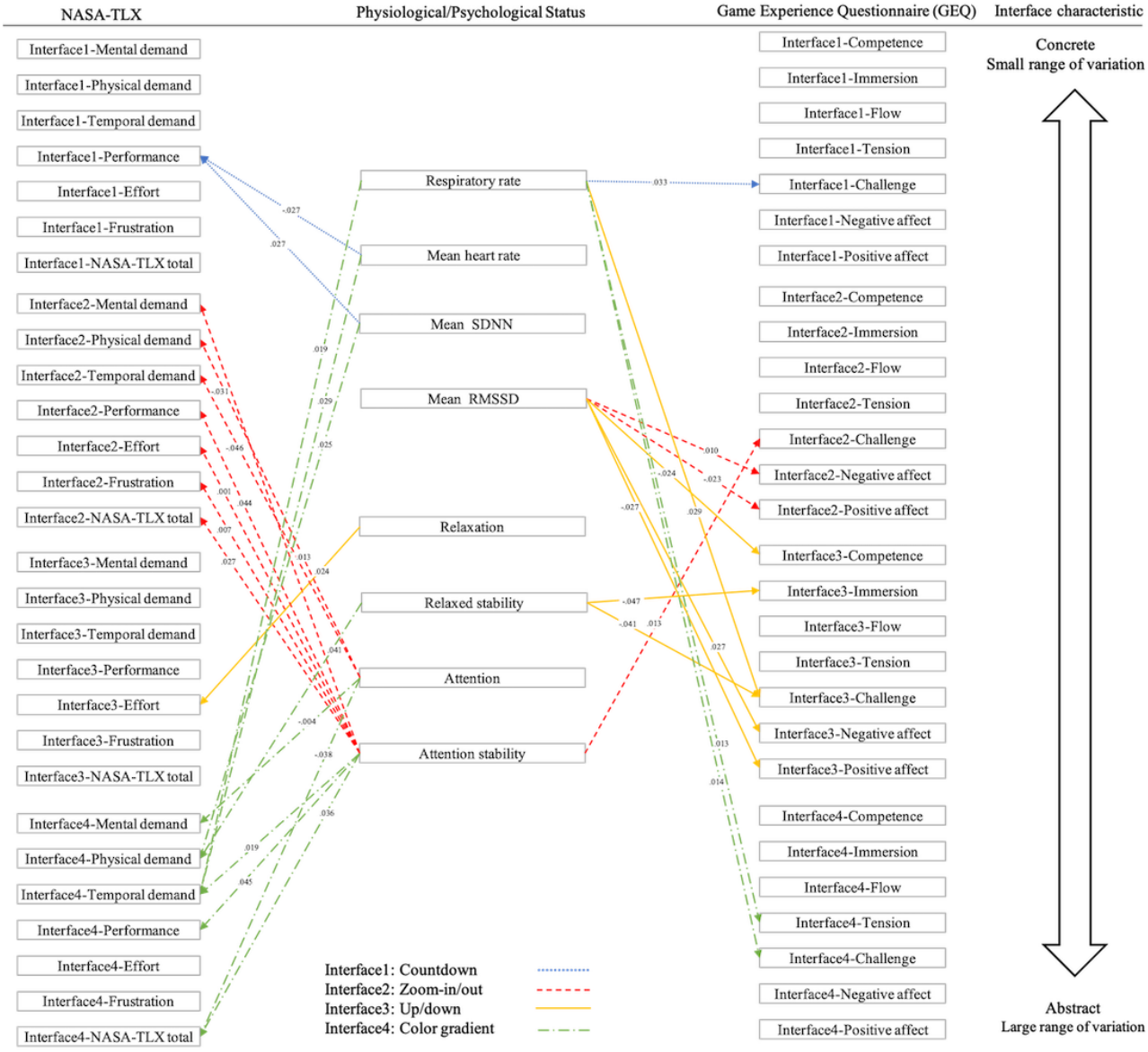


Physiological and Psychological Values

As shown in Figure 7, the countdown guidance mechanism had the most obvious effect on physiological states, particularly subjective performance as well as mean heart rate (r=–0.385, P=.03) and mean SDNN (r=0.385, P=.03). For the zoom-in/out interface, attention was negatively correlated with mental demand (r=–0.377, P=.03) and temporal demand (r=0.350, P=.046). Additionally, attention stability exhibited positive correlations with mental load (r=0.427, P=.01), effort (r=0.539, P=.001), frustration (r=0.458, P=.007), and overall score (r=0.385, P=.03). The limited range and specific guidance of this interface likely influenced attention. The effort and relaxation associated with the up/down interface were positively correlated (r=0.392, P=.02).

XSL•FO
RenderX

Figure 7. Correlations among physiological/psychological signals, cognitive load, and gaming experience for 4 interfaces (countdown, zoom-in/out, up/down, and color gradients). NASA-TLX: NASA Task Load Index; RMSSD: root mean square of successive differences; SDNN: standard deviation of normal to normal interval.



The color gradient interface was correlated with both physiological and psychological states, particularly in relation to temporal demand. Temporal demand was positively correlated with respiratory rate (r=0.413, P=.02), mean heart rate (r=0.380, P=.03), and attention stability (r=0.405, P=.02). Conversely, attention was negatively correlated with mental demand (r=–0.493, P=.004) and NASA-TLX total (r=–0.362, P=.04). Physical demand showed positive correlations with mean SDNN (r=0.389, P=.02) and relaxed stability (r=0.358, P=.04). While NASA-TLX total was negatively correlated with attention (r=–0.362, P=.04), it exhibited a positive correlation with attention stability (r=0.367, P=.04).

The challenge aspect of the game experience was highly correlated with the breathing rate. The challenge aspects of the countdown (r=0.378, P=.03), up/down (r=0.387, P=.03), and color gradient (r=0.440, P=.01) interfaces were all positively correlated with the breathing rate.

The up/down interface showed the most correlations with physiological/psychological signals. Relaxed stability was negatively correlated with immersion (r=–0.348, P=.047) and challenge (r=0.358, P=.04). Additionally, mean RMSSD was negatively correlated with competence (r=–0.391, P=.02).

The most significant game experiences affecting average RMSSD were positive and negative affect. The average RMSSD of the zoom-in/out interface was negatively correlated with positive affect (r=–0.395, P=.02) and positively correlated with negative affect (r=0.442, P=.01). The average RMSSD during the up/down interface was also negatively correlated with positive affect (r=–0.385, P=.03) and positively correlated with negative affect (r=0.384, P=.03).

Qualitative Findings and Implications

The study compiled interview results from 33 participants. Users felt that the countdown interface provided clear guidance. However, the rapid change of numbers disrupted their breathing

XSL•FO
RenderX

rhythm, and the prolonged inhalation and exhalation periods made it difficult to follow the guide. The zoom-in/out guidance mechanism is common in many apps, making it familiar and easy for users to understand and accept. However, care must be taken with the scale of the visual changes and the brightness of the colors, as excessive magnification or overly bright colors can cause visual strain. The up/down guidance, which mimics the natural rhythm of ocean waves, is visually soothing and facilitates relaxation, making it the preferred choice for most participants. When paired with music, it creates a sense of immersion, though the sensation of rising might cause some users to feel overwhelmed; therefore, it is crucial to consider the range of visual changes in the design. Users noted that the zoom-in/out and up/down guidance mechanisms align well with the body's natural breathing patterns, enhancing comprehension and effectiveness. This finding supports that of Chuanromanee and Metoyer [43]. The color gradients, which are a more abstract guidance method, offered a high tolerance for error, which some users liked as it allowed them to match their own breathing rate, resulting in less pressure. However, the lack of clarity in the variations when first using this interface might disrupt the breathing rhythm, and the brightening colors could cause visual stimulation.

The effectiveness of different interfaces varies with individual differences. The interviews revealed that even with the same breathing rate, different guidance mechanisms gave users different perceptions of the rate, as visual animations, colors, and changes in size also affected subjective pressure and relaxation effects. Most participants felt that the initial stages were visually guided; however, after prolonged use, they shifted to auditory guidance. Therefore, visual and auditory guidance should complement each other, allowing users to become familiar with the breathing rate and subsequently continue with auditory guidance. Future app designs should incorporate more personalized options, such as selecting the breathing guidance rate, as many participants reported difficulty keeping up with the 6 breaths per minute initially. Enhancing auditory and visual changes, like voice and other animation guidance, should also be considered. Additionally, introducing stories or reward mechanisms could enrich the overall app experience and enhance user compliance with relaxation training.

Discussion

Principal Findings

Overview

In our 2-stage experiment, the first stage revealed that the zoom-in/out interface demanded the lowest cognitive load and achieved the highest user experience score, while the color gradient interface demanded the highest cognitive load. However, in the second stage, the color gradient interface was reported to demand the lowest overall cognitive load, followed by the up/down interface, which was also associated with the lowest breathing rate and mean heart rate as well as the highest mean RMSSD. This indicates that the duration of the diaphragmatic breathing game affects cognitive load and relaxation. For short-term use, the zoom-in/out interface incurs a lower cognitive load due to its moderate variation range and specific guidance, making it more effective initially. However, for longer durations, the more abstract up/down and color gradient interfaces are more effective, as they allow users to breathe at their own pace, reducing cognitive load and enhancing relaxation.

The countdown had high cognitive load in both stages and low relaxation in the second stage, indicating that specific numerical guidance causes higher stress regardless of duration, leading to poor relaxation outcomes. According to the results, abstract guidance generated a lower cognitive load, aligning with our hypothesis. However, abstract guidance demonstrated a more significant effect on influencing breathing rate compared with concrete guidance, contradicting our hypothesis.

Countdown

This is the most specific guidance with the smallest range of variation, primarily related to physiological relaxation. Subjective performance with this interface is significantly correlated with mean heart rate and SDNN. Additionally, participants with lower sleep self-efficacy experienced higher temporal demands, which impacted relaxation. The immersion experience with this interface is negatively correlated with users troubled by waking too early. Therefore, the applicability of countdown guidance in relaxation training games should be carefully considered.

Zoom-In/Out

This interface strikes a balance between abstract and specific guidance with a smaller range of variation. It shows more positive correlations between cognitive load and insomnia severity. There is also a significant negative correlation between cognitive load and sleep self-efficacy, making it unsuitable for users with sleep problems. Cognitive load with this interface is significantly correlated with attention and attention stability, requiring higher attention stability to complete tasks, thus increasing overall cognitive load.

Up/Down

This abstract interface with a larger range of variation shows no significant correlation between cognitive load and insomnia severity, sleep self-efficacy, or physiological and psychological relaxation, indicating that it is less influenced by differences in user groups and is thus suitable for most users. There is a negative correlation between difficulty falling asleep and mental demand for this interface, which indicates that it is more effective for those with sleep onset difficulties. The negative affect triggered by this interface is positively correlated with mean RMSSD.

Color Gradients

This interface is the most abstract with the largest range of variation. There is a significant negative correlation between cognitive load and the sleep self-efficacy item "lie in bed with your thoughts turned off" (SSE-3), indicating that abstract and highly variable visual guidance requires more effort to understand and adapt to. Additionally, the cognitive load of this interface shows significant correlations with breathing rate, attention, and attention stability, particularly with a positive correlation between temporal demand and breathing rate. The

XSL•FO
RenderX

abstract nature of the guidance increases temporal pressure and breathing rate. At the same time, understanding the abstract changes demands sustained focus, affecting both attention and cognitive load.

In terms of game experience, a sense of competence, immersion, challenge, and positive affect influence cognitive load, particularly mental and temporal load, subjective performance, and frustration. Higher competence and immersion reduce subjective cognitive load. However, challenges are associated with increased effort and frustration. Therefore, relaxation training should enhance competence and immersion while reducing challenges to lower cognitive load. Furthermore, sleep self-efficacy is also strongly correlated with competence, immersion, and temporal demand. Therefore, attention should be given to the stress caused by the time requirements of the guidance mechanisms.

Mental and temporal demands significantly impact attention and breathing rate. Higher mental load decreases attention, while higher temporal demand increases breathing rate. Therefore, breathing guidance interfaces should minimize time pressure. Game-induced flow, tension, and negative affects impact various physiological and psychological relaxation indicators.

The correlation analysis revealed that the challenge was positively associated with breathing rate in multiple instances. Additionally, negative affect was significantly positively correlated with mean RMSSD. Because negative affects such as boredom or weariness also indicate lower stimulation in a low-stimulation environment, people tend to feel more relaxed. Conversely, positive affect includes fun, enjoyment, or pleasure, which may shift attention to the game interface rather than the act of diaphragmatic breathing itself, leading to reduced physiological relaxation. However, emotional responses themselves are quite complex, so future research could delve deeper into their impact on relaxation.

Limitations

There are several limitations to our study. First, it focused on analyzing 4 specific guidance mechanisms and considered only the dependent variables of cognitive load, game experience, relaxation, and attention, without including a direct comparison to traditional video-based teaching. The experiments were not restricted to being conducted exclusively during the day or prior to sleep, so the therapeutic effects on insomnia were not within the scope of this study. In future studies, diaphragmatic breathing games could be incorporated into brief behavioral treatments for insomnia, with experiments conducted closer to bedtime. Through long-term observation and follow-up, it would be possible to determine whether gamified relaxation training can enhance the effectiveness of insomnia treatment.

Another limitation of the study was the difficulty in recruiting insomniacs during the implementation of the study, resulting in a limited number of participants with severe insomnia. Future studies are recommended to include a larger sample of insomniacs to improve the generalizability and statistical power of the findings.

Finally, In phase 1 of the study, the order of the guidance mechanisms was not counterbalanced, and all participants followed the same sequence. Future studies should consider using counterbalancing techniques to control for potential order effects.

Conclusions

In the design of diaphragmatic breathing guidance for patients with insomnia, the study results indicate that the up/down interface is the most effective, as it results in lower cognitive load and higher relaxation. Both gaming experience and cognitive load significantly affect relaxation, with psychological and temporal demands having a notable impact on breathing rate. Positive and negative affect are critical factors influencing relaxation, with negative affect associated with higher relaxation effects. These findings offer valuable insights into the design of diaphragmatic breathing training, suggesting that appropriate guidance methods and gamified designs can enhance relaxation effects. Recommendations are made for the design of presleep breathing guidance and relaxation training apps.

Acknowledgments

This study was supported by the Taipei University System Academic Cooperation Project and represents a collaborative effort between the National Taipei University of Technology and Taipei Medical University, under project number USTP-NTUT-TMU-113-02.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Correlations among variables for four interfaces.
[DOCX File , 72 KB - games_v13i1e67000_app1.docx]

References

1. Roth T. Insomnia: definition, prevalence, etiology, and consequences. J Clin Sleep Med 2007;3(5 Suppl):S7-10 [FREE Full text] [Medline: 17824495]

XSL•FO
RenderX

2. Ampadu S, Jiang Y, Debrah E, Antwi CO, Amankwa E, Gyamfi SA, et al. Online personalized recommended product quality and e-impulse buying: a conditional mediation analysis. J Retail Consum Serv 2022;64:102789. [doi: 10.1016/j.jretconser.2021.102789]
3. Edinger JD, Means MK. Cognitive-behavioral therapy for primary insomnia. Clin Psychol Rev 2005;25(5):539-558. [doi: 10.1016/j.cpr.2005.04.003] [Medline: 15951083]
4. Edinger JD, Arnedt JT, Bertisch SM, Carney CE, Harrington JJ, Lichstein KL, et al. Behavioral and psychological treatments for chronic insomnia disorder in adults: an American Academy of Sleep Medicine clinical practice guideline. J Clin Sleep Med 2021;17(2):255-262 [FREE Full text] [doi: 10.5664/jcsm.8986] [Medline: 33164742]
5. Troxel WM, Germain A, Buysse DJ. Clinical management of insomnia with brief behavioral treatment (BBTI). Behav Sleep Med 2012;10(4):266-279 [FREE Full text] [doi: 10.1080/15402002.2011.607200] [Medline: 22946736]
6. Zaccaro A, Piarulli A, Laurino M, Garbella E, Menicucci D, Neri B, et al. How breath-control can change your life: a systematic review on psycho-physiological correlates of slow breathing. Front Hum Neurosci 2018;12:353 [FREE Full text] [doi: 10.3389/fnhum.2018.00353] [Medline: 30245619]
7. Chittaro L, Sioni R. Evaluating mobile apps for breathing training: the effectiveness of visualization. Comput Hum Behav 2014;40:56-63. [doi: 10.1016/j.chb.2014.07.049]
8. Faust-Christmann CA, Taetz B, Zolynski G, Zimmermann T, Bleser G. A biofeedback app to instruct abdominal breathing (Breathing-Mentor): pilot experiment. JMIR mHealth uHealth 2019;7(9):e13703 [FREE Full text] [doi: 10.2196/13703] [Medline: 31573919]
9. Rockstroh C, Blum J, Göritz AS. A mobile VR-based respiratory biofeedback game to foster diaphragmatic breathing. Virtual Reality 2020;25(2):539-552. [doi: 10.1007/s10055-020-00471-5]
10. Balaskas A, Schueller SM, Cox AL, Doherty G. Understanding users' perspectives on mobile apps for anxiety management. Front Digit Health 2022;4:854263 [FREE Full text] [doi: 10.3389/fdgth.2022.854263] [Medline: 36120712]
11. Zarour M, Alharbi M. User experience framework that combines aspects, dimensions, and measurement methods. Cogent Engineering 2018;4(1):1421006. [doi: 10.1080/23311916.2017.1421006]
12. Deterding S, Dixon D, Khaled R, Nacke L. From game design elements to gamefulness: defining 'gamification'. 2011 Presented at: MindTrek '11: Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments; 2011 September 28-30; USA. [doi: 10.1145/2181037.2181040]
13. Cechetti NP, Bellei EA, Biduski D, Rodriguez JPM, Roman MK, De Marchi ACB. Developing and implementing a gamification method to improve user engagement: a case study with an m-Health application for hypertension monitoring. Telematics Inform 2019;41:126-138. [doi: 10.1016/j.tele.2019.04.007]
14. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. Psychol Rev 1977;84(2):191-215. [doi: 10.1037//0033-295x.84.2.191] [Medline: 847061]
15. Backlund P, Engstrom H, Johannesson M, Lebram M, Sjoden B. Designing for self-efficacy in a game based simulator: an experimental study and Its implications for serious games design. 2008 Presented at: International Conference Visualisation; 2008 July 9-10; UK. [doi: 10.1109/vis.2008.26]
16. Weerdmeester J, van M, Harris O, Smit N, Engels RCME, Granic I. Exploring the role of self-efficacy in biofeedback video games. 2017 Presented at: CHI PLAY '17 Extended Abstracts: Extended Abstracts Publication of the Annual Symposium on Computer-Human Interaction in Play; 2017 October 15-18; Netherlands. [doi: 10.1145/3130859.3131299]
17. Ruiter Petrov ME, Lichstein KL, Huisingh CE, Bradley LA. Predictors of adherence to a brief behavioral insomnia intervention: daily process analysis. Behav Ther 2014;45(3):430-442. [doi: 10.1016/j.beth.2014.01.005] [Medline: 24680236]
18. Lacks P. Behavioral Treatment for Persistent Insomnia. London: Pergamon Press; 1987.
19. Rutledge CM, Guardia ACL, Bluestein D. Predictors of self-efficacy for sleep in primary care. J Clin Nurs 2013;22(9-10):1254-1261. [doi: 10.1111/jocn.12005] [Medline: 23574289]
20. Ghose SM, Dzierzewski JM, Dautovich ND. Sleep and self-efficacy: the role of domain specificity in predicting sleep health. Sleep Health 2023;9(2):190-195 [FREE Full text] [doi: 10.1016/j.sleh.2022.09.008] [Medline: 36333191]
21. Sweller J. Cognitive load during problem solving: effects on learning. Cognit Sci 1988;12(2):257-285. [doi: 10.1016/0364-0213(88)90023-7]
22. Wuyts J, De Valck E, Vandekerckhove M, Pattyn N, Bulckaert A, Berckmans D, et al. The influence of pre-sleep cognitive arousal on sleep onset processes. Int J Psychophysiol 2012;83(1):8-15. [doi: 10.1016/j.ijpsycho.2011.09.016] [Medline: 21963535]
23. Harvey AG. A cognitive model of insomnia. Behav Res Ther 2002;40(8):869-893. [doi: 10.1016/s0005-7967(01)00061-4] [Medline: 12186352]
24. Harvey AG. Pre-sleep cognitive activity: a comparison of sleep-onset insomniacs and good sleepers. Br J Clin Psychol 2000;39(3):275-286. [doi: 10.1348/014466500163284] [Medline: 11033749]
25. Sevcenko N, Ninaus M, Wortha F, Moeller K, Gerjets P. Measuring cognitive load using in-game metrics of a serious simulation game. Front Psychol 2021;12:572437 [FREE Full text] [doi: 10.3389/fpsyg.2021.572437] [Medline: 33841227]
26. Kaya N, Epps HH. Relationship between color and emotion: a study of college students. Coll Stud J 2004;38(3):396-405 [FREE Full text]

XSL•FO
RenderX

27. Saito M. A comparative study of color preferences in Japan, China and Indonesia, with emphasis on the preference for white. Percept Mot Skills 1996;83(1):115-128. [doi: 10.2466/pms.1996.83.1.115] [Medline: 8873184]
28. Grassini S, Revonsuo A, Castellotti S, Petrizzo I, Benedetti V, Koivisto M. Processing of natural scenery is associated with lower attentional and cognitive load compared with urban ones. J Environ Psychol 2019;62:1-11. [doi: 10.1016/j.jenvp.2019.01.007]
29. Farokhnezhad Afshar P, Bahramnezhad F, Asgari P, Shiri M. Effect of white noise on sleep in patients admitted to a coronary care. J Caring Sci 2016;5(2):103-109 [FREE Full text] [doi: 10.15171/jcs.2016.011] [Medline: 27354974]
30. Task Force of the European Society of Cardiology the North American Society of Pacing Electrophysiology. Heart rate variability. Circulation 1996;93(5):1043-1065. [doi: 10.1161/01.CIR.93.5.1043]
31. Bernardi L, Porta C, Gabutti A, Spicuzza L, Sleight P. Modulatory effects of respiration. Auton Neurosci 2001;90(1-2):47-56. [doi: 10.1016/S1566-0702(01)00267-3] [Medline: 11485292]
32. Tsai HJ, Kuo TBJ, Lee GS, Yang CCH. Efficacy of paced breathing for insomnia: enhances vagal activity and improves sleep quality. Psychophysiology 2015;52(3):388-396. [doi: 10.1111/psyp.12333] [Medline: 25234581]
33. Pham T, Lau ZJ, Chen SHA, Makowski D. Heart rate variability in psychology: a review of HRV indices and an analysis tutorial. Sensors (Basel) 2021;21(12):3998 [FREE Full text] [doi: 10.3390/s21123998] [Medline: 34207927]
34. Shaffer F, Ginsberg JP. An overview of heart rate variability metrics and norms. Front Public Health 2017;5:258 [FREE Full text] [doi: 10.3389/fpubh.2017.00258] [Medline: 29034226]
35. Katmah R, Al-Shargie F, Tariq U, Babiloni F, Al-Mughairbi F, Al-Nashash H. A review on mental stress assessment methods using EEG signals. Sensors (Basel) 2021;21(15):5043 [FREE Full text] [doi: 10.3390/s21155043] [Medline: 34372280]
36. Li TM, Chao HC, Zhang J. Emotion classification based on brain wave: a survey. Hum Cent Comput Inf Sci 2019;9(1). [doi: 10.1186/s13673-019-0201-x]
37. Lan Z, Sourina O, Wang L, Liu Y. Real-time EEG-based emotion monitoring using stable features. Vis Comput 2015;32(3):347-358. [doi: 10.1007/s00371-015-1183-y]
38. Giannakakis G, Grigoriadis D, Giannakaki K, Simantiraki O, Roniotis A, Tsiknakis M. Review on psychological stress detection using biosignals. IEEE Trans Affective Comput 2022;13(1):440-460. [doi: 10.1109/taffc.2019.2927337]
39. Bastien CH, Vallières A, Morin CM. Validation of the insomnia severity index as an outcome measure for insomnia research. Sleep Med 2001;2(4):297-307. [doi: 10.1016/s1389-9457(00)00065-4] [Medline: 11438246]
40. Hart SG, Staveland LE. In: Hancock PA, Meshkati N, editors. Development of NASA-TLX (Task Load Index): results of empirical theoretical research. North-Holland: Human Mental Workload; 1988.
41. Hassenzahl M, Burmester M, Koller F. AttrakDiff: Ein fragebogen zur messung wahrgenommener hedonischer und pragmatischer qualität [AttrakDiff: A questionnaire for measuring perceived hedonic and pragmatic quality]. In: Szwillus G, Ziegler J, editors. Mensch & Computer 2003: Interaktion in Bewegung. Berichte des German Chapter of the ACM [Mensch & Computer 2003: Interaction in Motion. Reports of the German Chapter of the ACM] [Book in German]. Germany: ACM. Vieweg+Teubner Verlag; 2003.
42. Xia H, Pan X, An W, Zhang Z. Can online rating reflect authentic customer purchase feelings? Understanding how customer dissatisfaction relates to negative reviews. J Comput Inf Syst 2019;61(4):314-327. [doi: 10.1080/08874417.2019.1647766]
43. Chuanromanee T, Metoyer R. Evaluation and comparison of four mobile breathing training visualizations. 2020 Presented at: IEEE International Conference on Healthcare Informatics (ICHI); 2020 December 3; Germany. [doi: 10.1109/ichi48887.2020.9374383]

Abbreviations

BBTI: brief behavioral treatment for insomnia
CBT-I: cognitive behavioral therapy for insomnia
EEG: electroencephalogram
GEQ: Game Experience Questionnaire
HQ-I: hedonic quality-identity
HQ-S: hedonic quality-stimulation
HRV: heart rate variability
ISI: Insomnia Severity Index
NASA-TLX: NASA Task Load Index
PQ: pragmatic quality
RMSSD: root mean square of successive differences
SDNN: standard deviation of normal to normal interval
SSE: sleep self-efficacy

XSL•FO
RenderX

Edited by A Coristine; submitted 28.09.24; peer-reviewed by J Roland, N Carmona; comments to author 20.11.24; revised version received 10.01.25; accepted 23.01.25; published 04.03.25.

Please cite as:
Lai YJ, Chiu HY, Wu KC, Chang CW
Diaphragmatic Breathing Interfaces to Promote Relaxation for Mitigating Insomnia: Pilot Study
JMIR Serious Games 2025;13:e67000
URL: https://games.jmir.org/2025/1/e67000
doi:10.2196/67000
PMID:40053714

©Yi-Jen Lai, Hsiao-Yean Chiu, Ko-Chiu Wu, Chun-Wei Chang. Originally published in JMIR Serious Games (https://games.jmir.org), 04.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Effects and Acceptability of a 1-Week Home-Based Virtual Reality Training for Supporting the Management of Stress and Anxiety: Randomized Pilot Trial

Federica Pallavicini[1], PhD; Eleonora Orena[2], MA; Lisa Arnoldi[3], MA; Federica Achille[3], MA; Stefano Stefanini[3], MA; Maddalena Cassa[2], MA; Alessandro Pepe[1], PhD; Guido Veronese[1], PhD; Luca Bernardelli[4], MA; Francesca Sforza[4], MA; Sara Fascendini[3], MA; Carlo Alberto Defanti[3], MA; Marco Gemma[2], MA; Massimo Clerici[5], PhD; Giuseppe Riva[6,7], PhD; Fabrizia Mantovani[1], PhD

[1]Department of Human Sciences for Education "Riccardo Massa", University of Milano-Bicocca, Milan, Italy

[2]IRCCS Neurological Institute Carlo Besta, Milan, Italy

[3]European Biomedical Research Foundation, Gazzaniga, Italy

[4]Become-Hub, Milan, Italy

[5]Department of Medicine and Surgery, University of Milano-Bicocca, Milan, Italy

[6]Humane Technology Lab, Department of Psychology, Catholic University of the Sacred Heart, Milan, Italy

[7]Applied Technology for Neuro-Psychology Lab, IRCCS Istituto Auxologico Italiano, Milan, Italy

Corresponding Author:
Federica Pallavicini, PhD
Department of Human Sciences for Education "Riccardo Massa"
University of Milano-Bicocca
Piazza dell'Ateneo Nuovo 1, 20126
Milan, 20100
Italy
Phone: 39 02 64484944
Email: federica.pallavicini@unimib.it

Related Article:

This is a corrected version. See correction statement: https://games.jmir.org/2025/1/e77791

Abstract

Background: Virtual reality (VR) is helpful for the management of stress and anxiety. However, current interventions have limitations related to location (ie, therapist's office or hospitals) and content (ie, virtual experiences only for relaxation).

Objective: This randomized pilot trial aims to investigate the efficacy and acceptability of a brief remote VR-based training for supporting stress and anxiety management in a sample of Italian health care workers.

Methods: A total of 29 doctors and nurses (n=21; 72% female; mean age 35.6, SD 10.3 years) were recruited and randomized to a VR intervention group or a control group in a passive control condition. Participants assigned to the VR intervention group received remote VR-based training consisting of 3 sessions at home delivered in 1 week using the VR psychoeducational experience "MIND-VR" and the 360° relaxing video "The Secret Garden." The primary outcome measures were stress, anxiety, depression, and the knowledge of stress and anxiety assessed at baseline and posttreatment. We also evaluated the immediate effect of the remote VR-based training sessions on the perceived state of anxiety and negative and positive emotions. The secondary outcome measure was the usability at home of the VR system and content.

Results: The VR intervention significantly reduced stress levels as assessed by the Perceived Stress Scale (6.46, 95% CI 2.77 to 10.5; *P*=.046) and increased the knowledge of stress and anxiety, as evaluated by the ad hoc questionnaire adopted (–2.09, 95% CI –3.86 to –0.529; *P*=.046). However, the home-based VR training did not yield similar reductions in stress, anxiety, and depression levels as assessed by the Depression, Anxiety, and Stress Scale-21 items or in trait anxiety as evaluated through the State-Trait Anxiety Inventory Form Y-1. After the home training sessions with VR, there was a significant decrease in anxiety, anger, and sadness and an increase in happiness levels. Analyses of the questionnaires on usability indicated that the health care

XSL•FO
RenderX

workers found using the VR system at home easy and without adverse effects related to cybersickness. Of 33 participants, 29 (88%) adhered to the protocol.

Conclusions: The results of this randomized pilot study suggest that a week-long home VR intervention, created with content created specifically for this purpose and available free of charge, can help individuals manage stress and anxiety, encouraging further research investigating the potential of remote VR interventions to support mental health.

Trial Registration: ClinicalTrials.gov NCT04611399; https://tinyurl.com/scxunprd

(JMIR Serious Games 2025;13:e50326) doi:10.2196/50326

KEYWORDS

virtual reality; relaxation; anxiety; depression; emotions; health care professionals; health care workers; hospital; randomized clinicial trial; hospitals

Introduction

Background

Virtual reality (VR) is defined as a set of technologies, including a head-mounted display (HMD), computer, and mobile devices, that allow users to interact with a 3D environment in real time [1]. This technology has been rapidly advancing in recent years, with various applications ranging from gaming and entertainment to education and health care [2-5]. In the last decade, VR has become a cutting-edge technology that could provide enhanced value for health care, representing a helpful and practical instrument for psychological support [6-8].

Several studies, reviews, and meta-analyses have shown that VR is useful for decreasing stress, anxiety, and negative emotions, generating positive emotional states [9-11].

Relaxing VR content, especially those created ad hoc by researchers and clinicians to mimic nature [9,12] and immersive video games [11,13], have been shown to help individuals learn and practice relaxation techniques, such as breath exercise [14], progressive muscle relaxation [15], biofeedback [16,17], and mindfulness [18,19]. Additionally, VR can also be an effective tool for psychoeducation, an intervention that provides individuals with information about their illness and teaches them coping skills [20-23].

The efficacy of VR to support the management of stress and anxiety has been shown since the 2000s to be useful in different populations of people, both healthy and enduring various mental and physical disorders [5,24]. This fact appears important because stress and anxiety are common mental health conditions that have become increasingly prevalent, especially since the COVID-19 pandemic [25,26], with a prevalence of 29.6% and 31.9% [26].

One group that is particularly affected by stress and anxiety is health care workers. Some of their most common causes of stress and anxiety include long working hours and constant exposure to illness and death [27,28]. The COVID-19 pandemic has also added an extra layer of stress and anxiety among health care workers [29,30]. The high prevalence of stress and anxiety among health care workers is a significant concern for public health. It affects the well-being of health care workers and has implications for patient care and safety [31,32].

Interestingly, recent studies showed the usefulness of VR-based interventions for diminishing stress and anxiety among health care workers. For example, watching 360° videos of calming natural environments was effective for relaxation in a sample of ICU [33]. Furthermore, a 3-minute immersive video of a nature scene was found to lower stress levels among frontline health care workers in COVID-19 treatment units [34].

Although the use of VR for supporting the management of stress and anxiety has an increasingly solid scientific basis, there are two limitations still limiting its adoption and usefulness, both among health care professionals and more broadly. First, in past studies, experienced clinicians only administered VR interventions in their offices or the hospital. However, recently the introduction of standalone and mobile VR systems, due to their high ease of use and low cost, has made this technology feasible for daily use at home [35]. Therefore, home VR training represents a promising new intervention for remote psychological support [36-38], including relaxation training [39,40]. However, little is currently known about the usefulness and acceptability of VR-based home programs.

Second, in previous studies that have used VR for stress and anxiety management, the proposed intervention involved using VR only for relaxation [33,34]. None of these also incorporated a part of psychoeducation. However, as suggested in the literature, psychoeducation is important for stress and anxiety management, as it could increase treatment adherence and reduce self-esteem [41]. By providing individuals with an understanding of the nature of anxiety and stress, along with practical coping strategies, psychoeducation empowers them to manage these challenges effectively [42,43]. It also plays a preventive role by raising awareness and reducing the stigma surrounding these issues, thus potentially averting more severe mental health problems [44,45].

Aims of This Study

Within the context described above, the main aim of the present randomized pilot trial was to evaluate the effectiveness and acceptability of a home-based VR intervention for managing stress and anxiety in a sample of Italian health care workers.

The primary objective of this randomized pilot trial focuses on the usefulness of the proposed remote VR-based intervention for decreasing stress and anxiety and enhancing the knowledge about these conditions compared to a passive group as the control condition. The second objective was to evaluate the immediate effect on stress, anxiety, and negative and positive emotions of at-home VR training sessions. Finally, the third

XSL•FO
RenderX

objective was to assess the usability at home of the VR system and content.

Methods

Study Design

This pilot trial followed a 2-arm, parallel-group, randomized controlled trial model with between-subjects repeated measures (pre and post) with VR intervention as the experimental condition and a passive group as the control condition. The introduction and methods sections are based on the published protocol [46].

Ethical Considerations

The research was approved by the Ethics Committee of the University of Milan-Bicocca (ie, approval 0061757/20 on September 25, 2020) and of the Foundation IRCCS Carlo Besta Neurological Institute Foundation (ie, approval number 75 on September 16, 2020). This study followed the ethical standards outlined in the Declaration of Helsinki (1964) and its subsequent revisions. Before data collection, the plan for this randomized pilot trial's design, data collection, and analysis was registered in the ClinicalTrials.gov database (NCT04611399) on February 11, 2020.

Before taking part in this study, participants filled out an informed consent form. The form explained the procedures, risks, and benefits of participating in the research, ensuring confidentiality of personal data, and adherence to ethical principles. The informed consent also stated that participation was voluntary, and the participants could withdraw at any time. All data gathered during this study were stored anonymously in a web central database repository. Every individual was assigned a unique participant number consisting of name and surname first letters and a consecutive number. This unique code identified all participant-specific data (eg, epidemiological and clinical study data). Data access and storage followed the data security, including password-protected access to all computers and folders. The participants received no incentives or rewards to participate.

The trial was conducted at 2 medical sites in Italy: the Foundation IRCCS Carlo Besta Neurological Institute Foundation and the Fondazione Europea Ricerca Biomedica. In each medical site, the participants were recruited from hospital units that treated patients with COVID-19 (ie, the emergency department, surgical units, and critical care units). Health care workers were informed about the possibility of participating in this study with oral communication and a formal email from the institutional study referent. The participants had to confirm their intention to participate in this study by responding to the email while the contact person set up an appointment for the screening interview, as described below.

Participants

The participants' eligibility was verified through a screening interview with the research psychologist, who explained the training's aims and methods and conducted an assessment to identify the participants who could enter this study. This study's inclusion criteria were the following: (1) being currently employed as a health care worker, (2) maximum age of 65 years, (3) absence of medical disorders (heart disease or blood pressure, neurological disorders, or epilepsy), (4) absence of pharmacotherapy that could interfere with the measured data (psychoactive drugs, antihypertensive, or antidepressants), and (5) no significant visual impairment (all with normal visual acuity or corrected to normal).

A total of 43 individuals were assessed for eligibility to participate in this study: 10 (23%) participants did not consent to participate, and 4 (9%) did not complete the posttreatment data and were considered dropouts. For the analyses of the long-term impact of the intervention, we used a complete case analysis. By focusing on the participants with complete data, we aimed to minimize potential bias while maximizing the use of available data.

Out of 33 participants, 29 (88%) adhered to the protocol: 21 (72%) females and 8 (28%) males; mean age of 35.6 (SD 10.3) years; mean years of professional seniority of 10.1 (SD 10.2) years. Further, 22 (76%) participants were from the Foundation IRCCS Carlo Besta Neurological Institute Foundation and 7 (24%) from the Fondazione Europea Ricerca Biomedica. Furthermore, 18 (62%) individuals were doctors and 11 (38%) were nurses or social health workers.

Sample Size Estimation

A formal sample size calculation was not feasible due to the absence of data that could serve as the foundation for the calculation. However, for pilot studies with an anticipated moderate standardized effect size, enrolling a minimum of 15 individuals per group is recommended to achieve a power of 90% for the subsequent main studies [47].

Randomization and Blinding

Before this study commenced, the participants were randomly assigned to the experimental or control group in a 1:1 manner. This study's coordinator created the randomized list with a block randomization procedure using a true random number generator [48]. During the screening interview, eligible individuals who met the inclusion criteria described above received written information about the procedure and were asked to sign the consent form to participate in this study. Only qualified individuals who provided informed consent were randomly assigned to the experimental or control group. The participants received information regarding the allocation result during the screening interview.

Interventions

VR Intervention

We describe the protocol in detail in the following section.

Session 1 was the intake session. Once the participants signed the informed consent and completed the online baseline questionnaire (ie, demographic questions, ad hoc questionnaire on technological solutions and VR, Perceived Stress Scale [PSS-10], State-Trait Anxiety Inventory Form Y-2 [STAI-Y2], Depression, Anxiety and Stress Scale-21 items [DASS-21], and an ad hoc questionnaire on knowledge of stress and anxiety), a research psychologist provided them with a detailed explanation

XSL•FO
RenderX

of the aims and methodology of the 1-week home-based VR program. For the intervention in the VR group, we used the Oculus Quest 2 (Facebook Technologies LLC), a consumer-grade standalone VR system that consists of an HMD and 2 controllers. The Oculus Quest 2 offers a single LCD panel and 1832 × 1920-pixel resolution per eye, with a 90 Hz refresh rate. The participants received a brief 15-minute training on using the VR system and were then given the Oculus Quest 2 to take home.

Training sessions (sessions 2, 3, and 4) consisted of 3 home sessions of approximately 30 minutes each, conducted in 1 week with a distance of 2 days between 1 session and another. The participants tried for about 15 minutes "MIND-VR," a VR-based psychoeducational experience on stress and anxiety created by a team from the University of Milano-Bicocca in collaboration with AnotheReality [21]. The user explored one of three areas within a virtual island in each session. Each focused on different aspects related to stress and anxiety (ie, definitions, causes, symptoms, and main treatments). Subsequently, the participants used the VR relaxation content "The Secret Garden," a 10-minute computer graphic 360° video developed by Riva et al [39,40] for relaxation training freely available on the website (Figure 1) [49]. At the beginning and the end of each session, the participants were asked to complete online the visual analog scale for anxiety (VAS-A) and the State-Trait Anxiety Inventory Form Y-1 (STAI-Y1).

Session 5 was the posttraining session. At the end of the training, the research psychologist met the participants for the postintervention assessment. On this occasion, the participants returned the Oculus Quest 2. The participants were asked to complete the online questionnaires within 1 week after completion of the intervention (ie, PSS-10, STAI-Y2, DASS-21, ad hoc questionnaire on knowledge of stress and anxiety, System Usability Score [SUS], Subjective Difficulty Measure [SDM], Net Promoter Score [NPS], and ad hoc questionnaire on difficulties and adverse effects).

Figure 1. Screenshots of (A) "The Secret Garden" and (B, C) "MIND-VR." VR: virtual reality.



Control Group

This group will undergo baseline and postintervention assessments without undergoing any training during the 1-week intervention period.

Measures

Primary Outcome Measure

At baseline and postintervention, the participants in the VR and the control group completed the following questionnaires:

- PSS-10 [50]: the PSS is a 10-item self-reported measure to assess the current stress level.
- STAI-Y2 [51]: a validated and widely used measure of trait anxiety.
- DASS-21 [52]: a set of 3 self-report scales designed to measure depression, anxiety, and stress symptoms.
- Ad hoc questionnaire on the knowledge of stress and anxiety: the participants completed 4 factual questions in a multiple-choice format (eg, "Which is the first phase of the General Adaptation Syndrome?") and 2 conceptual questions in a short-answer format (eg, "What are the three main categories of stress symptoms?"). The questions were based on the methodology used in a previous study [53]. The 4 factual questions are given a score of 0 if the answer is incorrect and 1 if correct, while the 2 conceptual questions are given a score of 0 to 3 depending on the number of correct items entered. The maximum total score is 10.

XSL•FO
RenderX

Secondary Outcome Measure

Postintervention, the participants in the experimental group filled out the following questionnaires that served as the secondary outcome measures:

- SUS [54]: a measure of usability aspects (ie, efficiency, clarity, or reliability). The participant's scores for each question are converted into a new number, added together, and then multiplied by 2.5 to convert the original scores from 0-40 to 0-100. Based on the research, an SUS score above 68 would be considered above average, while any score below 68 is below average (eg, [55,56]).
- SDM: a horizontal line 100 mm long, anchored by word descriptors at each end, from "not at all" to "very much." The users mark on the line the perceived difficulty in using the VR system, "MIND-VR," and "The Secret Garden" on a scale from 0 ("not at all") to 10 ("very much"). The lower the score, the lower the perceived difficulty.
- NPS [57,58]: it evaluates user satisfaction with a product, in this case: the home-based VR intervention in general, "MIND-VR," and "The Secret Garden." It requires the participants to indicate on a scale from 0 ("not at all") to 10 ("very much") how much they would recommend the products to a family member or friend. The scoring procedure involves categorizing respondents into 3 groups based on their ratings: promoters (scores of 9 or 10), passives (scores of 7 or 8), and detractors (scores of 0 to 6). Then, the percentages of respondents in each category are calculated by dividing the number of respondents in each group by the total number of respondents and multiplying by 100. Finally, the NPS is determined by subtracting the percentage of detractors from the percentage of promoters using the formula: NPS = % promoters – % detractors. The resulting NPS can range from –100 to +100. A positive score indicates that there are more promoters than detractors, while a negative score suggests the opposite. A higher NPS generally indicates higher customer satisfaction and loyalty [57,58].
- Ad hoc questionnaire on difficulties and adverse effects: individuals were asked to rate on a 7-point Likert scale (1="not at all" and 7="very much"): If they had difficulty following the experimental protocol; if they had nausea, headache, dizziness, and eyestrain while using the VR system in general; if they had nausea, headache, dizziness, and eyestrain while using "MIND-VR"; if they had nausea, headache, dizziness, and eyestrain experienced while using "The Secret Garden." The Cronbach α coefficient of the questionnaire was 0.96.

Other Measures

At baseline, the participants in the VR and the control group completed the following: (1) Demographic: genre, age, years of education, profession, hospital, work department, and years of professional seniority. (2) Ad hoc questionnaire on the use of VR: the individuals were asked to indicate whether they had ever tried VR before and their level of knowledge of this technology. (3) Besides, to measure changes in the affective states of the individuals during the intervention, the participants of the experimental group were asked to fill in before and after each of the training sessions the following self-report questionnaires: (a) STAI-Y1 [51], used to assess state anxiety (ie, a temporary emotional condition characterized by apprehension, tension, and fear about a particular situation or activity). (b) VAS-A [59], a horizontal line on a scale from 0 to 100, anchored by word descriptors at each end ("no anxiety" and "very severe anxiety"). The individuals mark on the line the point that they feel represents their perception of their current level of state anxiety. (c) Visual analog scale for emotions (VAS-E) [59]—the participants indicated on a scale from 0 to 100 their current experienced level of 6 primary emotions: anger, happiness, disgust, fear, sadness, and surprise. Several studies have confirmed its reliability and validity (eg, [60]).

Statistical Analyses

Data analyses were carried out through IBM SPSS Statistics (version 25.0 software package for Windows; SPSS Inc) software. ANOVA was used to evaluate baseline characteristics of the 2 groups involved in this study and the overall significance of improvement across primary outcome measures. Categorical variables were compared using Fisher or chi-square tests, and continuous variables using *t* tests (2-tailed) or Mann-Whitney tests, as appropriate.

We used repeated ANOVA for the primary end point outcome. Time was treated as a categorical variable, and the models included group, time, and group-by-time interaction as fixed effects. The conclusions about the usefulness of the VR intervention were based on between-session comparisons of change in PSS-10, STAY-Y2, and DASS-21, and in a questionnaire on the knowledge of stress and anxiety from baseline to posttreatment. Given the relatively small sample, bootstrapping (number of samples: 1000) was used to increase the robustness of statistical analyses. This approach allowed us to estimate the variability of results more accurately and to obtain more precise estimates of the parameters of interest.

A 2×3 repeated measures ANOVA was used to analyze the within-session change in the situational state anxiety (STAI-Y1 and VAS-A) and negative and positive emotions (VAS-E) scores before and after each home-based VR training session. Tests of statistical significance and CIs were 2-sided. A $P<.05$ was considered statistically significant. Descriptive methods were used to report the usability at homes of the VR system and content for the secondary outcome intervention (ie, SUS, SDM, NPS, and ad hoc questionnaire on difficulties and adverse effects).

Results

Descriptive Statistics and Baseline Characteristics

The trial flow chart is summarized in Figure 2. The presence of spurious associations among demographic variables and experimental groups was evaluated by using a chi-square test of independence and *t* test analyses for independent samples (Table 1). The results highlighted that no statistically significant associations were found among considered variables in the 2 groups.

Concerning other the participants' characteristics, baseline measures did not report statistically significant differences

XSL•FO
RenderX

between the 2 groups, meaning that variable scores were homogeneous (Table 2).

Figure 2. Trial flow chart. DASS-21: Depression, Anxiety, and Stress Scale-21 items; PSS: Perceived Stress Scale; STAI-Y1: State-Trait Anxiety Inventory Form Y-1; STAI-Y2: State-Trait Anxiety Inventory Form Y-2; VAS-A: visual analog scale for anxiety; VAS-E: visual analog scale for emotions.

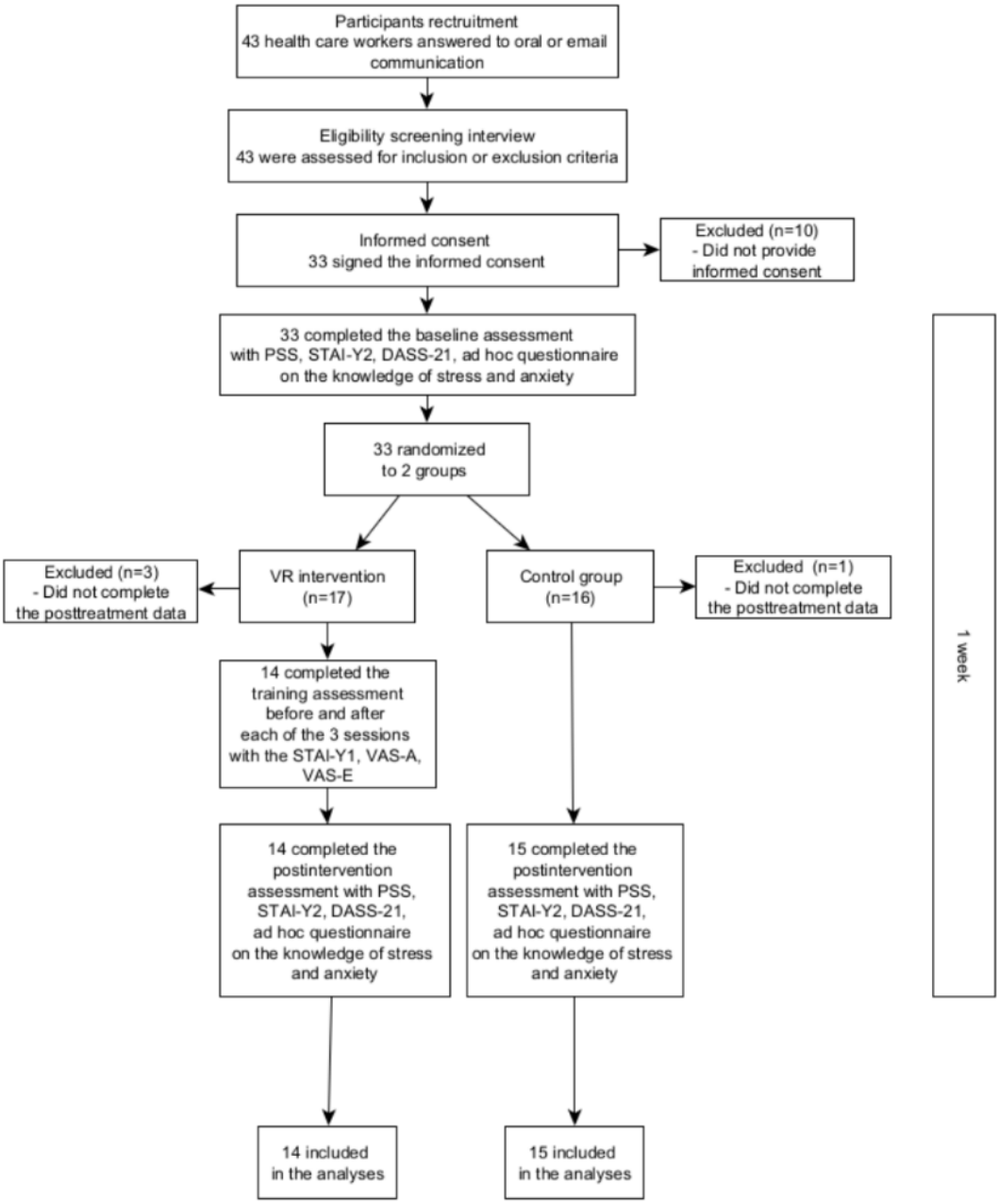


Table 1. Chi-square test of independence on gender and previous experience with VR in the 2 experimental groups.[a]

Variables	VR group (n=15)	Control group (n=14)	Chi-square (*df*)	*P* value
Gender			0.013 (1)	.91
Female	11	10		
Male	4	4		
Previous experience with VR			0.024 (1)	.87
Yes	6	6		
No	9	8		

[a]VR: virtual reality.

XSL•FO
RenderX

Table 2. Mean comparison of demographic and primary outcome measures assessed at baseline in the 2 experimental conditions (VR group: n=15; control group: n=14).[a]

Variables	Mean (SD)	*t* test (*df*)	*P* value
Age (years)		–1.3 (27)	.2
VR	33.2 (10)		
Control	38.1 (10.3)		
Years of professional seniority		–1.27 (27)	.21
VR	7.8 (9.4)		
Control	12.5 (10.7)		
Knowledge of VR		–0.087 (27)	.43
VR	1.53 (0.743)		
Control	1.86 (1.35)		
PSS-10[b]		0.005 (27)	.99
VR	20.8 (5.7)		
Control	20.8 (5.1)		
STAI-Y2[c]		0.121 (27)	.90
VR	46.8 (12.6)		
Control	46.2 (13.2)		
DASS-21[d] stress		0.614 (27)	.55
VR	28.1 (10.3)		
Control	26 (8.2)		
DASS-21 anxiety		0.263 (27)	.79
VR	20.1 (5.5)		
Control	19.5 (5.9)		
DASS-21 depression		1.196 (27)	.24
VR	23.1 (4.5)		
Control	20.5 (6.5)		
Knowledge of stress and anxiety		–0.475 (27)	.64
VR	4.8 (2.6)		
Control	5.2 (2.01)		

[a]VR: virtual reality.

[b]PSS-10: Perceived Stress Scale.

[c]STAI-Y2: State-Trait Anxiety Inventory Form Y-2.

[d]DASS-21: Depression, Anxiety, and Stress Scale-21 items.

Impact of the VR Intervention on Primary Outcomes

To investigate whether a home-based VR intervention can support health care workers in managing stress and anxiety, we compared the effects of the VR intervention and control condition on the primary outcome measures. The average change in primary outcome measures for the VR and control conditions from baseline to postintervention is depicted in Table 3.

The VR intervention led to a decrease in PSS-10 scores (mean 6.46, SD 7.41; 95% CI 2.77 to 10.5), which was significantly larger than the control condition (mean 0.785, SD 7.43; 95% CI –2.93 to 4.66; t_{27}=2.06; *P*=.046). The effect size calculated as Cohen *d* of the reported change in the VR intervention compared to the control condition was high, *d*=0.765. Figure 3 shows a significant difference between the VR and control conditions in their average change in perceived stress level from baseline to after the intervention as measured by the PSS-10.

However, results did not reveal similar reductions in stress, anxiety, and depression levels as measured by the DASS-21 or in trait anxiety as assessed by the STAI-Y1 (Table 3). Furthermore, results showed that the VR intervention led to an increase in the knowledge of stress and anxiety, as assessed by the ad hoc questionnaire adopted (mean –2.09, SD 3.28; 95% CI –3.86 to –0.529) significantly greater than in the control condition (mean 0.71, SD 2.67; 95% CI –1.33 to 1.38; t_{27}=–2.09;

XSL•FO
RenderX

P=.046; Figure 4). The effect size calculated as Cohen *d* of the reported change in the VR intervention compared to the control condition was high (*d*=–0.778).

Table 3. Comparison of the effect of the home-based VR[a] intervention and control condition on scores on PSS-10[b], STAI-Y2[c], DASS-21[d], and ad hoc questionnaire on the use of VR. All variables are calculated as change scores from the baseline to the second measurement.

Variables	Control (n=14), mean (SD)	VR intervention (n=15), mean (SD)	*t* test	*P* value	Bootstrap *P* value	Bootstrap 95% CI
PSS-10	0.785 (7.43)	6.46 (7.41)	2.06	.049	.046	0.29 to 10.8
STAI-Y2 (trait)	3.38 (15.1)	2.06 (15.9)	–0.211	.83	.83	–12 to 10.8
DASS-21 stress	–1.85 (16.1)	8.53 (10.6)	2.06	.049	.053	1.2 to 20.6
DASS-21 anxiety	–2 (10.6)	4 (5.34)	1.93	.06	.08	0.87 to 12.3
DASS-21 depression	–2.57 (13.39)	5.06 (5.11)	2.05	.05	.06	0.14 to 14.7
Knowledge of stress and anxiety	0.71 (2.67)	–2.26 (3.28)	–2.09	.046	.046	–4.4 to –1.2

[a]VR: virtual reality.

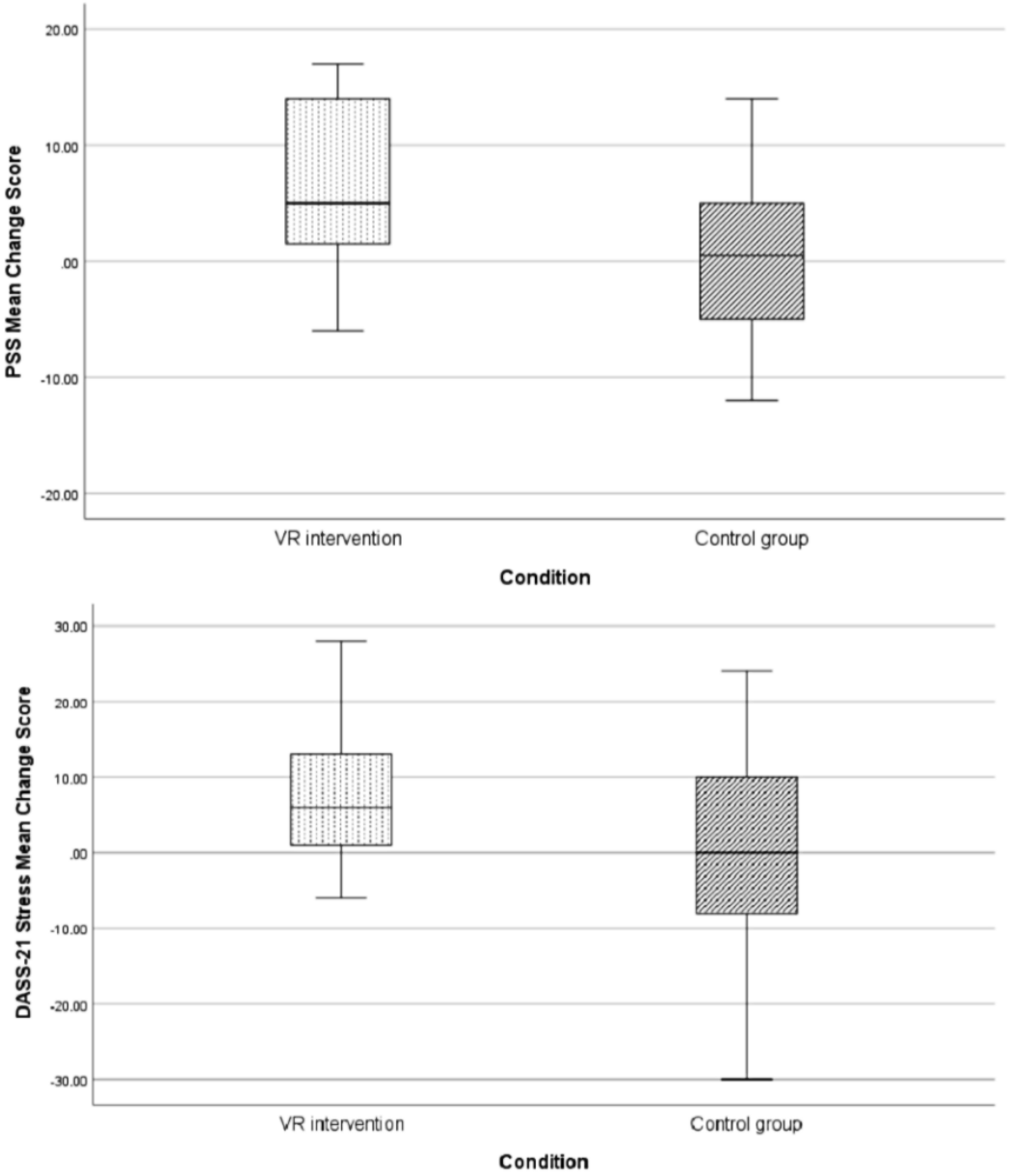
[b]PSS-10: Perceived Stress Scale.

[c]STAI-Y2: State-Trait Anxiety Inventory Form Y-2.

[d]DASS-21: Depression, Anxiety, and Stress Scale-21 items.

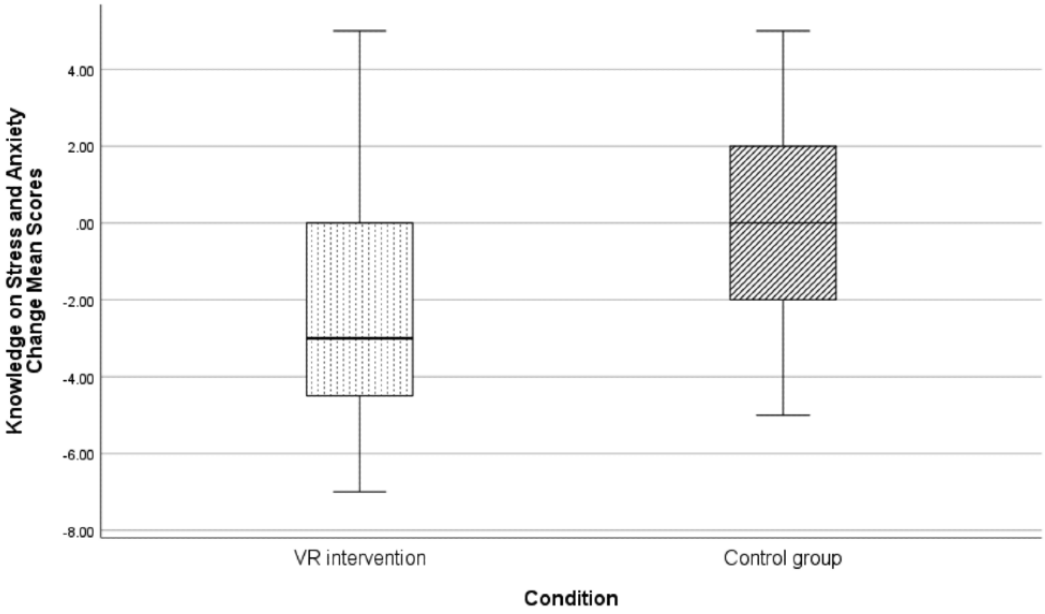
XSL•FO
RenderX

Figure 3. Change in the stress level as assessed with PSS-10 and DASS-21 from baseline measurement to posttreatment (1 week) for both conditions. Positive values indicate a decrease in stress levels. Error bars are 95% CIs. DASS-21: Depression, Anxiety, and Stress Scale-21 items; PSS-10: Perceived Stress Scale; VR: virtual reality.



XSL•FO
RenderX

Figure 4. Change in the knowledge of stress and anxiety as measured by the ad hoc questionnaire from baseline measurement to posttreatment (1 week) for both conditions. Negative values indicate an increase in the knowledge of stress and anxiety. Error bars are 95% CIs. VR: virtual reality.



Immediate Effects of the VR Intervention on Secondary Outcomes

A 2×3 repeated measures ANOVA revealed a significant effect of time in the STAI-Y1 ($F_{1,14}$=5.8; P=.03; η2=0.293) and VAS-A ($F_{1,14}$=5.33; P=.03; η2=0.276), indicating that the VR training sessions were able to decrease state anxiety. Besides, results on the VAS-E showed a statistically significant difference (P=.05) before and after the use of "MIND-VR" in the scores at VAS happiness ($F_{1,14}$=4.56; P=.05; VAS-A; η2=0.246), VAS anger ($F_{1,14}$=7.8; P=.10; VAS-A; η2=0.358), and VAS sadness ($F_{1,14}$=8.53; P=.01; η2=0.279), indicating that the training sessions using VR were useful for enhancing happiness and decreasing sadness and anger (Table 4).

XSL•FO
RenderX

Table 4. Mean changes in state anxiety as assessed with STAI-Y1[a] and VAS-A[b], and positive and negative emotions as measured through VAS-E[c] in the 3 training sessions of the home-based VR[d] intervention.

Variables	Pre mean (SD)	Post mean (SD)	Effects	*F* test (*df*)	*P* value	η2
STAI-Y1						
Session 1	41.4 (14.3)	33.4 (8.7)	Time	5.8 (1,14)	.03	0.293
Session 2	42.4 (13.4)	37.7 (13.9)	Session	0.45 (1,14)	.65	0.065
Session 3	40.4 (12)	36.8 (11.5)	Time × session	0.465 (1,14)	.64	0.067
VAS-A						
Session 1	36.7 (23.5)	20.7 (20.8)	Time	5.33 (1,14)	.04	0.276
Session 2	33.3 (20.9)	24 (23.8)	Session	0.069 (1,14)	.93	0.011
Session 3	32 (22.7)	28 (19.3)	Time × session	1.15 (1,14)	.35	0.151
VAS-DIS[e]						
Session 1	9.3 (19.8)	5.3 (20.7)	Time	0.723 (1,14)	.41	0.049
Session 2	12.7 (25.7)	9.3 (24.6)	Session	3.08 (1,14)	.08	0.322
Session 3	7.3 (15.7)	5.3 (15.5)	Time × session	0.072 (1,14)	.93	0.011
VAS-HP[f]						
Session 1	56.7 (23.1)	62.6 (11.7)	Time	4.56 (1,14)	.05	0.246
Session 2	51.3 (24.4)	62 (15.2)	Session	0.171 (1,14)	.84	0.026
Session 3	50.7 (24)	64 (15.5)	Time × session	0.638 (1,14)	.54	0.089
VAS-AG[g]						
Session 1	20.7 (27.9)	6.7 (16.3)	Time	7.8 (1,14)	.01	0.358
Session 2	16 (25.5)	10 (19.6)	Session	0.041 (1,14)	.96	0.006
Session 3	16.7 (22.8)	12 (17.4)	Time × session	0.528 (1,14)	.60	0.075
VAS-FE [h]						
Session 1	31.3 (30.6)	14 (22.6)	Time	2.83 (1,14)	.11	0.168
Session 2	15.3 (18.8)	12.7 (23.4)	Session	1.4 (1,14)	.28	0.177
Session 3	16 (23.5)	12.7 (23.1)	Time × session	0.962 (1,14)	.41	0.129
VAS-SP[i]						
Session 1	18 (20.7)	16 (20.9)	Time	0.003 (1,14)	.95	0.006
Session 2	8.7 (14)	11.3 (17.6)	Session	2.21 (1,14)	.15	0.254
Session 3	10.7 (18.7)	10.7 (18.7)	Time × session	0.958 (1,14)	.41	0.128
VAS-SD[j]						
Session 1	36 (37.5)	19.3 (27.6)	Time	8.53 (1,14)	.01	.379
Session 2	28.7 (26.7)	21.3 (20.9)	Session	0.773 (1,14)	.48	0.106
Session 3	26 (28.2)	15.3 (20.9)	Time × session	0.383 (1,14)	.69	0.056

[a]STAI-Y2: State-Trait Anxiety Inventory Form Y-2.

[b]VAS-A: visual analog scale for anxiety.

[c]VAS-E: visual analog scale for emotions.

[d]VR: virtual reality.

[e]VAS-DIS: visual analog scale for disgust.

[f]VAS-HP: visual analog scale for happiness.

[g]VAS-AG: visual analog scale for anger.

[h]VAS-FE: visual analog scale for fear.

[i]VAS-SP: visual analog scale for surprise.

XSL•FO
RenderX

[j]VAS-SD: visual analog scale for sadness.

Secondary Outcomes: Usability at Home of the VR System and Content

Regarding the usability at home of the VR system, the mean scores at the SUS were greater than 68 (mean 78, SD 12.8; 95% CI 70-86.5), indicating a high usability. Regarding SDM, the scores were low both in general (mean 2.5, SD 1.1; 95% CI 1.93-3.07) and concerning "MIND-VR" (mean 2.3, SD 1.03; 95% CI 1.7-2.7) and "The Secret Garden" (mean 2.3, SD 1.1; 95% CI 1.7-2.8), showing that the VR system and content were perceived as easy-to-use. In addition, results showed low scores in the adverse effects of VR use related to cybersickness symptoms, both overall (mean 2.9, SD 1.7; 95% CI 2.13-3.8) and concerning "MIND-VR" (mean 2.4, SD 1.4; 95% CI 1.8-3.2) and "The Secret Garden" (mean 2.4, SD 1.4; 95% CI 1.73-3.13), indicating that the use of VR had no effects related to cybersickness. As for the NPS, however, it was –13 concerning the program as a whole, –33 for "MIND-VR," and –6 for "The Secret Garden," suggesting varying levels of satisfaction among the participants.

Discussion

Principal Findings

This pilot trial delivered clinical and feasibility data on a home-based VR intervention for supporting stress and anxiety management in a sample of health care workers, one of the categories most affected by these conditions and their adverse effects.

Starting with the primary outcome measures, the results of this randomized pilot trial showed that, compared to the passive control condition, our 1-week remote VR-based intervention significantly reduced the level of perceived stress as assessed by the PSS-10. This reduction is underscored by the notable effect size of d=0.765, suggesting a meaningful impact of the intervention. However, the home-based VR training did not yield similar reductions in stress, anxiety, and depression levels as assessed by the DASS-21 or in trait anxiety as evaluated through the STAI-Y1.

The discrepancy in results, wherein a reduction was noticed in perceived stress levels as evaluated by the PSS-10 but not reflected in the outcomes from the DASS-21, suggests potential differences in the sensitivity or specificity of these measures in capturing changes following the VR intervention. It could also imply that the VR intervention primarily targeted aspects of stress that are more closely aligned with the constructs measured by the PSS-10, highlighting the importance of selecting appropriate outcome measures tailored to the specific objectives of the intervention.

Furthermore, the fact that the results showed no differences in anxiety and depression levels as assessed by the DASS-21 or in trait anxiety as evaluated through the STAI-Y1 could be explained by the short duration of the training. In support of this hypothesis, also in the study by Riva et al [40] on the use of VR to assist individuals in coping with the psychological burden related to the COVID-19 pandemic, the use at home for one week of a 360° video for relaxation (ie, "The Secret Garden," the same one used in our study), used through a low-cost cardboard HMD, had brought improvements in stress levels, but not for the perceived anxiety levels (as measured with the DASS-21) [40]. In contrast, a reduction in trait anxiety was observed after a program with VR lasting 5 weeks, consisting of 8 sessions based on biofeedback techniques, and delivered at the therapist's office [15]. Future studies must deeply investigate the optimal duration of remote training for VR-based stress and anxiety management to maximize its effectiveness.

Regarding the level of anxiety and stress knowledge, the results indicated a significant change between different time points (baseline vs end of intervention assessment) after the home VR intervention. Specifically, compared with the passive control condition, the VR intervention significantly increased knowledge of anxiety and stress, with an effect size high (d=–0.778).

This result, in line with previous literature [21,61,62], stresses the potential of VR as a valuable and innovative tool for promoting scientific and medical knowledge on stress and anxiety and other mental disorders such as depression [22,23]. Through immersion in the VR environment, individuals may feel secure and free from judgment, leading to a greater understanding of their mental health condition [22]. Greater awareness of symptoms can have significant implications for clinical outcomes, including treatment-seeking, adherence, and recovery [63,64], decreasing hospitalization rates, and saving long-term physician consultation costs. More studies are needed to expand the limited literature on VR-based psychoeducation. For example, it would be interesting to test whether inserting engaging elements and guiding cues before the VR psychoeducational experience can increase effectiveness in learning. Equally important is testing the total duration of the psychoeducation program in VR, trying to propose interventions of longer duration than the one proposed in this study (ie, 1-week).

Results for the secondary outcomes indicated that the home VR training sessions immediately decreased perceived stress and anxiety levels. Specifically, the results of the analyses on STAI-Y1, VAS-A, and VAS-E showed that after the use of the psychoeducational (ie, "MIND-VR") and relaxing (ie, "The Secret Garden") VR experience, a significant increase in the intensity of positive emotions (ie, happiness), a significant decrease after VR home training sessions in the levels of anxiety, anger, and sadness, and an increase in the levels of happiness was experienced. This fact appears important as it points out how even using VR at home for short periods can help induce a state of relaxation in the individual, with immediate positive psychological effects. As underlined by the broaden-and-build model [65], experiencing positive emotions can enhance interaction with others or engagement in creative challenges. Besides, a positive emotional state positively impacts learning processes, promoting acquiring the information provided [66].

This study's virtual content can be easily and affordably incorporated as supplementary treatment or ongoing support to

XSL•FO
RenderX

improve the efficacy of established evidence-based interventions for health care workers. It could also be readily modified to help other groups dealing with stress and anxiety. “MIND-VR” and “The Secret Garden” are available for free in multiple languages, including English. As the use of free content could decrease the cost and accessibility of care support, it seems essential that future studies study the topic in detail, identifying other content potentially beneficial for both stress and anxiety management and other conditions, such as VR video games or naturalistic experiences.

Finally, regarding the usability at home of the VR system and content, analyses of questionnaires (ie, SUS, SDM, or ad hoc questionnaires on adverse effects) indicated that health care workers found the VR system easy to use and without adverse effects related to cybersickness symptoms (eg, nausea and vertigo). Moreover, both the VR psychoeducational experience (“MIND-VR”) and the relaxation content (“The Secret Garden”) were reported as easy to use and did not induce cybersickness–related effects. These findings offer preliminary data supporting the ease of use and lack of adverse effects of VR used in the home, even in individuals without high knowledge of this technology and of different genders and ages, as in our sample case. As the use of VR in home settings can increase participation, involvement, and compliance to follow psychological support programs, increasing treatment adherence and lowering self-stigma [67,68], future studies should deeply investigate this topic.

However, it is important to highlight that from the results of the NPS analysis, a percentage of the individuals were not satisfied with the training or content, suggesting an opportunity for improvement. To gain valuable insights into areas where changes need to be made, it will be important to carefully examine the participants’ feedback also from a qualitative perspective, such as through interviews and focus groups.

The results of this study should be interpreted based on several additional potential limitations. First, in this study, a third control arm (ie, treatment as usual) was not included due to the limitation of cost and feasibility. While it has been suggested that there may not be a significant difference in measured outcomes between active and passive control groups [69], future research needs to compare the effectiveness of the VR intervention with other methods. For example, future studies could explore the comparative effectiveness of VR interventions with placebo interventions and with traditional therapy approaches such as cognitive behavioral therapy or exposure therapy, while also investigating the synergistic effects of combining VR with mindfulness techniques for stress reduction and anxiety management.

Second, in this study, we used a full case analysis instead of following intention-to-treat principles. The analysis was conducted on the complete data (per protocol), believing that this approach was more appropriate to the main objective of this study, which was to evaluate the feasibility, acceptability, and potential preliminary impact of the intervention in a small group of participants. Pilot studies are not intended to provide definitive estimates of efficacy, but rather to test methodologies and gather useful information for planning future larger studies. In this context, the priority was not to obtain generalizable results but to explore the practical applicability of the protocol and identify potential critical issues. In addition, the small sample and missing data (n=4, 12%, dropout) make the application of imputation techniques delicate, which could have introduced significant bias. Although the per-protocol analysis provided useful information to assess the potential effect of the intervention, future larger studies will need to adopt rigorous approaches based on intention-to-treat principles to ensure a more robust and balanced assessment of the effectiveness. Third, it is impossible to guarantee that the participants performed the remote VR-based training session or that another individual replaced them. To date, this drawback is common with most (if not all) home-based training programs. Fourth, in this study, health care workers who volunteered to participate, in addition to having specific demographic characteristics (ie, the majority were female adults with low knowledge of VR), may have been more stressed or receptive to a VR intervention, so the responses may be susceptible to selection bias. The presented results should be replicated in more diverse samples to improve the generalizability of our findings. Fifth, in this study, we used self-reported questionnaires to assess the perceived level of stress and anxiety. Future research should also assess physiological measures related to these conditions (eg, heart rate variability and cortisol levels). In addition, it might be worth investigating the clinical and learning outcomes at follow-up assessments to determine the longer-term effects of remote VR-based training. Finally, in this study, we used only quantitative data, but it could be useful to collect also qualitative feedback to provide deeper insights into the usability and acceptability of the technology, offering valuable perspectives from the participants’ subjective experiences.

Conclusions

This pilot trial provided clinical outcomes and feasibility results for a remote VR-based intervention for managing stress and anxiety in health care workers. The observed clinical outcomes showed greater improvement in perceived stress as assessed by the PSS-10 and the knowledge of stress and anxiety in the VR group than in the control group. Nevertheless, the findings did not indicate analogous declines in perceived stress, anxiety, and depression levels as gauged by the DASS-21 or in enduring anxiety as evaluated by the STAI-Y1. The home VR training sessions immediately decreased perceived levels of stress and anxiety. The participants found using the VR system and content at home easy to use and without cyber sickness effects. The findings should encourage further research and more extensive studies exploring the potential of VR interventions for delivering psychological support programs in the home, both for caregivers and for other categories of people who need support in managing stress and anxiety.

XSL•FO
RenderX

Data Availability

This study's datasets are available upon request from the authors.

Authors' Contributions

FP conceived the work and wrote the first draft of the manuscript. FP, EO, LA, FA, SS, and MC supervised the study in its clinical aspect. All authors contributed to the research activities related to this study and agreed to the published version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1178 KB - games_v13i1e50326_app1.pdf]

References

1. Kardong-Edgren S, Farra SL, Alinier G, Young HM. A call to unify definitions of virtual reality. Clin Simul Nurs 2019;31:28-34. [doi: 10.1016/j.ecns.2019.02.006]
2. Rojas-Sánchez MA, Palos-Sánchez PR, Folgado-Fernández JA. Systematic literature review and bibliometric analysis on virtual reality and education. Educ Inf Technol (Dordr) 2023;28(1):155-192 [FREE Full text] [doi: 10.1007/s10639-022-11167-5] [Medline: 35789766]
3. Pallavicini F, Pepe A, Minissi ME. Gaming in virtual reality: what changes in terms of usability, emotional response and sense of presence compared to non-immersive video games? Simul Gaming 2019;50(2):136-159. [doi: 10.1177/1046878119831420]
4. Izard SG, Juanes JA, Peñalvo FJG, Estella JMG, Ledesma MJS, Ruisoto P. Virtual reality as an educational and training tool for medicine. J Med Syst 2018;42(3):50. [doi: 10.1007/s10916-018-0900-2] [Medline: 29392522]
5. Pallavicini F, Pepe A, Clerici M, Mantovani F. Virtual reality applications in medicine during the COVID-19 pandemic: systematic review. JMIR Serious Games 2022;10(4):e35000 [FREE Full text] [doi: 10.2196/35000] [Medline: 36282554]
6. Freeman D, Reeve S, Robinson A, Ehlers A, Clark D, Spanlang B, et al. Virtual reality in the assessment, understanding, and treatment of mental health disorders. Psychol Med 2017;47(14):2393-2400 [FREE Full text] [doi: 10.1017/S003329171700040X] [Medline: 28325167]
7. Fodor LA, Cote CD, Cuijpers P, Szamoskozi, David D, Cristea IA. The effectiveness of virtual reality based interventions for symptoms of anxiety and depression: a meta-analysis. Sci Rep 2018;8(1):10323 [FREE Full text] [doi: 10.1038/s41598-018-28113-6] [Medline: 29985400]
8. Riva G. Virtual reality in clinical psychology. Compr Clin Psychol 2022:91-105. [doi: 10.1016/B978-0-12-818697-8.00006-6]
9. Lindner P, Miloff A, Hamilton W, Carlbring P. The potential of consumer-targeted virtual reality relaxation applications: descriptive usage, uptake and application performance statistics for a first-generation application. Front Psychol 2019;10:132 [FREE Full text] [doi: 10.3389/fpsyg.2019.00132] [Medline: 30778311]
10. Anderson-Hanley C, Maloney M, Barcelos N, Striegnitz K, Kramer A. Neuropsychological benefits of neuro-exergaming for older adults: a pilot study of an interactive physical and cognitive exercise system (iPACES). J Aging Phys Act 2017;25(1):73-83. [doi: 10.1123/japa.2015-0261] [Medline: 27337738]
11. Pallavicini F, Pepe A. Virtual reality games and the role of body involvement in enhancing positive emotions and decreasing anxiety: within-subjects pilot study. JMIR Serious Games 2020;8(2):e15635 [FREE Full text] [doi: 10.2196/15635] [Medline: 32554371]
12. Browning MHEM, Mimnaugh KJ, van Riper CJ, Laurent HK, LaValle SM. Can simulated nature support mental health? Comparing short, single-doses of 360-degree nature videos in virtual reality with the outdoors. Front Psychol 2020;10:2667 [FREE Full text] [doi: 10.3389/fpsyg.2019.02667] [Medline: 32010003]
13. Kim H, Kim DJ, Kim S, Chung WH, Park K, Kim JDK, et al. Effect of virtual reality on stress reduction and change of physiological parameters including heart rate variability in people with high stress: an open randomized crossover trial. Front Psychiatry 2021;12:614539 [FREE Full text] [doi: 10.3389/fpsyt.2021.614539] [Medline: 34447320]
14. Schrempf MC, Petzold J, Petersen MA, Arndt TT, Schiele S, Vachon H, et al. A randomised pilot trial of virtual reality-based relaxation for enhancement of perioperative well-being, mood and quality of life. Sci Rep 2022;12(1):12067 [FREE Full text] [doi: 10.1038/s41598-022-16270-8] [Medline: 35835944]
15. Gaggioli A, Pallavicini F, Morganti L, Serino S, Scaratti C, Briguglio M, et al. Experiential virtual scenarios with real-time monitoring (interreality) for the management of psychological stress: a block randomized controlled trial. J Med Internet Res 2014;16(7):e167 [FREE Full text] [doi: 10.2196/jmir.3235] [Medline: 25004803]
16. Maarsingh BM, Bos J, Van Tuijn CFJ, Renard SB. Changing stress mindset through Stressjam: a virtual reality game using biofeedback. Games Health J 2019;8(5):326-331 [FREE Full text] [doi: 10.1089/g4h.2018.0145] [Medline: 31539291]

XSL•FO
RenderX

17. van Rooij M, Lobel A, Harris O, Smit N, Granic I. DEEP: a biofeedback virtual reality game for children at-risk for anxiety. 2016 Presented at: Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems; 2016 May 07; California, San Jose, USA.
18. Navarro-Haro MV, López-Del-Hoyo Y, Campos D, Linehan MM, Hoffman HG, García-Palacios A, et al. Meditation experts try virtual reality mindfulness: a pilot study evaluation of the feasibility and acceptability of virtual reality to facilitate mindfulness practice in people attending a mindfulness conference. PLoS One 2017;12(11):e0187777 [FREE Full text] [doi: 10.1371/journal.pone.0187777] [Medline: 29166665]
19. Seabrook E, Kelly R, Foley F, Theiler S, Thomas N, Wadley G, et al. Understanding how virtual reality can support mindfulness practice: mixed methods study. J Med Internet Res 2020;22(3):e16106. [doi: 10.2196/16106] [Medline: 32186519]
20. Pallavicini F, Orena E, Achille F, Cassa M, Vuolato C, Stefanini S, et al. Psychoeducation on stress and anxiety using virtual reality: a mixed-methods study. App Sci 2022;12(18):9110. [doi: 10.3390/app12189110]
21. Pallavicini F, Orena E, di Santo S, Greci L, Caragnano C, Ranieri P, et al. MIND-VR: design and evaluation protocol of a virtual reality psychoeducational experience on stress and anxiety for the psychological support of healthcare workers involved in the COVID-19 pandemic. Front Virtual Real 2021;2:620225. [doi: 10.3389/frvir.2021.620225]
22. Migoya-Borja M, Delgado-Gómez D, Carmona-Camacho R, Porras-Segovia A, López-Moriñigo J, Sánchez-Alonso M, et al. Feasibility of a virtual reality-based psychoeducational tool (VRight) for depressive patients. Cyberpsychol Behav Soc Netw 2020;23(4):246-252. [doi: 10.1089/cyber.2019.0497] [Medline: 32207997]
23. Tielman ML, Neerincx MA, van Meggelen M, Franken I, Brinkman W. How should a virtual agent present psychoeducation? Influence of verbal and textual presentation on adherence. Technol Health Care 2017;25(6):1081-1096 [FREE Full text] [doi: 10.3233/THC-170899] [Medline: 28800346]
24. Riches S, Azevedo L, Bird L, Pisani S, Valmaggia L. Virtual reality relaxation for the general population: a systematic review. Soc Psychiatry Psychiatr Epidemiol 2021;56(10):1707-1727 [FREE Full text] [doi: 10.1007/s00127-021-02110-z] [Medline: 34120220]
25. Shah SMA, Mohammad D, Qureshi MFH, Abbas MZ, Aleem S. Prevalence, psychological responses and associated correlates of depression, anxiety and stress in a global population, during the coronavirus disease (COVID-19) pandemic. Community Ment Health J 2021;57(1):101-110 [FREE Full text] [doi: 10.1007/s10597-020-00728-y] [Medline: 33108569]
26. Salari N, Hosseinian-Far A, Jalali R, Vaisi-Raygani A, Rasoulpoor S, Mohammadi M, et al. Prevalence of stress, anxiety, depression among the general population during the COVID-19 pandemic: a systematic review and meta-analysis. Global Health 2020;16(1):57 [FREE Full text] [doi: 10.1186/s12992-020-00589-w] [Medline: 32631403]
27. Ruotsalainen JH, Verbeek JH, Mariné A, Serra C. Preventing occupational stress in healthcare workers. Cochrane Database Syst Rev 2015(4):CD002892 [FREE Full text] [doi: 10.1002/14651858.CD002892.pub5] [Medline: 25847433]
28. Johnson J, Hall LH, Berzins K, Baker J, Melling K, Thompson C. Mental healthcare staff well-being and burnout: a narrative review of trends, causes, implications, and recommendations for future interventions. Int J Ment Health Nurs 2018;27(1):20-32 [FREE Full text] [doi: 10.1111/inm.12416] [Medline: 29243348]
29. Spoorthy MS, Pratapa SK, Mahant S. Mental health problems faced by healthcare workers due to the COVID-19 pandemic-a review. Asian J Psychiatr 2020;51:102119 [FREE Full text] [doi: 10.1016/j.ajp.2020.102119] [Medline: 32339895]
30. Marvaldi M, Mallet J, Dubertret C, Moro MR, Guessoum SB. Anxiety, depression, trauma-related, and sleep disorders among healthcare workers during the COVID-19 pandemic: a systematic review and meta-analysis. Neurosci Biobehav Rev 2021;126:252-264 [FREE Full text] [doi: 10.1016/j.neubiorev.2021.03.024] [Medline: 33774085]
31. Søvold LE, Naslund JA, Kousoulis AA, Saxena S, Qoronfleh MW, Grobler C, et al. Prioritizing the mental health and well-being of healthcare workers: an urgent global public health priority. Front Public Health 2021;9:679397 [FREE Full text] [doi: 10.3389/fpubh.2021.679397] [Medline: 34026720]
32. Krystal JH. Responding to the hidden pandemic for healthcare workers: stress. Nat Med 2020;26(5):639. [doi: 10.1038/s41591-020-0878-4] [Medline: 32350461]
33. Nijland JWHM, Veling W, Lestestuiver BP, Van Driel CMG. Virtual reality relaxation for reducing perceived stress of intensive care nurses during the COVID-19 pandemic. Front Psychol 2021;12:706527 [FREE Full text] [doi: 10.3389/fpsyg.2021.706527] [Medline: 34659021]
34. Beverly E, Hommema L, Coates K, Duncan G, Gable B, Gutman T, et al. A tranquil virtual reality experience to reduce subjective stress among COVID-19 frontline healthcare workers. PLoS One 2022;17(2):e0262703 [FREE Full text] [doi: 10.1371/journal.pone.0262703] [Medline: 35139092]
35. Birckhead B, Eberlein S, Alvarez G, Gale R, Dupuy T, Makaroff K, et al. Home-based virtual reality for chronic pain: protocol for an NIH-supported randomised-controlled trial. BMJ Open 2021;11(6):e050545 [FREE Full text] [doi: 10.1136/bmjopen-2021-050545] [Medline: 34130965]
36. Garcia LM, Birckhead BJ, Krishnamurthy P, Sackman J, Mackey IG, Louis RG, et al. An 8-week self-administered at-home behavioral skills-based virtual reality program for chronic low back pain: double-blind, randomized, placebo-controlled trial conducted during COVID-19. J Med Internet Res 2021;23(2):e26292 [FREE Full text] [doi: 10.2196/26292] [Medline: 33484240]

XSL•FO
RenderX

37. Sampaio M, Haro MVN, De Sousa B, Melo WV, Hoffman HG. Therapists make the switch to telepsychology to safely continue treating their patients during the COVID-19 pandemic. Virtual reality telepsychology may be next. Front Virtual Real 2021;1:576421 [FREE Full text] [doi: 10.3389/frvir.2020.576421] [Medline: 33585834]
38. Pedram S, Palmisano S, Perez P, Mursic R, Farrelly M. Examining the potential of virtual reality to deliver remote rehabilitation. Comput Hum Behav 2020;105:106223. [doi: 10.1016/j.chb.2019.106223]
39. Riva G, Bernardelli L, Browning MHEM, Castelnuovo G, Cavedoni S, Chirico A, et al. COVID feel good-an easy self-help virtual reality protocol to overcome the psychological burden of coronavirus. Front Psychiatry 2020;11:563319 [FREE Full text] [doi: 10.3389/fpsyt.2020.563319] [Medline: 33173511]
40. Riva G, Bernardelli L, Castelnuovo G, Di Lernia D, Tuena C, Clementi A, et al. A virtual reality-based self-help intervention for dealing with the psychological distress associated with the COVID-19 lockdown: an effectiveness study with a two-week follow-up. Int J Environ Res Public Health 2021;18(15):8188 [FREE Full text] [doi: 10.3390/ijerph18158188] [Medline: 34360479]
41. Motlova LB, Balon R, Beresin EV, Brenner AM, Coverdale JH, Guerrero APS, et al. Psychoeducation as an opportunity for patients, psychiatrists, and psychiatric educators: why do we ignore it? Acad Psychiatry 2017;41(4):447-451. [doi: 10.1007/s40596-017-0728-y] [Medline: 28536990]
42. Van Daele T, Hermans D, Van Audenhove C, Van den Bergh O. Stress reduction through psychoeducation: a meta- analytic review. Health Educ Behav 2012;39(4):474-485. [doi: 10.1177/1090198111419202] [Medline: 21986242]
43. Donker T, Griffiths KM, Cuijpers P, Christensen H. Psychoeducation for depression, anxiety and psychological distress: a meta-analysis. BMC Med 2009;7:79 [FREE Full text] [doi: 10.1186/1741-7015-7-79] [Medline: 20015347]
44. Alonso M, Guillén AI, Muñoz M. Interventions to reduce internalized stigma in individuals with mental illness: a systematic review. Span J Psychol 2019;22:E27. [doi: 10.1017/sjp.2019.9] [Medline: 31084665]
45. Mughairbi FA, Abdulaziz Alnajjar A, Hamid A. Effects of psychoeducation and stress coping techniques on posttraumatic stress disorder symptoms. Psychol Rep 2020;123(3):710-724. [doi: 10.1177/0033294118825101] [Medline: 30760172]
46. Pallavicini F, Orena E, di Santo S, Greci L, Caragnano C, Ranieri P, et al. A virtual reality home-based training for the management of stress and anxiety among healthcare workers during the COVID-19 pandemic: study protocol for a randomized controlled trial. Trials 2022;23(1):451 [FREE Full text] [doi: 10.1186/s13063-022-06337-2] [Medline: 35655231]
47. Whitehead AL, Julious SA, Cooper CL, Campbell MJ. Estimating the sample size for a pilot randomised trial to minimise the overall trial sample size for the external pilot and main trial for a continuous outcome variable. Stat Methods Med Res 2016;25(3):1057-1073 [FREE Full text] [doi: 10.1177/0962280215588241] [Medline: 26092476]
48. Research Randomizer. URL: https://www.randomizer.org/ [accessed 2025-01-17]
49. Experience the magic of COVID Feel Good in Spatial Video with Vision Pro!. COVID Feel Good. URL: https://www.covidfeelgood.com/ [accessed 2025-01-17]
50. Cohen S, Kamarck T, Mermelstein R. Perceived stress scale. In: Measuring Stress: A Guide For Health and Social Scientists. Oxford: Oxford University Press; 1997.
51. Spielberger CD, Sydeman SJ. State-trait anxiety inventory and state-trait anger expression inventory. In: Maruish ME, editor. The Use of Psychological Testing for Treatment Planning and Outcome Assessment. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc; 1994:292-321.
52. Lovibond PF, Lovibond SH. The structure of negative emotional states: comparison of the Depression Anxiety Stress Scales (DASS) with the Beck depression and anxiety inventories. Behav Res Ther 1995;33(3):335-343. [doi: 10.1016/0005-7967(94)00075-u] [Medline: 7726811]
53. Parong J, Mayer RE. Learning science in immersive virtual reality. J Educ Psychol 2018;110(6):785-797. [doi: 10.1037/edu0000241]
54. Peres SC, Pham T, Phillips R. Validation of the System Usability Scale (SUS): SUS in the wild. Proc Hum Factors Ergon Soc Ann Meet 2013;57(1):192-196. [doi: 10.1177/1541931213571043]
55. Huang Y, Hu Y, Chan U, Lai P, Sun Y, Dai J, et al. Student perceptions toward virtual reality training in dental implant education. PeerJ 2023;11:e14857 [FREE Full text] [doi: 10.7717/peerj.14857] [Medline: 37168535]
56. Jelonek M, Fiala E, Herrmann T, Teizer J, Embers S, König M, et al. Evaluating virtual reality simulations for construction safety training: a user study exploring learning effects, usability and user experience. i-com 2022;21(2):269-281. [doi: 10.1515/icom-2022-0006]
57. Reichheld FF. The one number you need to grow. 2003. URL: https://hbr.org/2003/12/the-one-number-you-need-to-grow [accessed 2025-01-21]
58. Fisher N, Kordupleski R. Good and bad market research: a critical review of net promoter score. Appl Stoch Models Bus Ind 2018;35(1):138-151. [doi: 10.1002/asmb.2417]
59. Aitken RC. Measurement of feelings using visual analogue scales. Proc R Soc Med 1969 Oct;62(10):989-993 [FREE Full text] [doi: 10.1177/003591576906201005] [Medline: 4899510]
60. Flint A, Raben A, Blundell JE, Astrup A. Reproducibility, power and validity of visual analogue scales in assessment of appetite sensations in single test meal studies. Int J Obes Relat Metab Disord 2000;24(1):38-48. [doi: 10.1038/sj.ijo.0801083] [Medline: 10702749]

XSL•FO
RenderX

61. Riva G, Vigna C, Grassi A, Raspelli S, Cipresso P, Pallavicini F, et al. Learning island: the development of a virtual reality system for the experiential training of stress management. Stud Health Technol Inform 2012;173:369-371. [Medline: 22357020]
62. Rizzo AA, John B, Williams J, Newman B, Koenig S, Lange B, et al. Stress resilience in virtual environments: training combat relevant emotional coping skills using virtual reality. 2012 Presented at: 9th Conference on Disability, Virtual Reality & Associated Technologies; 2012 Sept 10-12; Laval, France p. 229-237.
63. Ulberg R, Amlo S, Dahl HJ, Høglend P. Does insight mediate treatment and enhance outcome? Psychoanal Inq 2017;37(3):140-152. [doi: 10.1080/07351690.2017.1285184]
64. Langley EL, Wootton BM, Grieve R. The utility of the health belief model variables in predicting help-seeking intention for anxiety disorders. Aust Psychol 2020;53(4):291-301. [doi: 10.1111/ap.12334]
65. Fredrickson BL. The role of positive emotions in positive psychology. The broaden-and-build theory of positive emotions. Am Psychol 2001;56(3):218-226 [FREE Full text] [doi: 10.1037//0003-066x.56.3.218] [Medline: 11315248]
66. Tyng CM, Amin HU, Saad MNM, Malik AS. The influences of emotion on learning and memory. Front Psychol 2017;8:1454 [FREE Full text] [doi: 10.3389/fpsyg.2017.01454] [Medline: 28883804]
67. Szekely R, Mason O, Frohlich D, Barley E. The use of virtual reality to reduce mental health stigma among healthcare and non-healthcare students: a systematic review. Behaviour & Information Technology 2023 Jul 04:1-18. [doi: 10.1080/0144929X.2023.2232049]
68. Baghaei N, Chitale V, Hlasnik A, Stemmet L, Liang HN, Porter R. Virtual Reality for Supporting the Treatment of Depression and Anxiety: Scoping Review. JMIR Ment Health 2021 Sep 23;8(9):e29681 [FREE Full text] [doi: 10.2196/29681] [Medline: 34554097]
69. Au J, Gibson BC, Bunarjo K, Buschkuehl M, Jaeggi SM. Quantifying the difference between active and passive control groups in cognitive interventions using two meta-analytical approaches. J Cogn Enhanc 2020;4(2):192-210 [FREE Full text] [doi: 10.1007/s41465-020-00164-6] [Medline: 34337311]

Abbreviations

DASS-21: Depression, Anxiety, and Stress Scale-21 items
HMD: head-mounted display
NPS: Net Promoter Score
PSS-10: Perceived Stress Scale
SDM: Subjective Difficulty Measure
STAI-Y1: State-Trait Anxiety Inventory Form Y-1
STAI-Y2: State-Trait Anxiety Inventory Form Y-2
SUS: System Usability Score
VAS-A: visual analog scale for anxiety
VAS-E: visual analog scale for emotions
VR: virtual reality

Edited by A Coristine; submitted 27.06.23; peer-reviewed by P Lindner, X Liu; comments to author 24.10.23; revised version received 01.03.24; accepted 20.11.24; published 06.03.25.

Please cite as:
Pallavicini F, Orena E, Arnoldi L, Achille F, Stefanini S, Cassa M, Pepe A, Veronese G, Bernardelli L, Sforza F, Fascendini S, Defanti CA, Gemma M, Clerici M, Riva G, Mantovani F
Effects and Acceptability of a 1-Week Home-Based Virtual Reality Training for Supporting the Management of Stress and Anxiety: Randomized Pilot Trial
JMIR Serious Games 2025;13:e50326
URL: https://games.jmir.org/2025/1/e50326
doi:10.2196/50326
PMID:40053782

©Federica Pallavicini, Eleonora Orena, Lisa Arnoldi, Federica Achille, Stefano Stefanini, Maddalena Cassa, Alessandro Pepe, Guido Veronese, Luca Bernardelli, Francesca Sforza, Sara Fascendini, Carlo Alberto Defanti, Marco Gemma, Massimo Clerici, Giuseppe Riva, Fabrizia Mantovani. Originally published in JMIR Serious Games (https://games.jmir.org), 06.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium,

XSL•FO
RenderX

provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Domain-Specific Evaluation of Exergame Metrics Among Older Adults With Mild Neurocognitive Disorder: Secondary Analysis of 2 Randomized Controlled Trials

Wanda Kaiser[1], MSc; Eling D de Bruin[1,2,3], PhD; Patrick Manser[1,3], Dr Sc

[1]Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Huddinge (Stockholm), Sweden

[2]Department of Health, OST - Eastern Swiss University of Applied Sciences, St.Gallen, Switzerland

[3]Department of Health Sciences and Technology, ETH Zurich, Zurich, Switzerland

Corresponding Author:
Patrick Manser, Dr Sc
Department of Neurobiology, Care Sciences and Society
Karolinska Institutet
Alfred Nobels Allé 23
Huddinge (Stockholm), 14183
Sweden
Phone: 46 70 923 74 45
Email: patrick.manser@ki.se

Abstract

Background: Exergame-based training enhances physical and cognitive performance in older adults, including those with mild neurocognitive disorder (mNCD). In-game metrics generated from user interactions with exergames enable individualized adjustments. However, there is a need to systematically investigate how well such game metrics capture true cognitive and motor-cognitive performance to provide a more robust basis for personalized training.

Objective: The primary objective was to identify valid game metrics as indicators for in-game domain-specific cognitive performance during exergaming in individuals with mNCD. We also aimed to explore game metric performance changes over time during exergame-based training.

Methods: Data were analyzed from individuals with mNCD who completed a 12-week home-based, exergame-based intervention following the Brain-IT training concept. A cross-sectional analysis was conducted by correlating game metrics with standardized neurocognitive reference assessments. To confirm the alternative hypothesis, we predetermined the following criteria: (1) statistically significant correlation ($P \leq .05$; uncorrected; 1-sided) with (2) a correlation coefficient (Pearson r or Spearman ρ) of ≥0.4. Visual and curve-fitting longitudinal analyses were conducted to explore game performance changes over time.

Results: Data were available from 31 participants (mean age 76.4, SD 7.5 y; n=9, 29% female). In total, 33% (6/18) of the game metrics were identified as valid indicators for in-game cognitive performance during exergaming. In the neurocognitive domain of learning and memory, these metrics included the mean reaction time (ρ=–0.747), the number of collected items (ρ=0.691), and the precision score (r=–0.607) for the game Shopping Tour (P<.001 in all cases), as well as the point rate (P=.008; r=0.471) for the game Simon. In addition, point rate was a valid indicator for executive function (P=.006; r=0.455) and visuospatial skills (P=.02; r=0.474) for the games Targets and Gears, respectively. The exploratory longitudinal analysis revealed high interindividual variability, with a general trend of the expected typical curvilinear curves of rapid initial improvements followed by a plateau in performance.

Conclusions: This study demonstrated that metrics reflecting the precision of responses generally performed better than metrics reflecting the speed of responses. These observations highlight the importance of selecting valid game metrics for implementation in exergame designs. Further research is needed to explore the potential of game metrics and identify factors contributing to individual variability in in-game performance and performance progression, as well as identifying and adopting strategies that facilitate individual learning success and thus promote effectiveness in improving health outcomes.

(JMIR Serious Games 2025;13:e65878) doi:10.2196/65878

XSL•FO
RenderX

KEYWORDS

exergaming; cognition; neurocognitive disorder; game metrics; older adults; performance progression

Introduction

Background

There is growing interest in interventions that aim to mitigate the physical and cognitive decline associated with aging in older adults [1]. Applying gamified digital health technologies with real-time user behavior and gameplay interactions such as computerized cognitive training games or exergame technology is becoming increasingly popular to address the challenges of physical and cognitive decline in the aging population [2-7]. Within this field, considerable attention has been directed toward exergame-based physical and motor-cognitive training, which is considered more promising than conventional physical, cognitive, or motor-cognitive training to improve cognitive and physical performance in diverse user populations [8-10]. Exergames are defined as "technology-driven physical activities, such as video game play, that require participants to be physically active or exercise in order to play the game" [11]. Exergame-based training typically focuses on incorporating cognitive tasks into motor tasks than can be designed to specifically train a variety of physical (eg, cardiorespiratory, motor, balance, strength, or multicomponent exercises) and cognitive (eg, learning and memory, complex attention, executive function, and visuospatial skills) domains [5]. The positive impact of exergaming on physical and cognitive functions has been observed in healthy older adults [5,8,12-14], as well as in older adults with cognitive impairments such as mild neurocognitive disorder (mNCD) or more severe forms such as dementia [5,15-17].

mNCD describes a critical stage of cognitive decline that goes beyond normal aging without affecting the capacity for independence in everyday activities [18]. Early interventions may effectively influence the course of cognitive decline, making those with mNCD an optimal target population for secondary prevention [19-22]. Several pharmacological and nonpharmacological methods have been proposed to slow down the progression of symptoms accompanying mNCD and improve overall quality of life [20,23]. According to current research, the most effective type of exercise to achieve this goal is to implement physical training, ideally with integrated cognitive tasks (simultaneous motor-cognitive training) [21,24-30]. This type of training has been proposed to operate through various disease-modifying mechanisms, including multisystem effects that reduce mNCD-related neuropathological damage [31,32] and promote synaptogenesis, neurogenesis, and angiogenesis [33,34]. It also helps maintain or increase cognitive reserve [26,35,36], empowering individuals to retain independence in everyday functioning despite neurodegenerative changes [33,34].

Physical and motor-cognitive training may represent the predominant nonpharmacological intervention that effectively mitigates cognitive decline in mNCD [27,30,37,38] and is recommended for secondary prevention of mNCD by a collaborative international guideline [39] as well as a global consensus on optimal exercise recommendations for enhancing healthy longevity in older adults [22]. In this context, exergames provide a promising solution to deliver this type of training [30]. Recent systematic reviews support the use of exergame-based training for improving both physical and cognitive performance in individuals with mNCD. However, studies in this field exhibit high heterogeneity, which weakens the strength of the evidence [15,16,40-42]. Furthermore, recent research suggests that, while exergame-based training may indeed improve cognitive performance, its superiority over traditional training remains uncertain, and further research should investigate the conditions in which exergames are most effective [12].

Several exergame systems have been investigated for older adults and have been shown to effectively promote engagement and motivation [43-46], which largely explains the superior adherence to exergame- or technology-enhanced training compared to adherence to conventional approaches of physical and motor-cognitive activities, exercise, or training [46,47]. Furthermore, exergames offer high value due to the technical adaptability of various elements. However, while many exergames are technically well designed for attractiveness and entertainment, they often lack systematic user adaptability and specificity [48]. Therefore, future exergames should be better tailored to specific target groups as well as individual goals and abilities [5,8,48,49].

To enhance the user adaptability of exergames, it has been recommended to implement real-time adaptive exergame mechanisms that support the dual-flow concept proposed by Sinclair et al [50]. According to this concept, an optimal exergame experience requires a balance between the challenge of the game and the player's skills, as well as between the intensity of the game and the player's fitness. A common method for implementing adaptive exergame mechanics is to use the data that are directly collected during exergaming [49]. These data are commonly referred to as in-game metrics or performance measures and capture the real-time interaction between the user and the serious game. Game metrics are quantitative measures that can range from simple measures such as point scores or reaction time to more complex measures that combine various parameters [51]. Game metrics have the potential to improve personalization and individualized progression of training according to users' performance, provide (real-time) feedback, and monitor gameplay [51-53]. However, there are currently no standards or guidelines for what data should be collected and for what specific purpose [52,54].

By investigating game metrics and understanding how well they can capture true cognitive and motor-cognitive performance, we can provide a robust basis for in-game adjustments and monitoring of performance progression [54,55]. While many studies emphasize the importance of individually tailored training and commonly use in-game metrics for this purpose, only a limited number have analyzed the psychometric properties of in-game metrics. Previous studies that have investigated game metrics from exergames for older adults with and without physical or cognitive impairments have found

XSL•FO
RenderX

moderate to strong correlations between some game metrics and clinical assessments [54,56,57]. This suggests that certain game metrics may be useful for measuring physical and cognitive performance. Furthermore, one study demonstrated that game metrics can distinguish between healthy older adults and those with mNCD [56]. In addition, another study explored a difference in exergame performance progression between older adults with and without cognitive impairment [57]. However, further studies are needed with domain-specific evaluation, larger participant samples, and inclusion of participants with a clinical diagnosis of mNCD, as stated by Guimarães et al [57].

Objectives

The primary objective was to identify valid game metrics as indicators for in-game domain-specific cognitive performance during exergaming in older adults with mNCD. We hypothesized statistically significant correlations between specific game metrics and reference assessments within their corresponding neurocognitive subdomains. As a secondary objective, we aimed to explore game metric performance change over time during a 12-week exergame-based training intervention for older adults with mNCD.

Methods

Overview

This study is a secondary analysis of data from 2 randomized controlled trials (RCTs) that aimed to evaluate the feasibility (feasibility study [FS]; NCT04996654) [58] and effectiveness (effectiveness study [ES]; NCT0538707) [59] of an individually tailored home-based exergame-based training concept specifically for the secondary prevention of mNCD developed within the Brain-IT project [60]. The training concept is rooted in years of iterative and user-centered co-design, development, testing, and evaluation with continuous patient and public involvement [61]. It serves as a guideline for the implementation of the training by providing algorithmic decision trees for the structure, content, and individualized tailoring of the training and can be implemented using different commercially available hardware [59].

The project was structured in 3 phases. In phase 1, we combined a comprehensive literature synthesis [61] with qualitative research, including primary end users (individuals with mNCD), secondary end users (multidisciplinary health and care professionals), other exergaming researchers, and experts from the exergaming industry [62], to elaborate a set of design requirements for the Brain-IT training concept. In phase 2, possible concepts were co-designed and developed based on the requirements defined in phase 1. The first prototype of the Brain-IT training concept [59,61] then entered the iterative cycle of feasibility, usability, safety, and acceptance testing and integrating of findings for further development until an "acceptable" solution was achieved (refer to the FS [58], the data from which are analyzed in this study). This process resulted in a novel intervention type specifically targeting various relevant mechanisms of action to alleviate the pathological state of mNCD by combining, for the first time, exergame-based motor-cognitive training with biofeedback-guided resonance breathing as an adjunct neuromodulatory intervention [59]. In phase 3, we confirmed the effectiveness (refer to the ES [59], the data from which are analyzed in this study) of the Brain-IT training in improving global cognitive performance as well as immediate and delayed verbal recall. We observed that the training not only could effectively slow down cognitive decline in comparison to usual care but also resulted in 55% of participants showing clinically relevant improvements in cognitive performance [59].

We compiled this manuscript according to the latest version of the CONSORT (Consolidated Standards of Reporting Trials) Statement for Randomized Trials of Nonpharmacologic Treatments (Tables S1 in Multimedia Appendix 1). As this is a secondary analysis of 2 RCTs, not all original trial design elements may be fully applicable or reported in this paper as they were not relevant for the specific analysis conducted (for full reproducibility, please also refer to the published FS [58], the study protocol for the ES [63], the ES itself [59], and the training concept—Supplementary File 2 of the ES [59]).

Ethical Considerations

All original study procedures were carried out in accordance with the Declaration of Helsinki. The original study protocols were approved by the ETH Zurich Ethics Commission (EK 2021-N-79) for the FS and the Ethics Committees of Zurich and Eastern Switzerland (EK-2022-00386) for the ES.

All study participants were fully informed about the study procedures in person (at the interested person's home or at one of the study centers depending on their preference) through verbal explanations and an information sheet. After sufficient time for consideration (ie, at least 24 hours after handing out the study information sheet [which was approved by the corresponding ethics committee] but, on average, approximately 1 week), suitable patients willing to participate in the study provided written informed consent in a second in-person meeting with one of the trained investigators of the study team at the home of the interested person or at one of the study centers. The original informed consent forms allowed for the secondary analyses conducted in this study as all study participants provided consent for (1) original datasets being made available in a publicly accessible repository in deidentified form and (2) the transfer of encrypted data for research purposes. Only deidentified data were analyzed for this study (for more detail, see the Data Availability section).

No compensation was granted to participants, but detailed feedback on individual performance as well as the study outcomes in general was provided at the end of the studies.

Study Design and Participants

Recruitment for the 2 parallel-group and single-blinded (the outcome evaluator of pre- and postintervention measurements was blinded to group allocation) RCTs took place between July 2021 and October 2023 in collaboration with health care facilities in the larger area of Zurich, Switzerland. The eligibility criteria for study participants have been published previously [58,59,63] and are listed in Textbox 1 [64-68].

XSL•FO
RenderX

Textbox 1. Eligibility criteria for participation in the 2 randomized controlled trials, the data from which were analyzed in this study. The eligibility criteria were checked by a trained investigator of the study team. Information about the clinical diagnosis and comorbidities was obtained from the collaborating recruitment partners.

Inclusion criteria

- Clinical diagnosis of mild neurocognitive disorder according to the *International Classification of Diseases, 11th Revision* [64], or the *Diagnostic and Statistical Manual of Mental Disorders, 5th Edition* [65], OR patients screened for mild cognitive impairment (MCI) according to the following criteria: (1) informant (ie, health care professional)-based suspicion of MCI confirmed by (2) an objective screening of MCI based on the validated German version [66] of the Quick Mild Cognitive Impairment screen [67], with a recommended cutoff score for cognitive impairment (or dementia) of <63/100 [68] while not falling below the cutoff score for dementia (ie, <45/100) [68]
- German speaking
- Able to stand for at least 10 minutes without assistance

Exclusion criteria

- Mobility impairments (ie, gait and balance) that prevented experiment participation
- Presence of additional, clinically relevant (ie, acute or symptomatic or both) neurological disorders (ie, epilepsy, stroke, multiple sclerosis, Parkinson disease, brain tumors, or traumatic disorders of the nervous system)
- Presence of any other unstable or uncontrolled diseases (eg, uncontrolled high blood pressure and progressing or terminal cancer)

After successful recruitment, participants underwent a premeasurement (see the Reference Assessments subsection in the Outcomes section) at the study center located at ETH Hönggerberg and were randomly assigned in a 1:2 (FS) or 1:1 (ES) ratio to either the control or intervention group. This secondary analysis investigated only the data from participants allocated to the intervention groups who engaged with the exergame-based training.

There were no important changes to the trial design and study setting after commencement that were relevant to the analyses presented in this paper. More details can be found in the published FS [58] and ES [59].

Training Intervention

Following the premeasurement and group allocation, a study investigator installed the exergame system in the participants' homes, where the training sessions took place. A standardized familiarization session was conducted to introduce the exergame system (ie, 4 games from the first scheduled training session [at level 1] were introduced and played for 2 minutes each). The exergame system Senso Flex (Dividat AG; hardware: prototype version 2; software: version 22.4.0-360-gf9df00d5b) was used [69]. The participants were instructed to follow the Brain-IT training concept [59] for the 12-week intervention, which consisted of completing an instructed minimum of 5 training sessions per week. Each training session had a duration of 24 minutes and included 6 motor-cognitive games and 3 resonance breathing exercises. Approximately one-third of the sessions were supervised by a study investigator, with more supervision provided at the beginning of the intervention and less toward the end [59]. The intervention was personalized and individually adapted based on several predefined parameters and included games for the neurocognitive domains of learning and memory, executive function, complex attention, and visuospatial skills [59,70]. The relevant parameters for this study were (1) the focus on participants' main cognitive deficits through an individual constellation of games and (2) the individual adaptation of the standardized difficulty levels (levels 1-10), with each participant starting at level 1 [59].

The parameters for determining the 10 game levels were defined through agreement between an experienced neuropsychologist and the research team and were extensively tested [71]. According to the Brain-IT training concept, level 1 should refer to an introductory level (ie, even most impaired patients with mNCD should be able to play the game at the first trial), and in level 10 (ie, "healthy functioning" level), the game demands are expected to be challenging but doable for an average healthy older adult. The remaining levels are defined to increase neurocognitive demands consecutively and regularly from level to level. In the definition of these 10 game levels, we integrated the methodology by Huber et al [72], who established an extended taxonomy supporting motor-cognitive learning. This methodology provides a framework for structuring training to specifically integrate models of skill acquisition [73,74] that has also been described to apply to cognitive skill learning and relearning [75].

For more details on the specific exergames, as well as all algorithmic decisions for the personalized design and individualized progression of the training, including a description of all games and game level settings, we kindly refer interested readers to the published Brain-IT training concept [59]. To ensure replicability, the Brain-IT training concept was planned and reported according to the Consensus on Exercise Reporting Template [76]. The training concept provides specific instructions on how to adapt the training when implemented with alternative hardware and software solutions (Supplementary File 2 of the ES [59]).

Outcomes

Demographics

During the premeasurement, participants provided self-reported information on sociodemographic variables, including age, sex, BMI, and years of education. The characterization of the etiology of mNCD was derived from the diagnostic information provided by the collaborating recruitment partners. Global cognitive performance was assessed using the validated German

XSL•FO
RenderX

version [66] of the Quick Mild Cognitive Impairment screen [77,78].

Game Metrics

A total of 14 different exergames were implemented in the Brain-IT training concept. During the development phase of the training concept, neuropsychologists assigned each exergame to its primary targeted neurocognitive function to ensure its content validity [61]. However, it should be mentioned that exergames typically involve and train multiple neurocognitive functions. The exergames were assigned to the primary targeted neurocognitive domains of (1) learning and memory, (2) executive function, (3) complex attention, and (4) visuospatial skills. In addition, they were subcategorized into neurocognitive subdomains [61]. For this study, we analyzed 8 exergames, with 2 games representing each neurocognitive domain, as listed in Table 1. These games were selected based on data availability (ie, games that were frequently played during the intervention and were introduced to most, if not all, participants to ensure robust analyses). We collected all game metrics that were computed by the exergame system and, therefore, available for analyses (ie, no selection or exclusion of specific metrics). A detailed description of the game's content, parameters to adapt, and game metrics (referred to as "performance measures") can be found in the Brain-IT training concept [59].

XSL•FO
RenderX

Table 1. Overview of the analyzed games and their outcome variables categorized by neurocognitive domains and subdomains[a].

Neurocognitive domain and subdomain	Exergame		Neuropsychological assessment—outcome variable
	Name	Outcome variable	
Learning and memory			
Free recall	Shopping Tour	• Mean reaction time (ms) • Number of collected items • Number of mistakes • Precision score (%)	WMS-IV-LM 1[b]—free recall (total point score)
Serial recall	Simon (color and number)	• Mean reaction time (ms) • Point rate	PEBL[c] DSF[d] (total point score)
Executive function			
Working memory	Nomis (color and number)	• Mean reaction time (ms) • Point rate	PEBL DSB[e] (total point score)
Planning	Targets	• Hits (number per min) • Misses (number per min) • Point rate	HOTAP (points × min^{-1})
Complex attention			
Selective attention	Habitats	• Mean reaction time (ms) • Point rate	TAP[f] Go-No go (median reaction time [ms])
Processing speed	Simple	• Mean reaction time (ms) • Point rate	PEBL TMT-A[g] (completion time [s])
Visuospatial skills			
Visual perception	Gears	• Mean reaction time (ms) • Point rate	PEBL MRT[h] (performance score)
Visuoconstructional reasoning	Tetris	• Game score (min–1)	PEBL MRT (performance score)

[a]The categorization of exergames into the neurocognitive domains and subdomains was derived from the published Brain-IT training concept. Accordingly, each game was categorized into the primary neurocognitive domain and subdomain being trained (and secondary subdomain in the case of a game that focuses on more than one neurocognitive domain or subdomain). The categorization was made through agreement between an experienced neuropsychologist and the research team to ensure the content validity of the exergames used to train each neurocognitive domain or subdomain. The variables listed in the Outcome variables column under Exergame describe all game metrics that were computed by the exergame system and, therefore, available for analyses (ie, no selection or exclusion of specific metrics). For each of these games, the most appropriate reference assessment was selected from the available published datasets. Specifically, we used the neurocognitive domains or subdomains that the specific games primarily trained (as defined in Table 1 of the published Brain-IT training concept [59]) as references and selected the reference assessment that measured the same primary neurocognitive domain or subdomain from the available published datasets.

[b]WMS-IV-LM 1: Wechsler Memory Scale–Fourth Edition–Logical Memory subtest part 1.

[c]PEBL: Psychology Experiment Building Language.

[d]DSF: digit span forward.

[e]DSB: digit span backward.

[f]TAP: Test of Attentional Performance.

[g]TMT-A: Trail Making Test part A.

[h]MRT: Mental Rotation Task.

Reference Assessments

Data from several domain-specific neuropsychological assessments collected during the premeasurement were used as reference. For each assessment, the participant received standardized instructions from a study investigator, who also ensured that the participant understood the task and the procedure. The assessments and the outcome variables used for this analysis are listed in Table 1. The study protocol of the ES [63] provides detailed information about the assessments. For each of the games, the most appropriate reference assessment was selected from the available published datasets. Specifically, we used the neurocognitive domains or subdomains that the specific games primarily trained (as defined in Table 1 of the published Brain-IT training concept [59]) as references and selected the reference assessment that measured the same primary neurocognitive domain or subdomain from the available published datasets, as listed in Table 1 of the published Brain-IT

XSL•FO
RenderX

training concept [59]. A brief description of the assessments and their function is provided in this section.

For the neurocognitive domain of learning and memory, the German version of the Logical Memory subtest part 1 from the Wechsler Memory Scale–Fourth Edition [79,80] was used to measure the neurocognitive subdomain of free recall. Data for the subdomain of serial recall were obtained from the digit span forward test of the Psychology Experiment Building Language (PEBL) [81], a computer-based software platform for psychological experiments. Executive function was assessed using the HOTAP picture-sorting test part A [82] to evaluate planning skills and the PEBL digit span backward test [81] to evaluate working memory. For the neurocognitive domain of complex attention, data from the PEBL Trail Making Test part A [81] for processing speed and data from the Go-No go task 1 of 2 from the Test of Attentional Performance [83,84] for selective attention were used. Finally, we considered data from the computerized PEBL Mental Rotation Task [81] that is based on the classic mental rotation task to assess visuospatial skills.

Data Analysis and Statistical Methods

Overview

Statistical analysis was conducted using R (version 2023.12.0+369; R Foundation for Statistical Computing). Data are reported as mean and SD for interval scales or continuous data and as counts and percentages for categorical variables. Descriptive statistics were calculated for all outcome variables. The assumption of normality distribution was checked through the Shapiro-Wilk test alongside a visual examination of the data [85].

Primary Objective: Cross-Sectional Analysis

For the primary objective, we used a cross-sectional analysis. Only the game metrics from the first time that each participant played the first level of difficulty were analyzed. There were 2 main reasons for this. First, not all participants played the same number of repetitions of a game level due to individual adaptations, and second, we aimed to minimize the influence of learning effects in this cross-sectional analysis. This analysis included all participants who played the corresponding game at least once. The training concept underwent minor adjustments between the FS and the ES. One such adjustment involved extending the duration of certain games from 2 to 3 minutes. Metrics that were affected by this change and are time dependent were documented as metrics over time (min^{-1}). The remaining minor adjustments to the training concept did not affect the analyses or the interpretation of our results and included factors such as the introduction of newly available games, which were not analyzed in this study.

One-way bivariate correlation analyses were conducted to investigate the relationship between game metrics and performance in clinical assessment within the corresponding neurocognitive subdomains. Depending on the distribution of the data, either Pearson (parametric statistical analyses) or Spearman (nonparametric statistical analyses) correlation analyses were conducted. *P* values and correlation coefficients (Pearson *r* or Spearman ρ), including 95% CIs, were calculated (using bootstrap for nonparametric analyses) [85]. To confirm the alternative hypothesis, we predetermined the following criteria: (1) statistically significant correlation ($P \leq .05$; uncorrected; 1-sided) with (2) a correlation coefficient of $\rho \geq 0.4$, as proposed by Streiner et al [86] for studies in which perfect correlations are rare and many influencing factors are outside of the researchers' control, which is the case in this analysis due to the complex nature of exergames and various influences such as task design, external factors, and personal characteristics. The correlation coefficients were interpreted as weak ($\rho < 0.3$), moderate ($\rho = 0.3$-0.5), or strong ($\rho > 0.5$) [85,87].

We decided against applying a Bonferroni correction routinely for multiple comparisons considering our hypothesis and methodology. This is recommended because our study did not imperatively require the avoidance of type-I errors and we had preplanned hypotheses [88]. In addition, correlations were assessed only within their predefined neurocognitive subdomains, making our data dependent. Sullivan and Feinn [89] noted that applying the Bonferroni correction to dependent data may hinder the detection of relevant associations and increase the risk of a type-II error.

In this study, we did not conduct an a priori sample size calculation as we analyzed existing datasets from studies conducted as part of the Brain-IT project. Therefore, the sample size for these analyses was determined based on the number of datasets available from the 2 RCTs and was fixed and could not be altered. To ensure appropriate interpretation of our results, we conducted a post hoc power analysis using G*Power (version 3.1) [90,91]. As the Spearman rank correlation coefficient is computationally identical to the Pearson product-moment coefficient, power analyses were conducted using the same methodology as for estimating the power of a Pearson correlation for nonparametric analyses.

Secondary Objective: Longitudinal Analysis

For the secondary analysis, we only used game metrics that met the predefined criteria of the primary objective to ensure the validity of this analysis. For this analysis, data were collected longitudinally over the 12-week intervention period, excluding participants who dropped out during the intervention. The definition of each of our 10 game levels followed the principles described in the extended taxonomy supporting motor-cognitive learning by Huber et al [72]. As this taxonomy specifically integrates the model of skill acquisition [73] that has also been described to apply to cognitive skill learning and relearning [75], we based our longitudinal analysis on the model of skill acquisition [73], which integrates the 3-stage model of human skill acquisition by Fitts [74].

Accordingly, the performance data (game metrics) were presented as a function of time. As participants did not play each game an equal number of times, we expressed the variable "time" as a percentage of game completion time (ie, the first time that a participant played the game was 1%, and the last time they played the game was 100%). Full completion (100%) represented either the introduction of a new, slightly more challenging game or the end of the intervention. In addition, a second graph was generated displaying the performance changes within specific game levels. This was done to provide further insights into the data and is particularly relevant for accuracy

XSL•FO
RenderX

metrics (eg, precision score), for which we expected different characteristics of the performance curves. More specifically, we generally expected either the typical sigmoid curve or curvilinear curves of rapid initial improvements followed by a plateau in performance. However, for accuracy metrics, we expected that this curve would only be visible within a specific game level, with a slight drop in accuracy each time the game level was increased.

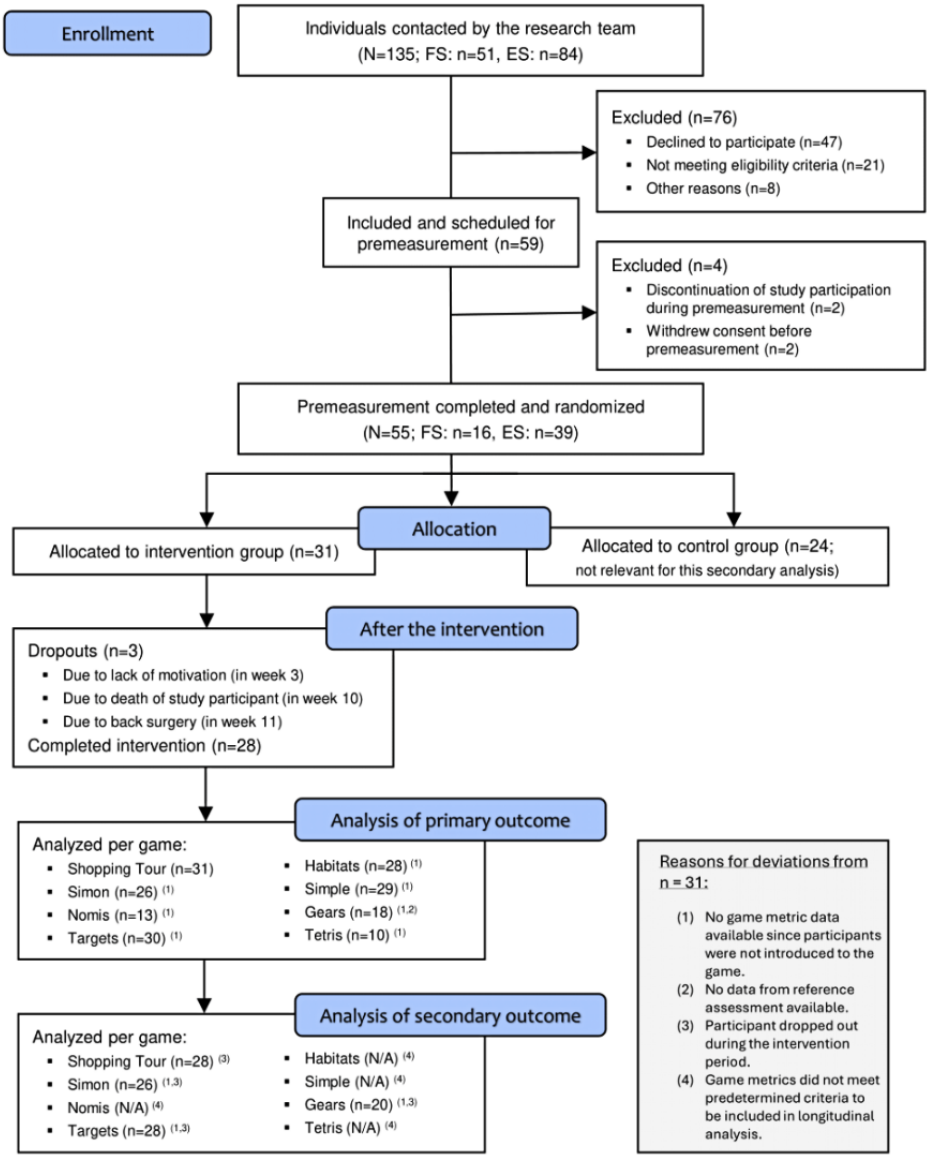
Performance change curves were visually represented for each individual participant along with an overall curve that reflected the average performance across all participants [92]. However, we refrained from making any a priori assumptions on the type of relationships (eg, linear vs curvilinear relationship) as, on the one hand, our algorithmic decisions on individualized progression of exergame demands aimed to "break through" the expected (repeated) plateau in performance, which would be expected to create repeated curvilinear curves with a slight drop in performance after each plateau once the game difficulty was increased; however, on the other hand, there are no previous analyses available for the exergame systems and metrics used to base this theory-derived assumption on, whereas similar previous research has observed a more linear progression change over time in patients with mNCD in contrast to the typical curvilinear curves observed in healthy older adults [57]. Therefore, we chose to use the locally estimated scatterplot smoothing regression model to fit smooth curves, which provides a flexible approach to visualize data. This model is recommended for exploratory analysis because there is no need for a priori specification of relationships [93]. Finally, we measured and reported the average interquartile ranges number of times that participants played the game.

Results

Overview

A summary of the participant flow is provided in Figure 1. Data were available for a total of 31 participants who were allocated to the intervention group in the 2 RCTs analyzed. The time between the initial premeasurement with the group allocation and the start of the intervention ranged from 1 to 2 weeks.

Figure 1. Summary of the participant flow throughout the study. FS: feasibility study; ES: effectiveness study; N/A: not applicable.



XSL•FO
RenderX

Demographic Data

The demographic and clinical characteristics of all participants are listed in Table 2.

Table 2. Demographic and clinical characteristics of the participants in the intervention groups (N=31)[a].

Participant characteristic	Values
Age (y), mean (SD)	76.4 (7.5)
Sex (female), n (%)	9 (29)
Education (y), mean (SD)	14.9 (4.0)
BMI (kg/m^2), mean (SD)	24 (2.4)
Global cognitive performance (QMCI[b] total score), mean (SD)	57.4 (14.0)
Etiology of mNCD[c], n (%)	
mNCD due to Alzheimer disease	19 (61)
Mild frontotemporal NCD[d]	4 (13)
Mild vascular NCD	5 (16)
Unclear or not yet determined	3 (10)

[a]Demographic information was assessed based on self-report of the study participants. Global cognitive performance was assessed using the validated German version [66] of the Quick Mild Cognitive Impairment screen total score [77,78]. The characterization of the etiology of mild neurocognitive disorder was derived from the diagnostic information provided by the collaborating recruitment partners.

[b]QMCI: Quick Mild Cognitive Impairment screen.

[c]mNCD: mild neurocognitive disorder.

[d]NCD: neurocognitive disorder.

Primary Objective: Cross-Sectional Analysis

Table 3 presents the results of the correlation analyses between game metrics and neurocognitive reference assessments. In the neurocognitive domain of learning and memory, 3 metrics of the game Shopping Tour (ie, mean reaction time, precision score, and collected items) and 1 metric of the game Simon (ie, point rate) met the alternative hypothesis criteria by exhibiting statistically significant moderate to strong correlations with the corresponding reference assessments (see Table 3 for all *P* values). In the neurocognitive domains of executive function and visuospatial skills, the metric point rate of the games Targets and Gears met the alternative hypothesis criteria by exhibiting statistically significant moderate correlations with the corresponding reference assessments. None of the metrics of the games targeting the neurocognitive domain of complex attention or of the remaining 2 games, Nomis and Tetris, met the alternative hypothesis criteria. Descriptive statistics for all outcome variables are provided in Tables S2 and S3 in Multimedia Appendix 1.

XSL•FO
RenderX

Table 3. Correlation between game metrics and reference assessments within their neurocognitive subdomains (N=31)[a].

Neurocognitive domain, game and reference assessment, and outcome	Sample size, n (%)	Statistics					
		Type of analysis	*P* value	Spearman ρ (95% CI)	Pearson *r* (95% CI)	H_A[b] met?	Statistical power (post hoc)
Learning and memory							
Shopping Tour (WMS-IV-LM 1[c] score)							
Mean reaction time (ms)	31 (100)	Nonparametric	<.001[d]	*–0.747*[e] (–0.876 to 0.548)	—[f]	Yes	0.979
Number of mistakes	31 (100)	Parametric	.15	—	–0.264 (–0.565 to 0.100)	No	0.669
Number of collected items	31 (100)	Nonparametric	<.001[d]	*0.691* (0.478 to 0.830)	—	Yes	0.923
Precision score (%)	31 (100)	Parametric	<.001[d]	—	*0.607* (0.374 to 1)	Yes	0.739
Simon (PEBL[g] DSF[h]score)							
Mean reaction time (ms)	26 (84)	Nonparametric	.04[d]	–0.344 (–0.680 to 0.145)	—	No	0.516
Point rate	26 (84)	Parametric	.008[d]	—	*0.471* (0.166 to 1)	Yes	0.528
Executive function							
Nomis (PEBL DSB[i] score)							
Mean reaction time (ms)	13 (42)	Nonparametric	.40	–0.079 (–0.575 to 0.572)	—	No	0.505
Point rate	13 (42)	Nonparametric	.60	–0.081 (–0.590 to 0.579)	—	No	0.705
Targets (HOTAP combined score)							
Number of hits	30 (97)	Nonparametric	.03[d]	0.353 (–0.062 to 0.626)	—	No	0.514
Number of misses	30 (97)	Nonparametric	.08	–0.264 (–0.597 to 0.098)	—	No	0.512
Point rate	30 (97)	Nonparametric	.006[d]	*0.455* (0.144 to 0.715)	—	Yes	0.523
Complex attention							
Habitats (TAP[j] Go-No go reaction time)							
Mean reaction time (ms)	28 (90)	Parametric	.25	—	–0.131 (–1 to 0.195)	No	0.506
Point rate	28 (90)	Nonparametric	.36	0.072 (–0.371 to 0.448)	—	No	0.503
Simple (PEBL TMT-A[k] completion time)							
Mean reaction time (ms)	29 (94)	Nonparametric	.58	0.039 (–0.361 to 0.474)	—	No	0.719
Point rate	29 (94)	Nonparametric	.92	–0.264 (–0.541 to 0.079)	—	No	0.997
Visuospatial skills							
Gears (PEBL MRT[l] performance score)							
Mean reaction time (ms)	18 (58)	Nonparametric	.63	0.085 (–0.402 to 0.663)	—	No	0.752

XSL•FO
RenderX

Neurocognitive domain, game and reference assessment, and outcome	Sample size, n (%)	Statistics					
		Type of analysis	*P* value	Spearman ρ (95% CI)	Pearson *r* (95% CI)	H_A[b] met?	Statistical power (post hoc)
Point rate	18 (58)	Nonparametric	.02[d]	*0.474* (0.018 to 0.787)	—	Yes	0.520
Tetris (PEBL MRT performance score)							
Point score	10 (32)	Nonparametric	.06	–0.043 (–0.783 to 0.089)	—	No	0.069

[a]The categorization of exergames into the neurocognitive domains and subdomains was derived from the published Brain-IT training concept, and for each game, we selected the most appropriate reference assessment from the available published datasets (Table 1). The variables listed in the first column describe all game metrics that were computed by the exergame system and, therefore, available for analyses (ie, no selection or exclusion of specific metrics).

[b]H_A: alternative hypothesis.

[c]WMS-IV-LM 1: Wechsler Memory Scale–Fourth Edition–Logical Memory subtest part 1.

[d]Statistically significant at $P<.05$.

[e]Italics indicate a correlation coefficient of >0.4.

[f]Not applicable.

[g]PEBL: Psychology Experiment Building Language.

[h]DSF: digit span forward.

[i]DSB: digit span backward.

[j]TAP: Test of Attentional Performance.

[k]TMT-A: Trail Making Test part A.

[l]MRT: Mental Rotation Task.

Secondary Objectives: Longitudinal Analysis

This analysis included the 6 game metrics that met the criteria for the primary objective. Table 4 presents the results for the number of times that the participants played each game. Figure 2 shows the performance progression curves over gameplay time in percentages. For all game metrics, there was an initial increase in average performance progression, or a decrease in the case of the metric "mean reaction time." This was followed by a plateau (see Simon, Targets, and the mean reaction time metric of Shopping Tour) or even a slight decline (see Gears and the precision score and collected items metrics of Shopping Tour) in performance. Interindividual differences in performance progression were observed, particularly for the metric "point rate." While some individuals exhibited a flatter curve, others had a more pronounced initial increase or a sigmoidal curve pattern. Figure 3 shows the performance progression within difficulty levels. In addition to the findings shown in Figure 2, we observed that, in the game Shopping Tour, the metrics "precision score" and "collected items" increased within most levels, whereas the metric of mean reaction time initially declined and then leveled off. For the metric of point rate in the games Gears and Simon, we observed considerable drops in performance between certain levels of difficulty.

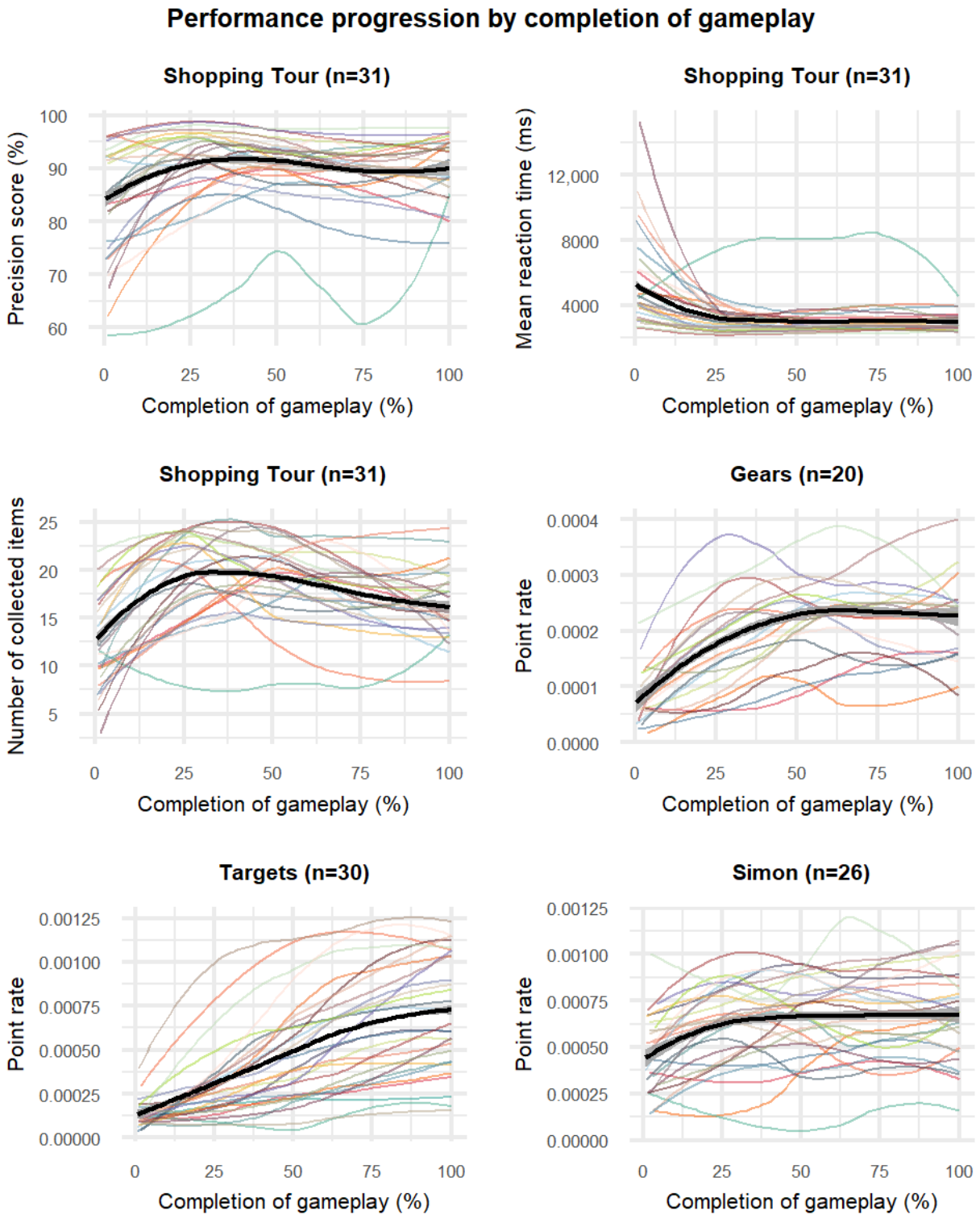
Table 4. Sessions played per participant that were analyzed for the secondary objective[a].

Game	Sessions played, median (IQR)
Shopping Tour	132 (99)
Simon	72 (23)
Targets	95 (44)
Gears	64 (25)

[a]The descriptive statistics presented refer to the median and the IQR of the number of sessions that the study participants (individuals with mild neurocognitive disorder) played in each of the exergames analyzed in this study over the 12-week intervention period.

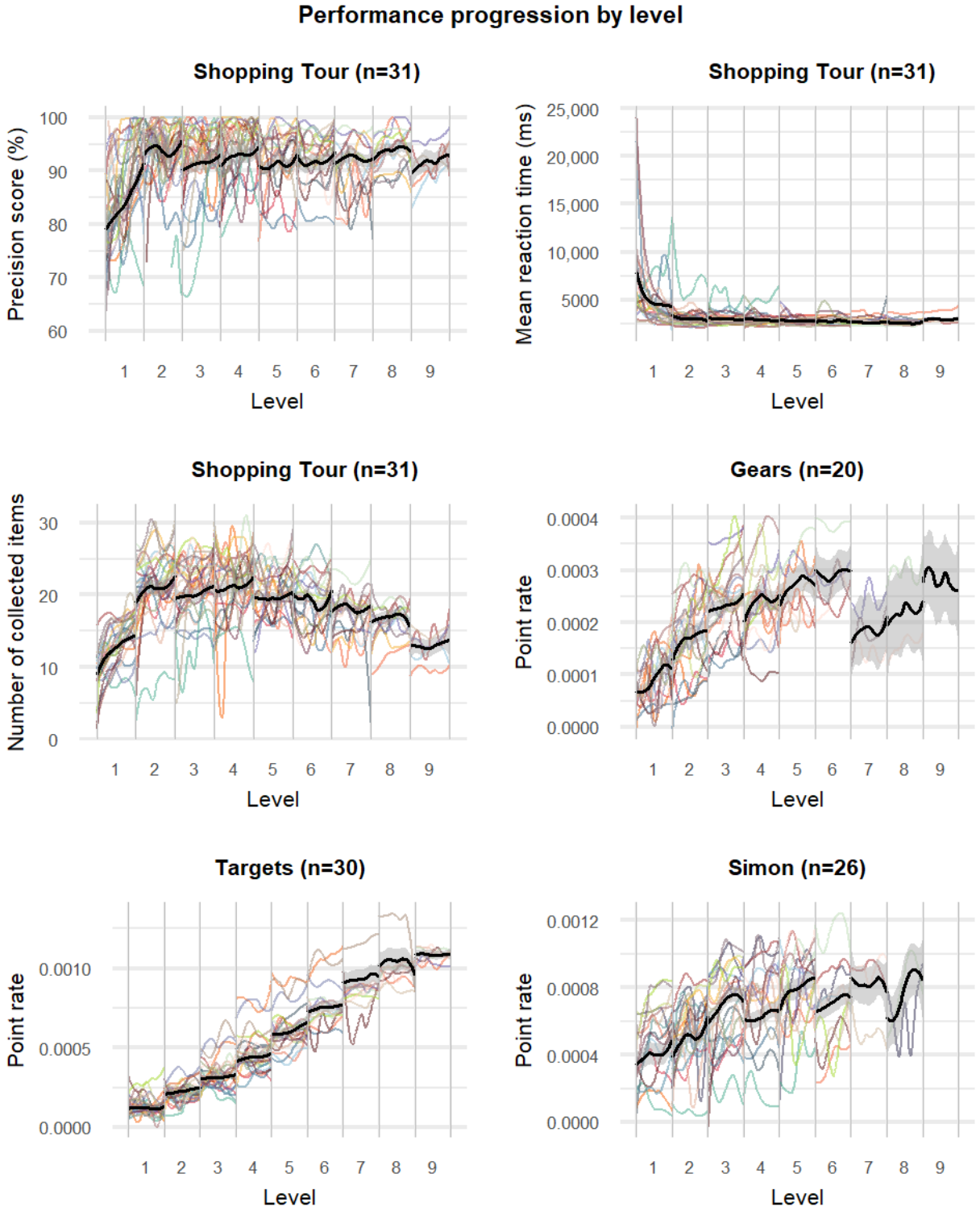
XSL•FO
RenderX

Figure 2. Performance progression over gameplay time.



XSL•FO
RenderX

Figure 3. Performance progression over difficulty levels.



Discussion

Principal Findings

Overview

This study aimed to identify valid game metrics as indicators for in-game domain-specific cognitive performance during exergaming in individuals with mNCD by analyzing the correlation between game metrics and neurocognitive assessments. Key findings were that 33% (6/18) of the game metrics within the neurocognitive domains of learning and memory, executive function, and visuospatial skills were identified as valid indicators for in-game domain-specific cognitive performance during exergaming. This study's secondary objective was to visually explore game metric performance changes over time and within difficulty levels. The results revealed high interindividual variability and overall

XSL•FO
RenderX

trends of the expected typical curvilinear curves of rapid initial improvements followed by a plateau in performance.

Primary Outcome: Cross-Sectional Analysis

In total, 4 games (Shopping Tour, Simon, Targets, and Gears) had at least one metric that met the predefined criteria and, therefore, was identified as a valid indicator for in-game domain-specific cognitive performance during exergaming. We observed that 83% (5/6) of the game metrics that met our alternative hypothesis reflected the precision of response (ie, point rate or precision score), whereas measures of the speed of response (ie, mean reaction time) met the criteria in only 17% (1/6) of cases even in the neurocognitive domain of "complex attention," which specifically trains processing speed. Furthermore, the FS of the Brain-IT training concept has already shown that reaction times had a high inter- and intraindividual variability, making it difficult to initiate training adaptations based on these game metrics [58]. These observations suggest that game metrics reflecting the precision of response may be a better indicator of cognitive performance during exergaming than metrics reflecting the speed of response. The game Shopping Tour exhibited strong correlations in 75% (3/4) of the metrics, whereas other games exhibited weak correlations in all collected metrics. This suggests that the results may depend on two main factors: (1) the validity of game metrics as an indicator of performance and (2) how well the game content aligns with its assigned neurocognitive domain or subdomain. The second factor, the content validity of the game, was confirmed by neuropsychologists during the design phase of the Brain-IT training concept [61]. However, it is important to note that exergames typically target different neurocognitive functions as well as motor skills. Although each game focuses on a specific neurocognitive domain or subdomain, other domains are often indirectly trained, which may confound the domain-specific analyses. Therefore, the observation that certain games displayed no metric that met the alternative hypothesis criteria is likely related to the validity of the game metrics as indicators of performance or to the content of the games not being specific enough to target the corresponding neurocognitive domains or subdomains.

While we have established that game metrics and game content may contribute to measuring domain-specific exergame performance, the complex nature of exergames indicates that additional factors more closely related to the training itself should also be considered. These factors are typically minimized in clinical assessments, yet they are a significant aspect of exergaming and may have influenced the outcomes of this study. An important consideration is the influence of the exergame design, including both hardware and software, on player performance. One such task design aspect is the scoring system. If certain tasks within the game are not appropriately scored relative to their difficulty, this could result in an inaccurate reflection of performance measures. Another consideration is the involvement of different difficulty levels. In our analysis, we correlated game metrics from only the first level of difficulty, which was considered an introductory level and may not have been challenging enough to show the full capabilities of the participants. In line with this consideration, Litz et al [94] conducted a similar study correlating exergame metrics with clinical assessments but across various difficulty levels and found some variable outcomes for the same game in different levels, with a tendency toward more consistent relationships between in-game metrics and standardized cognitive assessments at higher difficulty levels or when averaging scores over all game levels. Furthermore, exergames often include feedback mechanisms that reward correct decision-making (ie, precision of response) more than decision speed, which may explain the low correlation with reaction time metrics in our results. External factors such as physical environment or external support, as well as personal factors such as emotional status or self-efficacy [95], should also be considered as potential influencers of performance. In relation to this, Sajjadi et al [49] conducted a systematic review on adapting serious games to users with the aim of increasing engagement and effectiveness. This review proposed the use of game metrics for in-game adaptions but also highlighted the importance of considering additional factors such as physiological states and the personal traits of the user [49].

Our findings align with those of previous studies that have analyzed correlated metrics from exergames with clinical assessments of cognitive performance [54,56,57,94]. Specifically, Guimarães et al [57] correlated game metrics measured on the first gameplay against standard cognitive and mobility assessments, including the Montreal Cognitive Assessment (MoCA), in 13 older adults with mobility limitations. They observed that 2 of the 16 analyzed game metrics (from 4 different games) showed a statistically significant correlation with the MoCA total score, both related to scores on precision or accuracy of responses. They also correlated these game metrics with all the MoCA subdomain scores and found some (ie, 8 of the 112 analyses) statistically significant correlations with the visuospatial, executive functioning, and abstraction components of the MoCA. Similar results were observed for the physical performance outcomes, with several (17 of the 48 analyses) step characteristic outcomes during exergaming showing statistically significant correlations with physical test scores [57]. Petsani et al [54] computed Pearson correlations between the in-game metrics of 13 individuals with Parkinson disease with the Symbol Digit Modalities Test using a feature vector containing the mean value of all sessions for each in-game metric for each participant. They observed that 1 of the 11 analyzed game metrics (from 4 different games) showed a strong correlation with the Symbol Digit Modalities Test score [54]. Konstantinidis et al [56] computed Pearson correlations between an aggregated score derived from various in-game metrics of 113 participants (ie, 38 cognitively healthy, 64 with mild cognitive impairment, and 14 with mild dementia) with the Mini-Mental State Examination (MMSE), MoCA, and Trail Making Test parts A and B. They found consistent, statistically significant, moderate to strong correlations for all 4 analyses [56]. Finally, Litz et al [94] computed Spearman rank correlations between in-game metrics and 5 outcomes from 4 cognitive assessments in independently living older adults who were cognitively intact and multimorbid. They observed consistent moderate to strong correlations for most (88 out of 130) analyses, with similar observations for a range of physical assessments [94].

XSL•FO
RenderX

These observations indicate that only a selection of game metrics may be valid indicators for in-game cognitive performance. However, identifying specific game metrics that performed better across these different studies was challenging due to the high heterogeneity in study design, analyzed game metrics, and reference assessments used. Notably, while our study was based on a domain-specific approach, these previous studies correlated game metrics with a range of cognitive (and physical) assessments without initial categorization of game specificity. In addition, they frequently used broad assessments, such as the MoCA and MMSE, which are tests that cover several cognitive domains in the outcome score [96,97]. The correlation analyses between different game metrics and the MoCA or MMSE ranged from weak [57,94] to moderate [56,57] and strong [56,57], indicating diverse outcomes. When investigating specific neurocognitive domains, we also found diverse findings. In contrast to our findings, Litz et al [94] found moderate to strong correlations for the metrics scores and reaction time within the domain of complex attention. On the other hand, we found some moderate to strong correlations within the domain of learning and memory, whereas Guimarães et al [57] did not find such correlations in similar games.

Some studies have also used game metrics to predict cognitive or physical performance as measured using standardized clinical assessments. Petsani et al [54] used a decision tree classifier with the Gini index to predict, based on in-game metrics, whether participants belonged to groups with better or worse physical and cognitive states. This categorization was derived from clustering analysis based on 18 outcomes from a set of neuropsychological and physical assessments. They observed a high classification accuracy at 84.6% [54]. Konstantinidis et al [56] used a multilayer feedforward neural network to assess the predictive value of their aggregated game score and found an overall accuracy of 70.69% to distinguish between cognitively healthy, mild cognitive impairment, and mild dementia [56].

In summary, previous studies have shown mixed results but were able to identify certain game metrics that appear as valid indicators for physical, cognitive, or motor-cognitive performance in exergaming and may be reflective of standardized neuropsychological or physical assessments. However, while the use of such game metrics to individually tailor and adjust the games is widespread, these metrics have evidently rarely been scientifically investigated as we only identified a handful of similar previous studies [54,56,57,94]. Therefore, it appears important to carefully select and prevalidate game metrics in future studies, especially when they are used for purposes such as in-game adaptations or performance monitoring. Given the current advancements in computing fields such as artificial intelligence, it is likely that real-time personalized adaptation will play an even more important role in the future of exergames [98]. This development requires valid metrics as indicators of physical and cognitive in-game performance characteristics. The challenge for researchers and game developers is to identify or develop valid game metrics and integrate them strategically into exergame designs, preferably with a strong background in sports science or neuroscience to ensure construct validity, as previously proposed by our research group [58].

Secondary Outcome: Longitudinal Analysis

The results of the secondary objective indicated a common trend across all games, with the expected typical curvilinear curves of rapid initial improvements suggesting an early phase of skill acquisition followed by a plateau in performance. This finding is consistent with those of the existing literature [74,99,100] as well as with models of skill acquisition [73,74] that have also been described to apply to cognitive skill learning and relearning [75]. Specifically, the model of human performance by Fitts and Posner [74] identifies 3 stages: the cognitive phase, in which the user learns what to do and experiences a rapid performance gain; the associative phase, in which the user refines their actions and increases efficiency; and, finally, the autonomous phase, in which the skill becomes automatic and the user becomes proficient.

In our analysis, certain game metrics exhibited a more pronounced plateau (ie, Simon) or even a decline in performance (ie, Gears and the collected items metric of Shopping Tour), which would not be expected from the literature and suggests that players may not have experienced the associative phase. However, upon examining the data across specific difficulty levels, we observed that these plateaus did not necessarily indicate a halt in performance improvement but, rather, an increase in task difficulty that was not appropriately rewarded by the game metrics. In some games, the task difficulty significantly increased between certain levels (eg, Gears with more complex figures or Simon with longer sequences to remember; see the published training concept in Supplementary File 2 of the ES [59]). Between these levels of increasing task difficulty, a clear drop in performance was observed (eg, in Gears after levels 3 and 6 or in Simon after levels 1, 3, 5, and 7). This indicates that, for these games, the scoring system failed to account for varying levels of difficulty by awarding the same number of points for correctly completed tasks of varying difficulty. In contrast, the performance progression of Targets and the precision score metric of Shopping Tour demonstrated a consistent improvement with individual dips and leaps as is expected [101]. This indicates that the adaptation mechanisms were functioning well for these games and the game metrics can be used to monitor performance progression. Furthermore, for the game Shopping Tour, we observed inconsistent curves for different metrics. The mean reaction time decreased as the game level increased, reached an early plateau, and remained low. This was a somewhat unexpected finding as we expected reaction times to increase slightly and then show the typical reciprocal curve with repeated practice within the same game level, similar to what we expected from the precision metric. This observation suggests that average reaction speed may play a subordinate role in this game. In contrast, the precision score showed continuous improvement, with the expected temporary drops immediately after increasing the game levels (with the exception of the increase from level 1 to 2), which was similar to what was found in the study by Guimarães et al [57]. The exceptional observation that the precision score in the game Shopping Tour did not temporarily drop when the game level increased from level 1 to level 2 may be explained by (1) early

XSL•FO
RenderX

habituation and learning effects in the use of and interaction with the exergame device and the specific game and (2) the fact that the increment between these levels may have been too small to induce a temporary drop in performance.

Overall, these observations may indicate that measures of precision of response are also better in reflecting performance changes over time compared to measures of speed of response. In general, both visual representations of performance progression (over gameplay time in Figure 1 and across difficulty levels in Figure 2) provided valuable insights. If the scoring system can account for varying levels of difficulty (with scores that increase with task difficulty), the first option (Figure 1) may be recommended because it enables the analysis of an individual's overall performance progression regardless of the difficulty level they finally reached. However, this approach does not apply to accuracy metrics (eg, precision score). Accuracy metrics should exhibit a plateau curve toward maximum precision followed by a sharp initial drop after each level increase in the game followed by the typical reciprocal curves of improvement followed by a plateau in performance. This pattern is consistent with the results obtained in this study. However, it can only be observed in the visual representation that shows different levels of difficulty (Figure 2). In summary, our findings indicate the need for adaptable game metrics that can also account for varying tasks and levels of difficulty within the game.

Our data also demonstrate high interindividual variability in performance progression. These findings could be associated with the high variability in adherence (total sessions played per participant), which was mainly because some participants played the games much more often than instructed. In particular, the total number of completed sessions ranged from 33 to 159, with an average number of completed sessions of 54.4 (SD 13.0) in the FS and 71.5 (SD 26.2) in the ES [58,59]. However, it could also be associated with clinical or demographic characteristics of the participants. Research does confirm that individuals adapt their skills at different rates, which might be attributed to different learning styles [92,99,102]. Furthermore, previous studies have indicated that cognitively impaired groups have been observed to exhibit a flatter performance progression with a less prominent initial increase compared to healthy controls [57,103]. Other studies have suggested that age has an influence on individual performance progress in serious games [104,105]. Therefore, the observed interindividual variability in our analysis could be considered normal and an expected outcome as our group was relatively heterogeneous in clinical and demographic characteristics. Additional research is required to explore what individual characteristics may lead to different performance progressions; identify predictors for steeper performance curves in this field; and apply these findings to identify and adopt strategies that facilitate individual learning success and, thus, effectiveness in improving cognitive (and physical) performance.

Strengths and Limitations

One of the main strengths of this study was that we conducted the correlation analyses between game metrics and clinical assessments only within specific preassigned neurocognitive domains or subdomains. This approach was enabled by the already available content validation of the games to their primary trained neurocognitive domain or subdomain within the Brain-IT training concept [61]. Therefore, the primary objective was hypothesis driven, with predefined criteria for interpretation of the results. In addition, we used data from clinically validated and well-established neuropsychological assessments across all neurocognitive domains.

There are also some limitations that need to be discussed. First and most importantly, the sample size was small, and we did not conduct an a priori sample size calculation as we analyzed existing datasets from studies conducted as part of the Brain-IT project. To ensure appropriate interpretation of our results, we conducted a post hoc power analysis using G*Power (version 3.1) [90,91], which revealed that most of our analyses were underpowered. This might have resulted in potential false-negative or false-positive results, affects the robustness and generalizability of our findings, and warrants confirmatory studies with a priori sample size calculations to ensure more reliable and robust conclusions. Considering our criteria to confirm the alternative hypotheses as assumptions for a power analysis, a sample size of 37 or 50 would be required to show (1) a statistically significant correlation ($P \leq .05$; uncorrected; 1-sided) with (2) a correlation coefficient of $\rho \geq 0.4$ with sufficient power of $\beta > .80$ or $\beta > .90$, respectively, which may be used as a reference for planning future studies on this topic.

Second, this study was a secondary analysis, and the original studies as well as the intervention were not originally designed a priori to investigate the objectives of this study, which restricted our methodological choices in this secondary analysis. Although the game sessions were standardized in terms of game design and intervention delivery, there were interindividual differences in gameplay due to personalization and individual progression of the intervention.

Specifically, these restrictions limited the following methodological aspects for the primary research question. According to the Brain-IT training concept, the games were individually assigned considering the participants' baseline neuropsychological performance, with all participants starting at level 1 of the game demands in the first training session. Thereafter, either the games progressed or more difficult games were introduced as soon as (1) a plateau in performance was reached on predetermined game metrics, (2) a predetermined target score was reached on a predetermined game metric for progression to the next level, (3) the participant requested an increase in task demands, or (4) the staff supervising the participants deemed the increase in task demands feasible [59]. Given these rules for individualized progression of exergame demands, the path through these games and game levels was different for each participant, with large differences in the time to progress to the next levels and the maximum level reached after the 12-week intervention (as shown in Figure 3). For example, one participant might reach level 5 after only 2 weeks of training, whereas another might not reach this level until after 12 weeks. As the first research objective was tied to baseline cognitive performance as measured using standardized neuropsychological assessments, only the first level provided the same standardized conditions for all study participants that allowed the cross-sectional analyses to address this research

XSL•FO
RenderX

objective. This methodological limitation may have resulted in participants with higher baseline functioning being underchallenged. However, we did not observe any floor or ceiling effects in any of the game metrics except for number of misses in the game Targets and point score in the game Tetris (Figures S1-S8 in Multimedia Appendix 1), whereas there was a good amount of variability in all game metrics (Table S3 in Multimedia Appendix 1). Therefore, it can be assumed that this limitation did not have a significant impact on the observed results. Nevertheless, analysis of higher levels of difficulty would undoubtedly provide additional insights. As an example, the findings of Litz et al [94] observed a tendency toward more consistent relationships between in-game metrics and standardized cognitive assessments at higher difficulty levels or when averaging scores over all game levels [94]. However, given the individual paths through the game levels and the corresponding time differences to reach a given level in this study, it would be impossible to disentangle learning effects (for participants who had already played the game several times) and performance improvements over time during the intervention from "true" cognitive performance as measured at baseline. Furthermore, not all participants played the same games, resulting in smaller sample sizes in some games. Games that were considered more difficult, such as Nomis and Tetris, had small sample sizes as they were only played by participants with a higher baseline cognitive performance in the neurocognitive domains of executive functioning and visuospatial skills, which informed the allocation of the games in the first session. This could have biased the analysis. Therefore, future studies should be designed to specifically address the research question of this study rather than relying on existing datasets that might restrict the methodological choices in the analyses of the data. This entails a cross-sectional design with a priori sample size calculations rather than the use of data from longitudinal studies and applying different game levels to investigate whether the game levels induce the appropriate changes in perceived task demands (eg, as has been investigated in the work by Manser and de Bruin [71]) and whether and how performance at higher difficulty levels might differentially relate to cognitive performance assessed using standardized neuropsychological assessments.

The personalization and individualized progression of the training also presented a challenge for the analysis of our secondary research question (ie, performance progression) as players varied in the number of times they played the games and difficulty levels were individually adapted. To account for these interindividual differences, we found a method to present performance progression across gameplay by normalizing time to game completion as a percentage. Using this method, we were able to account for variability in adherence and difficulty levels at a group level. However, this method may result in loss of some information, and individual curves were compressed or stretched. These methodological constraints also prevented us from conducting more in-depth analyses of the observed high interindividual variability in performance progression (see Figures 2 and 3 with color coding of individual participant progression curves), such as exploratory analyses to identify potential factors contributing to the variability (eg, age, years of education, and baseline cognitive status) as these factors (particularly baseline cognitive status) influenced individualized tailoring (personalized allocation as well as individualized progression of games and game levels). Therefore, future studies should be specifically designed to account for these aspects and allow for more in-depth analysis to disentangle the factors that influence the high variability in performance progression over time and derive evidence-based recommendations on the implications of accounting for this variability in the methodological aspects of individually tailoring exergame-based training.

Third, although each exergame focused on 1 specific neurocognitive domain or subdomain, they often indirectly trained other domains, which may have confounded the domain-specific analysis. Fourth, the absence of data from specific motor assessments precluded their inclusion in our analyses. Finally, this study only included older adults with mNCD, and the results cannot be generalized to other populations.

Implications for Research

The personalization of training is becoming increasingly popular in the current development of technology-supported interventions, including exergames. Within this field, research should aim to identify valid parameters that can be used for in-game adaptation. The potential of game metrics should be further investigated through studies with structured interventions while also considering the impact of game task design as well as personal and external factors. Given the proposal by Netz [106] that (adherence to) optimal exercise intensity and motor-cognitive task complexity are driving mechanisms for global and task-specific neuroplasticity, respectively, and as discussed in previous research, personalized training adaptations should go beyond game metrics and consider additional aspects of players related to physiological states (e.g., attention, stress) or personal characteristics (e.g., learning style, intelligence) to provide an optimal exercise intensity and motor-cognitive task complexity [49]. Different biofeedback systems, such as biocybernetic adaptation loops, which continuously adjust game difficulty or game content in response to real-time physiological data from the user [107-109], are currently being investigated in this area and should be further explored. Furthermore, previous reviews have identified several features of fully immersive virtual reality (VR) that can support skill learning [110], whereas fully immersive VR or augmented reality has a fundamental advantage in delivering ecologically valid game scenarios via recreating a certain level of naturalistic sensory-motor interaction between the user and graphical user interface [111]. Such features include but are not limited to multisensory interactivity with haptic feedback in a high-fidelity VR with 3D representations and the presence of avatars [110]. Given that specially designed and customized VR interventions have been more effective in neurorehabilitation compared to the use of commercially available VR systems [112], these features should be more thoroughly investigated regarding their capacity to improve tailoring of the exercises to provide optimal exercise intensity and motor-cognitive task complexity.

The variability in performance progression across participants presents further aspects that need to be investigated. Research

XSL•FO
RenderX

should try to identify factors that can help understand and identify how individuals differ in their cognitive and motor-cognitive skill acquisition and performance progression. Different factors that may contribute to a steeper progression curve should be investigated. This should include subgroup analyses, which we suggest should include assessing (1) the baseline physical and cognitive state of the individuals [57,103,105], (2) age [104,105], (3) adherence, and (4) self-efficacy [95]. Moreover, game metrics obtained in the "game module" only provide information on physical and cognitive performance in a specific task (game). However, solely analyzing game metrics does not allow for any conclusions on the efficacy or effectiveness of the training as the habituation and learning effects cannot be disentangled from "true" physical and motor-cognitive performance improvements. Such "true" performance adaptations would either be assessed using standardized physical, motor, or neuropsychological assessments or obtained in a separate evaluation module that integrates scientifically validated gamified assessments to regularly verify whether the exercise or training stimulus is sufficient to induce the desired skill-related changes or near- and far-transfer effects. Therefore, future longitudinal studies should also seek to investigate to which extent performance improvements measured using game metrics within the "game" module translate to physical, motor, or cognitive performance enhancements measured using validated assessments and elucidate factors for promoting such transference effects.

It has been proposed that (adherence to) optimal intensity and complexity of exercises are the driving mechanisms for global neuroplasticity (ie, exercise intensity) and task-specific neuroplasticity (ie, cognitive and motor-cognitive task complexity) [106]. The ability of technology-enhanced training to individually progress exercises according to performance metrics in real time offers a potential advantage over conventional exercises and is considered a key advantage of serious video games (such as exergames or cognitive training games) [113]. There is initial empirical evidence supporting this assumption by showing that training with adaptive difficulty algorithms may be superior to generic games to improve both cognitive and noncognitive outcomes in individuals with mNCD in the context of computerized cognitive training [114]. Most studies on exergame-based training [5,115] and technology-supported physical activity interventions in general [46] seem to also have recognized this potential benefit and individually tailored or progressed the exercises by adjusting the difficulty or complexity of the task or games according to the users' ability or performance [115]. However, only 42% [5] or 14% [46] of studies have reported or provided sufficient details on exercise intensity, whereas details on cognitive demands were neither assessed nor reported in any of the analyzed studies in the review by Manser et al [5]. Moreover, while a recent review observed that tailored technology-enhanced physical activity interventions (including exergame-based training) resulted in significantly higher adherence than control interventions [46], a recent systematic review observed that the effects on cognitive performance were moderated not by individualization of training (ie, tailored vs generic [one size fits all]) for exergame-based training in middle-aged to older adults but by other variables (eg, "body position" variable, favoring stepping movements in a standing position, or "exercise intensity" variable, favoring moderate exercise intensity) [5]. Therefore, it remains to be investigated in more detail whether individualization of training translates to moderating the efficacy or effectiveness of technology-enhanced training and how such individualization may be optimally designed to maximize efficacy or effectiveness in exergame-based training [5], as well as the broader context of technology-enhanced physical and motor-cognitive training. Ultimately, these investigations should seek to improve the personalization of technology-enhanced interventions by identifying and adopting strategies that facilitate individual learning success and, thus, promote effectiveness in improving cognitive (and physical) performance.

Conclusions

This secondary analysis of 2 RCTs provided valuable insights into exergame metrics. This study demonstrated that a selection of game metrics can serve as valid indicators of in-game domain-specific cognitive performance during exergaming. Moreover, metrics that reflect the precision of responses overall performed better than metrics reflecting the speed of responses, suggesting that metrics for precision may be better indicators of cognitive performance and performance progression during exergaming. Therefore, it is recommended to carefully select and prevalidate game metrics for purposes beyond game performance measurement and monitoring, such as informing personalized adjustments of the difficulty and complexity of exergames. Incorporating valid game metrics could increase their value in exergame designs, which, in turn, could increase exergame engagement and effectiveness. Due to the complex nature of exergames, in-game performance may depend on additional factors such as game design, external conditions, and personal characteristics of the player. Furthermore, the analysis of performance progression revealed high interindividual differences, underlining the importance of personalized training and suggesting the need for further research to explore characteristics explaining these differences and identify and adopt strategies that facilitate individual learning success and, thus, promote effectiveness in improving health- or disease-related outcomes.

Acknowledgments

The authors would like to thank all collaborating recruitment partners of the original randomized controlled trials for establishing contact between individuals who had mild neurocognitive disorder and the study team and all participants in this study for their participation and valuable contribution to this project. They would also like to thank all the master's students in our study team who, in addition to the first author, were involved in the data collection and supervision of the participants, namely, Patricia

XSL•FO
RenderX

Groth, Kathrin Rohr, Chiara Bassi, Nadine Decher, Anna Riedler, Enis Ljatifi, Julia Müller, and Julia Czopek-Rowinska. The studies that collected the data analyzed in this study were funded by Dementia Research Switzerland – Synapsis Foundation (grant 2019-PI06) and the Gebauer Foundation and were financially supported by the Dalle Molle Foundation (Quality of life prize 2020). The Senso Flex training systems were provided free of charge by Dividat AG for the duration of these studies. Neither the funders nor Dividat AG played any role in the design of this study, nor did they play any role in the collection, management, analysis, and interpretation of data; writing of the report; or decision to submit the report for publication.

Data Availability

Publicly available datasets for the feasibility study [116] and the effectiveness study [117] were analyzed in this study. These datasets include the demographic and clinical information analyzed, as well as the raw data for the cognitive assessments. Descriptive statistics on the data on the game metrics are included in this paper and its supplementary files. Further inquiries can be directed to the corresponding author.

Authors' Contributions

WK, PM, and EDdB conceptualized the study. WK was responsible for data management, statistical analysis, and writing the manuscript under the supervision of PM and EDdB. All authors reviewed and revised the manuscript and approved the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT 2017 checklist for nonpharmacologic treatments and supplementary tables and figures, including descriptive statistics for clinical assessments and game metric data, along with scatterplots of correlation analyses between exergame metrics with neuropsychological assessments.
[PDF File (Adobe PDF File), 373 KB - games_v13i1e65878_app1.pdf]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1086 KB - games_v13i1e65878_app2.pdf]

References

1. Risk reduction of cognitive decline and dementia: WHO guidelines. World Health Organization. 2019. URL: https://www.who.int/publications/i/item/risk-reduction-of-cognitive-decline-and-dementia [accessed 2024-04-29]
2. Ge S, Zhu Z, Wu B, McConnell E. Technology-based cognitive training and rehabilitation interventions for individuals with mild cognitive impairment: a systematic review. BMC Geriatr 2018 Sep 15;18(1):213 [FREE Full text] [doi: 10.1186/s12877-018-0893-1] [Medline: 30219036]
3. Valenzuela T, Okubo Y, Woodbury A, Lord S, Delbaere K. Adherence to technology-based exercise programs in older adults: a systematic review. J Geriatr Phys Ther 2018;41(1):49-61. [doi: 10.1519/JPT.0000000000000095] [Medline: 27362526]
4. Chan AT, Ip R, Tran J, Chan JY, Tsoi K. Computerized cognitive training for memory functions in mild cognitive impairment or dementia: a systematic review and meta-analysis. NPJ Digit Med 2024 Jan 03;7(1):1 [FREE Full text] [doi: 10.1038/s41746-023-00987-5] [Medline: 38172429]
5. Manser P, Herold F, de Bruin ED. Components of effective exergame-based training to improve cognitive functioning in middle-aged to older adults - a systematic review and meta-analysis. Ageing Res Rev 2024 Aug;99:102385 [FREE Full text] [doi: 10.1016/j.arr.2024.102385] [Medline: 38914262]
6. Hernandez-Martinez J, Ramos-Espinoza F, Muñoz-Vásquez C, Guzman-Muñoz E, Herrera-Valenzuela T, Branco B, et al. Effects of active exergames on physical performance in older people: an overview of systematic reviews and meta-analysis. Front Public Health 2024;12:1250299 [FREE Full text] [doi: 10.3389/fpubh.2024.1250299] [Medline: 38655514]
7. Li R, Geng J, Yang R, Ge Y, Hesketh T. Effectiveness of computerized cognitive training in delaying cognitive function decline in people with mild cognitive impairment: systematic review and meta-analysis. J Med Internet Res 2022 Oct 27;24(10):e38624 [FREE Full text] [doi: 10.2196/38624] [Medline: 36301590]
8. Stojan R, Voelcker-Rehage C. A systematic review on the cognitive benefits and neurophysiological correlates of exergaming in healthy older adults. J Clin Med 2019 May 23;8(5):63 [FREE Full text] [doi: 10.3390/jcm8050734] [Medline: 31126052]
9. Temprado JJ. Can exergames be improved to better enhance behavioral adaptability in older adults? An ecological dynamics perspective. Front Aging Neurosci 2021;13:670166 [FREE Full text] [doi: 10.3389/fnagi.2021.670166] [Medline: 34122047]

XSL•FO
RenderX

10. Torre MM, Temprado JJ. A review of combined training studies in older adults according to a new categorization of conventional interventions. Front Aging Neurosci 2021;13:808539 [FREE Full text] [doi: 10.3389/fnagi.2021.808539] [Medline: 35177975]
11. Witherspoon L. ACSM Information on Exergaming. American College of Sports Medicine. URL: https://healthysd.gov/wp-content/uploads/2014/11/exergaming.pdf [accessed 2024-04-29]
12. Torre MM, Temprado JJ. Effects of exergames on brain and cognition in older adults: a review based on a new categorization of combined training intervention. Front Aging Neurosci 2022;14:859715 [FREE Full text] [doi: 10.3389/fnagi.2022.859715] [Medline: 35431905]
13. Gallou-Guyot M, Mandigout S, Bherer L, Perrochon A. Effects of exergames and cognitive-motor dual-task training on cognitive, physical and dual-task functions in cognitively healthy older adults: an overview. Ageing Res Rev 2020 Nov;63:101135 [FREE Full text] [doi: 10.1016/j.arr.2020.101135] [Medline: 32768594]
14. Jiang J, Guo W, Wang B. Effects of exergaming on executive function of older adults: a systematic review and meta-analysis. PeerJ 2022;10:e13194 [FREE Full text] [doi: 10.7717/peerj.13194] [Medline: 35433124]
15. Cai Z, Ma Y, Li L, Lu GZ. Effects of exergaming in older individuals with mild cognitive impairment and dementia: a systematic review and meta-analysis. Geriatr Nurs 2023;51:351-359 [FREE Full text] [doi: 10.1016/j.gerinurse.2023.03.028] [Medline: 37099867]
16. Zhao Y, Feng H, Wu X, Du Y, Yang X, Hu M, et al. Effectiveness of exergaming in improving cognitive and physical function in people with mild cognitive impairment or dementia: systematic review. JMIR Serious Games 2020 Jun 30;8(2):e16841 [FREE Full text] [doi: 10.2196/16841] [Medline: 32602841]
17. Swinnen N, Vandenbulcke M, de Bruin ED, Akkerman R, Stubbs B, Firth J, et al. The efficacy of exergaming in people with major neurocognitive disorder residing in long-term care facilities: a pilot randomized controlled trial. Alzheimers Res Ther 2021 Mar 30;13(1):70 [FREE Full text] [doi: 10.1186/s13195-021-00806-7] [Medline: 33785077]
18. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (DSM-5-TR). Washington, DC: American Psychiatric Association; 2013.
19. Kasper S, Bancher C, Eckert A, Förstl H, Frölich L, Hort J, et al. Management of mild cognitive impairment (MCI): the need for national and international guidelines. World J Biol Psychiatry 2020 Oct;21(8):579-594 [FREE Full text] [doi: 10.1080/15622975.2019.1696473] [Medline: 32019392]
20. Lissek V, Suchan B. Preventing dementia? Interventional approaches in mild cognitive impairment. Neurosci Biobehav Rev 2021 Mar;122:143-164 [FREE Full text] [doi: 10.1016/j.neubiorev.2020.12.022] [Medline: 33440197]
21. Livingston G, Huntley J, Liu KY, Costafreda SG, Selbæk G, Alladi S, et al. Dementia prevention, intervention, and care: 2024 report of the Lancet standing Commission. Lancet 2024 Aug 10;404(10452):572-628 [FREE Full text] [doi: 10.1016/S0140-6736(24)01296-0] [Medline: 39096926]
22. Izquierdo M, de Souto Barreto P, Arai H, Bischoff-Ferrari H, Cadore E, Cesari M, et al. Global consensus on optimal exercise recommendations for enhancing healthy longevity in older adults (ICFSR). J Nutr Health Aging 2025 Jan;29(1):100401 [FREE Full text] [doi: 10.1016/j.jnha.2024.100401] [Medline: 39743381]
23. Rodakowski J, Saghafi E, Butters MA, Skidmore ER. Non-pharmacological interventions for adults with mild cognitive impairment and early stage dementia: an updated scoping review. Mol Aspects Med 2015;43-44:38-53 [FREE Full text] [doi: 10.1016/j.mam.2015.06.003] [Medline: 26070444]
24. Yang C, Moore A, Mpofu E, Dorstyn D, Li Q, Yin C. Effectiveness of combined cognitive and physical interventions to enhance functioning in older adults with mild cognitive impairment: a systematic review of randomized controlled trials. Gerontologist 2020 Nov 23;60(8):633-642 [FREE Full text] [doi: 10.1093/geront/gnz149] [Medline: 31697831]
25. Meng Q, Yin H, Wang S, Shang B, Meng X, Yan M, et al. The effect of combined cognitive intervention and physical exercise on cognitive function in older adults with mild cognitive impairment: a meta-analysis of randomized controlled trials. Aging Clin Exp Res 2022 Feb;34(2):261-276 [FREE Full text] [doi: 10.1007/s40520-021-01877-0] [Medline: 34383248]
26. Gómez-Soria I, Marin-Puyalto J, Peralta-Marrupe P, Latorre E, Calatayud E. Effects of multi-component non-pharmacological interventions on cognition in participants with mild cognitive impairment: a systematic review and meta-analysis. Arch Gerontol Geriatr 2022;103:104751 [FREE Full text] [doi: 10.1016/j.archger.2022.104751] [Medline: 35839574]
27. Gavelin HM, Dong C, Minkov R, Bahar-Fuchs A, Ellis KA, Lautenschlager N, et al. Combined physical and cognitive training for older adults with and without cognitive impairment: a systematic review and network meta-analysis of randomized controlled trials. Ageing Res Rev 2021 Mar;66:101232 [FREE Full text] [doi: 10.1016/j.arr.2020.101232] [Medline: 33249177]
28. Barnes DE, Yaffe K. The projected effect of risk factor reduction on Alzheimer's disease prevalence. Lancet Neurol 2011 Sep;10(9):819-828 [FREE Full text] [doi: 10.1016/S1474-4422(11)70072-2] [Medline: 21775213]
29. Livingston G, Huntley J, Sommerlad A, Ames D, Ballard C, Banerjee S, et al. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. Lancet 2020 Aug;396(10248):413-446 [FREE Full text] [doi: 10.1016/s0140-6736(20)30367-6]

XSL•FO
RenderX

30. Herold F, Hamacher D, Schega L, Müller NG. Thinking while moving or moving while thinking - concepts of motor-cognitive training for cognitive performance enhancement. Front Aging Neurosci 2018;10:228 [FREE Full text] [doi: 10.3389/fnagi.2018.00228] [Medline: 30127732]
31. Lu Y, Bu FQ, Wang F, Liu L, Zhang S, Wang G, et al. Recent advances on the molecular mechanisms of exercise-induced improvements of cognitive dysfunction. Transl Neurodegener 2023 Feb 27;12(1):9 [FREE Full text] [doi: 10.1186/s40035-023-00341-5] [Medline: 36850004]
32. Huuha AM, Norevik CS, Moreira JB, Kobro-Flatmoen A, Scrimgeour N, Kivipelto M, et al. Can exercise training teach us how to treat Alzheimer's disease? Ageing Res Rev 2022 Mar;75:101559 [FREE Full text] [doi: 10.1016/j.arr.2022.101559] [Medline: 34999248]
33. Smith PJ. Pathways of prevention: a scoping review of dietary and exercise interventions for neurocognition. Brain Plast 2019 Dec 26;5(1):3-38 [FREE Full text] [doi: 10.3233/BPL-190083] [Medline: 31970058]
34. Kivipelto M, Mangialasche F, Ngandu T. Lifestyle interventions to prevent cognitive impairment, dementia and Alzheimer disease. Nat Rev Neurol 2018 Nov;14(11):653-666 [FREE Full text] [doi: 10.1038/s41582-018-0070-3] [Medline: 30291317]
35. Wang YY, Wang XX, Chen L, Liu Y, Li YR. A systematic review and network meta-analysis comparing various non-pharmacological treatments for older people with mild cognitive impairment. Asian J Psychiatr 2023 Aug;86:103635 [FREE Full text] [doi: 10.1016/j.ajp.2023.103635] [Medline: 37270875]
36. Wang YQ, Jia RX, Liang JH, Li J, Qian S, Li JY, et al. Effects of non-pharmacological therapies for people with mild cognitive impairment. A Bayesian network meta-analysis. Int J Geriatr Psychiatry 2020 Jun;35(6):591-600 [FREE Full text] [doi: 10.1002/gps.5289] [Medline: 32119152]
37. Fabel K, Kempermann G. Physical activity and the regulation of neurogenesis in the adult and aging brain. Neuromolecular Med 2008;10(2):59-66 [FREE Full text] [doi: 10.1007/s12017-008-8031-4] [Medline: 18286387]
38. Kempermann G, Fabel K, Ehninger D, Babu H, Leal-Galicia P, Garthe A, et al. Why and how physical activity promotes experience-induced brain plasticity. Front Neurosci 2010;4:189 [FREE Full text] [doi: 10.3389/fnins.2010.00189] [Medline: 21151782]
39. Veronese N, Soysal P, Demurtas J, Solmi M, Bruyère O, Christodoulou N, Alzheimer Europe, European College of Neuropsychopharmacology, European Geriatric Medicine Society (Lead Society), European Interdisciplinary Council on Ageing, European Society of ClinicalEconomic Aspects of OsteoporosisOsteoarthritis, International Association of Gerontology and Geriatrics-European Region, Scottish Brain Health ARC, World Psychiatry Association-Preventive Psychiatry Section, endorsed by the European Academy of Neurology. Physical activity and exercise for the prevention and management of mild cognitive impairment and dementia: a collaborative international guideline. Eur Geriatr Med 2023 Oct;14(5):925-952 [FREE Full text] [doi: 10.1007/s41999-023-00858-y] [Medline: 37768499]
40. Voinescu A, Papaioannou T, Petrini K, Stanton Fraser D. Exergaming for dementia and mild cognitive impairment. Cochrane Database Syst Rev 2024 Sep 25;9(9):CD013853 [FREE Full text] [doi: 10.1002/14651858.CD013853.pub2] [Medline: 39319863]
41. Chan JY, Liu J, Chan AT, Tsoi KK. Exergaming and cognitive functions in people with mild cognitive impairment and dementia: a meta-analysis. NJP Digit Med 2024 Jun 15;7(1):CD013853. [doi: 10.1038/S41746-024-01142-4]
42. Cai Z, Ma Y, Li L, Lu G. Effects of exergaming in older individuals with mild cognitive impairment and dementia: a systematic review and meta-analysis. Geriatr Nurs 2023;51:351-359 [FREE Full text] [doi: 10.1016/j.gerinurse.2023.03.028] [Medline: 37099867]
43. Marques LM, Uchida PM, Barbosa SP. The impact of Exergames on emotional experience: a systematic review. Front Public Health 2023 Sep 7;11:1209520 [FREE Full text] [doi: 10.3389/fpubh.2023.1209520] [Medline: 37744509]
44. Marshall J, Linehan C. Are exergames exercise? A scoping review of the short-term effects of exertion games. IEEE Trans. Games 2021 Jun;13(2):160-169 [FREE Full text] [doi: 10.1109/TG.2020.2995370]
45. Tan X, Wang K, Sun W, Li X, Wang W, Tian F. A review of recent advances in cognitive-motor dual-tasking for Parkinson's disease rehabilitation. Sensors 2024 Sep 30;24(19):6353 [FREE Full text] [doi: 10.3390/S24196353]
46. Dubbeldam R, Stemplewski R, Pavlova J, Cyma-Wejchenig M, Lee S, Esser P, et al. Technology-assisted physical activity interventions for older people in their home-based environment: a scoping review (Preprint). JMIRP Preprint posted online August 25, 2024 [FREE Full text] [doi: 10.2196/65746]
47. Rüth M, Schmelzer M, Burtniak K, Kaspar K. Commercial exergames for rehabilitation of physical health and quality of life: a systematic review of randomized controlled trials with adults in unsupervised home environments. Front Psychol 2023 Jun 2;14:1155569 [FREE Full text] [doi: 10.3389/fpsyg.2023.1155569] [Medline: 37333591]
48. Perrot A, Maillot P. Factors for optimizing intervention programs for cognition in older adults: the value of exergames. NJP Aging 2023 Mar 29;9(1):4 [FREE Full text] [doi: 10.1038/S41514-023-00103-7]
49. Sajjadi P, Ewais A, De Troyer O. Individualization in serious games: a systematic review of the literature on the aspects of the players to adapt to. Entertain Comput 2022 Mar;41:100468 [FREE Full text] [doi: 10.1016/j.entcom.2021.100468]
50. Sinclair J, Hingston P, Masek M. Considerations for the design of exergames. In: Proceedings of the 5th International Conference on Computer Graphics and Interactive Techniques in Australia and Southeast Asia. 2007 Presented at: GRAPHITE '07; December 1-4, 2007; Perth, Australia p. 289-295 URL: https://dl.acm.org/doi/10.1145/1321261.1321313 [doi: 10.1145/1321261.1321313]

XSL•FO
RenderX

51. Bamparopoulos G, Konstantinidis E, Bratsas C, Bamidis PD. Towards exergaming commons: composing the exergame ontology for publishing open game data. J Biomed Semantics 2016;7:4 [FREE Full text] [doi: 10.1186/s13326-016-0046-4] [Medline: 26865947]
52. Serrano-Laguna Á, Martínez-Ortiz I, Haag J, Regan D, Johnson A, Fernández-Manjón B. Applying standards to systematize learning analytics in serious games. Comput Stand Interfaces 2017:116-123 [FREE Full text] [doi: 10.1016/j.csi.2016.09.014]
53. Petsani D, Konstantinidis EI, Zilidou VI, Bamidis PD. Exploring health profiles from physical and cognitive serious game analytics. In: Proceedings of the 2nd International Conference on Technology and Innovation in Sports, Health and Wellbeing. 2018 Presented at: TISHW '18; June 20-22, 2018; Thessaloniki, Greece p. 1-6 URL: https://ieeexplore.ieee.org/document/8559562 [doi: 10.1109/tishw.2018.8559562]
54. Petsani D, Konstantinidis E, Katsouli AM, Zilidou V, Dias S, Hadjileontiadis L, et al. Digital biomarkers for well-being through exergame interactions: exploratory study. JMIR Serious Games 2022 Sep 13;10(3):e34768 [FREE Full text] [doi: 10.2196/34768] [Medline: 36099000]
55. Sevcenko N, Ninaus M, Wortha F, Moeller K, Gerjets P. Measuring cognitive load using in-game metrics of a serious simulation game. Front Psychol 2021;12:572437 [FREE Full text] [doi: 10.3389/fpsyg.2021.572437] [Medline: 33841227]
56. Konstantinidis EI, Bamidis PD, Billis A, Kartsidis P, Petsani D, Papageorgiou SG. Physical training in-game metrics for cognitive assessment: evidence from extended trials with the fitforall exergaming platform. Sensors (Basel) 2021 Aug 26;21(17):5756 [FREE Full text] [doi: 10.3390/s21175756] [Medline: 34502647]
57. Guimarães V, Sousa I, de Bruin ED, Pais J, Correia M. Using shoe-mounted inertial sensors and stepping exergames to assess the motor-cognitive status of older adults: a correlational study. Digit Health 2023;9:20552076231167001 [FREE Full text] [doi: 10.1177/20552076231167001] [Medline: 37009304]
58. Manser P, Poikonen H, de Bruin ED. Feasibility, usability, and acceptance of "Brain-IT"-a newly developed exergame-based training concept for the secondary prevention of mild neurocognitive disorder: a pilot randomized controlled trial. Front Aging Neurosci 2023;15:1163388 [FREE Full text] [doi: 10.3389/fnagi.2023.1163388] [Medline: 37810620]
59. Manser P, de Bruin ED. "Brain-IT": exergame training with biofeedback breathing in neurocognitive disorders. Alzheimers Dement 2024 Jul;20(7):4747-4764 [FREE Full text] [doi: 10.1002/alz.13913] [Medline: 38809948]
60. Targeting the brain using information technology for secondary prevention of mild neurocognitive disorder. Patrick Manser. URL: https://www.patrick-manser.com/projects/4278-brain-it-08-2020-04-2024 [accessed 2025-05-29]
61. Manser P, de Bruin ED. Making the best out of IT: design and development of exergames for older adults with mild neurocognitive disorder - a methodological paper. Front Aging Neurosci 2021;13:734012 [FREE Full text] [doi: 10.3389/fnagi.2021.734012] [Medline: 34955806]
62. Manser P, Adcock-Omlin M, de Bruin ED. Design considerations for an exergame-based training intervention for older adults with mild neurocognitive disorder: qualitative study including focus groups with experts and health care professionals and individual semistructured in-depth patient interviews. JMIR Serious Games 2023 Jan 05;11:e37616 [FREE Full text] [doi: 10.2196/37616] [Medline: 36602851]
63. Manser P, Michels L, Schmidt A, Barinka F, de Bruin ED. Effectiveness of an individualized exergame-based motor-cognitive training concept targeted to improve cognitive functioning in older adults with mild neurocognitive disorder: study protocol for a randomized controlled trial. JMIR Res Protoc 2023 Feb 06;12:e41173 [FREE Full text] [doi: 10.2196/41173] [Medline: 36745483]
64. ICD-11 international classificiation of disease 11th revision the global standard for diagnostic health information. World Health Organization. URL: https://icd.who.int/en [accessed 2020-07-20]
65. American Psychiatric Association. Diagnostic and Statistical Manual Of Mental Disorders: DSM-5™, 5th edition. Washington, DC: American Psychiatric Association; 2013.
66. Manser P, de Bruin ED. Diagnostic accuracy, reliability, and construct validity of the German quick mild cognitive impairment screen. BMC Geriatr 2024 Jul 18;24(1):613 [FREE Full text] [doi: 10.1186/s12877-024-05219-3] [Medline: 39026157]
67. O'Caoimh R, Molloy DW. The quick mild cognitive impairment screen (Qmci). In: Larner AJ, editor. Cognitive Screening Instruments: A Practical Approach. Cham, Switzerland: Springer; 2015:255-272.
68. O'Caoimh R, Gao Y, Svendovski A, Gallagher P, Eustace J, Molloy D. Comparing approaches to optimize cut-off scores for short cognitive screening instruments in mild cognitive impairment and dementia. J Alzheimers Dis 2017 Feb 10;57(1):123-133 [FREE Full text] [doi: 10.3233/JAD-161204] [Medline: 28222528]
69. Alles über den Dividat Senso. Dividat Senso. URL: https://dividat.com/senso [accessed 2024-04-29]
70. Sachdev PS, Blacker D, Blazer D, Ganguli M, Jeste D, Paulsen JS, et al. Classifying neurocognitive disorders: the DSM-5 approach. Nat Rev Neurol 2014 Nov;10(11):634-642 [FREE Full text] [doi: 10.1038/nrneurol.2014.181] [Medline: 25266297]
71. Manser P, de Bruin ED. Test-retest reliability and validity of vagally-mediated heart rate variability to monitor internal training load in older adults: a within-subjects (repeated-measures) randomized study. BMC Sports Sci Med Rehabil 2024 Jun 27;16(1):141 [FREE Full text] [doi: 10.1186/s13102-024-00929-y] [Medline: 38926777]
72. Huber SK, Manser P, de Bruin ED. PEMOCS: theory derivation of a concept for PErsonalized MOtor-Cognitive exergame training in chronic Stroke-a methodological paper with an application example. Front Sports Act Living 2024;6:1397949 [FREE Full text] [doi: 10.3389/fspor.2024.1397949] [Medline: 38915297]

XSL•FO
RenderX

73. Gentile AM. Skill acquisition: action, movement, and neuromotor processes. In: Carr JH, Shepherd RD, editors. Movement Science: Foundations for Physical Therapy in Rehabilitation. Rockville, MA: Aspen Publishers; 2000:111-187.
74. Fitts PM. Human Performance. New York, NY: Praeger Publishers; 1967.
75. VanLehn K. Cognitive skill acquisition. Annu Rev Psychol 1996;47:513-539 [FREE Full text] [doi: 10.1146/annurev.psych.47.1.513] [Medline: 15012487]
76. Slade SC, Dionne CE, Underwood M, Buchbinder R. Consensus on exercise reporting template (CERT): explanation and elaboration statement. Br J Sports Med 2016 Dec;50(23):1428-1437 [FREE Full text] [doi: 10.1136/bjsports-2016-096651] [Medline: 27707738]
77. O'Caoimh R, Gao Y, McGlade C, Healy L, Gallagher P, Timmons S, et al. The quick mild cognitive impairment (Qmci) screen: a new screening tool for mild cognitive impairment. Ir J Med Sci 2012;181:S228-S229. [doi: 10.1037/T83417-000]
78. O'Caoimh R, Molloy DW. The quick mild cognitive impairment screen (Qmci). In: Larner AJ, editor. Cognitive Screening Instruments: A Practical Approach. Cham, Switzerland: Springer; 2017:255-272.
79. Petermann F, Lepach A. Wechsler memory scale® – fourth edition. Testzentrale. 2012. URL: https://www.testzentrale.de/shop/wechsler-memory-scaler-fourth-edition.html [accessed 2024-04-29]
80. Wechsler D. Wechsler memory scale. 4th edition. Pearson. URL: https://www.pearsonassessments.com/en-us/Store/Professional-Assessments/Cognition-&-Neuro/Wechsler-Memory-Scale-%7C-Fourth-Edition/p/100000281 [accessed 2024-04-29]
81. Mueller ST, Piper BJ. The psychology experiment building language (PEBL) and PEBL test battery. J Neurosci Methods 2014 Jan 30;222:250-259 [FREE Full text] [doi: 10.1016/j.jneumeth.2013.10.024] [Medline: 24269254]
82. Menzel-Begemann A. Testverfahren zur Erfassung der Planungsfähigkeit im Alltag. Testzentrale. URL: https://www.testzentrale.de/shop/handlungsorganisation-und-tagesplanung.html [accessed 2024-04-29]
83. Pflüger M, Gschwandtner U. Testbatterie zur Aufmerksamkeitsprüfung (TAP) Version 1.7. Z Klin Psychol Psychother 2003 Apr 1;32(2):155-157. [doi: 10.1026//1616-3443.32.2.155]
84. Pflüger M, Gschwandtner U. Testbatterie zur Aufmerksamkeitsprüfung (TAP) Version 1.7. Z Klin Psychol Psychother 2003 Apr 1;32(2):155-157 [FREE Full text] [doi: 10.1026//1616-3443.32.2.155]
85. Field A, Miles J, Field Z. Discovering Statistics Using R. Thousand Oaks, CA: Sage Publications; 2012.
86. Streiner DL, Norman GR, Cairney J. Health Measurement Scales: A Practical Guide to Their Development and Use. Oxford, UK: Oxford Academic Press; 2015.
87. Cohen J. Statistical power analysis. Curr Dir Psychol Sci 1992 Jun 01;1(3):98-101. [doi: 10.1111/1467-8721.ep10768783]
88. Armstrong RA. When to use the Bonferroni correction. Ophthalmic Physiol Opt 2014 Sep;34(5):502-508 [FREE Full text] [doi: 10.1111/opo.12131] [Medline: 24697967]
89. Sullivan GM, Feinn RS. Facts and fictions about handling multiple comparisons. J Grad Med Educ 2021 Aug;13(4):457-460 [FREE Full text] [doi: 10.4300/JGME-D-21-00599.1] [Medline: 34434505]
90. Faul F, Erdfelder E, Buchner A, Lang A. Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. Behav Res Methods 2009 Nov;41(4):1149-1160 [FREE Full text] [doi: 10.3758/brm.41.4.1149]
91. Faul F, Erdfelder E, Lang A, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007 May;39(2):175-191 [FREE Full text] [doi: 10.3758/bf03193146] [Medline: 17695343]
92. Anderson DI, Lohse KR, Lopes TC, Williams AM. Individual differences in motor skill learning: past, present and future. Hum Mov Sci 2021 Aug;78:102818 [FREE Full text] [doi: 10.1016/j.humov.2021.102818] [Medline: 34049152]
93. Jacoby WG. Loess: a nonparametric, graphical tool for depicting relationships between variables. Elect Stud 2000 Dec;19(4):577-613 [FREE Full text] [doi: 10.1016/S0261-3794(99)00028-1]
94. Litz E, Ball C, Jansen C, Werner C, de Bruin E, Hauer K. Validation of a motor-cognitive assessment for a stepping exergame in older adults: use of game-specific, internal data stream. Games Health J 2020 Apr;9(2):95-107 [FREE Full text] [doi: 10.1089/g4h.2019.0081] [Medline: 31651191]
95. Wulf G, Lewthwaite R. Optimizing performance through intrinsic motivation and attention for learning: the OPTIMAL theory of motor learning. Psychon Bull Rev 2016 Oct 29;23(5):1382-1414. [doi: 10.3758/s13423-015-0999-9] [Medline: 26833314]
96. Julayanont P, Nasreddine ZS. Montreal cognitive assessment (MoCA): concept and clinical review. In: Larner AJ, editor. Cognitive Screening Instruments: A Practical Approach. Cham, Switzerland: Springer; 2017:139-195.
97. Arevalo-Rodriguez I, Smailagic N, Roqué I Figuls M, Ciapponi A, Sanchez-Perez E, Giannakou A, et al. Mini-Mental State Examination (MMSE) for the detection of Alzheimer's disease and other dementias in people with mild cognitive impairment (MCI). Cochrane Database Syst Rev 2015 Mar 05;2015(3):CD010783 [FREE Full text] [doi: 10.1002/14651858.CD010783.pub2] [Medline: 25740785]
98. Streicher A, Smeddinck JD. Personalized and adaptive serious games. In: Proceedings of the 2015 Conference on Entertainment Computing and Serious Games. 2015 Presented at: ECSG '15; July 5-10, 2015; Dagstuhl Castle, Germany p. 332-377 URL: https://link.springer.com/chapter/10.1007/978-3-319-46152-6_14 [doi: 10.1007/978-3-319-46152-6_18]
99. Schmidt RA, Lee TD, Winstein C, Wulf G, Zelaznik H. Motor Control and Learning: A Behavioral Emphasis. 5th edition. New York, NY: Human Kinetics; 2018.

XSL•FO
RenderX

100. Newell A, Rosenbloom PS. Mechanisms of skill acquisition and the law of practice. In: Anderson JR, editor. Cognitive Skills and Their Acquisition. New York, NY: Psychology Press; 2013:1-55.
101. Gray WD, Lindstedt JK. Plateaus, dips, and leaps: where to look for inventions and discoveries during skilled performance. Cogn Sci 2017 Sep;41(7):1838-1870 [FREE Full text] [doi: 10.1111/cogs.12412] [Medline: 27766673]
102. Haider H, Frensch PA. Why aggregated learning follows the power law of practice when individual learning does not: Comment on Rickard (1997, 1999), Delaney et al. (1998), and Palmeri (1999). J Exp Psychol Learn Mem Cogn 2002;28(2):392-406. [doi: 10.1037/0278-7393.28.2.392]
103. Papp KV, Snyder PJ, Maruff P, Bartkowiak J, Pietrzak R. Detecting subtle changes in visuospatial executive function and learning in the amnestic variant of mild cognitive impairment. PLoS One 2011;6(7):e21688 [FREE Full text] [doi: 10.1371/journal.pone.0021688] [Medline: 21779333]
104. Bonnechère B, Bier J, Van Hove O, Sheldon S, Samadoulougou S, Kirakoya-Samadoulougou F, et al. Age-associated capacity to progress when playing cognitive mobile games: ecological retrospective observational study. JMIR Serious Games 2020 Jun 12;8(2):e17121. [doi: 10.2196/17121] [Medline: 32530432]
105. Fernandez-Ballesteros R, Botella, Zamarron, Cabras, Molina, Schettini, Tarraga. Cognitive plasticity in normal and pathological aging. Clin Interv Aging 2012 Jan:15. [doi: 10.2147/cia.s27008]
106. Netz Y. Is there a preferred mode of exercise for cognition enhancement in older age?: a narrative review. Front Med (Lausanne) 2019;6:57 [FREE Full text] [doi: 10.3389/fmed.2019.00057] [Medline: 30984760]
107. Pope AT, Stephens CL, Gilleade K. Biocybernetic adaptation as biofeedback training method. In: Fairclough SH, Gilleade K, editors. Advances in Physiological Computing. Cham, Switzerland: Springer; 2014:91-115.
108. Fairclough S. A closed-loop perspective on symbiotic human-computer interaction. In: Proceedings of the 4th International Conference on Symbiotic Interaction. 2015 Presented at: Symbiotic Interactionth International Workshop, Symbiotic , Berlin, Germany, October , , Proceedings 4pringer; October 7-8, 2015; Berlin, Germany p. 57-67 URL: https://link.springer.com/chapter/10.1007/978-3-319-24917-9_6 [doi: 10.1007/978-3-319-24917-9_6]
109. Muñoz J, Gouveia E, Cameirão M, Bermúdez i Badia S. The biocybernetic loop engine: an integrated tool for creating physiologically adaptive videogames. In: Proceedings of the 4th International Conference on Physiological Computing Systems. 2017 Presented at: PCS '17; July 27-28, 2017; Madrid, Spain p. 45-47 URL: https://www.scitepress.org/Link.aspx?doi=10.5220/0006429800450054 [doi: 10.5220/0006429800450054]
110. Tusher HM, Mallam S, Nazir S. A systematic review of virtual reality features for skill training. Tech Know Learn 2024 Jan 05;29(2):843-878 [FREE Full text] [doi: 10.1007/S10758-023-09713-2]
111. Tieri G, Morone G, Paolucci S, Iosa M. Virtual reality in cognitive and motor rehabilitation: facts, fiction and fallacies. Expert Rev Med Devices 2018 Feb;15(2):107-117 [FREE Full text] [doi: 10.1080/17434440.2018.1425613] [Medline: 29313388]
112. Voinescu A, Sui J, Stanton Fraser D. Virtual reality in neurorehabilitation: an umbrella review of meta-analyses. J Clin Med 2021 Apr 02;10(7):23 [FREE Full text] [doi: 10.3390/jcm10071478] [Medline: 33918365]
113. Sokolov AA, Collignon A, Bieler-Aeschlimann M. Serious video games and virtual reality for prevention and neurorehabilitation of cognitive decline because of aging and neurodegeneration. Curr Opin Neurol 2020 Apr;33(2):239-248 [FREE Full text] [doi: 10.1097/WCO.0000000000000791] [Medline: 32073439]
114. Bahar-Fuchs A, Webb S, Bartsch L, Clare L, Rebok G, Cherbuin N, et al. Tailored and adaptive computerized cognitive training in older adults at risk for dementia: a randomized controlled trial. J Alzheimers Dis 2017 Oct 03;60(3):889-911 [FREE Full text] [doi: 10.3233/JAD-170404] [Medline: 28922158]
115. Maier M, Rubio Ballester B, Duff A, Duarte Oller E, Verschure PF. Effect of specific over nonspecific VR-based rehabilitation on poststroke motor recovery: a systematic meta-analysis. Neurorehabil Neural Repair 2019 Feb;33(2):112-129 [FREE Full text] [doi: 10.1177/1545968318820169] [Medline: 30700224]
116. Data for project 'feasibility, usability and acceptance of a newly developed exergame-based training concept for older adults with mild neurocognitive disorder - a pilot randomized controlled trial'. zenodo. URL: https://zenodo.org/records/7428378 [accessed 2024-04-29]
117. Data for Project "Brain-IT' - exergame training with biofeedback breathing in neurocognitive disorders. zenodo. URL: https://zenodo.org/records/10695988 [accessed 2024-04-29]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
ES: effectiveness study
FS: feasibility study
MMSE: Mini-Mental State Examination
mNCD: mild neurocognitive disorder
MoCA: Montreal Cognitive Assessment
PEBL: Psychology Experiment Building Language
RCT: randomized controlled trial

XSL•FO
RenderX

VR: virtual reality

Edited by A Coristine; submitted 28.08.24; peer-reviewed by P Makmee, X Cheng; comments to author 18.12.24; revised version received 22.01.25; accepted 12.03.25; published 21.05.25.

Please cite as:
Kaiser W, de Bruin ED, Manser P
Domain-Specific Evaluation of Exergame Metrics Among Older Adults With Mild Neurocognitive Disorder: Secondary Analysis of 2 Randomized Controlled Trials
JMIR Serious Games 2025;13:e65878
URL: https://games.jmir.org/2025/1/e65878
doi:10.2196/65878
PMID:

©Wanda Kaiser, Eling D de Bruin, Patrick Manser. Originally published in JMIR Serious Games (https://games.jmir.org), 21.05.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Positive Effect of Video-Game Play on College Students' Anxiety and Depression Symptoms During the COVID-19 Pandemic Shelter-in-Place Lockdowns: Mixed Methods Study

Fengbin Hu[1], PhD; Zixue Tai[2], PhD; Jianping Liu[3], PhD

[1]Research Center for Cyberspace Governance, MOE Laboratory for National Development and Intelligent Governance, Fudan University, Shanghai, China

[2]School of Journalism and Media, University of Kentucky, Lexington, KY, United States

[3]School of Journalism and Communication, Minjiang University, Fuzhou, China

Corresponding Author:
Fengbin Hu, PhD
Research Center for Cyberspace Governance
MOE Laboratory for National Development and Intelligent Governance
Fudan University
West Main Building, Guanghua Tower
220 Handan Road, Yangpu District
Shanghai, 200433
China
Phone: 86 21 55664202
Email: hufb@fudan.edu.cn

Abstract

Background: A growing body of research has examined the connection between video game play and relief from anxiety and depressive symptoms during the COVID-19 pandemic years. However, evidence has been limited in dissecting the role of video games in moderating personal health outcomes such as anxiety and depression. This research aimed to provide insights into this area by investigating Chinese college students living under difficult conditions during the COVID-19 pandemic shelter-in-place lockdowns.

Objective: This study investigated the effect of video games on mental disorders among Chinese college students during the COVID-19 pandemic shelter-in-place lockdown mandates.

Methods: A mixed methods approach was adopted. The quantitative portion included a cross-sectional survey of a national sample of 2818 (n=1396, 49.54% male vs n=1422, 50.46% female) college students from 8 provinces spanning 16 geographic regions during the extended COVID-19 pandemic lockdowns in late 2021 and early 2022. The qualitative portion encompassed 17 semistructured in-depth interviews of (9/17, 53% female vs 8/17, 47% male) students on their experiences, perceptions, and evaluations of playing video games during the lockdowns. Anxiety symptoms were measured using the 7-item self-administered Generalized Anxiety Disorder scale, while depression was assessed with the 9-item Patient Health Questionnaire scale. Multiple regression analyses were used to examine whether demographic variables (ie, sex and year in college), modality and content of play, and play time accounted for the outcomes of anxiety and depression. ANOVA tests were performed on overall playtime and the prelockdown-to-lockdown period change in game time on the severity of anxiety and depression symptoms. Thematic analysis of qualitative data provided additional perspectives on gaming dynamics in relation to anxiety and depression.

Results: No significant sex effect was detected in video game play outcomes for anxiety or depression. At the level of the study population, a significant difference between gamers and nongamers was observed to moderate anxiety (t_{2816}=−0.42, P=.02) but not depression (t_{2816}=−0.12, P=.60) symptoms, controlling for the effect of sex. Playing more online games, spending more time gaming, and an increase in play time were linked to heightened anxiety and depression symptoms. Conversely, playing with friends was connected to lowered anxiety (β=−0.05, P=.04) and depression (β=−0.08, P=.003) scores.

Conclusions: The buffer effects of video games may be strongest among routine players with moderate to low anxiety and depression symptoms, while excessive gaming, as shown in the overall amount of game time as well as the prelockdown-to-lockdown period increase in game time, may have detrimental consequences among those struggling with high

XSL•FO
RenderX

anxiety and depression. Social play was an effective mechanism in mitigating anxiety and depressive tendencies. Future research should address game content and modality as well as the long-term impacts of video game engagement during crises.

(JMIR Serious Games 2025;13:e58857) doi:10.2196/58857

KEYWORDS

anxiety; depression; video games; online games; mobile games; college students; COVID-19; COVID-19 pandemic lockdowns; China

Introduction

Background

The Global Burden of Diseases, Injuries, and Risk Factors Study has consistently shown that mental disorders are among the leading causes of global health-related burden, with no evidence of reduction since 1990, while the 2019 Global Burden of Diseases, Injuries, and Risk Factors Study showed that anxiety and depressive disorders were the 2 leading mental disorders in the world [1,2]. The World Health Organization, which advocates mental health as a basic human right, noted in its survey of member nations that, despite increasing interest in and understanding of mental health in the past 20 years, mental health care systems and services are ill-equipped in most countries, and nonexistent in some low-income regions, to handle people's needs [3].

Globally, the university student population is a high-risk demographic in developing mental health problems [4,5], and the number of students with serious emotional and mental health problems has been on the rise across college campuses in the 21st century [6,7]. In particular, there have been substantial increases in the prevalence rates of symptoms of depression, anxiety, and suicidal ideation among college students in recent decades [8-10]. The global crisis created by the COVID-19 pandemic has posed a monumental threat to the physical and mental well-being of the international community, and college students are especially vulnerable to elevated rates of anxiety, depression, and stress during moments of heightened uncertainties [11,12]. Moreover, self-isolation, stay-in-shelter lockdowns, and closures of college campuses have been enforced at times in most countries as an effort to contain the spread of the coronavirus. College students under lockdown reported spikes in stress, anxiety, and depressive symptoms as well as a lack of motivation, restless sleep, and loneliness [13,14]. In China, varying degrees of lockdowns have been implemented, ranging from neighborhood units to cities and whole provinces, at different times based on the severity of local spread of infections as a fundamental national strategy toward a zero-COVID policy [15]. Lockdowns have been found to exacerbate devastating psychological consequences and complicate interventions and coping at both the communal and individual levels [16].

Meanwhile, video games have been a leading sector of global growth and innovation in recent decades and have become a central part of people's cultural lives [17]. Besides their entertainment value, games have been associated with a variety of positive and negative impacts on individual cognitive, motivational, emotional, social, and other skills. For example, systematic reviews and meta-analytic studies of existing research have found that exergames work as effectively as conventional exercise in reducing anxiety [18], and commercial games show promise to ameliorate the symptoms of a variety of mental conditions and may produce positive effects on players' overall mental health [19]. On the opposing end, a meta-analysis of 101 studies focusing on children and adolescents found that playing video games is related to increased aggression and antisocial behavior as well as reduced academic performance [20]. With regard to mental well-being, one longitudinal study over a 3-week period in 7 English-speaking countries produced little to no evidence on the impact of time spent playing games on adult players' well-being [21], whereas a natural experiment involving Japanese console players extending over a 2-year period found a positive causal path linking gaming and mental well-being [22]. In particular, the interaction among individual, environmental, and gaming-related factors may result in problematic behavior often classified as gaming disorder [23], for which the prevalence rate was found to be higher in East Asian countries than other regions in the world [24].

Video games can provide ways of coping during difficult lifetime experiences and unusual times of personal difficulty [25]. The COVID-19 pandemic, especially during shelter-in-place lockdowns, created a circumstance of uncertainty and disruption. A systematic review of research spanning diverse demographics across different national settings during the COVID-19 pandemic or living through lockdowns indicated that gaming had a beneficial outcome in diminishing stress and anxiety in general but it had a detrimental effect on certain at-risk individuals [26]. Moreover, limited research evidence with people caught in lockdowns revealed that playing video games provided a compensatory response to the stressors of the COVID-19 pandemic and served as a viable coping mechanism as a short-term distraction [27,28]. However, the broad-strokes nature of the wording in the survey questions [27] and the small number of participants [28] in these studies left many important questions unanswered.

Therapeutic Use of Video Games

The digital revolution spearheaded by the smartphone, information networks, and personal communication technologies opens up unprecedented opportunities to tackle both the size and scale of the global mental health crisis, ranging from self-monitoring and self-management to online health interventions [29]. As video games become a ubiquitous part of global popular culture and mass entertainment, mounting attention has been paid by researchers and practitioners alike to the various roles video games play in the care of mental health for people (especially youth) with certain conditions in clinical and other settings, such as the therapeutic treatment of children and adolescents with attention-deficit/hyperactivity disorder,

XSL•FO
RenderX

autism, anxiety disorders, and older adults with cognitive impairment [30-32]. As a result, gaming for mental well-being has been a rather well-trodden area of academic research and practical explorations [33,34].

The pervasiveness and relatively low cost of video games make them a primary candidate for consideration and adoption in clinical use. Emerging (albeit limited) evidence-based studies have shown that game-based interventions are useful and effective in treating patients with depressive disorders [35-37], and playing games has shown promise in leading to a clinically measurable reduction in anxiety disorder symptoms [38-40]. On the other hand, a systematic review of existing research shows that there is not much conclusive evidence thus far demonstrating the effectiveness or cost-effectiveness of digital mental health interventions, including those using video games, in the treatment of anxiety and depression disorders [41,42]. This is clearly an area that calls for more rigorous scrutiny by the academic community. It is worth highlighting that most of the research efforts examining the mental health implications of using video games as an intervention have been focused on serious games and games specifically designed to meet the needs of certain psychological and mental conditions, and these were not commercially available to the public [36,37,43-45]. However, there is growing realization that off-the-shelf games, thanks to their immense popularity and pervasive reach, have untapped potential for preventive and therapeutic applications [46,47]. Meanwhile, research investigating the applicability and efficacy of commercial games for diverse therapeutic use in recent years has gained increasing traction [48-52]. This is especially relevant during public health crisis situations such as the COVID-19 pandemic, when gaming technology assumes multiple functions in reconfiguring people's lives [28,53] and evidence-based assessment is pivotal for understanding the potential roles of video games as coping mechanisms for diverse populations and demographics.

Research Questions and Objectives

A review of recent studies on the therapeutic use of video games revealed a striking dearth of research aimed at the college-age cohort who not only are susceptible to depression and anxiety disorders but also are the most likely consumers of video games [40]. The purpose of this research was to examine the interplay of general-purpose video game play (ie, commercial games) and the mental health risks of anxiety and depression symptoms during periods of stringent shelter-in-place lockdowns among a national sample of college students in China. Lockdowns marked moments of drastic disruption of daily routines and provided an unusual opportunity to study improvised coping through readily available video games, and in this case, the role of video games in moderating anxiety and depression symptoms among college students. Specifically, our research questions were posited to investigate the association between several explanatory variables such as sex, year in college, platform (online PC vs mobile), amount of gameplay, modality of play, type of games, and the outcome variables of the severity of depressive and generalized anxiety disorder symptoms in college students living under conditions of shelter-in-place lockdowns in China. Moreover, the research also sought to test if there was significant difference in symptoms of depressive and generalized anxiety disorders between college students who started to play games and those who did not use video games during lockdown periods.

Methods

Participants and Procedure

College students were recruited from 16 geographic regions encompassing 8 provincial areas (ie, Beijing, Shanghai, Fujian, Henan, Hebei, Hunan, Hubei, and Zhejiang) across China through the service of a professional survey company. Data collection took place in December 2021 and January 2022, during an extended period of elevated infections across multiple regions in China during the COVID-19 pandemic. Many local government authorities enforced stringent shelter-in-place mandates in residential areas with diagnosed cases of patients with COVID-19, with public transportations suspended and checkpoints added at neighborhood units to curb human movement. A purposive sample of currently registered college students from 36 campuses who were under lockdown mandates at the time were recruited to participate in the web-based survey. To exclusively target college students who were caught under concurrent shelter-in-place lockdown mandates, 3 layers of filter were adopted by the survey company: first, only regions under lockdown mandates were chosen; second, a confirmation was obtained from respective sampled universities that a lockdown was in place; third, students chosen to participate were asked to start their survey with answering the question "Are you currently under COVID-19 lockdowns," and only these who answered "yes" were directed to continue the survey, which was completed via Wenjuanxing [54], China's most popular online survey platform. Most of the participants completed the questions on their smartphones, while the rest finished on a PC. Participants were reached via mobile phone, WeChat (Tencent Holdings Limited) account, or email and were first briefed with the overall purpose of the study and were assured anonymity; filling out the survey was entirely voluntary, and students could terminate at any time. For the sake of comparison, data from both gamers and nongamers were collected, and 2818 valid responses (n=1396, 49.53% male participants vs n=1422, 50.46% female participants) were received after incomplete questionnaires were discarded.

The quantitative survey data were supplemented by in-depth interviews of 17 informants (9 female participants vs 8 male participants; 14 undergraduate vs 3 graduate students). For both the survey and in-depth interviews, question items were initially developed based on feedback from 2 informal web-based focus groups (6 and 7 participants each) we conducted with college students under lockdowns; the wording of these items was then distributed to a dozen college students for comments and suggestions. Interviewees were recruited from multiple campuses using a snowballing sampling approach through soliciting recommendations from friends of friends based on a predetermined set of demographic attributes. Only video game users from students who were living under conditions of shelter-in-place lockdowns during December 2021 and January 2022 were included. A set of 8 open-ended questions were followed in the interviews pertinent to the practices, rationale,

XSL•FO
RenderX

and perceived benefits or outcomes of video game engagement during shelter-in-place lockdowns (Multimedia Appendix 1). The interviews, which typically took 30 to 40 minutes, were completed in January 2022 via WeChat, China's most popular instant messaging social media app, through audio or text chatting. The qualitative data collection was terminated once our iterative processing of the interview data led us to conclude that saturation was reached [55]. The interviews were recorded and transcribed for subsequent analysis.

Ethical Considerations

The study protocol was approved by the Social Sciences Ethics Committee of Minjiang University (MJU.CJC.20210011). Informed consent was solicited electronically before the web-based survey, and participants were provided with detailed information regarding the research objectives, procedures, potential risks, and benefits of the study. Students aged <18 years were excluded from the research. At the beginning of the survey, each participant was asked to read and sign an informed consent form before answering the survey questions; a similar process was followed for the virtual interviews. Participants could opt out at any time, and survey takers who completed the questionnaire received ¥10 RMB (approximately US $1.40) cash payment at the end. All data collected were deidentified when being merged into a single data corpus.

Measures

Since our primary goal was to investigate the association between video game play and the scale of college students' anxiety and depression symptoms, we used the 7-item Generalized Anxiety Disorder (GAD-7) questionnaire for the screening of anxiety [56,57] and the self-administered 9-item Patient Health Questionnaire (PHQ-9) for the assessment of individual depression [58,59]. The questionnaires were initially developed in primary care and have now been widely adopted as reliable and valid measures in diagnosing depression and anxiety disorders, respectively, in clinical settings. The Chinese versions we used in the survey have been validated among diverse Chinese patients as well as the general population for their validity and reliability [60-62]. The instrument demonstrated high reliability (internal consistency) for both the GAD-7 and the PHQ-9 items in our dataset: Cronbach α was 0.953 for anxiety and 0.932 for depression. We followed the cutoff points recommended by clinical diagnosis in both the West [56,57,63] and among Chinese demographics [60-62] to categorize generalized anxiety disorder into minimal (score 0-4), mild (score 5-9), moderate (score 10-14), and severe (score 15-21) levels. Video game play was measured using a set of questions about whether and, if yes, how often survey participants played mobile or online games, whether they thought playing games was effective in reducing anxiety or depression, the modality of play (be it single person vs multiperson), daily hours of game time, and the type of games they played the most. Demographic variables included year in college and sex (Multimedia Appendix 2).

Statistical Analysis

Statistical analysis was performed using SPSS for Windows (version 29; IBM). Descriptive statistics were reported as percentages for nominal variables and as means and SDs for ordinal or ranking variables. Multiple linear regression analysis was run to determine whether demographic variables and gameplay patterns accounted for the outcome of anxiety and depression, respectively. Preliminary diagnostic analyses were performed to test the assumptions for multiple regression modeling using anxiety disorder and depressive symptoms as respective outcome variables. Low multilinearity and appropriate homoscedasticity were confirmed for each regression model. The criterion variable was the composite score of the GAD-7 items for the anxiety regression model and the cumulative score of the PHQ-9 items for depression. Before being included in the regression model, an omnibus test was performed on each explanatory variable using a 1-way ANOVA model on the respective outcomes of anxiety and depressive symptoms, with all tests approaching statistical significance at the $P<.05$ level. The explanatory (ie, independent) variables were year in college (dummy coded; first-year student=0), whether gameplay was perceived to be effective (dummy coded; female=0), how often one played mobile games, how often one played online games, how many hours one spent playing games every day, change in daily playtime (from nonlockdown to lockdown days), modality of play (dummy coded; play alone=0), and type of games one played the most (dummy coded; single player/console games=0). Following research findings that playing video games during the COVID-19 pandemic might have a detrimental impact on certain players [26], we were interested in examining if the amount of game time played an important moderating role and therefore conducted ANOVA tests to compare groupwise variations based on the severity of anxiety and depressive disorder symptoms among college students during lockdowns.

For the qualitative interview data, we followed the logic of grounded theory coding as explained by Charmaz [64] in a 2-step analytic procedure. In the initial phase of "open-ended coding," 2 researchers independently traversed the data corpus to pinpoint semantic categories and thematic patterns, which were then cross-examined, compared, and crystallized into coherent topic summaries. The second phase of "focused coding" allowed us to relate the interview data to meaningful gameplay experiences and salient themes. This structured approach aligns closely with the codebook thematic analysis paradigm highlighted by Clarke and Braun [65]. In the initial stage of data familiarization, the 3 researchers first worked independently and next worked together to reach consensus in the deductive development of coding frames; the subsequent inductive (ie, structured) mapping of the whole interview data corpus was analyzed by following the established coding scheme.

Results

Demographic Data and Descriptive Statistics

Of the 2818 survey respondents, there were 2669 (94.5%) undergraduate students, while there were 149 (5.3%) graduate and other students. Percentagewise, the first-year student cohorts dominated at 39.3% (1108/2818) of the total, followed by second-year student (811/2818, 28.8%), third-year student

XSL•FO
RenderX

(486/2818, 17.2%), and fourth-year student (264/2818, 9.4%) demographics. Participants were almost evenly distributed along sex lines (1422/2818, 49.5% male participants vs 1396/2818, 50.5% female participants). Most of the surveyed students (1894/2818, 67.2%) reported playing games amid the COVID-19 pandemic shelter-in-place lockdowns, and male students were much more likely than their female counterparts to opt to play games during lockdowns (1030/1422, 73.8% vs 864/1396, 60.8%; χ^2_1=54.21, P<.001). For those who played games during lockdowns, the typical tendency was to play mobile games (mean 2.89, SD 0.95) and online games (mean 2.58, SD 1.02) rather frequently. For the sake of disambiguation, mobile games in China refer to those games that are typically played on smartphones, and online games are those that are played on a computer or handheld device. On average, student gamers played 2 to 3 (mean 2.62, SD 1.36) hours per day, and the amount of game time noticeably increased during lockdown moments than on nonlockdown days (mean 3.70, SD 0.84 hours). It should be acknowledged here that our question concerning increase in time included both mobile and PC online games in a single measurement, and it did not make separate inquiries. As for modality, most students played with their friends, colleagues, and other people they knew (1241/1894, 65.5%), and the types of games they played the most were massively multiplayer online games (674/1894, 35.6%); combat, competitive, or sports games (517/1894, 27.3%); and casual, chess, or card games (425/1894, 22.4%). Detailed statistics on the survey responses are outlined in Table 1.

XSL•FO
RenderX

Table 1. Participant characteristics.

Variables	Played games	
	Yes (n=1894)	No (n=924)
Year in college, n (%)		
1	781 (41.2)	327 (35.4)
2	576 (30.4)	235 (25.4)
3	285 (15)	201 (21.8)
4	165 (8.7)	99 (10.7)
Graduate student and other	87 (4.6)	62 (6.7)
Sex, n (%)		
Male	1030 (54.4)	366 (39.6)
Female	864 (45.6)	558 (60.4)
Symptoms, mean (SD)		
Anxiety (GAD-7[a]) composite score (0-21)	4.47 (4.38)	4.87 (4.26)
Depression (PHQ-9[b]) composite score (0-27)	7 (5.46)	7.11 (5.25)
During the lockdowns, do you feel like playing video games is effective in reducing anxiety and depression? n (%)		
Yes	1047 (55.3)	N/A[c]
No	847 (44.7)	N/A
During the lockdowns, how often do you play mobile games? (1=rarely; 2=occasionally; 3=often; 4=everyday), mean (SD)	2.89 (0.95)	N/A
During the lockdowns, how often do you play online games? (1=rarely; 2=occasionally; 3=often; 4=everyday), mean (SD)	2.58 (1.02)	N/A
On an average day, approximately how many hours do you play games during the lockdowns? (1=0-1 h; 2=1-2 h; 3=2-3 h; 4=3-4 h; 5=4-5 h; 6=>5 h), mean (SD)	2.62 (1.36)	N/A
Compared with nonlockdown days, is there any change in the amount of time you spend playing games during the lockdowns? (1=much less; 2=somewhat less; 3=About the same; 4=somewhat more; 5=much more), mean (SD)	3.70 (0.84)	N/A
During the COVID-19 pandemic lockdowns, what was the most common modality of your gameplay? n (%)		
Play alone	556 (29.4)	N/A
Play with friends (people I know)	1241 (65.5)	N/A
Play with strangers	97 (5.1)	N/A
Type of games played the most during the lockdowns, n (%)		
Single-player or console games	187 (9.9)	N/A
Massively multiplayer online	674 (35.6)	N/A
Casual, chess, and card games	425 (22.4)	N/A
Combat, sports, or competitive games	517 (27.3)	N/A
Strategy games	91 (4.8)	N/A

[a]GAD-7: 7-item Generalized Anxiety Disorder.

[b]PHQ-9: 9-item Patient Health Questionnaire.

[c]N/A: not applicable.

Population-Level Differences Between Gamers and Nongamers

On average, survey respondents reported mild symptoms of anxiety (mean 4.60, SD 4.34) and depression (mean 7.03, SD 5.39), although there was a wide-ranging variance for both, as reflected in the SDs. Among the 1125 students who acknowledged playing games, 622 (55.3%) said doing this was effective in helping them deal with anxiety and depression. At the aggregate level, because a higher proportion of male students opted to play games than their female counterparts (as reported in the Demographic Data and Descriptive Statistics section), we ran a univariate, general linear model procedure using gameplay (responses=yes vs no) as the independent variable and anxiety and depression symptom scores as the respective dependent variable, with sex as the covariate to control for its

XSL•FO
RenderX

potential cofounding effect. A significant video game effect was established between gamers and nongamers with regard to their overall anxiety disorder symptoms (mean 4.47, SD 4.38 vs mean 4.88, SD 4.27; $F_{1,2}$=6.21; *P*=.01; η^2=0.003) but not their depressive symptoms (mean 7, SD 5.46 vs mean 7.11, SD 5.25; $F_{1,2}$=0.11; *P*=.75; η^2=0.000).

Video Game Play and Anxiety

Table 2 reports the coefficients of explanatory variables on the composite anxiety score. Being a male or female gamer did not play a significant role in accounting for whether gaming was impactful in mitigating anxiety symptoms. Compared with a first-year student, a second-year student (β=0.05, *P*=.04), fourth-year student (β=0.12, *P*<.001), or graduate or other student (β=0.06, *P*=.02) playing video games was more likely to experience elevated anxiety during lockdowns. Playing more online games (β=0.07, *P*=.03) and spending more time gaming in terms of the total daily hours (β=0.08, *P*=.005) and increase in time (β=0.06, *P*=.045) during lockdowns were associated with heightened anxiety symptoms. Compared with gaming alone, playing more casual, chess, or card games (β=0.08, *P*=.04) was linked to higher anxiety. On the other hand, playing with friends (β=−0.05, *P*=.04) was significantly linked to alleviated anxiety scores, and self-perceived efficacy of gameplay (β=−0.10, *P*<.001) contributed significantly to a lowered level of anxiety.

XSL•FO
RenderX

Table 2. Multiple linear regression coefficients toward anxiety disorder symptoms (N=1894).

Demographics and game behaviors	Unstandardized coefficients B (SE)	Standardized β coefficients (95% CI)	*t* ratio	*P* value
Constant	1.47 (0.56)	—[a] (0.37-2.57)	2.61	.009
Sex[b] (male)	−0.16 (0.21)	−0.02 (−0.58 to 0.26)	−0.77	.44
Year in college[c]				
2	0.49 (0.24)	0.05 (0.03-0.95)	2.07	.04
3	0.53 (0.30)	0.04 (−0.06 to 1.12)	1.78	.08
4	1.88 (0.37)	0.12 (1.16-2.60)	5.09	<.001
Graduate and other	1.19 (0.49)	0.06 (0.235-2.15)	2.44	.02
EF[d,e] (yes)	−0.85 (0.22)	−0.10 (−1.27 to −0.42)	−3.90	<.001
MG[f] (rating 1-4)	0.22 (0.13)	0.05 (−0.30 to 0.46)	1.72	.09
OG[g] (rating 1-4)	0.28 (0.13)	0.07 (0.03-0.53)	2.23	.03
HR[h] (rating 1-6)	0.26 (0.09)	0.08 (0.08-0.44)	2.79	.005
CH[i] (rating 1-5)	0.26 (0.14)	0.06 (−0.04 to 0.52)	1.85	.04
MD[j,k]				
With friends	−0.49 (0.23)	−0.05 (−0.94 to −0.02)	−2.08	.04
With strangers	−0.60 (0.48)	−0.03 (−1.54 to 0.34)	−1.25	.21
TE[l,m]				
Massively multiplayer online	0.47 (0.37)	0.05 (−0.25 to 1.19)	1.29	.20
Casual, chess, or card games	0.79 (0.38)	0.08 (0.05-1.54)	2.10	.04
Combat or sports	0.42 (0.38)	0.04 (−0.32 to 1.15)	1.10	.30
Strategy	0.63 (0.55)	0.03 (−0.45 to 1.72)	1.14	.25

[a]Not available.

[b]Dummy coding; female=0.

[c]Dummy coding; first-year student=0.

[d]EF: during the lockdowns, do you feel like playing video games is effective in reducing anxiety and depression?

[e]Dummy coding; effective (no)=0.

[f]MG: during the lockdowns, how often do you play mobile games?

[g]OG: during the lockdowns, how often do you play online games?

[h]HR: on an average day, approximately how many hours do you play games during the lockdowns?

[i]CH: compared with nonlockdown days, is there any change in the amount of time you spend playing games during the lockdowns?

[j]MD: during the COVID-19 pandemic lockdowns, what was the most common modality of your gameplay?

[k]Dummy coding; play alone=0.

[l]TE: type of games played the most during the lockdowns.

[m]Dummy coding: single-player or console games=0.

Video Game Play and Depression

Regarding depressive disorder symptoms (Table 3), no significant sex effect was detected, whereas being a second-year student (vs a first-year student; β=0.07, *P*=.003), third-year student (β=0.05, *P*=.03), or fourth-year student (β=0.09, *P*<.001) increased the likelihood of feeling depressed. Similar to the anxiety regression model, increased online gameplay (β=0.07, *P*=.02) and more playtime as measured in the total daily hours (β=0.07, *P*=.01) and surge in time (β=0.06, *P*=.02) were significantly related to increases in depression levels. Conversely, social play with friends (β=−0.08, *P*=.003) was a significant contributor to reduced depressive disorder. As with the anxiety model, self-perception of video game efficacy (β=−0.08, *P*<.001) significantly accounted for mitigated depressive symptoms.

XSL•FO
RenderX

Table 3. Multiple linear regression coefficients toward depression disorder symptoms (N=1894).

Demographics and game behaviors	Unstandardized coefficients, B (SE)	Standardized β coefficients (95% CI)	t-ratio	*P* value
Constant	3.43 (0.70)	—[a] (2.06-4.81)	4.90	<.001
Sex[b] (male)	−0.41 (0.27)	−0.04 (−0.93 to 0.11)	−1.55	.12
Year in college[c]				
2	0.89 (0.30)	0.07 (0.31-1.47)	3.01	.003
3	0.79 (0.37)	0.05 (0.06-1.52)	2.12	.03
4	1.75 (0.46)	0.09 (0.84-2.65)	3.78	<.001
Graduate and other	0.60 (0.61)	0.02 (−0.60-1.79)	0.97	.33
EF[d,e] (yes)	−0.92 (0.27)	−0.08 (−1.46 to −0.39)	−3.41	<.001
MG[f] (1-4)	0.28 (0.16)	0.05 (−0.02 to 0.59)	1.81	.07
OG[g] (1-4)	0.36 (0.16)	0.07 (0.05-0.67)	2.28	.02
HR[h] (1-6)	0.30 (0.12)	0.07 (0.07-0.52)	2.54	.01
CH[i] (1-5)	0.41 (0.18)	0.06 (0.07-0.76)	2.33	.02
MD[j,k]				
With friends	−0.87 (0.29)	−0.08 (−1.44 to −0.30)	−2.99	.003
With strangers	−0.39 (0.60)	−0.02 (−1.56 to 0.79)	−0.64	.52
TE[l,m]				
Massively multiplayer online	0.16 (0.46)	0.01 (−0.74 to 1.06)	0.34	.73
Casual, chess, or card games	0.55 (0.47)	0.04 (−0.38 to 1.48)	1.17	.24
Combat or sports	0.25 (0.47)	0.02 (−0.68 to 1.17)	0.52	.60
Strategy	0.35 (0.69)	0.01 (−1.01 to 1.71)	0.51	.61

[a]Not available.

[b]Dummy coding; female=0.

[c]Dummy coding; first-year student=0.

[d]EF: during the lockdowns, do you feel like playing video games is effective in reducing anxiety and depression?

[e]Dummy coding; effective (no)=0.

[f]MG: during the lockdowns, how often do you play mobile games?

[g]OG: during the lockdowns, how often do you play online games?

[h]HR: on an average day, approximately how many hours do you play games during the lockdowns?

[i]CH: compared with nonlockdown days, is there any change in the amount of time you spend playing games during the lockdowns?

[j]MD: during the COVID-19 pandemic lockdowns, what was the most common modality of your gameplay?

[k]Dummy coding; play alone=0.

[l]TE: type of games played the most during the lockdowns.

[m]Dummy coding: single-player or console games=0.

Association Between Video Game Play and Severity of Anxiety and Depression

In terms of severity levels in gamer versus nongamer (n=924) distribution, the prevalence rate was 51.53% (n=976) versus 48.38% (n=447) for the minimal level, 38.17% (n=723) versus 39.83% (n=368) for the mild level, 8.39% (n=159) versus 9.85% (n=91) for the moderate level, and 1.9% versus 1.95% (n=18) for the severe level for anxiety symptoms. Similarly, following standard clinical practice [58,59,66], total scores of <5, 5, 10, 15, and 20 were used as cutoff points for the minimal, mild, moderate, moderately severe, and severe levels of depression, respectively. Comparing gamers with nongamers (n=924), the prevalence rate for depressive symptoms was 35.01% (n=663) versus 33.98% (n=314) for the minimal level, 41.55% (n=787) versus 43.40% (n=401) for the mild level, 13.41% (n=254) versus 14.50% (n=134) for the moderate level, 7.39% (n=140) versus 5.52% (n=51) for the moderately severe level, and 2.64% (n=50) versus 2.6% (n=24) for the severe level.

As the number of daily hours of gametime and the degree of change in playtime from before the lockdown to during the lockdown in the COVID-19 pandemic were 2 prominent intervening factors influencing video game use on one hand and both anxiety and depressive disorders on the other hand,

XSL•FO
RenderX

we further analyzed the respective role of these 2 factors in modulating the levels of anxiety and depressive symptoms.

A 1-way ANOVA test using daily playtime as the dependent variable and anxiety severity as the grouping variable indicated significant groupwise differences ($F_{3,1890}$=14.10; P<.001; η^2=0.022; 95% CI 0.011-0.036). As outlined in Table 4, students with severe anxiety and moderate anxiety departed significantly from all other cohorts in their daily playtime, with pairwise 2-tailed t tests displaying statistical significance at the P<.01 level. In terms of prelockdown to lockdown playtime change during the COVID-19 pandemic, it can be seen from Table 5 that students with severe anxiety symptoms differed significantly from all other groups ($F_{3,1890}$=9.29; P<.001; η^2=0.016; 95% CI 0.006-0.028). In other words, students experiencing severe anxiety symptoms also registered the biggest increase in daily gametime (P<.001 against all other cohorts).

Table 4. Pairwise comparisons and Cohen d values for the association between the amount of daily playtime[a] and anxiety severity[b].

	Mean difference (95% CI)	Cohen d	t test[c] (df)	P value
Severe anxiety versus...				
Moderate anxiety	0.78 (0.24-1.32)	0.53	2.87 (193)	.005
Mild anxiety	1.21 (0.76-1.66)	0.90	5.26 (757)	<.001
Minimal anxiety	1.22 (0.78-1.67)	0.91	5.36 (1010)	<.001
Moderate anxiety versus...				
Mild anxiety	0.43 (0.19-0.66)	0.32	3.61 (880)	<.001
Minimal anxiety	0.44 (0.21-0.67)	0.33	3.82 (1133)	<.001
Mild anxiety versus minimal anxiety	0.01 (−0.11 to 0.14)	0.01	0.21 (1697)	.83

[a]Daily playtime is measured using a 1 to 6 scale. Table 1 provides further details (referenced as HR).

[b]Descriptive statistics of daily playtime amount per anxiety severity levels: severe anxiety (n=36; mean 3.78, SD 1.66); moderate anxiety (n=159; mean 2.99, SD 1.44); mild anxiety (n=723; mean 2.57, SD 1.33); and minimal anxiety (n=976; mean 2.55, SD 1.33).

[c]All t tests were 2-tailed.

Table 5. Pairwise comparisons and Cohen d values for the association between the change in daily playtime[a] and anxiety severity[b].

	Mean difference (95% CI)	Cohen d	t test[c] (df)	P value
Severe anxiety versus...				
Moderate anxiety	0.59 (0.27 to 0.92)	0.67	3.62 (193)	<.001
Mild anxiety	0.63 (0.36 to 0.90)	0.79	4.61 (757)	<.001
Minimal anxiety	0.71 (0.43 to 0.99)	0.85	5.01 (1010)	<.001
Moderate anxiety versus...				
Mild anxiety	0.04 (−0.10 to 0.18)	0.05	0.53 (880)	.60
Minimal anxiety	0.12 (−0.03 to 0.26)	0.14	1.59 (1133)	.11
Mild anxiety versus minimal anxiety	0.08 (−0.002 to 0.16)	0.09	1.91 (1697)	.06

[a]Change in daily playtime is measured using a 1 to 5 scale. Table 1 provides further details (referenced as CH).

[b]Descriptive statistics of change in daily playtime per anxiety severity levels: severe anxiety (n=36; mean 4.36, SD 0.72); moderate anxiety (n=159; mean 3.77, SD 0.92); mild anxiety (n=723; mean 3.73, SD 0.81); and minimal anxiety (n=976; mean 3.66; SD 0.84).

[c]All t tests were 2-tailed.

A 1-way ANOVA test identified a significant effect of daily game time on the level of depression severity ($F_{4,1889}$=9.79; P<.001; η^2=0.021; 95% CI 0.009-0.035). Post hoc pairwise tests pinpointed the severe depressive cohort as deviating from all other groups (P<.01), as reported in Table 6. While the 1-way ANOVA omnibus test showed overall significance with regard to change in playtime ($F_{4,1889}$=7.03; P<.001; η^2=0.016; 95% CI 0.006-0.026), post hoc groupwise comparisons again pointed to the severe depressive group being consistently distanced from all other groups (P<.001; Table 7).

XSL•FO
RenderX

Table 6. Pairwise comparisons and Cohen d values for the association between the amount of daily playtime[a] and depression severity[b].

	Mean difference (95% CI)	Cohen *d*	*t* test[c] (*df*)	*P* value
Severe depression versus...				
Moderately severe depression	0.68 (0.17-1.19)	0.44	2.64 (188)	.009
Moderate depression	0.89 (0.46-1.31)	0.64	4.11 (302)	<.001
Mild depression	0.98 (0.60-1.36)	0.73	5.04 (835)	<.001
Minimal depression	1.10 (0.70-1.49)	0.80	5.43 (711)	<.001
Moderately severe depression versus...				
Moderate depression	0.21 (−0.08 to 0.49)	0.15	1.43 (392)	.16
Mild depression	0.30 (0.06 to 0.54)	0.23	2.46 (925)	.01
Minimal depression	0.42 (0.17 to 0.67)	0.30	3.27 (801)	.001
Moderate depression versus...				
Mild depression	0.09 (−0.09 to 0.28)	0.07	1.00 (1039)	.32
Minimal anxiety	0.21 (0.02-0.40)	0.16	2.13 (915)	.03
Mild depression versus minimal depression	0.12 (−0.02 to 0.25)	0.09	1.66 (1448)	.10

[a]Daily playtime is measured using a 1 to 6 scale. Table 1 provides further details (referenced as HR).

[b]Descriptive statistics of daily playtime amount per depression severity levels: severe depression (n=50; mean 3.58, SD 1.75); moderately severe depression (n=140; mean 2.90, SD 1.49); moderate depression (n=254; mean 2.69, SD 1.32); mild depression (n=787; mean 2.60, SD 1.31); and minimal depression (n=663; mean 2.48, SD 1.35).

[c]All *t* tests were 2-tailed.

Table 7. Pairwise comparisons and Cohen d for the association between change in daily playtime[a] and depression severity[b].

	Mean difference (95% CI)	Cohen *d*	*t* test[c] (*df*)	*P* value
Severe depression versus...				
Moderately severe depression	0.38 (0.10-0.66)	0.44	2.70 (188)	.008
Moderate depression	0.42 (0.16-0.67)	0.50	3.25 (302)	.001
Mild depression	0.51 (0.28-0.75)	0.62	4.25 (835)	<.001
Minimal depression	0.57 (0.32-0.81)	0.67	4.58 (711)	<.001
Moderately severe depression versus...				
Moderate depression	0.04 (−0.13 to 0.21)	0.05	0.44 (392)	.66
Mild depression	0.13 (−0.01 to 0.28)	0.16	1.77 (925)	.08
Minimal depression	0.19 (0.04-0.34)	0.23	2.42 (801)	.02
Moderate depression versus...				
Mild depression	0.10 (−0.02 to 0.21)	0.12	1.62 (1039)	.11
Minimal anxiety	0.15 (0.03-0.27)	0.18	2.46 (915)	.01
Mild depression versus minimal depression	0.06 (−0.03 to 0.14)	0.07	1.26 (1448)	.21

[a]Change in daily playtime is measured using a 1 to 5 scale. Table 1 provides further details (referenced as CH).

[b]Descriptive statistics of change in daily playtime per depression severity levels: severe depression (n=50; mean 4.20; SD=0.88); moderately severe depression (n=140; M=3.82; SD 0.84); moderate depression (n=254; mean 3.78, SD 0.82); mild depression (n=787; mean 3.69, SD 0.82); minimal depression (n=663; mean 3.63, SD 0.84).

[c]All *t* tests were 2-tailed.

Thematic Patterns of Qualitative Interview Data

An analysis of our in-depth interview data revealed several salient themes in the context of game use under COVID-19 pandemic lockdowns. First, the interviewees consistently pointed out a few circumstantial factors leading them to gravitate toward more gameplay during shelter-in-place moments: excessive exposure to the computer monitor and other digital devices where video games are easily available (15/17, 88%); suddenly gained extra hours that could lead to boredom if not occupied (13/17, 76%); escapism found in video games that could offer a temporary mental release from the grim realities of the

XSL•FO
RenderX

COVID-19 pandemic and the lockdown (12/17, 71%); and isolation due to blockage of mobility that boosted the urge for sociality (11/17, 65%). Besides its entertainment value, gaming was cited by about one-third (7/17, 41%) of the interviewees as a preferred way to get together with friends and colleagues to chat while playing. Playing with family members was mentioned by some (5/17, 29%) to be an effective means in creating distractions from the lockdowns and in collectively coping with the challenging living conditions through bonding and emotional support. This overall sentiment is aptly illustrated by a quote from a graduate student interviewee:

> *Playing videogames gives me the opportunity to hook up with my good buddies that I otherwise would not be able to do during normal days when everybody is busy. I know my friends also play games, and we would first talk to one another on social media and find a common time. Once we are in the game, we turn the audio chat feature on and talk during the whole gameplay. It is a great way to catch up with one another and release myself emotionally.* [Participant A, female participant, second year of doctoral studies]

Besides socializing with friends, video game was also mentioned by a few interviewees as serving the role of a virtual companion, as imparted in this quote from a male third-year student participant:

> *Having the game on is a way to have a company with me, even when I am doing other things by myself...I use the auto-play calling feature to let NBA2K play on my computer. It is like a spiritual pal sitting by me.* [Participant B, male participant, third year of doctoral studies]

While a slight majority of the informants (10/17, 59%) indicated playing games helped them in handling their anxiety and pressure, others acknowledged its complicated nature as a way of coping with mental well-being. Some interviewees mentioned the transient nature of the euphoric feelings they obtained in the game-on environment, as demonstrated from this quote:

> *At times of boredom, playing games fills the void and brings joy to me. But after the gameplay, my downcast feeling kicks back in when the reality of COVID-19 and the lockdown returns to mind.* [Participant C, female participant, third-year student undergraduate student]

Another third-year college student described her game experience as bringing her "fleeting indulgence but durable angst" while living in the shadows of the lockdown. Moreover, 3 interviewees acknowledged that their bad mood during lockdowns often spilled over to gameplay, resulting in verbal confrontations, cursing, and other conflicts with other players. Playing under these conditions sometimes could further exacerbate their mental state. Out of 17 interviewees, 4 (24%) highlighted the double-edged nature of gaming during lockdowns: always being close to the computer, coupled with excessive leisure time, created the hard-to-resist temptation to enter the game world and stay there more than they wanted. In those cases, the joy of gaming was acknowledged to be transient, but this habitual behavior could eventually elevate one's anxiety and depressive symptoms.

Discussion

Principal Findings

First, we would like to foreground a discussion of the particularity of the environmental setting of the COVID-19 pandemic and its lockdowns, which is pivotal for extrapolating our findings to other contexts. The prolonged COVID-19 pandemic disrupted routine habits and behaviors and has taken a profound toll on mental well-being. Stay-at-home lockdowns were adopted in most countries in the early stages of the COVID-19 pandemic as a measure to hinder the spread of the virus. However, China's zero-COVID policy and its enforcement differentiated the Chinese shelter-in-place lockdowns from those in the rest of the world due to the extent of closures and reduction in human mobility and contact. The draconian nature of the COVID-19 pandemic lockdown measures triggered widespread protests (aka the "A4 Blank Paper Movement") in 21 provinces and over 200 college campuses in late 2022, which led to the end of the regime's zero-COVID policy [67]. It is useful to take the environmental circumstances of the Chinese lockdown into consideration while evaluating the findings and the analysis.

Research in cross-national settings has found that video games are helpful in reducing stress, anxiety, and depression among college students and other demographics during the stay-at-home period following the COVID-19 outbreak [26,28,53]. Findings in this research showed that, in comparison with nonplayers, video game players were linked to overall lower anxiety symptoms but not depressive disorder symptoms at the level of the study population, controlling for the effect of sex. Because our research design was cross-sectional and correlational in nature, we cannot draw conclusions on causation concerning the efficacy of video games based on the regression models. Nonetheless, ancillary evidence in support of the benefits of video games in mitigating anxiety and depression came from the survey responses directed to the question asking the informants whether playing video games produced such effects; 55.3% of those who played games answered in the affirmative. The same pattern was corroborated in the in-depth interviews. However, it must be pointed out that several pivotal individual and environmental factors that were absent from this research are needed for a full understanding of the behavioral tendencies and their associated mental health outcomes. For example, physical exercise, social connections and support, living conditions, and previous gaming experience could play an important moderating role in the process. In this regard, control groups and longitudinal data would be most valuable in dissecting the specific role video games play in coping with anxiety and depression under stringent life moments.

A sizable proportion of the surveyed students did not perceive video games to be effective in coping with anxiety and depression. This probably comes as no surprise, because gaming-related behavior could result in counterproductive and detrimental consequences in a certain segment of players, especially problematic ones, at moments of uncertainty and

XSL•FO
RenderX

disruption such as the COVID-19 pandemic and in the face of drastic measures of the lockdowns [26]. Longitudinal data show that avoidant coping (tendency to retreat or divert attention from the pandemic) and excessive gaming time (which happened at the expense of social networking time and perceived family support) were associated with greater distress during the extended COVID-19 pandemic stay-at-home period [68]. Our study reveals that college students who spent an extended amount of time playing video games were associated with severe levels of anxiety and depression, and students in the severe anxiety and depression cohorts were also the most likely to increase gaming time during shelter-in-place lockdowns. This pattern is highly congruent with findings in the extant research literature in this area [26]. This suggests that certain at-risk individuals who are susceptible to gaming disorders should be closely monitored at difficult times, as coping with video games has the potential to aggravate, not mitigate, the mental health state.

The role of video games goes beyond a mere entertainment product, as playing with one's social circle of friends, colleagues, and other close connections facilitates casual catch-ups and informal chitchats, which in turn would aid in dealing with personal anxiety and depression challenges. Self-determination and modality of play are important factors in determining outcomes [27,69]. Findings presented in this research show that social playing (with family and friends) was significantly correlated with both reduced generalized anxiety and depressive disorder symptoms. This is corroborated with evidence from the in-depth interviews, as many participants indicated that simply connecting with friends and intimate circles in the game sphere had the effects of helping them feel connected and distracted from on-hand difficulties and distress. Loneliness, which was exacerbated by social isolation, was found to be strongly linked to anxiety and depressive symptoms during the COVID-19 outbreak [70]. Meanwhile, social motivations for gaming could potentially have the benefits of reducing loneliness, but they could accentuate gaming disorders and therefore negatively impact mental health [71]. However, the lack of survey tools in measuring diverse motivational factors from this study limits our perspectives in further understanding these dynamics. Nonetheless, the complex relationships and interplays of these pivotal factors will be worthy of further exploration under diverse circumstances.

Although the questionnaire used in this study did not measure internet gaming disorder, an accumulated body of empirical evidence indicates that the primary predictor of this disorder rests with excessive game time, and gaming disorder shows a high comorbidity with depressive and anxiety disorders [72,73]. On the basis of triangulating evidence from diverse settings, there is reason to speculate that the buffering effects of videogaming may be the most asserting among routine players with moderate to low anxiety and depression symptoms, while this impact is compounded by disorderly gaming behaviors and may incur adverse consequences on those who are already struggling with high anxiety and depression [26,49,69,72,73]. Compared with other types of players, only problematic gamers among Chinese adolescents showed a positive association with anxiety severity [74]. Other research has indicated that students experiencing life stressors display fewer depressive symptoms while using adaptive coping strategies compared with maladaptive ones [75-77]. The divergent paths of adaptive (ie, proactive) and maladaptive (ie, reactive or avoidant) coping styles including the use of video games toward outcomes of alleviating anxiety and depression should warrant further scrutiny from researchers and practitioners alike. This points to the important role of intentionality in modulating the impact of video game adoption during times of unforeseen, heightened mental distress and uneasiness, which may offer useful guidance for college counselors and advisers as well as parents and college students themselves in favor of adaptive (ie, proactive) rather than maladaptive (ie, reactive) strategies in using video games to cope with excruciating and unanticipated stressors. Moreover, the comorbid condition of gaming disorder with anxiety and depression disorder symptoms among certain at-risk individuals remains a vital area of empirical scrutiny.

Promising Directions for Future Research

Gameplay modality and game content are important moderating factors to contemplate in examining the role of gaming as an intervention for anxiety and depression. The socializing aspects of video games have been claimed to boast great potential in accruing social support and rapport and thus boost well-being [33,78], with limited support from emerging research on video game use during the COVID-19 pandemic [26]. Our research pinpointed a prominent association of social playing (ie, with friends, colleagues, and people one knows) with lowered anxiety and depressive symptoms, corroborated in both the quantitative survey and the qualitative interviews. Systematic and scoping reviews of extant research investigating the effects of video games on anxiety and depression have consistently detected a surprising paucity of focus on this particular (ie, socializing) aspects of gameplay [30,31,34,37,39,41,43,48,50,69], and this is a fruitful area of future research for both theory testing and practical guidance. Regarding game content, evidence is inconclusive from our survey to interpret the role of engagement with specific game types. This may be more of an issue of a lack of precise or appropriate classification and typology in our study and calls for improvement in future research.

Limitations

Several limitations should be acknowledged surrounding the design and accomplishment of this research. The use of self-reported data might involve inaccuracies in important measurements, which formed the basis of our analysis. Selection bias in both the survey and in-depth interview participants could have tainted the results in certain directions. The survey tools only focused on some aspects and did not cover many potentially influential factors at the individual, social, and ecological levels that could have played essential roles in determining video game effects. In addition, preexisting mental health conditions and previous habitual video game experience may be pivotal in moderating engagement outcomes at moments of public health crises such as the COVID-19 outbreak. Therefore, diverse and expanded assessment tools on pivotal parameters from symptom diagnosis to individual traits and behavioral tendencies are desired. The single-shot, cross-sectional, and correlational design creates challenges for causal inferences and generalizations.

XSL•FO
RenderX

There was a high level of fluidity and unpredictability throughout the COVID-19 outbreak, and lockdown policies and measures showed intense variations from region to region as well as from country to country. Constrained by the particular timing of the survey, the results reported in the study could only shed light on the short-term impact and would not be able to address long-term ramifications. The survey was conducted at a particular moment in China's zero-COVID crusade across multiple geographic regions, which might have introduced peculiar environmental conditions and unusual gameplay dynamics hard to account for in the data and the analysis. As a result, sequential or causal inferences cannot be drawn between video game use and mental health outcomes as related to anxiety and depression among the respondents; nor can the findings be easily generalizable to the large population. Nonetheless, the findings in this study offer much-needed insight on empirical evidence on the interplay of different modalities of video game play and various manifestations of generalized anxiety and depressive disorders at an unusual public health crisis moment.

Conclusions

This study investigated college students' video game play during the draconian COVID-19 shelter-in-place lockdowns in China. Overall, compared with nonuse, playing video games during lockdowns was linked to significantly lowered anxiety symptoms but not sizable differences in depressive disorder symptoms. There were nuanced variations at the individual level with regard to those who resorted to playing video games as a way to deal with mental well-being. First, it is worth highlighting that the self-perceived efficacy of video games in mitigating the downsides of anxiety and depressive disorders served as a significant predictor for lowered symptoms. Playing video games did not affect college students at all stages, as second-year students and fourth-year students were much more likely to have both anxiety and depression than their counterparts, when game playing was held constant. Compared with mobile games, online games were more effective at alleviating anxiety and depressive symptoms. Playing in moderation (ie, for a modest amount of time) was found to be positively associated with reduced anxiety and depressive symptoms. Conversely, excessive engagement with gaming and a dramatic increase in playtime during lockdowns were linked to high levels of anxiety and depressive disorders. Social play with friends and family members (but not strangers) was positively correlated with alleviated anxiety and depression scores. These findings could offer actionable and pragmatic tips for guidance with college counselors and mental health professionals in their effort to help students cope with turmoil and disruption.

Acknowledgments

The authors are grateful to all the study participants for completing the web-based survey and the in-depth interviews during the COVID-19 pandemic lockdowns.

Data Availability

The datasets generated and analyzed in this study are available from the corresponding author upon reasonable request.

Authors' Contributions

All authors contributed to the conceptualization and design of the web-based survey. FH and ZT designed the in-depth interviews and FH completed the in-depth interviews and conducted an analysis of the interview data. FH and JL coordinated and conducted the survey data collection. FH and ZT performed the statistical analysis, and ZT finished the first draft of the paper. All authors made contributions to the final draft.

Conflicts of Interest

None declared.

Multimedia Appendix 1
In-depth interview guide.
[DOCX File , 15 KB - games_v13i1e58857_app1.docx]

Multimedia Appendix 2
Wenjuanxing online survey questionnaire.
[DOCX File , 23 KB - games_v13i1e58857_app2.docx]

References

1. GBD 2019 Mental Disorders Collaborators. Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet Psychiatry 2022 Feb;9(2):137-150 [FREE Full text] [doi: 10.1016/S2215-0366(21)00395-3] [Medline: 35026139]

XSL•FO
RenderX

2. Yang X, Fang Y, Chen H, Zhang T, Yin X, Man J, et al. Global, regional and national burden of anxiety disorders from 1990 to 2019: results from the Global Burden of Disease Study 2019. Epidemiol Psychiatr Sci 2021 May 06;30:e36 [FREE Full text] [doi: 10.1017/S2045796021000275] [Medline: 33955350]
3. World mental health report: transforming mental health for all. World Health Organization. URL: https://www.who.int/publications/i/item/9789240049338 [accessed 2025-05-29]
4. Patel V, Saxena S, Lund C, Thornicroft G, Baingana F, Bolton P, et al. The Lancet Commission on global mental health and sustainable development. Lancet 2018 Oct 27;392(10157):1553-1598. [doi: 10.1016/S0140-6736(18)31612-X] [Medline: 30314863]
5. Jenkins PE, Ducker I, Gooding R, James M, Rutter-Eley E. Anxiety and depression in a sample of UK college students: a study of prevalence, comorbidity, and quality of life. J Am Coll Health 2021 Jan 29;69(8):813-819. [doi: 10.1080/07448481.2019.1709474] [Medline: 31995452]
6. Storrie K, Ahern K, Tuckett A. A systematic review: Students with mental health problems--a growing problem. Int J Nurs Pract 2010 Feb;16(1):1-6. [doi: 10.1111/j.1440-172X.2009.01813.x] [Medline: 20158541]
7. Wynaden D, Wichmann H, Murray S. A synopsis of the mental health concerns of university students: results of a text-based online survey from one Australian university. High Educ Res Dev 2013 Oct;32(5):846-860. [doi: 10.1080/07294360.2013.777032]
8. Lipson SK, Lattie EG, Eisenberg D. Increased rates of mental health service utilization by U.S. college students: 10-year population-level trends (2007-2017). Psychiatr Serv 2019 Jan 01;70(1):60-63 [FREE Full text] [doi: 10.1176/appi.ps.201800332] [Medline: 30394183]
9. Mofatteh M. Risk factors associated with stress, anxiety, and depression among university undergraduate students. AIMS Public Health 2021;8(1):36-65 [FREE Full text] [doi: 10.3934/publichealth.2021004] [Medline: 33575406]
10. Casey SM, Varela A, Marriott JP, Coleman CM, Harlow BL. The influence of diagnosed mental health conditions and symptoms of depression and/or anxiety on suicide ideation, plan, and attempt among college students: findings from the Healthy Minds Study, 2018-2019. J Affect Disord 2022 Feb 01;298(Pt A):464-471. [doi: 10.1016/j.jad.2021.11.006] [Medline: 34774646]
11. Wang C, Wen W, Zhang H, Ni J, Jiang J, Cheng Y, et al. Anxiety, depression, and stress prevalence among college students during the COVID-19 pandemic: a systematic review and meta-analysis. J Am Coll Health 2023 Oct 01;71(7):2123-2130. [doi: 10.1080/07448481.2021.1960849] [Medline: 34469261]
12. Chang JJ, Ji Y, Li YH, Pan HF, Su PY. Prevalence of anxiety symptom and depressive symptom among college students during COVID-19 pandemic: a meta-analysis. J Affect Disord 2021 Sep 01;292:242-254 [FREE Full text] [doi: 10.1016/j.jad.2021.05.109] [Medline: 34134022]
13. Birmingham WC, Wadsworth LL, Lassetter JH, Graff TC, Lauren E, Hung M. COVID-19 lockdown: impact on college students' lives. J Am Coll Health 2023 Apr 22;71(3):879-893. [doi: 10.1080/07448481.2021.1909041] [Medline: 34292141]
14. Elmer T, Mepham K, Stadtfeld C. Students under lockdown: comparisons of students' social networks and mental health before and during the COVID-19 crisis in Switzerland. PLoS One 2020 Jul 23;15(7):e0236337 [FREE Full text] [doi: 10.1371/journal.pone.0236337] [Medline: 32702065]
15. Keng S, Zhong L, Xie F. Why did China's Zero-COVID policy persist? Decision urgency, regime type, and political opportunity structures. J Contemp China 2023 Mar 22;33(146):206-222. [doi: 10.1080/10670564.2023.2193143]
16. Gan Y, Ma J, Wu J, Chen Y, Zhu H, Hall BJ. Immediate and delayed psychological effects of province-wide lockdown and personal quarantine during the COVID-19 outbreak in China. Psychol Med 2022 May;52(7):1321-1332 [FREE Full text] [doi: 10.1017/S0033291720003116] [Medline: 32787981]
17. Muriel D, Crawford G. Video Games as Culture: Considering the Role and Importance of Video Games in Contemporary Society. New York, NY: Routledge; 2018.
18. Abd-Alrazaq A, Al-Jafar E, Alajlani M, Toro C, Alhuwail D, Ahmed A, et al. The effectiveness of serious games for alleviating depression: systematic review and meta-analysis. JMIR Serious Games 2022 Jan 14;10(1):e32331 [FREE Full text] [doi: 10.2196/32331] [Medline: 35029530]
19. Boldi A, Rapp A. Commercial video games as a resource for mental health: a systematic literature review. Behav Inf Technol 2021 Jul 02;41(12):2654-2690. [doi: 10.1080/0144929X.2021.1943524]
20. Ferguson CJ. Do angry birds make for angry children? A meta-analysis of video game influences on children's and adolescents' aggression, mental health, prosocial behavior, and academic performance. Perspect Psychol Sci 2015 Sep 17;10(5):646-666. [doi: 10.1177/1745691615592234] [Medline: 26386002]
21. Vuorre M, Johannes N, Magnusson K, Przybylski AK. Time spent playing video games is unlikely to impact well-being. R Soc Open Sci 2022 Jul 27;9(7):220411 [FREE Full text] [doi: 10.1098/rsos.220411] [Medline: 35911206]
22. Egami H, Rahman MS, Yamamoto T, Egami C, Wakabayashi T. Causal effect of video gaming on mental well-being in Japan 2020-2022. Nat Hum Behav 2024 Oct 19;8(10):1943-1956. [doi: 10.1038/s41562-024-01948-y] [Medline: 39160286]
23. Király O, Koncz P, Griffiths MD, Demetrovics Z. Gaming disorder: a summary of its characteristics and aetiology. Compr Psychiatry 2023 Apr;122:152376 [FREE Full text] [doi: 10.1016/j.comppsych.2023.152376] [Medline: 36764098]
24. Liao Z, Chen X, Huang Q, Shen H. Prevalence of gaming disorder in East Asia: a comprehensive meta-analysis. J Behav Addict 2022 Sep 26;11(3):727-738 [FREE Full text] [doi: 10.1556/2006.2022.00050] [Medline: 35932469]

XSL•FO
RenderX

25. Iacovides I, Mekler ED. The role of gaming during difficult life experiences. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 Presented at: CHI '19; May 4-9, 2019; Glasgow, UK p. 1-12 URL: https://dl.acm.org/doi/10.1145/3290605.3300453 [doi: 10.1145/3290605.3300453]
26. Pallavicini F, Pepe A, Mantovani F. The effects of playing video games on stress, anxiety, depression, loneliness, and gaming disorder during the early stages of the COVID-19 pandemic: PRISMA systematic review. Cyberpsychol Behav Soc Netw 2022 Jun 01;25(6):334-354 [FREE Full text] [doi: 10.1089/cyber.2021.0252] [Medline: 35639118]
27. Türkay S, Lin A, Johnson D, Formosa J. Self-determination theory approach to understanding the impact of videogames on wellbeing during COVID-19 restrictions. Behav Inf Technol 2022 Jul 01;42(11):1720-1739. [doi: 10.1080/0144929x.2022.2094832]
28. Barr M, Copeland-Stewart A. Playing video games during the COVID-19 pandemic and effects on players' well-being. Games Cult 2021 May 06;17(1):122-139. [doi: 10.1177/15554120211017036]
29. Bucci S, Schwannauer M, Berry N. The digital revolution and its impact on mental health care. Psychol Psychother 2019 Jun 28;92(2):277-297. [doi: 10.1111/papt.12222] [Medline: 30924316]
30. Primack BA, Carroll MV, McNamara M, Klem ML, King B, Rich M, et al. Role of video games in improving health-related outcomes: a systematic review. Am J Prev Med 2012 Jun;42(6):630-638 [FREE Full text] [doi: 10.1016/j.amepre.2012.02.023] [Medline: 22608382]
31. Sharifzadeh N, Kharrazi H, Nazari E, Tabesh H, Edalati Khodabandeh M, Heidari S, et al. Health education serious games targeting health care providers, patients, and public health users: scoping review. JMIR Serious Games 2020 Mar 05;8(1):e13459 [FREE Full text] [doi: 10.2196/13459] [Medline: 32134391]
32. David OA, Costescu C, Cardos R, Mogoa'e C. How effective are serious games for promoting mental health and health behavioral change in children and adolescents? A systematic review and meta-analysis. Child Youth Care Forum 2020;49:817-838 [FREE Full text]
33. Bowman ND, Rieger D, Tammy Lin JH. Social video gaming and well-being. Curr Opin Psychol 2022 Jun;45:101316. [doi: 10.1016/j.copsyc.2022.101316] [Medline: 35313181]
34. Snodgrass JG, Lacy MG, Cole SW. Internet gaming, embodied distress, and psychosocial well-being: a syndemic-syndaimonic continuum. Soc Sci Med 2022 Feb;295:112728 [FREE Full text] [doi: 10.1016/j.socscimed.2019.112728] [Medline: 31879045]
35. Bergmann M, Wollbrandt I, Gittel L, Halbe E, Mackert S, Philipsen A, et al. Effects of a video game intervention on symptoms, training motivation, and visuo-spatial memory in depression. Front Psychiatry 2023 Aug 24;14:1173652 [FREE Full text] [doi: 10.3389/fpsyt.2023.1173652] [Medline: 37692312]
36. Dias LP, Barbosa JL, Vianna HD. Gamification and serious games in depression care: a systematic mapping study. Telemat Inform 2018 Apr;35(1):213-224. [doi: 10.1016/j.tele.2017.11.002]
37. Ruiz M, Moreno M, Girela-Serrano B, Díaz-Oliván I, Muñoz LJ, González-Garrido C, et al. Winning the game against depression: a systematic review of video games for the treatment of depressive disorders. Curr Psychiatry Rep 2022 Jan 03;24(1):23-35 [FREE Full text] [doi: 10.1007/s11920-022-01314-7] [Medline: 35113313]
38. Barnes S, Prescott J. Empirical evidence for the outcomes of therapeutic video games for adolescents with anxiety disorders: systematic review. JMIR Serious Games 2018 Feb 28;6(1):e3 [FREE Full text] [doi: 10.2196/games.9530] [Medline: 29490893]
39. Staiano AE, Flynn R. Therapeutic uses of active videogames: a systematic review. Games Health J 2014 Dec;3(6):351-365 [FREE Full text] [doi: 10.1089/g4h.2013.0100] [Medline: 26192642]
40. Gliosci R, Barros Pontes E Silva T. Therapeutic interventions with videogames in treatments for depression: a systematic review. Games Health J 2023 Aug 01;12(4):269-279. [doi: 10.1089/g4h.2022.0094] [Medline: 36946750]
41. Lehtimaki S, Martic J, Wahl B, Foster KT, Schwalbe N. Evidence on digital mental health interventions for adolescents and young people: systematic overview. JMIR Ment Health 2021 Apr 29;8(4):e25847 [FREE Full text] [doi: 10.2196/25847] [Medline: 33913817]
42. Wang GY, Simkute D, Griskova-Bulanova I. Neurobiological link between stress and gaming: a scoping review. J Clin Med 2023 Apr 25;12(9):3113 [FREE Full text] [doi: 10.3390/jcm12093113] [Medline: 37176554]
43. Shah A, Kraemer KR, Won CR, Black S, Hasenbein W. Developing digital intervention games for mental disorders: a review. Games Health J 2018 Aug;7(4):213-224. [doi: 10.1089/g4h.2017.0150] [Medline: 30106642]
44. Syed UE, Kamal A. Video game-based and conventional therapies in patients of neurological deficits: an experimental study. Disabil Rehabil Assist Technol 2021 Apr 05;16(3):332-339. [doi: 10.1080/17483107.2019.1679266] [Medline: 31684776]
45. Zayeni D, Raynaud JP, Revet A. Therapeutic and preventive use of video games in child and adolescent psychiatry: a systematic review. Front Psychiatry 2020;11:36 [FREE Full text] [doi: 10.3389/fpsyt.2020.00036] [Medline: 32116851]
46. Colder Carras M, Van Rooij AJ, Spruijt-Metz D, Kvedar J, Griffiths MD, Carabas Y, et al. Commercial video games as therapy: a new research agenda to unlock the potential of a global pastime. Front Psychiatry 2017 Jan 22;8:300 [FREE Full text] [doi: 10.3389/fpsyt.2017.00300] [Medline: 29403398]

XSL•FO
RenderX

47. Phillips C, Klarkowski M, Frommel J, Gutwin C, Mandryk RL. Identifying commercial games with therapeutic potential through a content analysis of steam reviews. Proc ACM Hum Comput Interact 2021 Oct 06;5(CHI PLAY):1-21. [doi: 10.1145/3474682]
48. Pine R, Fleming T, McCallum S, Sutcliffe K. The effects of casual videogames on anxiety, depression, stress, and low mood: a systematic review. Games Health J 2020 Aug;9(4):255-264. [doi: 10.1089/g4h.2019.0132] [Medline: 32053021]
49. Kowal M, Conroy E, Ramsbottom N, Smithies T, Toth A, Campbell M. Gaming your mental health: a narrative review on mitigating symptoms of depression and anxiety using commercial video games. JMIR Serious Games 2021 Jun 16;9(2):e26575 [FREE Full text] [doi: 10.2196/26575] [Medline: 34132648]
50. Pallavicini F, Bouchard S. Editorial: assessing the therapeutic uses and effectiveness of virtual reality, augmented reality and video games for emotion regulation and stress management. Front Psychol 2019 Dec 11;10:2763 [FREE Full text] [doi: 10.3389/fpsyg.2019.02763] [Medline: 31920816]
51. LaFleur LB, Hebert ZJ, Dupuy AS. Leveling up your game: the use of video games as a therapeutic modality. J Creat Ment Health 2017 Jul 07;13(1):58-67. [doi: 10.1080/15401383.2017.1328293]
52. Pallavicini F, Pepe A, Mantovani F. Commercial off-the-shelf video games for reducing stress and anxiety: systematic review. JMIR Ment Health 2021 Aug 16;8(8):e28150 [FREE Full text] [doi: 10.2196/28150] [Medline: 34398795]
53. Boldi A, Rapp A. Making sense of the pandemic: multiple functions of commercial video games during one year and a half of COVID-19 crisis. Int J Hum Comput Stud 2023 Dec;180:103141. [doi: 10.1016/j.ijhcs.2023.103141]
54. WJX can meet your various questionnaire/form needs. Changsha Ranxing Information Technology. URL: https://www.wjx.cn/ [accessed 2025-05-29]
55. Hennink M, Kaiser BN. Sample sizes for saturation in qualitative research: a systematic review of empirical tests. Soc Sci Med 2022 Jan;292:114523 [FREE Full text] [doi: 10.1016/j.socscimed.2021.114523] [Medline: 34785096]
56. Plummer F, Manea L, Trepel D, McMillan D. Screening for anxiety disorders with the GAD-7 and GAD-2: a systematic review and diagnostic metaanalysis. Gen Hosp Psychiatry 2016 Mar;39:24-31. [doi: 10.1016/j.genhosppsych.2015.11.005] [Medline: 26719105]
57. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
58. Manea L, Gilbody S, McMillan D. A diagnostic meta-analysis of the Patient Health Questionnaire-9 (PHQ-9) algorithm scoring method as a screen for depression. Gen Hosp Psychiatry 2015 Jan;37(1):67-75. [doi: 10.1016/j.genhosppsych.2014.09.009] [Medline: 25439733]
59. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
60. Chen S, Chiu H, Xu B, Ma Y, Jin T, Wu M, et al. Reliability and validity of the PHQ-9 for screening late-life depression in Chinese primary care. Int J Geriatr Psychiatry 2010 Nov 22;25(11):1127-1133. [doi: 10.1002/gps.2442] [Medline: 20029795]
61. Zeng QZ, He YL, Liu H, Miao JM, Chen JX, Xu HN, et al. Reliability and validity of Chinese version of the Generalized Anxiety Disorder 7-item (GAD-7) scale in screening anxiety disorders in outpatients from traditional Chinese internal department. Chin Ment Health J 2022 Nov;27(3):163-168 [FREE Full text] [doi: 10.3969/j.issn.1000-6729.2013.03.001] [Medline: 35584970]
62. Chiu B, Chin W. Systematic review and meta-analysis on the patient health questionnare-9 (PHQ-9) for depression screening in Chinese primary care patients. Fam Med Care 2018;1(1):66. [doi: 10.15761/FMC.1000105]
63. Williams N. The GAD-7 questionnaire. Occup Med 2014 Apr 03;64(3):224. [doi: 10.1093/occmed/kqt161]
64. Charmaz K. Constructing Grounded Theory. 2nd edition. Thousand Oaks, CA: Sage Publication; 2014.
65. Clarke V, Braun V. Thematic Analysis: A Practical Guide. Thousand Oaks, CA: Sage Publication; 2021.
66. Kroenke K, Spitzer RL. The PHQ-9: a new depression diagnostic and severity measure. Psychiatric Annals 2002 Sep 01;32(9):509-515. [doi: 10.3928/0048-5713-20020901-06]
67. Thornton P. The A4 movement: mapping its background and impact. Chin Lead Mon 2023;75(75):31 [FREE Full text]
68. Cheng C, Lau YC, Luk JW. Social capital-accrual, escape-from-self, and time-displacement effects of internet use during the COVID-19 stay-at-home period: prospective, quantitative survey study. J Med Internet Res 2020 Dec 24;22(12):e22740 [FREE Full text] [doi: 10.2196/22740] [Medline: 33320824]
69. Villani D, Carissoli C, Triberti S, Marchetti A, Gilli G, Riva G. Videogames for emotion regulation: a systematic review. Games Health J 2018 Apr;7(2):85-99 [FREE Full text] [doi: 10.1089/g4h.2017.0108] [Medline: 29424555]
70. Savolainen I, Vuorinen I, Sirola A, Oksanen A. Gambling and gaming during COVID-19: the role of mental health and social motives in gambling and gaming problems. Compr Psychiatry 2022 Aug;117:152331 [FREE Full text] [doi: 10.1016/j.comppsych.2022.152331] [Medline: 35772368]
71. Palgi Y, Shrira A, Ring L, Bodner E, Avidor S, Bergman Y, et al. The loneliness pandemic: loneliness and other concomitants of depression, anxiety and their comorbidity during the COVID-19 outbreak. J Affect Disord 2020 Oct 01;275:109-111 [FREE Full text] [doi: 10.1016/j.jad.2020.06.036] [Medline: 32658811]

XSL•FO
RenderX

72. Stavropoulos V, Vassallo J, Burleigh TL, Gomez R, Colder Carras M. The role of internet gaming in the association between anxiety and depression: a preliminary cross-sectional study. Asia Pac Psychiatry 2022 Jun 20;14(2):e12474. [doi: 10.1111/appy.12474] [Medline: 34013597]
73. Pearcy BT, McEvoy PM, Roberts LD. Internet gaming disorder explains unique variance in psychological distress and disability after controlling for comorbid depression, OCD, ADHD, and anxiety. Cyberpsychol Behav Soc Netw 2017 Feb;20(2):126-132. [doi: 10.1089/cyber.2016.0304] [Medline: 28085490]
74. Chang CI, Fong Sit H, Chao T, Chen C, Shen J, Cao B, et al. Exploring subtypes and correlates of internet gaming disorder severity among adolescents during COVID-19 in China: a latent class analysis. Curr Psychol 2022 Apr 29;42(23):1-12 [FREE Full text] [doi: 10.1007/s12144-022-03133-8] [Medline: 35505828]
75. Terrell KR, Stanton BR, Hamadi HY, Merten JW, Quinn N. Exploring life stressors, depression, and coping strategies in college students. J Am Coll Health 2024 Apr 15;72(3):923-932. [doi: 10.1080/07448481.2022.2061311] [Medline: 35427463]
76. Mahmoud JS, Staten RT, Hall LA, Lennie TA. The relationship among young adult college students' depression, anxiety, stress, demographics, life satisfaction, and coping styles. Issues Ment Health Nurs 2012 Mar 24;33(3):149-156. [doi: 10.3109/01612840.2011.632708] [Medline: 22364426]
77. Tran AW, Lumley MN. Internalized stigma and student well-being: the role of adaptive and maladaptive coping. Soc Work Ment Health 2019 Jan 06;17(4):408-425. [doi: 10.1080/15332985.2018.1563023]
78. Halbrook YJ, O'Donnell AT, Msetfi RM. When and how video games can be good: a review of the positive effects of video games on well-being. Perspect Psychol Sci 2019 Nov;14(6):1096-1104. [doi: 10.1177/1745691619863807] [Medline: 31672105]

Abbreviations

GAD-7: 7-item Generalized Anxiety Disorder
PHQ-9: 9-item Patient Health Questionnaire

Edited by A Coristine, T Rashid Soron; submitted 26.03.24; peer-reviewed by HF Sit, A Bakhtiari; comments to author 13.07.24; revised version received 30.09.24; accepted 06.05.25; published 30.05.25.

Please cite as:
Hu F, Tai Z, Liu J
The Positive Effect of Video-Game Play on College Students' Anxiety and Depression Symptoms During the COVID-19 Pandemic Shelter-in-Place Lockdowns: Mixed Methods Study
JMIR Serious Games 2025;13:e58857
URL: https://games.jmir.org/2025/1/e58857
doi:10.2196/58857
PMID:

©Fengbin Hu, Zixue Tai, Jianping Liu. Originally published in JMIR Serious Games (https://games.jmir.org), 30.05.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Nature Engagement Outcomes of Viewing Nature Through a 360° Video or a Tablet Screen: Randomized Trial

Elena Brambilla[1], MSc; Karen Stendal[2], Prof Dr; Vibeke Sundling[3], Prof Dr; Giovanna Calogiuri[1], Prof Dr

[1]Department of Nursing and Health Sciences, Faculty of Health and Social Sciences, University of South-Eastern Norway, Drammen, Norway

[2]Department of Business, Marketing and Law, USN School of Business, University of South-Eastern Norway, Ringerike, Norway

[3]Department of Optometry, Radiography and Lighting Design, Faculty of Health and Social Sciences, University of South-Eastern Norway, Kongsberg, Norway

Corresponding Author:
Elena Brambilla, MSc
Department of Nursing and Health Sciences, Faculty of Health and Social Sciences
University of South-Eastern Norway
Grønland 58, 3045
Drammen, 3045
Norway
Phone: 47 35952549
Email: elena.brambilla@usn.no

Abstract

Background: Nature engagement, including affective and physical interactions with nature, is linked to a multitude of health and well-being benefits. Unfortunately, opportunities for engaging with nature are decreasing worldwide. University students, especially, are a demographic group that tends to engage little with nature. Immersive virtual nature (IVN; ie, digital nature content delivered through immersive devices, such as head-mounted displays) has been proposed as a medium to facilitate nature experiences and engagement. In recent years, 360° nature videos have emerged as an accessible way to create IVN content, although it is still unclear whether they can elicit presence and increase nature engagement to a greater extent than nature videos delivered through nonimmersive media.

Objective: We aimed to investigate the effectiveness of nature videos as a medium to promote nature engagement among university students, comparing devices with different levels of immersion. Specifically, 2 experimental conditions were tested: a 360° nature video delivered through a head-mounted display (IVN) and a matching video displayed on a tablet screen (nonimmersive virtual nature).

Methods: In total, 38 students were recruited at the library of a university campus and invited to participate in an organized hiking tour at the location displayed during the virtual nature experience. They were then randomized, using a random number generator, to either the IVN (n=20, 53%) or the nonimmersive virtual nature condition (n=18, 47%). Pre- and postexposure assessments of nature connectedness, intention to perform green exercise, intention to visit the hiking location, and intention to participate in the organized hiking tour were collected. Presence, cybersickness, and actual attendance on the tour were also assessed.

Results: A mixed ANOVA showed statistically significant pre- to postexposure assessment increases in nature connectedness ($F_{1,36}$=33.49; P<.001; η_p^2=0.48); intention to perform green exercise ($F_{1,36}$=5.55; P=.02); intention to visit the hiking location ($F_{1,36}$=15.34; P<.001; η_p^2=0.26); and intention to participate in the hiking tour ($F_{1,36}$=12.45; P=.001; η_p^2=0.30). Both conditions were associated with medium to high ratings of "being there" and "sense of reality" but low ratings of "realism." The cybersickness levels were generally low. Of the 38 students, 6 (16%) participated in the organized tour. The mixed ANOVA found no statistical differences between the two conditions for any of the outcomes. The participants' changes in nature connectedness (ρ=0.35; P=.03) and attendance on the hiking tour (ρ=.37; P=.02) correlated with the presence item "being there."

Conclusions: This study provides novel evidence on the potential of virtual nature as a medium to improve nature engagement among university students, adding to the current debate on the effectiveness of 360° videos. These findings can inform future research as well as initiatives seeking to promote nature engagement.

(JMIR Serious Games 2025;13:e63424) doi:10.2196/63424

XSL•FO
RenderX

KEYWORDS

virtual reality; immersive virtual nature; nature connectedness; nature engagement; nature-based interventions; outdoor; green exercise

Introduction

Background

The World Health Organization highlights the global need to foster mental health, reporting that 12% and 15% of the European and American populations experience psychological challenges, respectively [1]. College and university students are recognized to be a group at higher risk of mental health challenges, such as anxiety [2] and stress [3]. Recent reports show that the number of students reporting mental health issues has increased during the last decade from 11% to 31% [4]. Regular and recurrent nature engagement is a known salutogenic factor supporting the population's general health and psychological well-being [5]. However, outdoor time is dropping, especially in younger generations [6], and college and university students often prefer exercising indoors (eg, in indoor gyms and sports facilities) rather than out in nature [4,7]. Meanwhile, the use of technology and screen time is rapidly increasing, contributing to further disconnection from nature [8,9]. From a different perspective, technology may be used as a tool to reconnect people to nature and improve human-nature interactions [10]. In this respect, virtual reality (VR) is gaining increasing attention due to its rapid advancement and economic accessibility for consumers [10,11].

Benefits of Nature Engagement

According to the World Health Organization, direct exposure to nature (eg, contemplating the view of nature or spending time in, interacting with, or being physically active in a natural environment) provides mental and physical health benefits across socioeconomic strata and genders [12]. Emotional well-being, improved positive affect and mood, stress reduction, and social contact are just some of the benefits of human-nature interactions [13-15]. One important factor facilitating contact with nature and contributing to the health and well-being benefits associated with nature interactions is physical activity [16,17]. In this respect, Pretty et al [18] coined the concept of "green exercise," referring to any physical activity in the presence of nature.

Not only physical contact with nature, but also cognitive and affective engagement with nature is linked to health and well-being outcomes [13,19-22]. As a relevant part of human-nature interactions, nature connectedness, referring to an individual's sense of being emotionally connected to the natural world [23], plays an important role in human health. Studies link nature connectedness to various health outcomes including social, psychological, physical, and general well-being [19,24-26]. Moreover, evidence shows a positive correlation between nature connectedness and different well-being indicators, including happiness, mood, and life satisfaction, as well as reduced symptoms of stress-related mental disorders, and enhanced self-reported vitality, independently from a cultural and climatic context [20,27-29]. Nature connectedness can be classified as a personality trait or a temporary state [30]. However, the trait level of nature connectedness was found to be dynamic and varies across the life span [31,32].

It is important to highlight that, although being two different and largely independent ways of engaging with nature, a reciprocal influence exists between nature connectedness and physical contact with nature. Frequent experiences of nature or living near natural areas during childhood are important to strengthening people's feelings about nature and fostering a trait level of nature connectedness [33-35]. Changes in temporary state levels of nature connectedness can be elicited in adults through direct contact with nature, including both passive and active engagement with nature, such as viewing nature while sitting in a park and brief walks in a forest [27,36,37]. A systematic review [36] suggests that interacting with nature on a regular basis can lead to sustained increases in nature connectedness. Furthermore, individuals with higher levels of nature connectedness are more likely to visit and spend time in a natural environment and engage in outdoor recreation [33,38]. Therefore, nature connectedness both supports and is supported by physical human-nature interactions [18,39,40]. Unfortunately, with more than half of the global population living in urban areas [41], the opportunities for human-nature interactions are decreasing worldwide [14]. The progressive disconnection from nature may have negative consequences on human health [14,42], including physical and mental health challenges, such as sedentary behavior, obesity, and reduced happiness and life satisfaction [43].

Immersive Virtual Nature: The Case of 360° Videos

VR refers to a technology that provides the illusion of being in another place by substituting the primary sensory input with data generated by a computer [44]. VR can include different technologies with different levels of *immersion*, that is, "the extent to which the computer displays are capable of delivering an inclusive, extensive, surrounding and vivid illusion of reality to the senses of a human participant" [45]. For instance, head-mounted displays (HMDs), devices with a motion sensor that allow a 360° range of vision of a simulated environment while excluding the view of external surroundings, are commonly considered a form of immersive VR technology [46,47]. In contrast, computer screens are deemed as a form of nonimmersive technology. In this regard, the user's sense of *presence* (ie, the subjective sense of being in the simulated environment [45]) has been indicated as a key factor in defining immersive experiences [45,48-53] capable of eliciting powerful psychophysiological responses [54]. While it is generally assumed that a user's sense of presence is influenced by the levels of immersion of a given VR device, the extent to which immersive technology alone is sufficient to deliver an immersive experience, characterized by a strong feeling of presence, is still debated [50]. In this respect, *interactiveness* is considered paramount for delivering truly immersive VR experiences. Another factor is *cybersickness*, a specific type of visually induced motion sickness [55,56], which tends to be particularly

XSL•FO
RenderX

problematic in highly immersive VR devices and can negatively impact the VR experience.

Immersive virtual nature (IVN) refers to installations that use immersive VR technology to provide an illusory perception of being surrounded by and interacting with a natural environment [11]. IVN has been found effective in eliciting increased nature engagement outcomes, including nature connectedness [57] and intention to visit naturalistic locations [58-61]. In recent years, 360° videos (spherical videos that can be viewed through HMD) emerged as an accessible technique to generate IVN contents [62]. However, concerns have been raised regarding the effectiveness of 360° videos compared with computer-generated interactive IVN experiences [57]. Moreover, the extent to which 360° nature videos can increase nature engagement to a greater extent than videos displayed through nonimmersive technologies (eg, viewing nature through pictures, videos, or different screens) is not clear [57]. While some studies indicate that 360° nature videos tend to be more effective in eliciting nature connectedness than traditional nature videos [47,63], others suggest less consistent results [64,65]. Moreover, to the best of our knowledge, to date, only a few studies have investigated the effects of 360° nature video experiences on people's physical engagement with nature, including measurements of actual behavior, comparing experiences delivered through immersive versus nonimmersive technologies.

This Study

The purpose of this study was to investigate the extent to which a single virtual nature experience could elicit increased nature engagement outcomes in university students, and whether this effect varies depending on the level of immersion of the devices through which they are delivered. Specifically, in this study, the IVN condition consisted of viewing a 360° video featuring a naturalistic location through an HMD, while the nonimmersive virtual nature (NIVN) condition consisted of viewing the same video in rectangular format on a tablet screen. In this study, nature engagement is used as an umbrella term to enclose affective (operationalized as nature connectedness) and physical interactions with nature (operationalized as behavioral intentions and actual behavior relative to nature visits).

The following research question was outlined: What is the effectiveness of viewing a video featuring a naturalistic location either through an immersive (IVN) or a nonimmersive device (NIVN) in promoting nature engagement among university students?

To address the research question, we proposed the following hypotheses:

1. There will be a significant pre- to postexposure assessment increment in all the nature engagement outcomes in both conditions.
2. Compared with the NIVN condition, the IVN condition will elicit larger increments of all nature engagement outcomes, as well as greater levels of presence and cybersickness.
3. All nature engagement outcomes will significantly and positively correlate with the participants' level of presence and negatively correlate with the severity of cybersickness symptoms.

Methods

Research Design and Randomization

The study was designed as a randomized trial with parallel groups and pre- and postexposure assessments, and this paper is outlined following the CONSORT (Consolidated Standards of Reporting Trials) guidelines [66].

First, an ID code list was created, and each ID code was allocated to 1 of the 2 experimental conditions (IVN and NIVN) using a random number generator. Each participant was linked to a unique ID code just before undergoing the experiment. Anonymity was assured by storing participants' identifying information (ie, name, family name, and phone number) only in paper form separately from the collected data and by deleting it after completion of the project. The allocation sequence was concealed until participants were enrolled and assigned to the intervention. However, the trial was not completely blinded, as the participants were aware of the two experimental conditions.

Participants and Sample Size

The study took place from March 16, 2022, to April 28, 2022, at the library of the University of South-Eastern Norway, which offers bachelor and master programs in nursing, computer sciences, teacher education, management, and political sciences. University students were purposely targeted as a specific group of interest, as they are a population that tends to prefer indoor activities [4] and engage less in outdoor recreation [33]. Therefore, they could particularly benefit from exposure to the virtual nature. Participants were recruited through flyer distribution, advertisements posted on social media, and presentations of the study during classes. Eligible participants were university students living in Drammen, speaking Norwegian or English, and able to walk outdoors without aid. No other restrictions were applied. Participants were informed in verbal and written form about the general purpose of the study and its associated risks.

The sample size was estimated based on the primary outcome, nature connectedness. Two studies [47,67] investigating the effects of experiencing an IVN found a large and a medium effect size for nature connectedness. Hence, for this study, a medium effect size was selected to ensure sufficient power to detect meaningful effects without unnecessarily burdening a larger number of participants. The sample size calculation was performed in G*Power (Heinrich Heine University) for mixed between-within subject ANOVA with 2 groups and 2 repeated measurements, setting an expected power of 0.80, a precision of 0.05, and a moderate effect size (f=0.25), which resulted in an optimal sample size of 34. In total, 54 participants were recruited and assessed for eligibility, of whom 7 did not meet the eligibility criteria. The other 9 participants were enrolled but had to be excluded because of technical issues (ie, malfunctioning of the equipment). The final sample (participants included in the analyses) consisted of 38 participants.

XSL•FO
RenderX

Virtual Nature Experiences

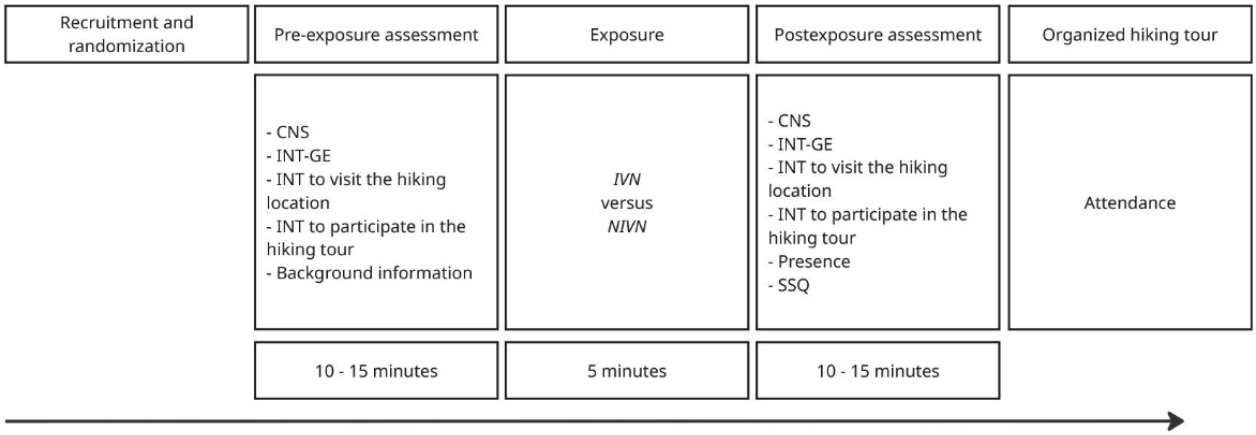
The IVN condition consisted of a 360° video of nature, delivered through a stand-alone HMD (Meta Quest 2, Reality Labs). Participants viewed the 360° video while sitting on a chair. In the NIVN condition, participants watched the same video on a tablet screen (a 26.7 cm, full high-definition, 1200-pixel LCD). The version of the video used in either condition was equal in duration. The video was filmed in a local forest named Dronningslepet and composed of a sequence of static clips (15 seconds each). The total length of the video was 5 minutes, which is generally deemed sufficient to elicit psychological responses while reducing the risk of cybersickness [10,57]. The video was filmed using a GoPro MAX (3840×2160 resolution; 25 frames per second) and edited in DaVinci Resolve 18 (Blackmagic Design) using sounds retrieved from the YouTube Audio Library [68]. A demo of the video is available on the web [69].

Procedures

Before undergoing either the IVN or the NIVN condition, participants were informed about the safe use of the technology. Subsequently, participants were invited to join in an organized hiking tour taking place at the same location where the video, displayed during the experiment, was filmed (ie, a local forest named Dronningslepet). Information regarding the tour duration, path length, and intensity was provided both verbally and in written form. Within 5 weeks from the experiment, participants could attend the tour (3 dates were made available), which had a duration of approximately 3 hours and included a local guide, food, and transportation. The participants were then asked to fill in the pre-exposure questionnaire.

At the completion of each experimental condition, the participants filled in the postexposure questionnaire. A reminder about the tour was sent to them via phone before each date. The participants who were allocated to the NIVN condition were invited to try the IVN condition after all the hiking tours were completed. The total duration of the procedure varied from 20 to 35 minutes. The experimental procedure is summarized in Figure 1. Data collections were handled by questionnaires created with the Nettskjema, a web-based survey solution developed and hosted by the University of Oslo. The questionnaires were provided in Norwegian and English to allow both international and local students to participate.

Figure 1. Experimental procedure, data gathering organization, and procedure duration. Background information included age, gender, faculty, residence, green exercise habits, and previous visits to the hiking location. CNS: Connectedness to Nature Scale; INT: intention; INT-GE: Intention to Perform Green Exercise questionnaire; IVN: immersive virtual nature; NIVN: nonimmersive virtual nature; SSQ: Simulator Sickness Questionnaire.



Outcomes

Nature Engagement Outcomes

Overview

The main outcomes included self-reported nature engagement collected by questionnaire before and after each experimental condition. In addition, attendance on the organized hiking tour was recorded. The questionnaires were available in both English and Norwegian. The English versions of the questionnaires were all validated. The questionnaires were translated from English to Norwegian by native speakers in collaboration with academics with expertise in environmental psychology, ensuring consistency with the cultural context.

The main outcomes related to both affective and physical engagement with nature and operationalized through the concept of nature connectedness, intention to visit nature, and actual visitation of nature. In total, 3 types of nature visits were considered: generic green exercise, independent visit to the location of the hiking tour (behavioral intentions), and participation in the organized hiking tour (behavioral intention and actual behavior). The questionnaires are provided in Multimedia Appendix 1.

Nature Connectedness

The state version of the Connectedness to Nature Scale (CNS) [27] was used to assess the extent to which individuals felt emotionally attached to the natural world. CNS was selected for its comprehensive assessment of the relationship between humans and nature (including animals, plants, and a sense of equality between the self and nature) and its high reliability [27]. The CNS demonstrated strong internal consistency, as measured by Cronbach α (α=0.82), which is consistent with previous studies (α=0.80–0.82) [27,67]. Additionally, the CNS was selected for its convergent validity and its positive correlation with other measures of nature connection such as the Inclusion of Nature in Self scale [70]. The scale includes

XSL•FO
RenderX

13 items (eg, “Right now I’m feeling a sense of oneness with the natural world around me”), rated on a 7-point Likert scale (1=strongly disagree; 7=strongly agree). The internal consistency of the scale in this study, as measured by Cronbach α (α=0.81-0.85), was consistent with previous studies [27,67].

Intention to Perform Green Exercise

The Intention to Perform Green Exercise questionnaire (INT-GE) [71], assessing the extent of intention to engage in green exercise in the near future, was used. The INT-GE has demonstrated a strong factorial validity in the original validation study [71] and both the original English version [71] and translated Norwegian version [58] demonstrate good internal consistency, as measured by Cronbach α (α=0.94-0.95). The scale includes 5 items (eg, “I expect to do Green Exercise”), rated on a 7-point Likert scale (1=strongly disagree or very unlikely; 7=strongly agree or very likely). Examples of what was intended as “green exercise” were provided in the caption, which also provided a time reference (“over the next 5 weeks”), that approximately corresponded to the time span between the participant’s virtual nature experience and the last planned hiking tour. In this study, the internal consistency of the scale, as measured by Cronbach α (α=0.95-0.95), was in line with previous studies [58,72].

Intention to Visit the Hiking Location

This scale was developed following established guidelines for developing questionnaires based on the theory of planned behavior within health services researchers [73]. Specifically, the participants’ intention to visit the hiking location was assessed through 3 items (eg, “After I received information about Dronningslepet, I consider visiting it within the next 5 weeks”) rated on a 7-point Likert scale (1=strongly disagree; 7=strongly agree). The scale in this study showed an adequate internal consistency, as measured by Cronbach α (α=0.93-0.97).

Intention to Participate and Actual Participation in the Organized Hiking Tour

A single-item question about the participants’ intention to attend the organized hiking tour (ie, “After I received information about the organized trip to Dronningslepet, the probability that I will participate in the tour is...”), rated on an 11-point Likert scale (0=very low; 5=neither high or low; 10=very high), was used. In addition, the participants’ attendance on the organized hiking tour was registered in loco by one author.

Other Outcomes

Presence

Postexposure assessment of the sense of presence was performed based on the study by Slater et al [74], with some adjustments relative to the context of the trial and consistent with previous work by Calogiuri et al [75]. The final 3 items were rated on an 11-point Likert scale (0=absolutely disagree; 10=absolutely agree) and formulated as follows: (1) “In the video I had the sense of ‘being there’” (ie, “being there”); (2) “I thought of the video as equal to the real environment” (ie, “realism”); and (3) “The video became more real or present to me compared to the real world; by “real world” we mean the room where you were undergoing the test (ie, “sense of reality”). In line with previous works [75,76], the items were used individually for further analysis.

Cybersickness

The Simulator Sickness Questionnaire (SSQ) [77] was administered after both experimental conditions to measure the severity of cybersickness. The SSQ has demonstrated good factorial validity in the original validation study [77]. The SSQ includes 16 items describing different symptoms (eg, “nausea,” “headache,” and “eyestrain”) and rated with a 4-point Likert scale (0=not at all; 4=a lot). The scale in this study showed adequate internal consistency, as measured by Cronbach α (α=0.86), consistent with a previous study that used the same Norwegian version of the scale [78].

Background Information

Participants’ age, gender, and residence details were collected for descriptive purposes. Furthermore, previous visits to the hiking location (ie, “Have you been to the hiking location before?”) were assessed through a single-item question to which participants could answer “yes,” “no,” or “I don’t remember.” In addition, green exercise habits (ie, “How often do you do green exercise?”) were assessed through a single-item question to which participants could answer “more than 3 times a week,” “2 times a week,” “1 time a week,” or “less than 1 time a week.”

Statistical Analysis

The score of all instruments was computed in line with instructions provided by the respective authors [27,71,73,74,77,79]. Reliability analysis was performed to control the scales’ internal consistency. To detect possible missing values and outliers and estimate the variables’ frequency distribution, explorative analysis was conducted. No missing or extreme values were identified except for INT-GE values, which included one outlier that was detected and adjusted, in line with Tabachnick and Fidell [80], by replacing the outlier’s score with a raw score of one unit smaller than the most extreme score in the distribution. This approach avoided the score distorting the statistic while maintaining the outlier value deviant but not as deviant as it was [81]. The frequency distributions were explored based on the examination of histograms and values for skewness at both time points, showing an acceptable normal distribution for all the variables and moderately skewed distribution for INT-GE. The skewness values for pre and postexposure assessment were as follows: CNS=0.03 and −0.02; INT-GE=−0.99 and −0.95; intention to participate in the hiking tour=0.43 and −0.35; and intention to visit the hiking location=−0.34 and −0.87.

As data followed normal distribution and variances were homogenous, a parametric mixed ANOVA was conducted to assess the impact of the two experimental conditions (between-subjects effect) on the dependent variables (ie, CNS, INT-GE, intention to visit the hiking location, and intention to participate in the hiking tour) across the 2 time points (within-subjects effect). The within-subjects factor was “time point” (pre- and postexposure assessment) while “condition” (IVN or NIVN) was set as the between-subjects factor. The effect size was expressed as partial eta squared (η_p^2). A possible difference between the conditions for the presence items and

XSL•FO
RenderX

the SSQ score were tested through a Mann-Whitney *U* test. Spearman rank correlation coefficient (ρ) was used to explore possible associations among all variables. For the variables with repeated measurements, differential scores from pre- to postexposure assessment (delta values) were computed and used in the correlation analysis. The strength of the correlation was interpreted according to the guidelines proposed by Cohen [82].

All analyses were conducted using SPSS (version 28.0.0.0.1; IBM Corp).

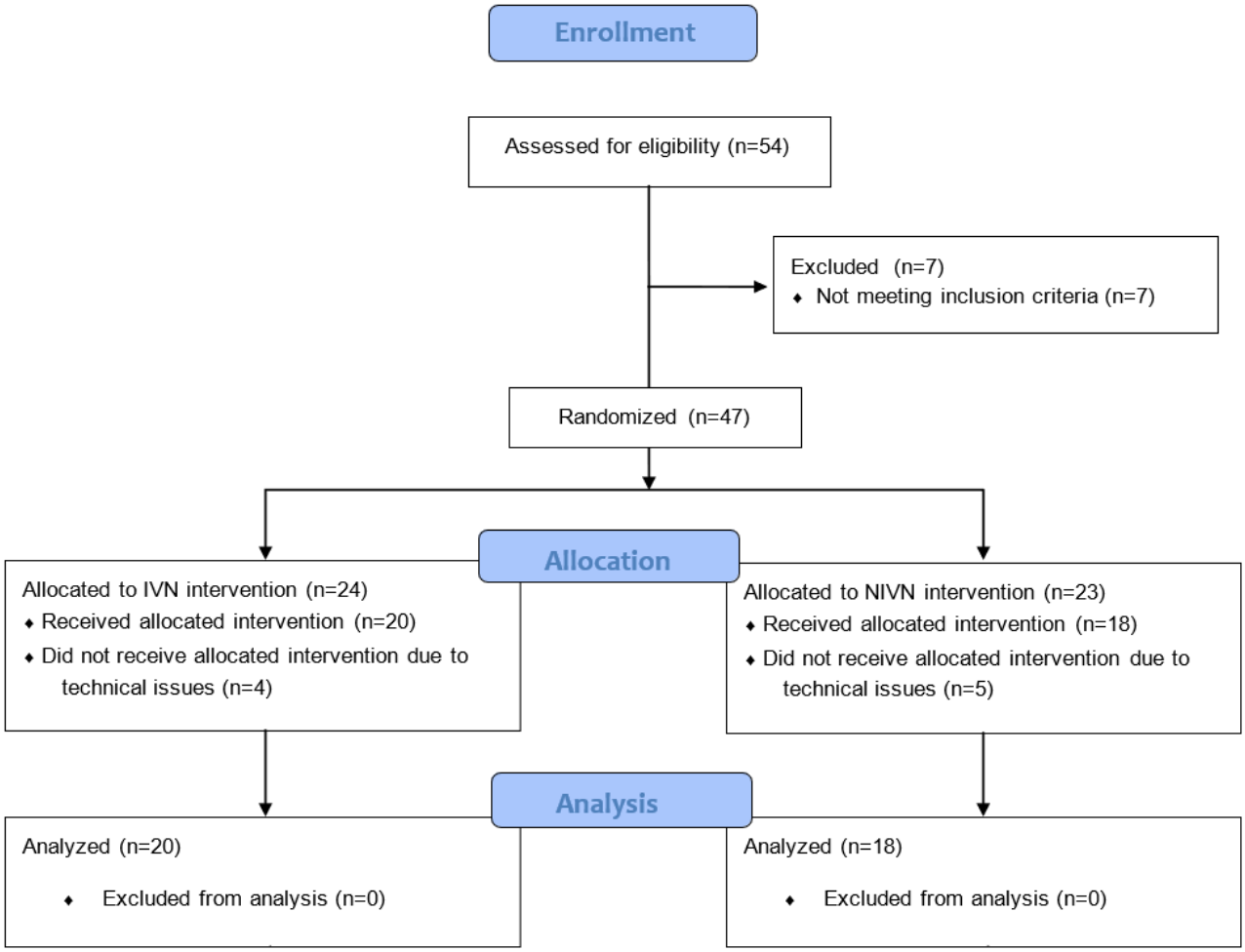
Ethical Considerations

A preliminary assessment by the Norwegian Regional Committees for Medical and Health Research Ethics determined that, in accordance with the Norwegian Health Research Act, the study did not require a full evaluation, as it did not involve medical or health interventions, the collection of sensitive personal data, or research on vulnerable populations. Accordingly, this randomized trial was not registered on a randomized controlled trial repository at the time of initiation. Subsequently, the trial was registered at the Norwegian Centre for Research Data, presently known as the Norwegian Agency for Shared Services in Education and Research (registration number 185017). Informed consent was obtained and delivered in written form both in English and Norwegian languages. It included information concerning the purpose of the study, contact details of both the project leader and the institution's data protection officer, legal basis for processing data, voluntary participation and the possibility to withdraw without having to give any reason for doing so, and anonymity. The data collected for the study have been deidentified to ensure participant confidentiality and privacy. A lottery prize consisting of an NOK 500 (US $55) gift card was offered. The study was conducted in compliance with the Guidelines for Research Ethics in the Social Sciences and the Humanities [83].

Results

A flow diagram of the experimental procedure and data collection process through the study phases is presented in Figure 2 [66].

Figure 2. CONSORT flow diagram of the study phases concerning participant enrollment, allocation, and analysis. IVN: immersive virtual nature; NIVN: nonimmersive virtual nature.



Hypothesis 1: Changes in Nature Engagement

The mixed between-within subjects ANOVA (Table 1) showed a significant pre- to postexposure assessment increase in both conditions, with a moderate to large effect size for INT-GE (Wilks Λ=0.87; $F_{1,36}$=5.55; P=.02; η_p^2=0.13) and a large effect size for CNS (Wilks Λ=0.52; $F_{1,36}$=33.49; P<.001; η_p^2=0.48); intention to visit the hiking location (Wilks Λ=0.70; $F_{1,36}$=15.34;

XSL•FO
RenderX

P<.001; η_p^2=0.30); and intention to participate in the hiking tour (Wilks Λ=0.74; $F_{1,36}$=12.450; P=.001; η_p^2=0.26).

Table 1. Descriptive results of pre and postexposure assessment of the main outcomes (N=38) and mixed between-within ANOVA.

Items	NIVN[a] (n=18), mean (SD)	IVN[b] (n=20), mean (SD)	Main effect of time		Time and condition interaction		Effect of condition	
			F test (*df*)	*P* value	*F* test (*df*)	*P* value	*F* test (*df*)	*P* value
CNS[c]			33.49 (1,36)	<.001	0.14 (1,36)	.71	0.85 (1,36)	.36
Before	4.65 (1.09)	4.42 (0.78)						
After	5.18 (1.04)	5.18 (1.04)						
INT-GE[d]			5.55 (1,36)	.02	0.31 (1,36)	.37	0.78 (1,36)	.38
Before	5.98 (1.20)	5.60 (1.66)						
After	6.24 (0.96)	5.79 (1.58)						
Intention to visit the hiking location			15.34 (1,36)	<.001	1.08 (1,36)	.31	0.06 (1,36)	.94
Before	4.48 (1.44)	4.28 (1.84)						
Before	5.07 (1.84)	5.02 (2.09)						
Intention to participate in the hiking tour			12.45 (1,36)	<.001	0.02 (1,36)	.89	0.55 (1,36)	.46
Before	6.72 (2.24)	6.25 (1.65)						
After	7.56 (1.82)	7.56 (1.82)						

[a]NIVN: nonimmersive virtual nature.

[b]IVN: immersive virtual nature.

[c]CNS: Connectedness to Nature Scale.

[d]INT-GE: Intention to Perform Green Exercise questionnaire.

Hypothesis 2: Differences Between Experimental Conditions

No statistically significant difference between baseline levels of the experimental conditions emerged for nature engagement outcomes and participants' characteristics (Table 2). No statistically significant condition by time point on effect was found for any of the main outcomes, indicating no difference in the pre- to postexposure assessment changes of the outcomes between the experimental conditions (Table 2). Of the 38 participants, 6 (16%) attended the hiking tour, 3 from each experimental condition (IVN: 3/20, 15% and NIVN: 3/18, 17%; possible differences between the conditions were not tested statistically). The Mann-Whitney U test for the items of presence (ie, "being there," "sense of reality," and "realism") revealed no statistically significant difference between the IVN and the NIVN conditions (Table 3). In total 35% (7/20) of participants in the IVN condition and 44% (8/18) of participants in the NIVN condition showed an SSQ total score higher than 20, which indicates severe symptoms of cybersickness, although the median values of the SSQ total score levels did not exceed the score of 20. The Mann-Whitney *U* test found no statistically significant differences between the experimental conditions for cybersickness.

XSL•FO
RenderX

Table 2. Participants' characteristics and evaluation of groups' baseline differences (N=38).

Characteristics	NIVN[a] (n=18)	IVN[b] (n=20)	Between-groups tests	*P* value
Age (y), mean (SD)	24.94 (7.07)	24.55 (3.41)	U=146.50	.33
Gender, n (%)			χ^2_1=0.05; ϕ=−0.09	.82
Men	7 (39)	6 (30)		
Women	11 (61)	14 (70)		
Nonbinary	0 (0)	0 (0)		
CNS[c], mean (SD)	4.65 (1.09)	4.42 (0.78)	Mean difference=−0.24 (95% CI −0.86 to 0.38); t_{38}=−0.78	.44
INT-GE[d], mean (SD)	5.98 (1.20)	5.60 (1.66)	Mean difference=−0.33 (95% CI −1.15 to 0.50); t_{38}=−0.80	.64
Intention to visit the hiking location, mean (SD)	4.48 (1.44)	4.28 (1.84)	Mean difference=−0.20 (95% CI −1.29 to 0.90); t_{38}=−0.37	.43
Intention to participate in the hiking tour, mean (SD)	6.72 (2.24)	6.25 (1.65)	Mean difference=−0.47 (95% CI −1.75 to 0.81); t_{38}=−0.74	.08
Previous visits to the hiking location, n (%)			χ^2_2=0.44; ϕ=0.11	.80
Yes	2 (11)	3 (15)		
No	13 (72)	15 (75)		
I do not remember	3 (17)	2 (10)		
Green exercise routine, n (%)			χ^2_1=0.66; ϕ=0.19	.42
Two times a week or more	10 (56)	8 (40)		
One time a week or less	7 (39)	12 (60)		

[a]NIVN: nonimmersive virtual nature.

[b]IVN: immersive virtual nature.

[c]CNS: Connectedness to Nature Scale.

[d]INT-GE: Intention to Perform Green Exercise questionnaire.

Table 3. Descriptive results of items of presence and cybersickness (N=38) stratified by groups and Mann-Whitney U test results.

Items	NIVN[a], median (IQR)	IVN[b], median (IQR)	Mann-Whitney *U* test	*P* value
Being there	8.00 (6.75-9.00)	7.50 (6.00-8.75)	203.00	.52
Realism	5.50 (1.75-7.00)	3.50 (1.50-4.75)	237.000	.10
Sense of reality	6.00 (3.00-8.00)	7.00 (5.25-8.75)	130.000	.15
SSQ[c] score	18.70 (7.48-46.75)	14.96 (4.67-33.66)	198.000	.60

[a]NIVN: nonimmersive virtual nature.

[b]IVN: immersive virtual nature.

[c]SSQ: Simulator Sickness Questionnaire.

Hypothesis 3: Association of Nature Engagement With Presence and Cybersickness

The correlational analysis performed among all outcomes from both IVN and NIVN conditions is presented in Table 4. The presence component "being there" showed a significant moderate and positive correlation with CNS (ρ=0.33; *P*=.03) and participation in the hiking tour (ρ=0.37; *P*=.02). However, it is important to stress that the validity of "being there" with participation in the hiking tour was limited by the small number of participants attending the hiking tour, hence caution is warranted in interpreting this outcome. Other significant correlations emerged among the indicators of nature engagement, specifically between intention to participate in the hiking tour and intention to visit the hiking location (ρ=0.39; *P*=.02) and between intention to visit the hiking location and INT-GE (ρ=0.34; *P*=.04). No other significant correlation emerged from the analysis.

XSL•FO
RenderX

Table 4. Spearman rank correlation coefficient of indicators of nature engagement in immersive virtual nature (IVN) and nonimmersive virtual nature (NIVN) conditions, expressed as differential scores from pre- to postexposure assessment (delta values), with presence and cybersickness (N=38).

Indicators of engagement with nature (delta values)	Being there	Realism	Sense of reality	SSQ[a] score	1	2	3	4
1. CNS[b]								
ρ	0.35	0.30	0.11	–0.21	—[c]	—	—	—
P value	.03[a]	.06	.51	.21	—	—	—	—
2. INT-GE[d]								
ρ	0.09	0.09	–0.20	–0.02	–0.14	—	—	—
P value	.58	.58	.23	.88	.41	—	—	—
3. Intention to visit the hiking location								
ρ	0.29	0.14	0.28	0.00	0.04	0.34	—	—
P value	.07	.40	.08	.99	.79	.04[e]	—	—
4. Intention to participate in the hiking tour								
ρ	0.21	0.31	0.22	0.24	0.07	0.25	0.39	—
P value	.21	.06	.19	.15	.70	.13	.02[e]	—
5. Participation in the hiking tour								
ρ	0.37	0.03	0.12	–0.23	0.23	0.29	0.06	0.23
P value	.02[a]	.87	.46	.16	.17	.12	.57	.71

[a]SSQ: Simulator Sickness Questionnaire.

[b]CNS: Connectedness to Nature Scale.

[c]Not applicable.

[d]INT-GE: Intention to Perform Green Exercise questionnaire.

[e]Correlation is significant at the .05 level (2-tailed test).

Discussion

Principal Findings

University students are often at a high risk for poor mental health [2-4] and tend to engage little with nature [4,7,33]. This randomized trial investigated the effectiveness of a single IVN or NIVN exposure in eliciting nature engagement among university students, examining whether the former can provide better effects compared to the latter. Possible associations of nature engagement outcomes with the participants' feelings of presence and cybersickness were also investigated. The findings showed an overall pre- to postexposure assessment increase in nature engagement outcomes for both conditions. Compared to NIVN, IVN did not provide additional benefits on nature engagement nor did it increase the likelihood of participation in an organized hiking tour. Furthermore, no statistically significant differences emerged between IVN and NIVN conditions concerning presence or cybersickness. Nature connectedness and participation in the hiking tour positively correlated with the presence item "being there." No significant correlation was found among any of the nature engagement outcomes and the other items of presence or cybersickness.

Promoting Nature Engagement Through Virtual Nature

Supporting prediction (hypothesis 1), when compared with pre-exposure, postexposure levels of nature engagement were significantly higher in both the IVN and the NIVN conditions. These results are largely in line with previous systematic reviews investigating the effects of direct and indirect contact with nature connectedness [36,57]. Different indirect nature exposure through audio-visual presentation devices, such as HMDs, computer screens, and nature images, can provide opportunities for fostering a person's sense of nature connection [10,36,47]. Furthermore, this study supports the previous research showing the effectiveness of virtual nature in fostering nature connectedness [27,36,84-86] and provides novel insights concerning the effectiveness of virtual nature as a tool for the promotion of physical engagement with nature, which has until now received little research attention [58].

Although the improvements observed in this study may seem relatively small, it should be noted that this is in line with previous studies [36,58,67], the effect size was moderate to large for INT-GE and large for the remaining outcomes. Furthermore, the average increment observed after just a single IVN exposure was considerable: 4% for INT-GE, 6% for CNS, 7% for the intention to visit the hiking location, and 8% for the intention to participate in the hiking tour. This suggests that even a short and relatively simple virtual nature experience

XSL•FO
RenderX

holds practical relevance in the pursuit of promoting nature engagement. As indicated by a previous systematic review [36], these outcomes may have long-term implications, especially when virtual nature experiences are repeatedly proposed over time. Moreover, the increased nature engagement elicited by the virtual nature experience may lead to more frequent contact with nature, thereby yielding additional long-term physical and mental health benefits. In this respect, although the actual participation in the hiking tour was low (6/38, 16% participants and 6/28, 21% participants who reported a high intention to participate in the organized hiking tour in the postexposure assessments), we cannot exclude that the virtual nature exposure may have elicited other forms of nature visits (eg, green exercise and independent visits to other naturalistic locations). Indeed, while participation in the hiking tour was bound to a relatively rigid time schedule, time flexibility and convenience are known to be important factors sustaining green exercise and frequent nature visits [87].

Immersive Versus Nonimmersive Experience?

Contrary to prediction (hypothesis 2), no statistical difference between the two conditions in the extent to which they elicited enhanced nature engagement outcomes emerged. This finding partly aligns with previous literature, although the evidence regarding this topic is still limited and largely mixed [57,88]. Indeed, while some studies suggest that IVN tends to elicit increased levels of nature connectedness to a greater extent than NIVN [47,63], other studies show no statistical differences [64,89]. While the findings of our study support the notion that immersion in itself has little or no impact on the nature engagement outcomes, other factors (eg, technology used, type of IVN scenarios, and degree of interactivity) may contribute to explaining the inconsistency of the findings in this field [10,47,58,90]. Therefore, it is important to stress the fact that our IVN condition was unable to generate greater feelings of presence compared with the NIVN condition. This may be explained by various factors: the relatively short duration of the 360° video, the lack of interactivity, the fact that the HMD was not paired with a supplementary high-quality headphone, and the low image resolution generally associated with consumer-oriented HMDs, especially if displaying 360° videos. Such an assumption finds support in the literature. For instance, a recent systematic review [57] investigating the effects of IVN on nature connectedness found consistent difference between IVN and NIVN only in studies using computer-generated scenarios (ie, computer graphics digitally recreating fictional or actual environments) but not among studies involving 360° videos. Moreover, it has been suggested that the levels of interactivity, as well as the quality and capability of the hardware used, as technological characteristics that enhance immersion, also play an important role in the extent to which IVN can effectively promote nature visits [58,67].

These findings prompt 2 reflections relative to the complex nature of immersion and the effectiveness of consumer-oriented VR technology. First, immersion is a fluid and complex construct [91], and despite that the levels of immersion are often linked to characteristics of the hardware such as frame rate and image resolution, it has been pointed out that software also plays a role in determining the immersion level [49]. Hence, having a 360° field of view and excluding the view of the real world may not be the sole technical aspect necessary to achieve a high level of immersion in IVN exposures. Second, consumer-oriented technology entails some limitations, especially regarding its inferior quality and capacity compared to state-of-the-art VR technology. Nevertheless, consumer-oriented technology bears practical advantages. For instance, 360° videos are often highly realistic, relatively easy to produce, and cost-effective. This is especially useful for replicating local natural environments in real-world promotional initiatives. Furthermore, VR technology represents an attractive form of entertainment for many individuals, with benefits in terms of public engagement. Hence, we envisage that with the ongoing improvement in the quality and capacity of consumer-oriented VR technology, its effectiveness in research and health-promotion initiatives will also increase.

Presence, More Than Immersion, as a Key Factor for Nature Engagement Promotion

Partially supporting hypothesis 3, nature connectedness and participation to the hiking tour were associated with the presence item “being there.” These findings suggest that, beyond the hardware used (HMD or tablet screen), disturbances (eg, cybersickness), and perceived fidelity of the scenario, the subjective sense of “being there” may be a pivotal element in linking virtual nature experience and nature engagement. It is important to stress that the validity of these correlations is limited by the small sample size, especially as only 6 participants attended the hiking tour. Nevertheless, this finding aligns with previous research. For instance, a recent study [89] showed that, while exposure to IVN did not enhance higher levels of nature connectedness compared to its NIVN counterpart, participants who experienced a higher sense of embodiment in the virtual world tended to reflect more about nature and their role toward it. Furthermore, other studies also found that “being there,” more than other aspects of presence, positively correlated with different psychological outcomes [10,75,79,92]. Indeed, this item (ie, in the computer-generated world I had the sense of “being there”), is considered a core component of presence [45], distinct from, although related to, “realism” (ie, the extent to which the participant perceived the virtual world as similar to the real world) and “sense of reality” (ie, the extent to which the virtual world became more present than the real world).

However, questions remain on how the feeling of “being there” may facilitate enhanced nature connectedness and motivation to visit nature. It has been suggested that anticipated psychological benefits [58] and conditioned responses [93] may play an important role in this respect. In line with such perspectives, a vivid sensation of “being in” a natural environment would trigger the recall of past experiences, alongside the psychological states associated with it. This would, in turn, foster positive attitudes toward nature as well as a desire of reenacting those positive experiences.

It is important to emphasize that a person’s sense of presence can be influenced by several subjective factors, regardless of the technology used to deliver a VR experience. Indeed, immersion and presence are interrelated but distinct concepts and, while it is believed that the former is a primary determinant

XSL•FO
RenderX

of the latter [94], their relationship is complex. As the feeling of presence is subjective and can, among other things, depend on individual personality traits [95], the exposure to the same type of technology may evoke varying levels of presence among different individuals. Hence, as highlighted by the findings of this study, some individuals may experience high levels of presence in NIVN exposures—although it is also important to consider the limitations of the proposed IVN condition and the extent to which it may be actually considered a fully immersive VR installation.

Cybersickness did not significantly correlate with measurements of nature engagement. However, it should be noted that the IVN condition did not generate noteworthy challenges regarding cybersickness. This is supported by the fact that, despite some participants reporting "severe" symptoms of cybersickness according to the cut-off provided by Kennedy et al [77], the median values for the SSQ score were considerably lower when compared with values reported in previous literature. Indeed, a systematic review [96] reported a pooled total SSQ mean score of 28 across 55 studies for VR experiences delivered through HMDs. Moreover, no significant difference in cybersickness symptoms was found between the IVN and the NIVN, indicating that the reported symptoms may be associated with pre-existing states. Indeed, several cybersickness symptoms, such as headache, fatigue, and difficulty focusing and concentrating, may be commonly experienced in daily life.

Strengths and Limitations

Overview

One of the strengths of this study lies in the novelty of investigating the effects of IVN compared with NIVN on outcomes of physical engagement with nature, both regarding behavioral intention and actual behavior. As highlighted by two recent systematic reviews [57,88], studies on this specific topic are sparse, and further evidence on nature connectedness outcomes is needed. Moreover, to the best of our knowledge, this is the only randomized study that compared the effects of IVN versus NIVN on such outcomes, exploring factors that may influence the effectiveness of IVN, such as presence and cybersickness. Furthermore, the paper is outlined following the CONSORT guidelines [97]. However, some limitations need to be considered.

Limitations Relative to Design

First, the lack of a control group (ie, without virtual stimulation) limits the extent to which this study can establish a causal relationship between the experimental conditions and the outcomes. Second, the proposed IVN condition, consisting of a 360° video delivered through a stand-alone HMD, may not be considered fully immersive. Different findings may have emerged if a different technology (eg, including a high-end HMD or an interactive experience) was used.

Limitations Relative to the Sample

Different factors may have influenced the results regarding the participation in green exercise and the organized hiking tour. For instance, we purposefully targeted a population that tends to prefer indoor activities. Moreover, the study was conducted during the academic semester and examination sessions. While this strengthens the ecologic validity of the study, the tour dates may have conflicted with participants' study-related commitments. In addition, the demanding intensity of the tour (approximately 3 hours duration) may have been an impediment for some participants. Information regarding the participants' course of study was not recorded. However, we know that the students dwelling on the campus are largely linked to courses in nursing, computer sciences, teacher education, management, and political sciences—the campus does not offer courses in sports sciences or outdoor recreation. The participants may have had a personal interest in nature-based activities, which we were able to compute through the preassessments. Nevertheless, this could still be a potential source of bias, as the individual's interest in nature may have influenced both attendance and responses to the study.

The sample size calculation was based on an expected medium effect size, as in line with previous studies [47,67]. However, these studies used different VR technology, which succeeded in eliciting stronger experiences of presence, possibly leading to larger changes in nature connectedness. Hence, the sample size calculation for our study may have been underpowered. In addition, it should be noted that the sample size calculation was performed based on nature connectedness and a mixed between-within subjects ANOVA (required to address hypotheses 1 and 2); hence, the sample may be underpowered for the other outcome variables and for the correlation analysis.

Limitations Relative to the Instruments

The CNS, INT-GE, and SSQ, as well as the presence items, were previously validated through psychometric studies and are generally accepted as valid instruments. The Norwegian translations of these scales were used in previous similar research [58,67], showing adequate internal consistency and patterns similar to their original counterparts. However, the Norwegian translations have not been validated through comprehensive psychometric investigations. Relative to the use of intention-based measurements, it should be noted that meta-analysis evidence found a significant, although weak, association between intention and behavior in various contexts [98-100]. This highlights the value that such assessment may hold in research focused on health- or well-being–related behavior. However, the measure of intention to visit the hiking location was developed based on established guidelines [73] but not validated through comprehensive psychometric investigations. Hence, caution is warranted when interpreting the findings relative to these outcomes. These instruments also showed very high Cronbach α values, which may indicate redundancy among items [101]. Finally, some considerations should be made regarding the cultural context of this study. In Norway, there is an abundance of natural environments, and this may have impacted the results of the study concerning intentions to visit nature.

Implications for Practice and Future Research

These findings have significant implications in public and private institutions and organizations' initiatives seeking to promote nature engagement among university students. Given the health benefits associated with nature engagement and the

XSL•FO
RenderX

progressive disconnection from nature among young adults [14,42], schools, universities, municipalities, and health-promoting organizations may use virtual nature interventions, either in the form of low-end IVN or NIVN, as a tool in health-promoting and environmental education initiatives. Additional practical implications involve using virtual nature in interventions and initiatives across diverse contexts and population groups. For instance, urban citizens also often experience challenges accessing nature daily and may particularly benefit from virtual nature interventions. Businesses may also use virtual nature as a health-promotion activity to foster nature engagement among employees as an adaptive coping strategy. The use of virtual nature in such contexts and target groups should be further investigated.

The evidence on the effects of virtual nature on physical engagement with nature is still in its infancy and warrants more research. This should be appraised through either follow-up assessments measuring participants' frequency of nature visitation or proposing opportunities for in-real-life nature experiences. Recommendation for future research includes comparisons of multiple experimental conditions and possibly a control condition. For instance, highly immersive VR technologies, such as high-end VR systems and interactive IVN experiences, should be considered as worthy of investigation. In addition, potential seasonal variations in nature engagement could be explored by diversified virtual nature environments. Moreover, while the ecologic validity of the study lies in it being conducted during the examination sessions, exploring the effects of virtual nature environments in different academic semesters, where students have varying study-related commitments, may also be worth investigating. Furthermore, little is known about repeated exposures and follow-up assessments exploring the longevity and stability of the acute change in nature engagement following virtual nature exposure. Given the rapid development of VR technology, further investigation may use or compare different hardware and software.

Conclusions

This study shows that a single exposure to a 5-minute 360° nature video, delivered through either a stand-alone HMD or a tablet screen, elicited significant improvements in nature engagement outcomes among university students. These outcomes were associated only to a limited extent with the individual's sense of presence perceived during the virtual nature experience. No association was found concerning cybersickness. These findings indicate that virtual nature can offer the opportunity to connect with actual nature among university students. Considering the mental health and well-being benefits associated with nature experiences, these results can inform future research design and intervention among other groups.

Acknowledgments

The authors thank the University of South-Eastern Norway for their support and resources throughout this research.

The authors disclosed receipt of financial support from their respective affiliations for the research, authorship, and publication of this paper. The University of South-East Norway founded this study as a component of the PhD project of EB, and the authors contribution to this study.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

EB was responsible for conceptualization; data curation; formal analysis; investigation; methodology; software; original draft, review, and editing of the manuscript. KS was responsible for conceptualization and original draft, review, and editing of the manuscript. VS was responsible for conceptualization and original draft, review, and editing of the manuscript. GC was responsible for conceptualization; supervision; data curation; methodology; and original draft, review, and editing of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Overall questionnaires.
[DOCX File , 54 KB - games_v13i1e63424_app1.docx]

Multimedia Appendix 2
CONSORT 2010 checklist.
[PDF File (Adobe PDF File), 174 KB - games_v13i1e63424_app2.pdf]

References

XSL•FO
RenderX

1. Depression and other common mental disorders: global health estimates. World Health Organization. 2017. URL: https://www.who.int/publications/i/item/depression-global-health-estimates [accessed 2025-05-29]
2. Auerbach RP, Alonso J, Axinn WG, Cuijpers P, Ebert DD, Green JG, et al. Mental disorders among college students in the World Health Organization World Mental Health surveys. Psychol Med 2016 Oct;46(14):2955-2970 [FREE Full text] [doi: 10.1017/S0033291716001665] [Medline: 27484622]
3. Leppink EW, Odlaug BL, Lust K, Christenson G, Grant JE. The young and the stressed: stress, impulse control, and health in college students. J Nerv Ment Dis 2016 Dec;204(12):931-938. [doi: 10.1097/NMD.0000000000000586] [Medline: 27575792]
4. Students living conditions, 2021. Statistics Norway. URL: https://www.ssb.no/en/utdanning/hoyere-utdanning/artikler/students-living-conditions-2021 [accessed 2025-05-29]
5. White MP, Alcock I, Grellier J, Wheeler BW, Hartig T, Warber SL, et al. Spending at least 120 minutes a week in nature is associated with good health and wellbeing. Sci Rep 2019 Jul 13;9(1):7730 [FREE Full text] [doi: 10.1038/s41598-019-44097-3] [Medline: 31197192]
6. Calogiuri G, Patil GG, Aamodt G. Is green exercise for all? A descriptive study of green exercise habits and promoting factors in adult Norwegians. Int J Environ Res Public Health 2016 Dec 23;13(11):1165 [FREE Full text] [doi: 10.3390/ijerph13111165] [Medline: 27886098]
7. Shores KA, West ST. Pursuing leisure during leisure-time physical activity. J Phys Act Health 2010 Sep;7(5):685-694. [doi: 10.1123/jpah.7.5.685] [Medline: 20864766]
8. Larson LR, Green GT, Cordell HK. Children's time outdoors: results from a national survey. J Park Recreat Adm 2011;29(2):1-20 [FREE Full text]
9. Larson LR, Szczytko R, Bowers EP, Stephens LE, Stevenson KT, Floyd MF. Outdoor time, screen time, and connection to nature: troubling trends among rural youth? Environ Behav 2018 Oct 20;51(8):966-991. [doi: 10.1177/0013916518806686]
10. Litleskare SE, MacIntyre TE, Calogiuri G. Enable, reconnect and augment: a new ERA of virtual nature research and application. Int J Environ Res Public Health 2020 Mar 06;17(5):1738 [FREE Full text] [doi: 10.3390/ijerph17051738] [Medline: 32155911]
11. Calogiuri G, Litleskare S, MacIntyre TE. Future-thinking through technological nature. In: Donnelly AA, MacIntyre TE, editors. Physical Activity in Natural Settings. New York, NY: Routledge; 2019:279-298.
12. Urban green spaces and health. WHO Regional Office for Europe. URL: https://www.euro.who.int/__data/assets/pdf_file/0005/321971/Urban-green-spaces-and-health-review-evidence.pdf [accessed 2025-05-29]
13. Bowler DE, Buyung-Ali LM, Knight TM, Pullin AS. A systematic review of evidence for the added benefits to health of exposure to natural environments. BMC Public Health 2010 Aug 04;10(1):456 [FREE Full text] [doi: 10.1186/1471-2458-10-456] [Medline: 20684754]
14. Bratman GN, Anderson CB, Berman MG, Cochran B, de Vries S, Flanders J, et al. Nature and mental health: an ecosystem service perspective. Sci Adv 2019 Jul 05;5(7):eaax0903 [FREE Full text] [doi: 10.1126/sciadv.aax0903] [Medline: 31355340]
15. Kemperman A, Timmermans H. Green spaces in the direct living environment and social contacts of the aging population. Landsc Urban Plan 2014 Sep 26;129(25):44-54 [FREE Full text] [doi: 10.1016/j.landurbplan.2014.05.003]
16. Hartig T, Mitchell R, de Vries S, Frumkin H. Nature and health. Annu Rev Public Health 2014 Mar 18;35(1):207-228. [doi: 10.1146/annurev-publhealth-032013-182443] [Medline: 24387090]
17. Rogerson M, Barton J, Pretty J, Gladwell V. The green exercise concept: two intertwining pathways to health and well-being. In: Donnelly AA, MacIntyre TE, editors. Physical Activity in Natural Settings: Green and Blue Exercise. New York, NY: Routledge; 2019:75-94.
18. Pretty J, Griffin M, Sellens M, Pretty C. Green exercise: complementary roles of nature, exercise and diet in physical and emotional well-being and implications for public health policy. University of Essex. 2003. URL: https://tinyurl.com/4s6a52ue
19. Capaldi CA, Dopko RL, Zelenski JM. The relationship between nature connectedness and happiness: a meta-analysis. Front Psychol 2014 Sep 08;5:976 [FREE Full text] [doi: 10.3389/fpsyg.2014.00976] [Medline: 25249992]
20. Cervinka R, Röderer K, Hefler E. Are nature lovers happy? On various indicators of well-being and connectedness with nature. J Health Psychol 2012 May 22;17(3):379-388. [doi: 10.1177/1359105311416873] [Medline: 21859800]
21. Lafortezza R, Carrus G, Sanesi G, Davies C. Benefits and well-being perceived by people visiting green spaces in periods of heat stress. Urban For Urban Green 2009 Jan;8(2):97-108. [doi: 10.1016/j.ufug.2009.02.003]
22. Priest P. The healing balm effect: using a walking group to feel better. J Health Psychol 2007 Jan 01;12(1):36-52. [doi: 10.1177/1359105307071734] [Medline: 17158839]
23. Mayer FS, Frantz CM. The connectedness to nature scale: a measure of individuals' feeling in community with nature. J Environ Psychol 2004 Dec;24(4):503-515. [doi: 10.1016/j.jenvp.2004.10.001]
24. Howell AJ, Dopko RL, Passmore HA, Buro K. Nature connectedness: associations with well-being and mindfulness. Pers Individ Dif 2011 Jul;51(2):166-171. [doi: 10.1016/j.paid.2011.03.037]
25. Martin L, White MP, Hunt A, Richardson M, Pahl S, Burt J. Nature contact, nature connectedness and associations with health, wellbeing and pro-environmental behaviours. J Environ Psychol 2020 Apr;68:101389. [doi: 10.1016/j.jenvp.2020.101389]

XSL•FO
RenderX

26. Pritchard A, Richardson M, Sheffield D, McEwan K. The relationship between nature connectedness and eudaimonic well-being: a meta-analysis. J Happiness Stud 2019 Apr 30;21(3):1145-1167. [doi: 10.1007/S10902-019-00118-6]
27. Mayer FS, Frantz CM, Bruehlman-Senecal E, Dolliver K. Why is nature beneficial: the role of connectedness to nature. Environ Behav 2008 Sep 05;41(5):607-643. [doi: 10.1177/0013916508319745]
28. Pálsdóttir AM, Grahn P, Persson D. Changes in experienced value of everyday occupations after nature-based vocational rehabilitation. Scand J Occup Ther 2014 Jan 17;21(1):58-68 [FREE Full text] [doi: 10.3109/11038128.2013.832794] [Medline: 24041155]
29. van den Berg M, van Poppel M, van Kamp I, Andrusaityte S, Balseviciene B, Cirach M, et al. Visiting green space is associated with mental health and vitality: a cross-sectional study in four european cities. Health Place 2016 Mar;38:8-15. [doi: 10.1016/j.healthplace.2016.01.003] [Medline: 26796323]
30. Tiscareno-Osorno X, Demetriou Y, Marques A, Peralta M, Jorge R, MacIntyre TE, et al. Systematic review of explicit instruments measuring nature connectedness: what do we know and what is next? Environ Behav 2023 Dec 28;55(8-10):551-608 [FREE Full text] [doi: 10.1177/00139165231212321]
31. Hughes J, Rogerson M, Barton J, Bragg R. Age and connection to nature: when is engagement critical? Frontiers in Ecol & Environ 2019 Apr 23;17(5):265-269 [FREE Full text] [doi: 10.1002/fee.2035] [Medline: 31080557]
32. Mena-García A, Olivos P, Loureiro A, Navarro O. Effects of contact with nature on connectedness, environmental identity and evoked contents (Efectos del contacto con la naturaleza en conectividad, identidad ambiental y contenidos evocados). PsyEcology 2020 Feb 01;11(1):21-36. [doi: 10.1080/21711976.2019.1643663]
33. Calogiuri G. Natural environments and childhood experiences promoting physical activity, examining the mediational effects of feelings about nature and social networks. Int J Environ Res Public Health 2016 May 21;13(4):439 [FREE Full text] [doi: 10.3390/ijerph13040439] [Medline: 27110802]
34. Rosa CD, Profice CC, Collado S. Nature experiences and adults' self-reported pro-environmental behaviors: the role of connectedness to nature and childhood nature experiences. Front Psychol 2018 Jun 26;9:1055 [FREE Full text] [doi: 10.3389/fpsyg.2018.01055] [Medline: 30013494]
35. Steg LE, Van Den Berg AE, De Groot JI. Environmental Psychology: An Introduction. Hoboken, NJ: Wiley–Blackwell; 2013.
36. Sheffield D, Butler CW, Richardson M. Improving nature connectedness in adults: a meta-analysis, review and agenda. Sustainability 2022 Sep 30;14(19):12494-12479. [doi: 10.3390/su141912494]
37. Whitburn J, Linklater WL, Milfont TL. Exposure to urban nature and tree planting are related to pro-environmental behavior via connection to nature, the use of nature for psychological restoration, and environmental attitudes. Environ Behav 2019;51(7):787-810 [FREE Full text] [doi: 10.1177/0013916519841392]
38. Thompson CW, Aspinall P, Montarzino A. The childhood factor: adult visits to green places and the significance of childhood experience. Environ Behav 2007 Aug 02;40(1):111-143 [FREE Full text] [doi: 10.1177/0013916507300119]
39. Calogiuri G, Chroni S. The impact of the natural environment on the promotion of active living: an integrative systematic review. BMC Public Health 2014 Aug 24;14(1):873 [FREE Full text] [doi: 10.1186/1471-2458-14-873] [Medline: 25150711]
40. Leung GY, Hazan H, Chan CS. Exposure to nature in immersive virtual reality increases connectedness to nature among people with low nature affinity. J Environ Psychol 2022 Oct 03;83(7):101863-101834 [FREE Full text] [doi: 10.1016/j.jenvp.2022.101863]
41. Ritchie H, Roser M. Urbanization. Our World Data. 2018. URL: https://ourworldindata.org/urbanization [accessed 2025-05-29]
42. Rajoo KS, Karam DS, Abdullah MZ. The physiological and psychosocial effects of forest therapy: a systematic review. Urban For Urban Green 2020 Oct;54:126744. [doi: 10.1016/j.ufug.2020.126744]
43. Louv R. Last Child in the Woods: Saving Our Children from Nature-Deficit Disorder. Berlin, Germany: Algonquin Books; 2008.
44. Heim M. Interactive design: tunnel or spiral. In: Heim M, editor. Virtual Realism. Oxford, UK: Oxford University Press; 1998.
45. Slater M, Wilbur S. A framework for immersive virtual environments (FIVE): speculations on the role of presence in virtual environments. Presence (Camb) 1997 Dec;6(6):603-616. [doi: 10.1162/pres.1997.6.6.603]
46. Browning MH, Mimnaugh KJ, van Riper CJ, Laurent HK, LaValle SM. Can simulated nature support mental health? Comparing short, single-doses of 360-degree nature videos in virtual reality with the outdoors. Front Psychol 2019 Jan 15;10:2667 [FREE Full text] [doi: 10.3389/fpsyg.2019.02667] [Medline: 32010003]
47. Yeo NL, White MP, Alcock I, Garside R, Dean SG, Smalley AJ, et al. What is the best way of delivering virtual nature for improving mood? An experimental comparison of high definition TV, 360° video, and computer generated virtual reality. J Environ Psychol 2020 Dec;72:101500 [FREE Full text] [doi: 10.1016/j.jenvp.2020.101500] [Medline: 33390641]
48. Botella C, Fernández-Álvarez J, Guillén V, García-Palacios A, Baños R. Recent progress in virtual reality exposure therapy for phobias: a systematic review. Curr Psychiatry Rep 2017 Jul;19(7):42. [doi: 10.1007/s11920-017-0788-4] [Medline: 28540594]
49. Bowman DA, McMahan RP. Virtual reality: how much immersion is enough? Computer 2007;40(7):36-43. [doi: 10.1109/MC.2007.257]

XSL•FO
RenderX

50. Grassini S, Laumann K, Rasmussen Skogstad M. The use of virtual reality alone does not promote training performance (but sense of presence does). Front Psychol 2020;11:1743 [FREE Full text] [doi: 10.3389/fpsyg.2020.01743] [Medline: 32765384]
51. Steuer J. Defining virtual reality: dimensions determining telepresence. J Commun 1992;42(4):73-93. [doi: 10.1111/j.1460-2466.1992.tb00812.x]
52. Triberti S, Repetto C, Riva G. Psychological factors influencing the effectiveness of virtual reality-based analgesia: a systematic review. Cyberpsychol Behav Soc Netw 2014 Jul;17(6):335-345. [doi: 10.1089/cyber.2014.0054] [Medline: 24892195]
53. Diemer J, Alpers GW, Peperkorn HM, Shiban Y, Mühlberger A. The impact of perception and presence on emotional reactions: a review of research in virtual reality. Front Psychol 2015;6:26 [FREE Full text] [doi: 10.3389/fpsyg.2015.00026] [Medline: 25688218]
54. Litleskare S. Virtual green exercise-developing a new concept for health promotion. University of South-Eastern Norway. URL: https://openarchive.usn.no/usn-xmlui/handle/11250/3018427#:~:text=Virtual%20green%20exercise%20seeks%20to,based%20experiments%20that%20simulate%20outdoor [accessed 2025-05-29]
55. Kennedy RS, Drexler J, Kennedy RC. Research in visually induced motion sickness. Appl Ergon 2010 Jul;41(4):494-503. [doi: 10.1016/j.apergo.2009.11.006] [Medline: 20170902]
56. Smith JW. Immersive virtual environment technology to supplement environmental perception, preference and behavior research: a review with applications. Int J Environ Res Public Health 2015 Oct 11;12(9):11486-11505 [FREE Full text] [doi: 10.3390/ijerph120911486] [Medline: 26378565]
57. Brambilla E, Petersen E, Stendal K, Sundling V, MacIntyre TE, Calogiuri G. Effects of immersive virtual nature on nature connectedness: a systematic review protocol. Digit Health 2022;8:20552076221120324 [FREE Full text] [doi: 10.1177/20552076221120324] [Medline: 36081751]
58. Calogiuri G, Keegan BJ, Birkheim SL, Rydgren TL, Flaten OE, Fröhlich F, et al. A mixed-methods exploration of virtual reality as a tool to promote green exercise. Sci Rep 2022 May 05;12(1):5715 [FREE Full text] [doi: 10.1038/s41598-022-09622-x] [Medline: 35383263]
59. Gibson A, O'Rawe M. Virtual reality as a travel promotional tool: insights from a consumer travel fair. In: Jung T, Claudia tom Dieck M, editors. Augmented Reality and Virtual Reality: Empowering Human, Place and Business. Cham, Switzerland: Springer; 2018:93-107.
60. Potter LE, Carter L, Coghlan A. Virtual reality and nature based tourism. In: Proceedings of the 28th Australian Conference on Computer-Human Interaction. 2016 Presented at: OzCHI '16; November 29-December 2, 2016; Launceston, Australian p. 652-654 URL: https://dl.acm.org/doi/10.1145/3010915.3011854 [doi: 10.1145/3010915.3011854]
61. Tussyadiah IP, Wang D, Jung TH, tom Dieck MC. Virtual reality, presence, and attitude change: empirical evidence from tourism. Tour Manag 2018 Jun;66:140-154. [doi: 10.1016/j.tourman.2017.12.003]
62. Calogiuri G, Litleskare S, Fröhlich F. Physical activity and virtual nature: perspectives on the health and behavioral benefits of virtual green exercise. In: Brymer E, Rogerson M, Barton J, editors. Nature and Health: Physical Activity in Nature. New York, NY: Routledge; 2021:127-146.
63. Ahn SJ, Bostick J, Ogle E, Nowak KL, McGillicuddy KT, Bailenson JN. Experiencing nature: embodying animals in immersive virtual environments increases inclusion of nature in self and involvement with nature. J Comput Mediat Comm 2016 Sep 14;21(6):399-419. [doi: 10.1111/jcc4.12173]
64. Filter E, Eckes A, Fiebelkorn F, Büssing AG. Virtual reality nature experiences involving wolves on YouTube: presence, emotions, and attitudes in immersive and nonimmersive settings. Sustainability 2020 May 08;12(9):3823. [doi: 10.3390/su12093823]
65. Soliman M, Peetz J, Davydenko M. The impact of immersive technology on nature relatedness and pro environmental behavior. J Media Psychol 2017 Jan;29(1):8-17. [doi: 10.1027/1864-1105/a000213]
66. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. J Pharmacol Pharmacother 2010 Jul;1(2):100-107 [FREE Full text] [doi: 10.4103/0976-500X.72352] [Medline: 21350618]
67. Calogiuri G, Petersen E, Haile A, Flaten OE, Fröhlich F, Litleskare S. The impact of visualization techniques of immersive virtual scenarios in promoting nature connectedness: a blind randomized controlled trial with mixed-methods approach. J Environ Psychol 2023 Sep;90:102102. [doi: 10.1016/j.jenvp.2023.102102]
68. Youtube audio library. Youtube. URL: https://www.youtube.com/audiolibrary [accessed 2025-05-29]
69. Dronningslepet_video_screen. Elena Brambilla YouTube page. URL: https://www.youtube.com/watch?v=GrbscDRl0WI [accessed 2025-05-29]
70. Kleespies MW, Braun T, Dierkes PW, Wenzel V. Measuring connection to nature—a illustrated extension of the inclusion of nature in self scale. Sustainability 2021 Feb 06;13(4):1761. [doi: 10.3390/su13041761]
71. Flowers EP, Freeman P, Gladwell VF. The development of three questionnaires to assess beliefs about green exercise. Int J Environ Res Public Health 2017 Oct 04;14(10):1172 [FREE Full text] [doi: 10.3390/ijerph14101172] [Medline: 28976924]

XSL•FO
RenderX

72. Poli L, Greco G, Gabriele M, Pepe I, Centrone C, Cataldi S, et al. Effect of outdoor cycling, virtual and enhanced reality indoor cycling on heart rate, motivation, enjoyment and intention to perform green exercise in healthy adults. J Funct Morphol Kinesiol 2024 Oct 02;9(4):183 [FREE Full text] [doi: 10.3390/jfmk9040183] [Medline: 39449477]
73. Francis J, Eccles MP, Johnston M, Walker A, Grimshaw JM, Foy R, et al. Constructing questionnaires based on the theory of planned behaviour: a manual for health services researchers. Centre for Health Services Research, University of Newcastle. 2004. URL: https://openaccess.city.ac.uk/id/eprint/1735/ [accessed 2025-05-29]
74. Slater M, Usoh M, Steed A. Depth of presence in virtual environments. Presence (Camb) 1994 Jan;3(2):130-144. [doi: 10.1162/pres.1994.3.2.130]
75. Calogiuri G, Litleskare S, Fagerheim KA, Rydgren TL, Brambilla E, Thurston M. Experiencing nature through immersive virtual environments: environmental perceptions, physical engagement, and affective responses during a simulated nature walk. Front Psychol 2017 Jan 23;8:2321 [FREE Full text] [doi: 10.3389/fpsyg.2017.02321] [Medline: 29410635]
76. Nichols S, Haldane C, Wilson JR. Measurement of presence and its consequences in virtual environments. Int J Hum Comput Stud 2000 Mar;52(3):471-491. [doi: 10.1006/ijhc.1999.0343]
77. Kennedy RS, Lane NE, Berbaum KS, Lilienthal MG. Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. Int J Aviat Psychol 1993 Jul;3(3):203-220. [doi: 10.1207/s15327108ijap0303_3]
78. Litleskare S. The relationship between postural stability and cybersickness: it's complicated - an experimental trial assessing practical implications of cybersickness etiology. Physiol Behav 2021 Jul 01;236:113422. [doi: 10.1016/j.physbeh.2021.113422] [Medline: 33839164]
79. Litleskare S, Calogiuri G. Camera stabilization in 360° videos and its impact on cyber sickness, environmental perceptions, and psychophysiological responses to a simulated nature walk: a single-blinded randomized trial. Front Psychol 2019 Nov 1;10:2436 [FREE Full text] [doi: 10.3389/fpsyg.2019.02436] [Medline: 31736832]
80. Tabachnick BG, Fidell LS. Using Multivariate Statistics. 6th edition. New York, NY: Pearson; 2013.
81. Pallant J. SPSS Survival Manual: A Step by Step Guide to Data Analysis using IBM SPSS. Oxford, UK: McGraw-Hill; 2020.
82. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd edition. New York, NY: Routledge; 1988.
83. Guidelines for research ethics in the social sciences and the humanities. National Research Ethics Committees (NESH). URL: https://www.forskningsetikk.no/en/guidelines/social-sciences-and-humanities/guidelines-for-research-ethics-in-the-social-sciences-and-the-humanities/ [accessed 2025-05-29]
84. Arbuthnott KD, Sutter GC, Heidt CT. Natural history museums, parks, and connection with nature. Mus Manag Curatorsh 2014 Feb 28;29(2):102-121. [doi: 10.1080/09647775.2014.888818]
85. Chan SH, Qiu L, Esposito G, Mai KP, Tam KP, Cui J. Nature in virtual reality improves mood and reduces stress: evidence from young adults and senior citizens. Virtual Real 2021 Dec 26:1-16 [FREE Full text] [doi: 10.1007/s10055-021-00604-4] [Medline: 34849087]
86. Nisbet EK, Zelenski JM. Underestimating nearby nature: affective forecasting errors obscure the happy path to sustainability. Psychol Sci 2011 Oct 09;22(9):1101-1106. [doi: 10.1177/0956797611418527] [Medline: 21828351]
87. Calogiuri G, Elliott LR. Why do people exercise in natural environments? Norwegian adults' motives for nature-, gym-, and sports-based exercise. Int J Environ Res Public Health 2017 May 04;14(4):377 [FREE Full text] [doi: 10.3390/ijerph14040377] [Medline: 28375192]
88. Spano G, Theodorou A, Reese G, Carrus G, Sanesi G, Panno A. Virtual nature, psychological and psychophysiological outcomes: a systematic review. J Environ Psychol 2023 Aug;89:102044. [doi: 10.1016/j.jenvp.2023.102044]
89. Spangenberger P, Geiger SM, Freytag SC. Becoming nature: effects of embodying a tree in immersive virtual reality on nature relatedness. Sci Rep 2022 Jan 25;12(1):1311 [FREE Full text] [doi: 10.1038/s41598-022-05184-0] [Medline: 35079051]
90. Nukarinen T, Rantala J, Korpela K, Browning MH, Istance HO, Surakka V, et al. Measures and modalities in restorative virtual natural environments: an integrative narrative review. Comput Human Behav 2022 Jan;126:107008. [doi: 10.1016/j.chb.2021.107008]
91. Nilsson NC, Nordahl R, Serafin S. Immersion Revisited: a review of existing definitions of immersion and their relation to different theories of presence. Hum Technol 2016 Nov 30;12(2):108-134. [doi: 10.17011/ht/urn.201611174652]
92. Litleskare S, Fröhlich F, Flaten OE, Haile A, Kjøs Johnsen S, Calogiuri G. Taking real steps in virtual nature: a randomized blinded trial. Virtual Real 2022;26(4):1777-1793 [FREE Full text] [doi: 10.1007/s10055-022-00670-2] [Medline: 35818369]
93. Egner LE, Sütterlin S, Calogiuri G. Proposing a framework for the restorative effects of nature through conditioning: conditioned restoration theory. Int J Environ Res Public Health 2020 Sep 17;17(18):53-71 [FREE Full text] [doi: 10.3390/ijerph17186792] [Medline: 32957693]
94. Cummings JJ, Bailenson JN. How immersive is enough? A meta-analysis of the effect of immersive technology on user presence. Media Psychol 2015 May 15;19(2):272-309. [doi: 10.1080/15213269.2015.1015740]
95. Kober SE, Neuper C. Personality and presence in virtual reality: does their relationship depend on the used presence measure? Int J Hum Comput Interact 2013 Jan;29(1):13-25. [doi: 10.1080/10447318.2012.668131]

XSL•FO
RenderX

96. Saredakis D, Szpak A, Birckhead B, Keage HA, Rizzo A, Loetscher T. Factors associated with virtual reality sickness in head-mounted displays: a systematic review and meta-analysis. Front Hum Neurosci 2020;14:96 [FREE Full text] [doi: 10.3389/fnhum.2020.00096] [Medline: 32300295]
97. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. BMJ 2010 Mar 23;340(5):c869-c706 [FREE Full text] [doi: 10.1136/bmj.c869] [Medline: 20332511]
98. Eccles MP, Hrisos S, Francis J, Kaner EF, Dickinson HO, Beyer F, et al. Do self- reported intentions predict clinicians' behaviour: a systematic review. Implement Sci 2006 Dec 21;1(1):28 [FREE Full text] [doi: 10.1186/1748-5908-1-28] [Medline: 17118180]
99. Rhodes RE, Dickau L. Experimental evidence for the intention-behavior relationship in the physical activity domain: a meta-analysis. Health Psychol 2012 Nov;31(6):724-727. [doi: 10.1037/a0027290] [Medline: 22390739]
100. Webb TL, Sheeran P. Does changing behavioral intentions engender behavior change? A meta-analysis of the experimental evidence. Psychol Bull 2006 Mar;132(2):249-268. [doi: 10.1037/0033-2909.132.2.249] [Medline: 16536643]
101. Tavakol M, Dennick R. Making sense of Cronbach's alpha. Int J Med Educ 2011 Jul 27;2:53-55 [FREE Full text] [doi: 10.5116/ijme.4dfb.8dfd] [Medline: 28029643]

Abbreviations

CNS: Connectedness to Nature Scale
CONSORT: Consolidated Standards of Reporting Trials
HMD: head-mounted display
INT-GE: Intention to Perform Green Exercise questionnaire
IVN: immersive virtual nature
NIVN: nonimmersive virtual nature
SSQ: Simulator Sickness Questionnaire
VR: virtual reality

Edited by A Coristine; submitted 20.06.24; peer-reviewed by H Liu, F Borghesi; comments to author 13.12.24; revised version received 05.03.25; accepted 02.05.25; published 16.06.25.

Please cite as:
Brambilla E, Stendal K, Sundling V, Calogiuri G
Nature Engagement Outcomes of Viewing Nature Through a 360° Video or a Tablet Screen: Randomized Trial
JMIR Serious Games 2025;13:e63424
URL: https://games.jmir.org/2025/1/e63424
doi:10.2196/63424
PMID:

©Elena Brambilla, Karen Stendal, Vibeke Sundling, Giovanna Calogiuri. Originally published in JMIR Serious Games (https://games.jmir.org), 16.06.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Effects of The Legend of Zelda: Breath of the Wild and Studio Ghibli Films on Young People's Sense of Exploration, Calm, Mastery and Skill, Purpose and Meaning, and Overall Happiness in Life: Exploratory Randomized Controlled Study

Annisa Arigayota[1], BSc; Barbara Duffek[2], MSc, PhD; Congcong Hou[3], MSc, PhD; Andreas Benedikt Eisingerich[1], BSc, MPhil, PhD

[1]Imperial College Business School, Imperial College London, London, United Kingdom

[2]Robinson College of Business, Georgia State University, Atlanta, GA, United States

[3]Faculty of Commerce, Kyushu Sangyo University, Fukuoka, Japan

Corresponding Author:
Andreas Benedikt Eisingerich, BSc, MPhil, PhD
Imperial College Business School
Imperial College London
South Kensington Campus
London, SW7 2AZ
United Kingdom
Phone: 44 740333 ext 6597
Email: a.eisingerich@imperial.ac.uk

Abstract

Background: Young people feel increasingly anxious and sad nowadays. Engaging with works of art and entertainment, such as playing open-world games or watching Studio Ghibli films, can be more than just a pastime. However, the extent to which, if at all, open-world games and feelings of nostalgia affect overall happiness in life remains unclear.

Objective: This study aimed to examine the extent to which open-world games, such as *The Legend of Zelda: Breath of the Wild*, and nostalgia evoked by Studio Ghibli films, such as Hayao Miyazaki's *My Neighbor Totoro* or *Kiki's Delivery Service*, affect postgraduate students' sense of exploration, calm, mastery and skill, purpose and meaning, and, ultimately, happiness in life.

Methods: A controlled laboratory experiment was conducted using a 2 (playing an open-world game vs no open-world game) × 2 (nostalgia vs no nostalgia) between-subject design. Study participants (N=518) were randomly assigned to the study's 4 conditions and answered a brief questionnaire, examining their sense of exploration, calm, mastery and skill, purpose and meaning, and, ultimately, happiness in life. As part of the study, we conducted univariate analysis and bootstrapping-based moderated mediation analysis with 5000 resamples.

Results: The results showed a significant and positive impact of playing an open-world game on overall life happiness (mean $[M]_{playedgame}$ 4.563, SD 0.072, vs $M_{notplayedgame}$ 3.170, SD 0.072; $F_{(1, 517)}$=117.246, P<.001). Furthermore, the positive impact of open-world games on overall life happiness was significantly enhanced by nostalgia evoked by watching Studio Ghibli films ($M_{nostalgia}$ 5.45, SD 0.102, vs $M_{nonostalgia}$ 3.58, SD 0.102; SE 0.144, 95% CI 1.332-1.900; P<.001). Moreover, exploratory moderated mediation with bootstrapped-based analyses and 5000 resamples demonstrated that the effect of playing open-world games on happiness is mediated by a sense of exploration (effect=0.11; SE 0.05, 95% CI 0.04-0.21), sense of calm (effect=0.32; SE 0.09, 95% CI 0.15-0.51), sense of skill and mastery (effect=0.08; SE 0.05, 95% CI 0.01-0.18), and sense of purpose and meaning (effect=0.32; SE 0.14, 95% CI 0.06-0.60).

Conclusions: This study shows that playing an open-world game, such as The *The Legend of Zelda: Breath of the Wild*, and nostalgia evoked by Studio Ghibli films significantly foster a sense of exploration and calm in life, as well as a feeling of mastery and skill, and purpose and meaning, hence ultimately contributing positively to one's overall happiness in life.

Trial Registration: ISRCTN ISRCTN14757739; https://www.isrctn.com/ISRCTN14757739

XSL•FO
RenderX

(JMIR Serious Games 2025;13:e76522) doi:10.2196/76522

KEYWORDS

open-world games; nostalgia; lab experimental study; postgraduate students; My Neighbor Totoro; Kiki's Delivery Service; Nintendo

Introduction

Overview of Young People's Anxiety and Reduced Happiness

Prior work indicates that young people feel increasingly stressed, anxious, and sad [1-5]. Many educational systems have moved toward heightened testing, earlier introduction of complex curriculum, and an emphasis on standardized, measurable achievements [6-8]. Young people can feel pressured to perform well on standardized exams from a young age, which can foster a fear of failure and chronic stress [6,9]. Even outside of formal testing, parents and others often place a strong emphasis on high achievements, such as getting top grades or excelling in extracurricular activities [6,10]. Although setting goals can be positive, an overemphasis on competition rather than joy in the activity per se can lead to children developing anxiety and burnout. In addition, a growing number of young people have daily schedules packed with lessons, sports, and various extracurricular activities [2]. Although these can be rewarding, they often leave little room for imaginative free play or downtime, which are crucial elements of healthy emotional development and stress relief [2].

Being constantly guided or supervised by adults can limit a young person's ability to develop decision-making skills and self-confidence [2]. As a result, they may feel anxious when they do not have control or when they encounter new situations without adult direction [2,11]. Furthermore, young people nowadays have unprecedented access to social media, where they often may face cyberbullying, peer pressure, information overload, or unrealistic portrayals of others' lives [2,12]. Constant comparison to curated online images can breed insecurity and anxiety [2,12,13]. Further, from a young age, many children also hear about worldwide crises, such as pandemics, natural disasters, extreme violence, and climate change, through online 24/7 news cycles, television, or social media [14-17]. This constant influx of alarming information can create a sense of fear, disorientation, hopelessness, and uncertainty about the world [17].

Moreover, young people often mirror the emotional climate of their household [18,19]. If parents are under significant financial or job-related stress, children may internalize those anxieties [20-23]. Moreover, shifts in the family structure, such as divorce, separation, or frequent relocations, can lead to insecurity and uncertainty, which contributes to elevated stress levels in young people [24-28]. Further, parents' best intentions, such as trying to protect children from hardship, can sometimes lead to "helicopter" or "tiger" parenting [10,29-33]. Often, children are expected to not only do well in school but also develop artistic talents, athletic achievements, and social popularity. The perceived need to "excel at everything" can be a significant source of anxiety [30,33] and existential worries in young people who, in essence, are still learning how to process big-picture problems [34].

Overview of Open-World Games, Nostalgia, and Happiness in Life

Playing open-world video games, such as *The Legend of Zelda: Breath of the Wild*, can have several positive effects on an individual's happiness and overall well-being. For instance, open-world games give players a high degree of control over their actions and how they experience the game world [35,36]. Moreover, the vast, interactive landscapes in *The Legend of Zelda: Breath of the Wild* encourage exploration, which can be deeply satisfying [35,37]. This mirrors the human desire for discovery and adventure, providing a sense of wonder and joy that can be rare in daily life.

Further, open-world video games offer an escape from everyday stressors [38-42]. The calming visuals of Hyrule, combined with the game's music and the freedom to roam, can be soothing, acting as a form of stress relief. In addition, the act of playing can require full attention, akin to flow, where players are fully engaged in the present moment, which can reduce anxiety and increase happiness. Moreover, playing open-world games, such as *The Legend of Zelda: Breath of the Wild* provides clear goals (eg, freeing the Divine Beasts or completing shrines, collecting Korok Seeds), which can give players a sense of purpose and progress, elements known to enhance life satisfaction. Furthermore, players can approach problems in *The Legend of Zelda: Breath of the Wild* in multiple ways. Players often create their own stories or ways to interact with the game world, giving them a platform for personal narrative and expression.

Nostalgia is often characterized as a bittersweet emotional experience that arises when reflecting on memories from the past—typically those that carry personal significance or recall "the good old days" [43-45]. Nostalgia can affect people's happiness and overall well-being in several ways [43,44,46]. For instance, nostalgia can help people remember meaningful and treasured episodes in their lives, reinforcing values, personal milestones, and who they are at their core. This may offer a sense of continuity and grounding in a world that is often seen as increasingly complex and rapidly changing [44,47]. In times of stress, reflecting on cherished memories can foster optimism and hope [48,49]. Nostalgia may, thus, also increase a sense of belonging and connection, as reminiscing about events that a group or community experienced can strengthen group identity and foster unity. When people use nostalgic memories to celebrate who they have become and to highlight their strengths and resources, it may help boost self-esteem and gratitude in times of personal challenges and difficulty [48]. Reflection can become a catalyst for personal growth and contentment if channelled in a way that supports forward momentum. That is, nostalgia can be a comforting, identity-affirming experience that deepens social ties and personal meaning. Therefore,

XSL•FO
RenderX

nostalgia can be a valuable tool for reaffirming one's sense of meaning and purpose [49].

Studio Ghibli films, such as works by Hayao Miyazaki (*My Neighbor Totoro* and *Kiki's Delivery Service*), may have a unique capacity to evoke nostalgia and foster a sense of happiness in viewers of all ages. The reasons behind this phenomenon range from the richly detailed animation and gentle storytelling to the deeper themes of community, family, and profound connection to nature. Films such as *My Neighbor Totoro* and *Kiki's Delivery Service* often focus on the small, everyday wonders of life, such as playing in the countryside or sharing mealtime with family. By shining a spotlight on ordinary moments and turning them into something magical, Miyazaki's works tap into a universal longing for the innocence and wonder of childhood. This kind of warmth can evoke nostalgia for times when people felt safe and cared for.

Although prior work shows heightened levels of anxiety, sadness, and depression among young people, questions surrounding how open-world games, such as *The Legend of Zelda: Breath of the Wild*, and nostalgia, evoked by Studio Ghibli films, such as Miyazaki's *My Neighbor Totoro* and *Kiki's Delivery Service*, may impact young people's happiness in life remain largely unanswered. Thus, in this study, we aimed to answer the following research question: How does playing an open-world game and nostalgia evoked by watching a Studio Ghibli film impact young people's sense of exploration, calm, mastery and skill, purpose and meaning in life, and, ultimately, overall life happiness?

Background on Open-World Games and Nostalgia Evoked by Studio Ghibli Films

Effects on Happiness in Life

One of the most praised elements of open-world games, such as *The Legend of Zelda: Breath of the Wild*, is their vast, open-world landscape. Being able to freely roam lush forests, climb mountains, or glide across canyons provides a relaxing form of escapism (Figure 1). The freedom to chart one's own path fosters a sense of agency and control, which can be especially soothing in times of real-life stress or uncertainty [35,50]. Furthermore, the environments of *The Legend of Zelda: Breath of the Wild* are filled with tranquil music, detailed natural landscapes, and a subdued color palette. Stopping by a quiet pond or cooking a meal under the stars in-game may momentarily transport the player away from daily stresses, letting them recharge mentally. In addition, although there is a main storyline, the game places few constraints on when or how one progresses. This encourages players to set their own pace. The act of wandering without immediate pressure to win can be liberating, as it allows for relaxation and personal goal setting rather than constant performance demands.

Figure 1. Open-world landscape in *The Legend of Zelda: Breath of the Wild* (from Nintendo).



At the core of many Studio Ghibli films is a message of uplifting harmony, whether it is living peacefully alongside nature (as in *My Neighbor Totoro*; Figure 2) or finding your place in a new community (as in *Kiki's Delivery Service*). These uplifting themes can be deeply comforting and can remind viewers of simple joys and quiet wonders in life. Studio Ghibli stories often highlight the extraordinary within the ordinary and, hence, put emphasis on everyday life's magic. This perspective can encourage viewers to find small delights and moments of beauty in their own routines.

XSL•FO
RenderX

Figure 2. A scene of nature in *My Neighbor Totoro* (with permission from Studio Ghibli).



Both open-world games, such as *The Legend of Zelda: Breath of the Wild*, and nostalgia evoked by Studio Ghibli films may inspire childlike wonder and a sense of discovery. For instance, *The Legend of Zelda: Breath of the Wild* frequently rewards curiosity. Stumbling upon hidden shrines, secrets, or quirky nonplayer characters creates a constant cycle of discovery. This taps into a childlike sense of wonder, where every cliff might hide a new vista and every cave might hold a surprise. Further, activities such as collecting ingredients, cooking new recipes, or just standing at a high vantage point to appreciate varied weather patterns and admire a scenic view encourage a slow and mindful approach to gameplay.

A hallmark of Studio Ghibli films is the breathtaking portrayal of nature—the rolling hills, forests, and gardens that appear onscreen (Figure 3). Nature in these films is not just a backdrop; it is a character in its own right. Seeing characters treat the environment with reverence and kindness fosters a sense of harmony. This connection can remind viewers of the comfort and grounding effect that comes from being in tune with the natural world—a feeling many associate with childhood adventures and freedom.

Figure 3. Depiction of fields and forests in *My Neighbor Totoro* (with permission from Studio Ghibli).



XSL•FO
RenderX

Effects on a Sense of Exploration

Hyrule in *The Legend of Zelda: Breath of the Wild* includes deserts, snow-capped mountains, tropical beaches, forests, and more. Each region has unique landscapes, weather patterns, and resources. The breathtaking vistas and day-night/weather cycles make exploration intrinsically rewarding. Standing on a mountaintop at sunrise to overlook Hyrule fosters a feeling of awe and accomplishment. Furthermore, the presence of 900 Korok Seeds rewards the smallest inklings of curiosity, whether it is lifting a seemingly random rock, diving off a peak into a hidden ring, or examining a peculiar formation of objects. These frequent, bite-size surprises reinforce an exploratory mindset, teaching players that small detours often lead to pleasant discoveries.

Studio Ghibli films open windows to worlds where magic coexists with ordinary life, inviting audiences to explore both physical and emotional landscapes with fresh eyes. By grounding fantastical elements in relatable, slice-of-life experiences, viewers are encouraged to look for wonder in the ordinary. Studio Ghibli films often center on young protagonists, such as Mei and Satsuki in *My Neighbor Totoro* and Kiki in *Kiki's Delivery Service*. Their openness to discovery and excitement about new surroundings convey a sense of unfiltered curiosity. This perspective invites audiences, adults included, to reconnect with childlike wonder and remember what it felt like to explore the unknown with fresh enthusiasm. Rather than focusing on large-scale spectacles, these films linger on small, intimate moments. By spotlighting the beauty in small details, Studio Ghibli's storytelling fosters an appreciation for the little wonders in day-to-day life, reminding viewers that exploration is not always about epic quests; indeed, it can be about noticing the world right in front of us.

Moreover, a hallmark of Miyazaki's work is the deep reverence for nature. Trees, fields, rivers, clouds, and skies are lovingly portrayed as living, breathing elements. Studio Ghibli films subtly remind us that adventure can be found anywhere if we approach life with curiosity. This focus on venturing into the unknown and embracing challenges conveys the idea that exploration is both external (navigating a new environment) and internal (discovering who you are). Critically, these gentle, wonder-filled narratives do not emphasize grandiose heroics. Instead, they celebrate quieter forms of exploration, including making friends or learning how to navigate a new space. It is a reminder that exploration is often less about the destination and more about an attitude of open-minded curiosity. In doing so, they can inspire viewers to become explorers in their own lives, seeking out the hidden pockets of wonder that exist all around us.

Effects on a Sense of Calm

Open-world games provide a structured yet boundless space to take a break from real-life stressors. When individuals are immersed in Link's journey and the vastness of Hyrule, daily frustrations may feel more distant. This timeout can give people a chance to process emotions and reset. Watching a sunrise in the game, sheltering from a rainstorm, or stargazing at night can feel almost meditative. These small, ambient details anchor players to calm moments and encourage them to pause and appreciate the virtual world. Moreover, *The Legend of Zelda: Breath of the Wild* allows individuals to tackle challenges in multiple ways, often rewarding creative problem solving more than raw skill. People can bypass certain obstacles or find clever solutions using runes. This flexibility reduces the frustration that can come from being stuck in a linear progression. In addition, the soft, often minimalistic soundtrack uses gentle piano and nature sounds, fostering a peaceful ambiance that can help soothe an angry or anxious mind.

Even when confronted with difficulties—a lost broom, a sick parent, or initial loneliness in a new town, Studio Ghibli's young heroes maintain hope and kindness. Seeing them overcome these challenges encourages a positive mind set in viewers, reminding them of their own capacity for resilience. Studio Ghibli narratives are often propelled by acts of empathy, such as neighbors helping neighbors, spirits offering aid, and strangers becoming friends. In *My Neighbor Totoro*, for instance, the forest spirits show concern for the sisters, and the father's understanding approach to parenting enriches the film's sense of warmth. This kindhearted environment fosters a safe, comforting world that puts emphasis not on overdramatization of emotions but on kindness as a driving force (Figure 4).

XSL•FO
RenderX

Figure 4. Depiction of Kiki looking at clouds in *Kiki's Delivery Service* (with permission from Studio Ghibli).



Effects on a Sense of Mastery and Skill

Earning better weapons, upgrading one's armor, and increasing one's "hearts" or stamina wheel visibly reflect the effort put in by a player. This tangible growth fosters a feeling of skill advancement, as players see Link literally become stronger. Finding a Korok Seed, cooking a successful new recipe, or completing a shrine are all microrewards. Collectively, these frequent small wins build a consistent loop of effort→reward, reinforcing a sense of competence. In *The Legend of Zelda: Breath of the Wild*, players face numerous challenges that require problem solving, perseverance, and creativity. The game normalizes "failing forward," such as dying to a Lynel, and then refining strategies. This mirrors Kiki's journey of overcoming self-doubt and the sisters' adventures in *My Neighbor Totoro* to find their mother. These narratives collectively teach resilience, adaptability, and the value of personal growth, which are integral to finding purpose and sustaining happiness. Both gaming and the films celebrate learning from one's mistakes. This can encourage individuals to view their life's challenges as opportunities for development rather than setbacks.

Effects on a Sense of Purpose and Meaning

Open-world games, such as *The Legend of Zelda: Breath of the Wild*, and Studio Ghibli films may do more than entertain; they can offer players a profound sense of purpose and meaning. *The Legend of Zelda: Breath of the Wild*'s classic "hero's journey" storyline resonates with deeply rooted myths about overcoming adversity. Engaging with these archetypal narratives can prompt a sense of mission: players feel they are contributing to a cause bigger than themselves, even if it is within a fictional world. This "bigger than self" feeling can be deeply meaningful. Furthermore, many players find calm and meditative moments in the game by simply riding through open fields, quietly gathering resources, or watching sunrises atop tall peaks. These peaceful interludes can trigger introspection and help players engage with inner thoughts. The sense of wonder generated can remind people that there is more to life than daily stressors, instilling renewed purpose.

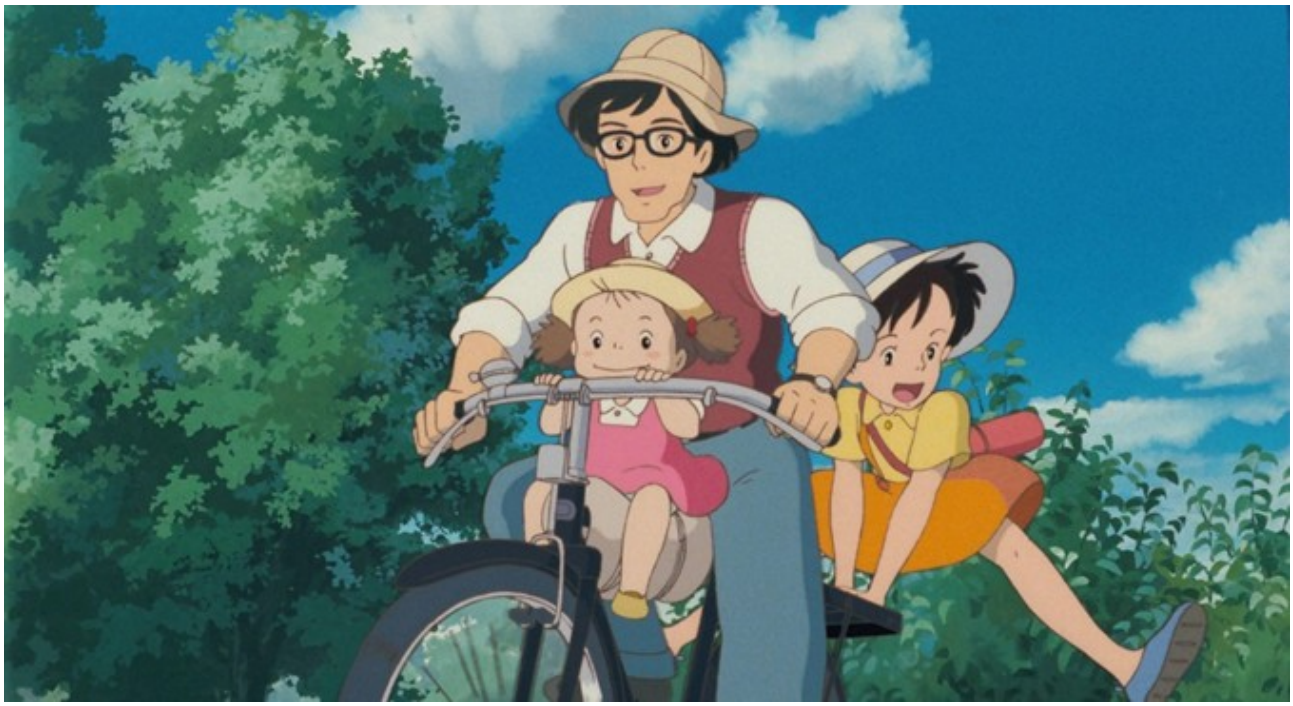
Instead of intense battles, Studio Ghibli films often focus on internal struggles (eg, Kiki's loss of confidence, Satsuki's and Mei's anxiety about their mother's health; Figure 5). In so doing, the stories highlight how purpose emerges from the personal growth that comes from confronting self-doubt, fear, or sorrow. Studio Ghibli protagonists find value in simple daily tasks (Figure 6), delivering goods, tidying up a house, or taking care of siblings. These ordinary moments become opportunities to strengthen relationships, contribute to the community, and experience quiet joy. This gentle reminder that purpose can be found in common routines, by embracing curiosity, wonder, and a respect for the unseen or unexplored aspects of life, may encourage viewers to reassess the significance of their own everyday actions.

XSL•FO
RenderX

Figure 5. Depiction of Satsuki looking for her young sister in *My Neighbor Totoro* (with permission from Studio Ghibli).



Figure 6. Depiction of ricing a bicycle together in *My Neighbor Totoro* (with permission from Studio Ghibli).



Hence, through intimate storytelling, focus on personal growth, and a compassionate depiction of community and nature, Studio Ghibli films gently guide viewers toward a greater awareness of what truly matters in life. By showcasing characters who discover meaning in acts of kindness, self-discovery, and a loving relationship with the world around them, these films inspire audiences to seek out or rekindle their own sense of purpose, whether through connection with others, appreciation of the every day, or a renewed embrace of their own unique gifts.

Methods

Study Design

The completed CONSORT-SPI (Consolidated Standards of Reporting Trials for Social and Psychological Interventions) 2018 checklist for reporting randomized trials of social and psychological interventions is shared in Multimedia Appendix 1. We registered our randomized control study with the International Standard Randomised Controlled Trial Number (ISRCTN) registry (ISRCTN 14757739). More specifically, as part of this study, we used a 2 (playing an open-world game vs

XSL•FO
RenderX

no open-world game) × 2 (nostalgia vs no nostalgia) between-subject experimental design. In spring 2025, posters and flyers were used on a university campus to recruit postgraduate university students to take part in a study on daily activities and well-being over the course of 3 weeks. All study participants were randomly allocated to 1 of the 4 study conditions using a random number generator:

- Condition 1: playing an open-world game + nostalgia
- Condition 2: playing an open-world game + no nostalgia
- Condition 3: no open-world game + nostalgia
- Condition 4 (control group): no open-world game + no nostalgia

We conducted the study in a lab experimental setting on a university campus. In condition 1, participants were invited to spend 30 minutes to play *The Legend of Zelda: Breath of the Wild* on the Nintendo Switch in handheld mode. After 30 minutes of video gameplay, the participants were invited to watch a brief 7-minute clip from a Studio Ghibli film (randomly assigned so that some participants watched *My Neighbor Totoro* and others watched *Kiki's Delivery Service*). Finally, the participants completed a brief questionnaire.

In condition 2, participants were invited to spend 30 minutes to play the open-world game as in Condition 1. After having played the game for 30 minutes, participants in condition 2 completed the brief questionnaire.

In condition 3, participants were randomly allocated to watch a brief 7-minute clip from either Studio Ghibli's *My Neighbor Totoro* or *Kiki's Delivery service* and subsequently complete a brief questionnaire.

Finally, in condition 4 (control group), participants simply completed the brief questionnaire.

Ethical Considerations

Ethical approval was received from Kyushu Sangyo University's Ethics Committee (approval number 2024-0017). All study participants were informed that the data collected will only be used to inform academic research. Further, they were reassured that data will be anonymized and treated with strict confidentiality. In addition, participants were told that they are free to stop and withdraw from the study at any time and without giving a reason. They were also informed that the data from participants who stop early will not be used as part of the study. Written consent forms were obtained from all study participants after they were offered an information leaflet, and time (minimum of 24 hours) for consideration of whether to take part in the study was allowed. Each study participant was thanked and received US $5 as a token of gratitude.

Participants and Procedure

A total of 530 university postgraduate students took part in this study. Twelve participants decided to stop and withdraw from the study during the data collection process. The data from the participants who decided to withdraw were not used as part of the study analyses. Hence, we obtained an effective total number of 518 (97.7%) participants for the study (condition 1: n=129, 24.9%; condition 2: n=130, 25.1%; condition 3: n=130, 25.1%; condition 4: n=129, 24.9%). Across all 4 conditions, participants answered a brief questionnaire, with responses scored on a 9-item Likert scale (from 1=strongly disagree to 9=strongly agree) capturing participants' overall life happiness; sense of exploration, calm, mastery and skill, and purpose and meaning in life; and the manipulation check questions for nostalgia.

As controls, we captured participants' enjoyment of the video game in conditions 1 and 2 ("I very much enjoyed playing the video game," "I found playing the video game to be interesting"; r=0.76), enjoyment of the Studio Ghibli film clip in conditions 1 and 3 ("I enjoyed watching the film clip," "I did not like the film clip very much" [reverse coding]; r=0.71), familiarity with the video game in conditions 1 and 2 ("I am very familiar with the video game," "I know the story of this video game very well"; r=0.81), and familiarity with the film in conditions 1 and 3 ("I am very familiar with the characters in the film," "I know the story of the characters in the film very well"; r=0.77), with anchors 1=strongly disagree to 9=strongly agree. The controls did not affect the results and thus are not discussed further.

Statistical Analysis

We checked measurement items and support for convergent validity and specifically examined whether the estimates for the average variance extracted (AVE) were indeed higher than the recommended threshold of 0.50 [51,52]. In addition, we examined the possibility that measurement errors may vary across items and calculated and compared the AVE for all pairs of constructs under investigation to the squared correlation between the 2 constructs of interest [53]. Moreover, we examined Cronbach α and composite reliability (CR) values. We conducted ANOVA as part of our manipulation checks for nostalgia and univariate analyses using IBM SPSS Statistics version 29.0. To examine the potential synergistic effect of nostalgia on the effects of open-world games on a sense of exploration, calm, skill and purpose, and overall happiness in life, we conducted an exploratory bootstrapping-based moderated mediation analysis (model 8) with 5000 resamples [54].

Results

Data Analysis

Figure 7 shows the participant selection flowchart.

Analysis results showed that the AVE exceeded the recommended threshold of 0.50 [51,52]. Moreover, the squared correlation between any pair of constructs was not higher than the respective AVE for each of the constructs in the pair, in support of discriminant validity [53]. In addition, Cronbach α and CR values were high. Detailed measurement items are listed in Table 1.

XSL•FO
RenderX

Figure 7. Study flowchart.

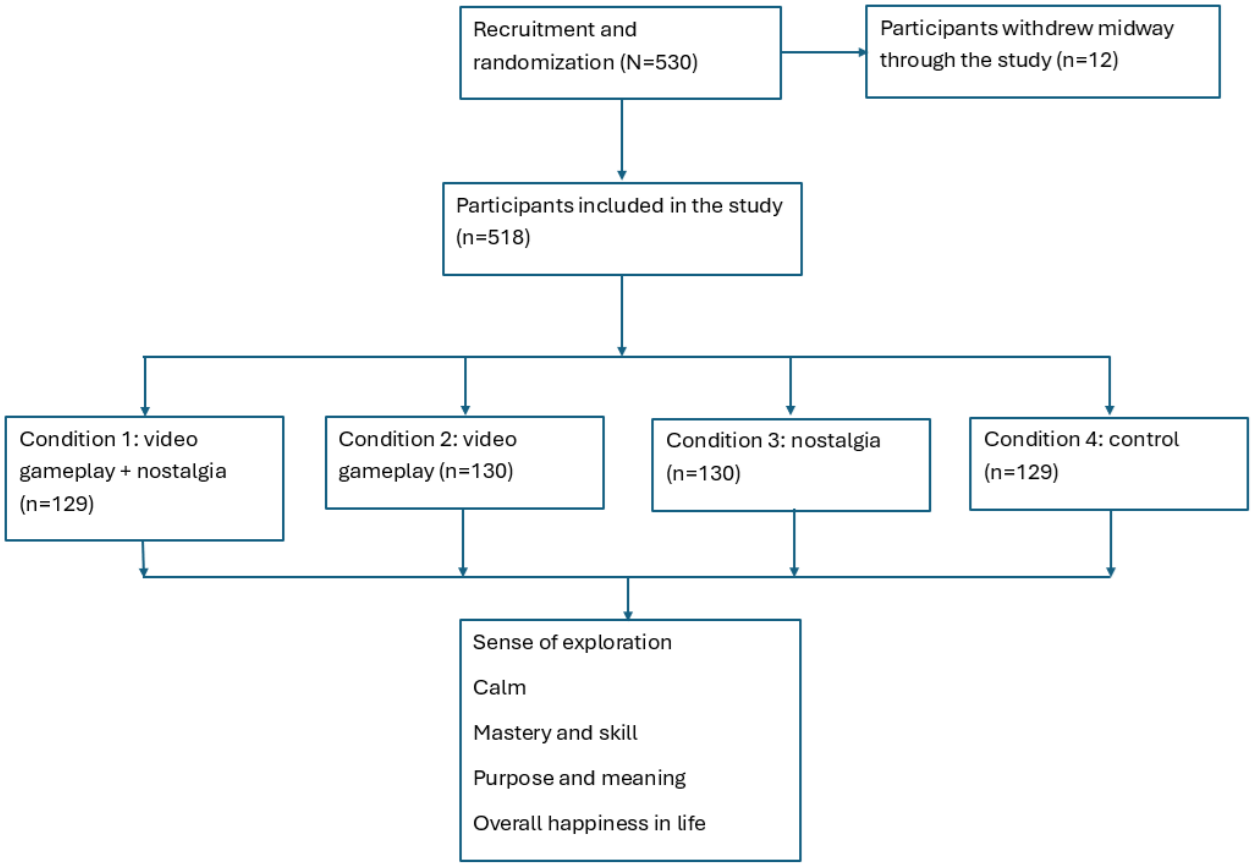


Table 1. Study measurement items and reliabilities.

Variable	Measure	Cronbach α	AVE[a]	CR[b]
Happiness in life	• "I feel grateful for the good things in my life." • "I am satisfied with the overall direction of my life and look forward to what lies ahead." • "Overall, I would describe myself as a happy person." • "I generally feel a sense of peace and contentment when I think about my life as a whole."	.90	0.59	0.85
Sense of exploration	• "I appreciate seeking out new experiences, even if they feel unfamiliar or slightly uncomfortable at first." • "I view my life as an adventure that lies ahead, filled with possibilities and opportunities." • "I view each day as a chance to discover something new."	.92	0.78	0.91
Sense of calm	• "I feel calm and collected in my daily life." • "I am feeling angry about a lot of things in my life." (Reverse-coded) • "I generally feel upset." (Reverse-coded)	.93	0.75	0.90
Sense of skill and mastery	• "I can find solutions when I face difficulties in my life." • "I believe I am well equipped to navigate life's day-to-day challenges." • "I have the necessary skills to master life."	.94	0.85	0.95
Sense of purpose and meaning	• "I feel my life contributes to something larger than myself." • "I experience a deep sense of fulfilment when I think about the path I am on." • "Overall, I view my life as meaningful and purposeful."	.95	0.86	0.95
Manipulation check for nostalgia	• "I feel a longing for how things used to be." • "Memories from my past bring me comfort and warmth." • "I get sentimental thinking about the 'good old days'."	.93	0.84	0.94

[a]AVE: average variance extracted.

[b]CR: composite reliability.

XSL•FO
RenderX

Manipulation Checks

A 2 × 2 ANOVA on the manipulation check confirmed that the manipulation of nostalgia was indeed successful. As expected, participants in the nostalgia-induced conditions indicated higher feelings of nostalgia (mean [M]$_{nostalgia}$ 6.61, SD 1.87) compared to those in conditions where nostalgia was not induced (M$_{nonostalgia}$ 3.77, SD 2.15; $F_{(1, 517)}$=257.38, P<.001). No significant effects of playing the video game ($F_{(1, 517)}$=0.124, P=.73) or game × nostalgia interaction ($F_{(1, 517)}$=0.582, P=.45) were observed.

Cell Means and Univariate Analyses

This study explored the positive effect of playing open-world games on overall happiness (M$_{playedgame}$ 4.563, SD 0.072, vs M$_{notplayedgame}$ 3.170, SD 0.072; $F_{(1, 517)}$=117.246, P<.001). The study results also showed a significant and positive impact of nostalgia evoked by watching Studio Ghibli films, such as Miyazaki's *My Neighbor Totoro* and *Kiki's Delivery Service* on happiness. Playing video games had a positive impact on happiness, and the effect was enhanced when nostalgia was induced (M$_{nostalgia}$ 5.45, SD 0.102, vs M$_{nonostalgia}$ 3.58, SD 0.102; SE=0.144, 95% CI 1.332-1.900, P<.001). Tables 2-5 present univariate analysis results and cell means comparing the effects.

Table 2. Results on happiness (dependent variable): study cell means estimates.

Video game played	Nostalgia induced	Mean (SD)	SE (95% CI)
No	No	2.409 (0.98)	0.102 (2.208-2.610)
No	Yes	3.931 (1.08)	0.102 (3.731-4.131)
Yes	No	3.579 (1.18)	0.102 (3.379-3.779)
Yes	Yes	5.547 (1.37)	0.102 (5.345-5.748)

Table 3. Effect of video gameplay on happiness (dependent variable): cell means comparisons (based on estimated marginal means).

Video game played	Nostalgia (I)	Nostalgia (J)	Mean difference (I – J)	SE (95% CI)	*P* value
No	No	Yes	–1.522[a]	0.144 (–1.806 to –1.238)	<.001
Yes	No	Yes	–1.968[a]	0.144 (–2.251 to –1.684)	<.001

[a]The mean difference was significant at the .05 level.

Table 4. Effect of nostalgia on happiness (dependent variable): cell means comparisons (based on estimated marginal means).

Nostalgia	Video game played (I)	Video game played (J)	Mean difference (I – J)	SE (95% CI)	*P* value
No	No	Yes	–1.170[a]	0.144 (–1.454 to –0.886)	<.001
Yes	No	Yes	–1.616[a]	0.144 (–1.900 to –1.332)	<.001

[a]The mean difference was significant at the .05 level.

XSL•FO
RenderX

Table 5. Results on happiness (dependent variable): univariate test results.

Condition	Sum of squares	Mean square	$F_{(1, 514)}$	*P* value
Video game not played				
Contrast	149.951	149.961	111.011	<.001
Error	694.344	1.351	—[a]	—
Video game played				
Contrast	250.689	250.689	185.577	<.001
Error	694.344	1.351	—	—
Nostalgia not induced				
Contrast	88.625	88.625	65.606	<.001
Error	694.344	1.351	—	—
Nostalgia induced				
Contrast	169.035	169.035	125.131	<.001
Error	694.344	1.351	—	—

[a]Not applicable.

Exploratory Moderated Mediation

To examine the effect of nostalgia on playing video games and happiness in life, we conducted an exploratory bootstrapping-based moderated mediation analysis with 5000 resamples (model 8) [54]. The analysis was conducted with happiness as the dependent variable; whether the participant played or did not play a video game as the independent variable; a sense of exploration, calm, skill and mastery, and purpose and meaning as mediating variables; and whether nostalgia was or was not induced as the moderator variable. The sample size was 518. Detailed results are presented in Tables 6-18. The detailed results showed that the effect of playing video games on happiness is mediated by a sense of exploration (effect=0.11, SE=0.05, 95% CI 0.04-0.21; Tables 6-8), a sense of calm (effect=0.32, SE=0.09, 95% CI 0.15-0.51; Tables 9-11), a sense of skill and mastery (effect=0.08, SE=0.05, 95% CI 0.01-0.18; Tables 12-14), and a sense of purpose and meaning (effect=0.32, SE=0.14, 95% CI 0.06-0.60; Tables 15-17). Table 18 shows the means for key constructs across the 4 study conditions. Together, these results highlight the importance of playing video games and nostalgia on happiness, through a sense of exploration, calm, skill and mastery, and purpose and meaning.

Table 6. Regression results with a sense of exploration as a mediator.

Dependent variable and predictors	B	SE (95% CI)	*t* Test (*df*)	*P* value
Sense of exploration (R=0.64, R^2=0.40, mean standard error=2.47, $F_{(3, 514)}$=116.28, *P*<.001)				
Constant	1.24	0.68 (–0.09 to 2.57)	1.83 (514)	.07
Video game played	0.39	0.43 (–0.44 to 1.24)	0.93 (514)	.36
Nostalgia	–0.12	0.43 (–0.97 to 0.72)	–0.29 (514)	.77
Video game played × nostalgia	1.06	0.27 (0.53 to 1.59)	3.93 (514)	<.001
Happiness (R=0.70, R^2=0.49, mean standard error=1.33, $F_{(4, 513)}$=124.87, *P*<.001)				
Constant	0.03	0.51 (–0.97 to 1.03)	0.06 (513)	.95
Video game played	0.68	0.32 (0.05 to 1.31)	2.13 (513)	.03
Sense of exploration	0.11	0.03 (0.04 to 0.17)	3.19 (513)	.002
Nostalgia	1.09	0.32 (0.46 1.72)	3.40 (513)	.001
Video game played × nostalgia	0.33	0.21 (–0.07 to 0.74)	1.63 (513)	.11

Table 7. Conditional effects of playing a video game on happiness at levels of nostalgia.

Nostalgia induced	Effect	SE (95% CI)	*t* Test (*df*)	*P* value
No	1.02	0.15 (0.72-1.31)	6.73 (513)	<.001
Yes	1.35	0.17 (1.03-1.68)	8.15 (513)	<.001

XSL•FO
RenderX

Table 8. Conditional indirect effects of playing a video game on happiness via a sense of exploration.[a]

Nostalgia induced	Effect	BootSE (95% CI)
No	0.15	0.05 (0.06-0.27)
Yes	0.27	0.08 (0.11-0.44)

[a]Index of moderated mediation: index=0.11, BootSE=0.0450 (95% CI 0.04-0.21).

Table 9. Regression results with a sense of calm as a mediator.

Dependent variable and predictors	B	SE (95% CI)	*t* Test (*df*)	*P* value
Sense of calm (R=0.71, R^2=0.51, mean standard error=2.62, $F_{(3, 514)}$=175.46, *P*<.001)				
Constant	0.09	0.71 (–1.31 to 1.48)	0.13 (514)	.90
Video game played	0.22	0.45 (–0.66 to 1.11)	0.49 (514)	.62
Nostalgia	1.03	0.45 (0.14 to 1.91)	2.78 (514)	.02
Video game played × nostalgia	1.08	0.28 (0.52 to 1.64)	3.79 (514)	<.001
Happiness (R=0.76, R^2=0.57, mean standard error=1.12, $F_{(4, 513)}$=172.01, *P*<.001)				
Constant	0.14	0.47 (–0.78 to 1.05)	0.29 (513)	.77
Video game played	0.66	0.29 (0.08 to 1.24)	2.24 (513)	.03
Sense of calm	0.30	0.03 (0.24 to 0.36)	10.37 (513)	<.001
Nostalgia	0.77	0.30 (0.19 to 1.35)	2.50 (513)	.01
Video game played × nostalgia	0.13	0.19 (–0.25, to 0.49)	0.65 (513)	.51

Table 10. Conditional effects of playing a video game on happiness at levels of nostalgia (mediator: sense of calm).

Nostalgia induced	Effect	SE (95% CI)	*t* Test (*df*)	*P* value
No	0.78	0.14 (0.51-1.05)	5.71 (513)	<.001
Yes	0.90	0.15 (0.61-1.19)	6.09 (513)	<.001

Table 11. Conditional indirect effects of playing a video game on happiness via a sense of calm.[a]

Nostalgia induced	Effect	BootSE (95% CI)
No	0.49	0.07 (0.26-0.54)
Yes	0.71	0.10 (0.52-0.93)

[a]Index of moderated mediation: index=0.32, BootSE=0. 0903 (95% CI 0.15-0.51).

Table 12. Regression results with a sense of skill and mastery as a mediator.

Dependent variable and predictors	B	SE (95% CI)	*t* Test (*df*)	*P* value
Sense of skill and mastery (R=0.50, R^2=0.25, mean standard error=3.28, $F_{(3, 514)}$=58.23, *P*<.001)				
Constant	2.73	0.80 (1.16 to 4.29)	3.42 (514)	<.001
Video game played	–0.54	0.50 (–1.53 to 0.45)	–1.07 (514)	.28
Nostalgia	–1.13	0.50 (–2.12 to –0.14)	–2.25 (514)	.03
Video game played × nostalgia	1.47	0.32 (0.84 to 2.09)	4.61 (514)	<.001
Happiness (R=0.70, R^2=0.49, mean standard error=1.34, $F_{(4, 513)}$=121.94, *P*<.001)				
Constant	0.01	0.51 (–1.01 to 1.02)	0.01 (513)	.99
Video game played	0.76	0.32 (0.12 to 1.39)	2.34 (513)	.02
Sense of calm	0.06	0.03 (0.00 to 0.11)	2.02 (513)	.04
Nostalgia	1.14	0.33 (0.50 to 1.78)	3.42 (513)	<.001
Video game played × nostalgia	0.36	0.21 (–0.04 to 0.77)	1.74 (513)	.08

XSL•FO
RenderX

Table 13. Conditional effects of playing a video game on happiness at levels of nostalgia (mediator: sense of skill and mastery).

Nostalgia induced	Effect	SE (95% CI)	*t* Test (*df*)	*P* value
No	1.12	0.15 (0.83-1.40)	7.63 (513)	<.001
Yes	1.48	0.16 (1.17-1.79)	9.30 (513)	<.001

Table 14. Conditional indirect effects of playing a video game on happiness via a sense of skill and mastery.[a]

Nostalgia induced	Effect	BootSE (95% CI)
No	0.05	0.03 (0.00-0.11)
Yes	0.14	0.07 (0.01-0.27)

[a]Index of moderated mediation: index=0.08, BootSE=0.0448 (95% CI 0.01-0.18).

Table 15. Regression results with a sense of purpose and meaning as a mediator.

Dependent variable and predictors	B	SE (95% CI)	*t* Test (*df*)	*P* value
Sense of purpose and meaning (R=0.39, R^2=0.16, mean standard error=4.02, $F_{(3, 514)}$=31.89, *P*<.001)				
Constant	1.64	0.88 (–0.09 to 3.73)	1.86 (514)	.06
Video game played	–0.21	0.56 (–1.31 to 0.88)	–0.38 (514)	.71
Nostalgia	0.18	0.56 (–0.92 to 1.28)	0.32 (514)	.75
Video game played × nostalgia	0.79	0.35 (0.10 to 1.49)	2.25 (514)	.02
Happiness (R=0.86, R^2=0.74, mean standard error=0.69, $F_{(4, 513)}$=358.40, *P*<.001)				
Constant	–0.50	0.37 (–1.01 to 1.02)	–1.37 (513)	.17
Video game played	0.81	0.23 (0.12 to 1.39)	3.51 (513)	<.001
Sense of calm	0.41	0.02 (0.00 to 0.11)	22.19 (513)	<.001
Nostalgia	1.00	0.23 (0.50 to 1.78)	4.35 (513)	<.001
Video game played × nostalgia	0.12	0.15 (–0.04 to 0.77)	0.84 (513)	.40

Table 16. Conditional effects of playing a video game on happiness at levels of nostalgia (mediator: sense of purpose and meaning).

Nostalgia induced	Effect	SE (95% CI)	*t* Test (*df*)	*P* value
No	0.94	0.10 (0.73-1.14)	9.01 (513)	<.001
Yes	1.06	0.11 (0.85-1.27)	9.97 (513)	<.001

Table 17. Conditional indirect effects of playing a video game on happiness via a sense of purpose and meaning.[a]

Nostalgia induced	Effect	BootSE (95% CI)
No	0.23	0.09 (0.06-0.42)
Yes	0.57	0.11 (0.34-0.78)

[a]Index of moderated mediation: index=0.32, BootSE=0.14 (95% CI 0.06-0.60).

XSL•FO
RenderX

Table 18. Means across study conditions (N=518).

Variable	Conditions, mean (SD)			
	Open-world game + Studio Ghibli film (nostalgia)	Open-world game	Studio Ghibli film (nostalgia)	Control group
Happiness in life	5.55[a] (1.37)	3.58[b] (1.18)	3.93[c] (1.08)	2.41[d] (.98)
Sense of exploration in life	6.04[a] (1.73)	4.03[c] (1.41)	3.51[b] (1.47)	2.57[d] (1.53)
Sense of calm in life	6.90[a] (1.85)	3.72[b] (1.45)	4.52[c] (1.73)	2.42[d] (1.39)
Sense of skill and mastery in life	5.26[a] (2.06)	3.45[c] (1.54)	2.86[b] (1.83)	2.53[b] (1.79)
Sense of purpose and meaning in life	4.74[a] (2.52)	2.98[c] (1.92)	3.37[c] (1.69)	2.40[b] (1.79)
Manipulation check for nostalgia	6.71[a] (1.73)	3.73[c] (2.22)	6.51[a] (2.01)	3.80[c] (2.08)

[a-d]Means with different superscripts are significantly different at $P<.05$.

Discussion

Principal Findings

Studio Ghibli films, such as *My Neighbor Totoro* and *Kiki's Delivery Service*, alongside open-world games, such as *The Legend of Zelda: Breath of the Wild*, offer immersive, interactive experiences that cultivate happiness by fostering wonder and exploration, calm, a sense of mastery and skill, and a sense of purpose and meaning in life. These works blend storytelling, exploration, and emotional engagement to nurture a sense of purpose and joy. These works are not mere distractions; our findings demonstrate that they may be viewed as "active escapism" that restores energy. This study shows how they can synergistically enhance young people's sense of exploration, as well as feelings of calm, mastery and skill, and purpose and meaning.

Comparison With Prior Work

Previous research has offered important insights into the role of video games in people's lives [55,56] and that multiplayer games can mitigate loneliness during social isolation [57,58], while excessive play may reduce mental health [59,60]; however, critical questions remain about how open-world games may affect people's overall happiness in life. Although prior studies have linked open-world games to cognitive escapism and relaxation benefits [35] and video games to well-being [61] and improved mood [62,63] and have noted nostalgia as a coping mechanism during stressful times [64-66] that enhances meaning and continuity [45,67-70], this study examined the synergistic effects of open-world games and nostalgia evoked by Studio Ghibli films on happiness in life. In addition, this study examined the process mechanism that helps explain the influence of open-world games, such as *The Legend of Zelda: Breath of the Wild*, and nostalgia evoked by Studio Ghibli films on life happiness. More specifically, this study makes a novel contribution by showing how open-world games and Studio Ghibli films can synergistically enhance people's overall happiness in life through a sense of exploration; feelings of calm, mastery, and skill; and a sense of purpose and meaning.

Specifically, this study shows that Studio Ghibli, such as *My Neighbor Totoro* and *Kiki's Delivery Service*, and open-world games, such as *The Legend of Zelda: Breath of the Wild*, do not just entertain, they offer frameworks for living well. They encourage us to seek wonder in the ordinary (eg, Totoro's acorns, Korok puzzles), embrace challenges as growth (eg, Kiki's flight, Link's battles), and find strength in community and solitude alike (eg, Satsuki's bond with Mei, Link's lone journey). By blending imaginative storytelling with interactive agency, Studio Ghibli films and open-world games can provide a holistic recipe for happiness: one part mindfulness and calm, one part mastery of skill and purpose, and endless curiosity and exploration. As Kiki learns to fly again and Link gazes at Hyrule's horizon, these stories whisper, "Happiness is not a destination, it is how we travel."

Limitations

This study has a number of limitations, which offer promising avenues for future research. First, we examined an expressed sense of exploration, calm, mastery and skill, purpose and meaning, and life happiness in a cross-sectional manner. Additional research that investigates the long-term effects of open-world games and nostalgia on actual life happiness, actual exploration, calm, skill, and purpose in life is richly deserving. Second, we accounted for participants' familiarity with and enjoyment of the open-world game in conditions 1 and 3 in our study, but we did not capture their immersion in the game and feelings of flow across our study conditions. We invite future research to account for immersion, flow, cognitive escapism, and relaxation in the study of open-world games and nostalgia. Third, we studied postgraduate students in one cultural setting. We encourage further study of playing open-world games and Studio Ghibli films across different countries and study participants to account for other potential confounding effects, including participants' gender, age, cultural background, and lifestyles. Fourth, we studied the effects of 2 Studio Ghibli films. To enhance confidence in the generalizability of the results, we encourage future research to examine the effects of other Studio Ghibli films, such as *Castle in the Sky* and *Howl's Moving Castle*, on people's sense of nostalgia and happiness in life.

XSL•FO
RenderX

Future Research

Prior research indicates that in addition to the mediators investigated in this study, there may be other avenues through which open-world gameplay and nostalgia can affect people's happiness in life [71-75]. For instance, people may get attached to the games they play and the stories they engage with [76-81], which may offer them comfort and enhance happiness in a world of turmoil. Moreover, feelings of mastery and skill, as well as elevation and inspiration, have been noted to positively affect people's lives [82-85]. We encourage future research to further unpack the roles of open-world gameplay and nostalgia in affecting people's satisfaction and long-term happiness. For instance, additional research that examines how open-world games and interaction with stories, such as Studio Ghibli films, may offer experiences seen as authentic, enrich people's lives, reduce feelings of loneliness and psychological pain, and help them walk through life with purpose, energy, joy, and fulfilment is richly deserving. By showing how open-world games, such as *The Legend of Zelda: Breath of the Wild*, and Studio Ghibli films enhance people's overall happiness in life in a synergistically manner, this study bridges gaming and film culture with mental health science. In so doing, the study challenges outdated stereotypes about gaming and engaging with storytelling and opens doors to innovative, accessible tools for well-being. We invite additional research to examine the potential of video gameplay and storytelling in affecting people's lives. This study indicates that what is so beautiful about Studio Ghibli films and *The Legend of Zelda: Breath of the Wild* is that they are not just tales or games but are also invitations—invitations to see the world with wider eyes, to stumble into hidden meadows of joy, and to remember that even when life feels like a tangled *Lost Woods*, there is always a path (or a paraglider) waiting to be discovered. They whisper, never forget, "You are the hero of your own Hyrule, the witch flying her own broom, and the kid who still knows how to wave at the Catbus." Future research exploring the potential influence of video gameplay and digital solutions in enhancing childlike wonder and the overall happiness in life and in reducing the risk of burnout is richly deserving.

Conclusion

This controlled laboratory experiment (N=518) shows the positive effects open-world gameplay and nostalgia can have by reducing postgraduate students' stress, anxiety, and burnout. The findings of this study show how participants playing open-world games, such as *The Legend of Zelda: Breath of the Wild*, and watching Studio Ghibli films, such as Miyazaki's *My Neighbor Totoro* and *Kiki's Delivery Service*, may report increased happiness, as well as increased feelings of exploration, calm, mastery and skill, and purpose and meaning in life.

Acknowledgments

We thank all study participants for their time and efforts to take part in the research. Furthermore, we thank Studio Ghibli, Tokyo, for permission to use their images of Totoro and Kiki as part of this manuscript. We also thank Nintendo, Kyoto, for allowing us to take pictures of their own Nintendo Switch Zelda game and use these as part of this manuscript. We have used DeepSeek and ChatGPT O1 for parts of the writing's grammar and spelling checks.

Data Availability

Data are available from the authors upon reasonable request.

Authors' Contributions

AA was responsible for writing, reviewing, and editing the manuscript. BD was responsible for methodology, formal analysis, investigation, and writing and reviewing the manuscript. CH was responsible for methodology, data collection, formal analysis, and investigation, as well as writing, reviewing, and editing the manuscript. ABE was responsible for conceptualization, methodology, formal analysis, and investigation, as well as writing, reviewing, and editing the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-SPI (Consolidated Standards of Reporting Trials for Social and Psychological Interventions) 2018 checklist.
[PDF File (Adobe PDF File), 152 KB - games_v13i1e76522_app1.pdf]

References

1. Pierceall EA, Keim MC. Stress and coping strategies among community college students. Commun Coll J Res Pract 2007 Oct 08;31(9):703-712. [doi: 10.1080/10668920600866579]
2. Haidt J. The Anxious Generation : How the Great Rewiring of Childhood Is Causing an Epidemic of Mental Illness. London, UK: Penguin Press; 2024.
3. Thompson D. Why American teens are so sad. The Atlantic. 2022. URL: https://www.theatlantic.com/newsletters/archive/2022/04/american-teens-sadness-depression-anxiety/629524/ [accessed 2025-07-05]

XSL•FO
RenderX

4. Osorio EK, Hyde E. The rise of anxiety and depression among young adults in the United States. Ballard Brief. 2021. URL: https://tinyurl.com/mr26px8s [accessed 2025-07-05]
5. Insights: youth mental health and wellbeing over time. headspace: National Youth Mental Health Survey Foundation. 2020. URL: https://headspace.org.au/our-impact/evaluation-research-reports/youth-mental-health-statistics/ [accessed 2025-07-05]
6. Kim J. Stress and coping mechanisms in South Korean high school students: academic pressure, social expectations, and mental health support. JRSSH 2024 May;3(5):45-54. [doi: 10.56397/jrssh.2024.05.09]
7. Segool NK, Carlson JS, Goforth AN, von der Embse N, Barterian JA. Heightened test anxiety among young children: elementary school students' anxious responses to high-stakes testing. Psychol Schools 2013 Mar 06;50(5):489-499. [doi: 10.1002/pits.21689]
8. Heissel J, Adam E, Doleac J, Figlio D, Meer J. Testing, stress, and performance: how students respond physiologically to high-stakes testing. Educ Finance Policy 2021;16(2):183-208. [doi: 10.1162/edfp_a_00306]
9. Wang Z, Li C, Xie Z, Hong O. Grit difference in the association between academic stress and adolescents' meaning in life: the roles of school burnout and self-compassion. Child Care Health Dev 2024 Nov 14;50(6):e70005. [doi: 10.1111/cch.70005] [Medline: 39540694]
10. La Rosa VL, Ching BH, Commodari E. The impact of helicopter parenting on emerging adults in higher education: a scoping review of psychological adjustment in university students. J Genet Psychol 2025 Oct 17;186(2):162-189. [doi: 10.1080/00221325.2024.2413490] [Medline: 39757795]
11. Shah EN, Szwedo DE, Allen JP. Parental autonomy restricting behaviors during adolescence as predictors of dependency on parents in emerging adulthood. Emerg Adulthood 2023 Feb;11(1):15-31 [FREE Full text] [doi: 10.1177/21676968221121158] [Medline: 36644228]
12. Oyinbo AG, Heavner K, Mangano KM, Morse B, El Ghaziri M, Thind H. Prolonged social media use and its association with perceived stress in female college students. Am J Health Educ 2024 Jan 23;55(3):189-198. [doi: 10.1080/19325037.2023.2295550]
13. Hawi NS, Samaha M. The relations among social media addiction, self-esteem, and life satisfaction in university students. Soc Sci Comput Rev 2016 Aug 10;35(5):576-586. [doi: 10.1177/0894439316660340]
14. Schor M, Clayton S, Cardoso MCV, Fernandes M, Frota MA, Surkan PJ. 'I am really scared for the future': climate distress among Brazilian youth. Environ Educ Res 2024 Sep 28:1-22. [doi: 10.1080/13504622.2024.2409946]
15. Hörter K. "It makes me sad when so much of the forest burns down!" Childhood and youth in the context of the climate crisis. Psychoanal Study Child 2024 Jan 23;77(1):389-401. [doi: 10.1080/00797308.2023.2287383]
16. Pearcey S, Burgess L, Shum A, Sajid E, Sargent M, Klampe M, et al. How the COVID-19 pandemic affected young people's mental health and wellbeing in the UK: a qualitative study. J Adolesc Res 2023 Feb 02;39(6):1573-1600. [doi: 10.1177/07435584231151902]
17. Gleason MM, Thompson LA. Depression and anxiety disorder in children and adolescents. JAMA Pediatr 2022 May 01;176(5):532. [doi: 10.1001/jamapediatrics.2022.0052] [Medline: 35344008]
18. Grinde B, Tambs K. Effect of household size on mental problems in children: results from the Norwegian Mother and Child Cohort study. BMC Psychol 2016 Jun 02;4(1):31 [FREE Full text] [doi: 10.1186/s40359-016-0136-1] [Medline: 27255744]
19. Bayer JK, Hiscock H, Ukoumunne OC, Price A, Wake M. Early childhood aetiology of mental health problems: a longitudinal population-based study. J Child Psychol Psychiatry 2008 Nov 27;49(11):1166-1174. [doi: 10.1111/j.1469-7610.2008.01943.x] [Medline: 18665879]
20. Cooper K, Stewart K. Does household income affect children's outcomes? A systematic review of the evidence. Child Ind Res 2020 Nov 04;14(3):981-1005. [doi: 10.1007/s12187-020-09782-0]
21. Crouter AC, Bumpus MF. Linking parents' work stress to children's and adolescents' psychological adjustment. Curr Dir Psychol Sci 2001 Oct 01;10(5):156-159. [doi: 10.1111/1467-8721.00138]
22. Repetti RL, Wood J. Effects of daily stress at work on mothers' interactions with preschoolers. J Fam Psychol 1997;11(1):90-108. [doi: 10.1037//0893-3200.11.1.90]
23. Strazdins L, Shipley M, Clements M, Obrien LV, Broom DH. Job quality and inequality: parents' jobs and children's emotional and behavioural difficulties. Soc Sci Med 2010 Jun;70(12):2052-2060. [doi: 10.1016/j.socscimed.2010.02.041] [Medline: 20382458]
24. Larson RW, Gillman S. Transmission of emotions in the daily interactions of single-mother families. J Marriage Fam 1999 Feb;61(1):21. [doi: 10.2307/353880]
25. Demo DH, Acock AC. The impact of divorce on children. J Marriage Fam 1988 Aug;50(3):619. [doi: 10.2307/352634]
26. van der Wal RC, Levelt L, Kluwer E, Finkenauer C. Exploring associations between children's forgiveness following parental divorce and psychological well-being. Fam Transit 2024 Feb 08;65(3):248-270. [doi: 10.1080/28375300.2024.2310432]
27. Frimmel W, Halla M, Winter-Ebmer R. How does parental divorce affect children's long-term outcomes? J Public Econ 2024 Nov;239:105201. [doi: 10.1016/j.jpubeco.2024.105201]
28. Zayas LH, Garcia Isaza A, Fuentes-Balderrama J, Rivera-Heredia ME. Well-being and contexts of development of U.S. citizen children in Mexico following parental deportation or voluntary relocation. Am J Orthopsychiatry 2024;94(5):508-517. [doi: 10.1037/ort0000734] [Medline: 38546563]

XSL•FO
RenderX

29. Cruciani G, Fontana A, Benzi IMA, Sideli L, Parolin LAL, Muzi L, et al. Mentalized affectivity, helicopter parenting, and psychopathological risk in emerging adults: a network analysis. Eur J Investig Health Psychol Educ 2024 Sep 18;14(9):2523-2541 [FREE Full text] [doi: 10.3390/ejihpe14090167] [Medline: 39329835]
30. Wang C, Shi H, Li G. Helicopter parenting and college student depression: the mediating effect of physical self-esteem. Front Psychiatry 2023 Jan 9;14:1329248 [FREE Full text] [doi: 10.3389/fpsyt.2023.1329248] [Medline: 38264635]
31. Türker M, Bahçeci F. New generation parenting attitude: helicopter parenting. Alanyazın 2024 May;5(1):45-56. [doi: 10.59320/alanyazin.1452153]
32. Kobakhidze MN, Ma (马莹) Y, Tsaloukidis A. Tiger parenting beyond cultural essentialism: discourses of class, culture, and competition in Hong Kong. ECNU Rev Educ 2023 Sep 06;7(3):531-553. [doi: 10.1177/20965311231198252]
33. Zhang R, Wang Z. Effects of helicopter parenting, tiger parenting and inhibitory control on the development of children's anxiety and depressive symptoms. Child Psychiatry Hum Dev 2024 Feb 24:Online ahead of print. [doi: 10.1007/s10578-024-01685-3] [Medline: 38401004]
34. Hughes B, Jones K. Young people's experiences of death anxiety and responses to the COVID-19 pandemic. Omega (Westport) 2022 Jun 17;90(2):656-673. [doi: 10.1177/00302228221109052]
35. Anto A, Basu A, Selim R, Foscht T, Eisingerich AB. Open-world games' affordance of cognitive escapism, relaxation, and mental well-being among postgraduate students: mixed methods study. J Med Internet Res 2024 Dec 17;26:e63760. [doi: 10.2196/63760]
36. Marques LM, Uchida PM, Aguiar FO, Kadri G, Santos RIM, Barbosa SP. Escaping through virtual gaming-what is the association with emotional, social, and mental health? A systematic review. Front Psychiatry 2023 Nov 8;14:1257685 [FREE Full text] [doi: 10.3389/fpsyt.2023.1257685] [Medline: 38025467]
37. Ryan RM, Rigby CS, Przybylski A. The motivational pull of video games: a self-determination theory approach. Motiv Emot 2006 Nov 29;30(4):344-360. [doi: 10.1007/s11031-006-9051-8]
38. Deci E, Ryan R. Intrinsic Motivation and Self-Determination in Human Behavior. Berlin: Springer Science & Business Media; 2013.
39. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000;55(1):68-78. [doi: 10.1037/0003-066X.55.1.68]
40. Deci EL, Ryan RM. Self-determination theory: a macrotheory of human motivation, development, and health. Can Psychol 2008 Aug;49(3):182-185. [doi: 10.1037/a0012801]
41. Huang T, Yeh J, Liao G, Cheng T, Chang Y, Teng C. How does escapism foster game experience and game use? Decis Supp Syst 2024 Jun;181:114207. [doi: 10.1016/j.dss.2024.114207]
42. Przybylski AK, Rigby CS, Ryan RM. A motivational model of video game engagement. Rev Gen Psychol 2010 Jun 01;14(2):154-166. [doi: 10.1037/a0019440]
43. Li B, Zhu Q, Li A, Cui R. Can good memories of the past instill happiness? Nostalgia improves subjective well-being by increasing gratitude. J Happiness Stud 2023 Jan 09;24(2):699-715 [FREE Full text] [doi: 10.1007/s10902-022-00616-0] [Medline: 36644477]
44. Layous K, Kurtz JL. Nostalgia: a potential pathway to greater well-being. Curr Opin Psychol 2023 Feb;49:101548 [FREE Full text] [doi: 10.1016/j.copsyc.2022.101548] [Medline: 36669249]
45. van Tilburg WAP, Wildschut T, Sedikides C. Nostalgia's place among self-relevant emotions. Cogn Emot 2018 Jun 24;32(4):742-759. [doi: 10.1080/02699931.2017.1351331] [Medline: 28738756]
46. Cox CR, Kersten M, Routledge C, Brown EM, Van Enkevort EA. When past meets present: the relationship between website‐induced nostalgia and well‐being. J Appl Soc Pyschol 2014 Dec 12;45(5):282-299. [doi: 10.1111/jasp.12295]
47. Hong EK, Sedikides C, Wildschut T. How does nostalgia conduce to global self-continuity? The roles of identity narrative, associative links, and stability. Pers Soc Psychol Bull 2022 May;48(5):735-749 [FREE Full text] [doi: 10.1177/01461672211024889] [Medline: 34151655]
48. Cheung W, Sedikides C, Wildschut T. Induced nostalgia increases optimism (via social-connectedness and self-esteem) among individuals high, but not low, in trait nostalgia. Pers Indiv Diff 2016 Feb;90:283-288. [doi: 10.1016/j.paid.2015.11.028]
49. Layous K, Kurtz JL, Wildschut T, Sedikides C. The effect of a multi-week nostalgia intervention on well-being: mechanisms and moderation. Emotion 2022 Dec;22(8):1952-1968. [doi: 10.1037/emo0000817] [Medline: 34591502]
50. Pallavicini F, Pepe A, Mantovani F. Commercial off-the-shelf video games for reducing stress and anxiety: systematic review. JMIR Ment Health 2021 Aug 16;8(8):e28150 [FREE Full text] [doi: 10.2196/28150] [Medline: 34398795]
51. Bagozzi RP, Yi Y. On the evaluation of structural equation models. JAMS 1988 Mar;16(1):74-94. [doi: 10.1007/bf02723327]
52. Anderson JC, Gerbing DW. Structural equation modeling in practice: a review and recommended two-step approach. Psychol Bull 1988;103(3):411-423. [doi: 10.1037//0033-2909.103.3.411]
53. Fornell C, Larcker DF. Evaluating structural equation models with unobservable variables and measurement error. J Mark Res 1981 Feb;18(1):39. [doi: 10.2307/3151312]
54. Hayes A. Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach. New York City, NY: Guilford Press; 2017.

XSL•FO
RenderX

55. Bocci F, Ferrari A, Sarini M. Putting the gaming experience at the center of the Therapy-The Video Game Therapy(®) approach. Healthcare (Basel) 2023 Jun 15;11(12):1767 [FREE Full text] [doi: 10.3390/healthcare11121767] [Medline: 37372884]
56. Ruiz M, Moreno M, Girela-Serrano B, Díaz-Oliván I, Muñoz LJ, González-Garrido C, et al. Winning the game against depression: a systematic review of video games for the treatment of depressive disorders. Curr Psychiatry Rep 2022 Feb 03;24(1):23-35. [doi: 10.1007/s11920-022-01314-7]
57. Nebel S, Ninaus M. Does playing apart really bring us together? Investigating the link between perceived loneliness and the use of video games during a period of social distancing. Front Psychol 2022 Feb 11;13:683842. [doi: 10.3389/fpsyg.2022.683842] [Medline: 35222171]
58. Mandryk RL, Frommel J, Armstrong A, Johnson D. How passion for playing world of warcraft predicts in-game social capital, loneliness, and wellbeing. Front Psychol 2020;11:2165 [FREE Full text] [doi: 10.3389/fpsyg.2020.02165] [Medline: 33071843]
59. Mikuška J, Vazsonyi AT. Developmental links between gaming and depressive symptoms. J Res Adolesc 2018 Sep 15;28(3):680-697. [doi: 10.1111/jora.12359] [Medline: 29139575]
60. Wei X, Zhou H, Zheng Q, Ren L, Chen N, Wang P, et al. Longitudinal interactions between problematic internet gaming and symptoms of depression among university students: differentiating anhedonia and depressed mood. Addict Behav 2025 Jan;160:108184 [FREE Full text] [doi: 10.1016/j.addbeh.2024.108184] [Medline: 39393293]
61. Granic I, Lobel A, Engels RCME. The benefits of playing video games. Am Psychol 2014 Jan;69(1):66-78. [doi: 10.1037/a0034857] [Medline: 24295515]
62. Russoniello C, O?Brien K, Parks J. The effectiveness of casual video games in improving mood and decreasing stress. J Cyber Ther Rehabil 2009;2(1):2-66.
63. Vuorre M, Johannes N, Magnusson K, Przybylski AK. Time spent playing video games is unlikely to impact well-being. R Soc Open Sci 2022 Jul 27;9(7):220411 [FREE Full text] [doi: 10.1098/rsos.220411] [Medline: 35911206]
64. Xia L, Roggeveen AL. How collective stress affects price fairness perceptions: the role of nostalgia. J Bus Res 2022 Nov;152:361-371. [doi: 10.1016/j.jbusres.2022.08.004]
65. Wildschut T, Sedikides C. Benefits of nostalgia in vulnerable populations. Eur Rev Soc Psychol 2022 Feb 26;34(1):44-91. [doi: 10.1080/10463283.2022.2036005]
66. Routledge C, Wildschut T, Sedikides C, Juhl J. Nostalgia as a resource for psychological health and well-being. Soc Pers Psychol Compass 2013 Nov 04;7(11):808-818. [doi: 10.1111/spc3.12070]
67. Gammon S, Ramshaw G. Distancing from the present: nostalgia and leisure in lockdown. Leisure Sci 2020 Jun 24;43(1-2):131-137. [doi: 10.1080/01490400.2020.1773993]
68. Batcho K. Nostalgia: retreat or support in difficult times? Am J Psychol 2013;126(3):355-367. [doi: 10.5406/amerjpsyc.126.3.0355] [Medline: 24027948]
69. Sedikides C, Wildschut T. Nostalgia: a bittersweet emotion that confers psychological health benefits. In: The Wiley Handbook of Positive Clinical Psychology. Hoboken, NJ: Wiley Blackwell; 2016:125-136.
70. Sedikides C, Wildschut T, Cheung W, Routledge C, Hepper EG, Arndt J, et al. Nostalgia fosters self-continuity: uncovering the mechanism (social connectedness) and consequence (eudaimonic well-being). Emotion 2016 Jun;16(4):524-539. [doi: 10.1037/emo0000136] [Medline: 26751632]
71. Anto A, Asif RO, Basu A, Kanapathipillai D, Salam H, Selim R, et al. Exploring the impact of social media on anxiety among university students in the United Kingdom: qualitative study. JMIR Form Res 2023 Jun 16;7:e43037 [FREE Full text] [doi: 10.2196/43037] [Medline: 37327030]
72. Hartnup B, Dong L, Eisingerich AB. How an environment of stress and social risk shapes student engagement with social media as potential digital learning platforms: qualitative study. JMIR Med Educ 2018 Jul 13;4(2):e10069 [FREE Full text] [doi: 10.2196/10069] [Medline: 30006324]
73. Anto A, Basu A, Selim R, Zaman J, Kanapathipillai D, Salam H, et al. Exploring the facilitators and barriers to addressing social media's impact on anxiety within primary care: a qualitative study. BJGP Open 2024 Jan 25;8(2):BJGPO.2023.0190. [doi: 10.3399/bjgpo.2023.0190]
74. El-Hilly AA, Iqbal SS, Ahmed M, Sherwani Y, Muntasir M, Siddiqui S, et al. Game on? Smoking cessation through the gamification of mHealth: a longitudinal qualitative study. JMIR Serious Games 2016 Oct 24;4(2):e18 [FREE Full text] [doi: 10.2196/games.5678] [Medline: 27777216]
75. Lin Y, Tudor-Sfetea C, Siddiqui S, Sherwani Y, Ahmed M, Eisingerich AB. Effective behavioral changes through a digital mHealth app: exploring the impact of hedonic well-being, psychological empowerment and inspiration. JMIR Mhealth Uhealth 2018 Jun 15;6(6):e10024 [FREE Full text] [doi: 10.2196/10024] [Medline: 29907557]
76. Eisingerich AB, Marchand A, Fritze MP, Dong L. Hook vs. hope: how to enhance customer engagement through gamification. Int J Res Mark 2019 Jun;36(2):200-215. [doi: 10.1016/j.ijresmar.2019.02.003]
77. Tudor-Sfetea C, Rabee R, Najim M, Amin N, Chadha M, Jain M, et al. Evaluation of two mobile health apps in the context of smoking cessation: qualitative study of cognitive behavioral therapy (CBT) versus non-CBT-based digital solutions. JMIR Mhealth Uhealth 2018 Apr 18;6(4):e98 [FREE Full text] [doi: 10.2196/mhealth.9405] [Medline: 29669708]

XSL•FO
RenderX

78. Eisingerich AB, Kretschmer T. In e-commerce, more is more. Harvard Business Review. 2008 Mar. URL: https://hbr.org/2008/03/in-e-commerce-more-is-more [accessed 2025-07-05]
79. Eisingerich AB, MacInnis DJ, Park CW. A theoretical framework exploring three foundational benefits of brand attachment. In: A Research Agenda for Brand Management in a New Era of Consumerism. Cheltenham, England: Edward Elgar Publishing; 2023:161-175.
80. Fritze MP, Eisingerich AB, Benkenstein M. Digital transformation and possession attachment: examining the endowment effect for consumers' relationships with hedonic and utilitarian digital service technologies. Electron Commer Res 2018 Jun 11;19(2):311-337. [doi: 10.1007/s10660-018-9309-8] [Medline: 26811110]
81. Fritze MP, Marchand A, Eisingerich AB, Benkenstein M. Access-based services as substitutes for material possessions: the role of psychological ownership. J Serv Res 2020 Mar 13;23(3):368-385. [doi: 10.1177/1094670520907691]
82. Bell SJ, Auh S, Eisingerich AB. Unraveling the customer education paradox: when, and how, should firms educate their customers? J Serv Res 2017 Feb 15;20(3):306-321. [doi: 10.1177/1094670517691847] [Medline: 22973420]
83. Foscht T, Lin Y, Eisingerich AB. Blinds up or down? The influence of transparency, future orientation, and CSR on sustainable responsible behavior. EJM 2018 Feb 02;52(3/4):476-498. [doi: 10.1108/EJM-10-2016-0576]
84. Eisingerich AB, MacInnis DJ, Whan Park C. Do CSR efforts that focus on helping the environment influence brand purchase more than other forms of CSR? J Bus Res 2023 Nov;168:114240. [doi: 10.1016/j.jbusres.2023.114240]
85. Lin Y, MacInnis DJ, Eisingerich AB. Strong anxiety boosts new product adoption when hope is also strong. J Mark 2020 Jul 10;84(5):60-78. [doi: 10.1177/0022242920934495]

Abbreviations

AVE: average variance extracted
CR: composite reliability

Edited by A Coristine; submitted 25.04.25; peer-reviewed by MJ Stroud; comments to author 13.06.25; revised version received 17.06.25; accepted 30.06.25; published 01.08.25.

Please cite as:
Arigayota A, Duffek B, Hou C, Eisingerich AB
Effects of The Legend of Zelda: Breath of the Wild and Studio Ghibli Films on Young People's Sense of Exploration, Calm, Mastery and Skill, Purpose and Meaning, and Overall Happiness in Life: Exploratory Randomized Controlled Study
JMIR Serious Games 2025;13:e76522
URL: https://games.jmir.org/2025/1/e76522
doi:10.2196/76522
PMID:

©Annisa Arigayota, Barbara Duffek, Congcong Hou, Andreas Benedikt Eisingerich. Originally published in JMIR Serious Games (https://games.jmir.org), 01.08.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Effects of a Board Game on Tic Management and Psychosocial Functioning in Adolescents With Tourette Syndrome: Randomized Controlled Trial

Mei Yin Lee[1], PhD; Huei Shyong Wang[2], MD; Chung Yueh Lien[3], PhD; Zhi Hong Chen[4], PhD

[1]School of Nursing, National Taipei University of Nursing and Health Sciences, Taipei City, Taiwan

[2]Linkou Chang Gung Memorial Hospital, Taoyuan City, Taiwan

[3]Department of Information Management, National Taipei University of Nursing and Health Sciences, Taipei City, Taiwan

[4]Graduate Institute of Information and Computer Education, National Taiwan Normal University, Taipei city, Taiwan

Corresponding Author:
Mei Yin Lee, PhD
School of Nursing
National Taipei University of Nursing and Health Sciences
No.365, Ming-Te Road, Peitou District
Taipei City, 112303
Taiwan
Phone: 886 228227101 ext 3108
Fax: 886 228213233
Email: meiyin0304@ntunhs.edu.tw

Abstract

Background: Tics and comorbidities significantly impact the social interactions and mental health of adolescents with Tourette syndrome (TS). Psychoeducation is an initial intervention for TS. Gamification is a common psychoeducational intervention for youths with chronic conditions. However, the effectiveness of board games in improving tic severity and mental health in adolescents with TS remains underexplored.

Objective: We developed a serious board game to investigate its effects on tic severity, mental health, social adjustment, and depression in adolescents with TS.

Methods: A single-blinded, 2-arm, parallel randomized controlled study was conducted. From September 2022 to July 2024, participants were recruited from a medical center in northern Taiwan. Seventy-nine adolescents with TS aged 12 to 18 years were randomly assigned to either a control group (n=39) or an intervention group (n=40). Both groups received care as usual (daily pyridoxine [50 mg] and psychoeducation), while the intervention group additionally participated in a weekly 60-minute board game session over a 4-week period. Outcome measures included the Yale Global Tic Severity Scale, Positive Mental Health Scale, Social Adjustment Scale for Adolescents with TS, and Beck Youth Inventory II – Depression scale.

Results: Generalized estimation equation results showed that, compared to the control group, the intervention group demonstrated significant improvements in positive mental health at the postintervention (β=5.19, 95% CI 0.36 to 10.02, *P*=.04) and follow-up (β=7.14, 95% CI 2.15 to 12.14, *P*=.005), with time-dependent effects. The intervention group also showed significant improvements in social adjustment (β=4.24, 95% CI 1.79 to 6.69, *P*<.001) and depression (β=–3.06, 95% CI –6.04 to –0.11, *P*=.04) at follow-up. No significant differences were observed between the 2 groups in tic severity.

Conclusions: The serious board game developed in this study significantly enhanced psychosocial functioning in adolescents with TS. As an alternative to verbal and written health communication, the board game serves as an innovative psychoeducational instrument for health care professionals to help adolescents with TS in tic management and mental health promotion. Future studies can develop and validate the feasibility of a digital version of the board game.

Trial Registration: ClinicalTrials.gov NCT05566236; https://clinicaltrials.gov/study/NCT05566236

(JMIR Serious Games 2025;13:e76208) doi:10.2196/76208

KEYWORDS

Tourette syndrome; adolescents; board game; mental health; psychosocial functioning

XSL•FO
RenderX

Introduction

Background

Tourette syndrome (TS) is a chronic neurodevelopmental disorder characterized by persistent, nonrhythmic motor tics and vocal tics that begin before the age of 18 years [1]. The prevalence of TS in youths is approximately 0.7% higher than in other age groups [2], lighting the need for care services tailored to their specific demands. Around 85% of individuals with TS have at least one psychiatric comorbidity, most commonly attention deficit hyperactivity disorder and obsessive-compulsive disorder [3]. Tics and comorbidities significantly impact the psychosocial functioning [4-7] and mental health of young people [8,9]. Compared to their typical (tic-free) peers, youths with TS exhibit poorer social adjustment [6,7] and higher rates of bullying [10]. Gamified strategies are interventions developed by health care providers to enhance physical and mental health in young people with chronic conditions [11,12]. The effectiveness of board games in improving tic management and psychosocial functioning in adolescents with TS remains unknown.

Due to the stigmatization of TS stemming from social framing and cultural beliefs, young people with TS often feel misunderstood and struggle with interpersonal interactions [9,13,14]. Adolescents with TS who experience higher levels of psychosocial stress demonstrate poorer social adjustment [5]. Depression is a known risk factor for psychosocial maladaptation in adolescents with TS [4]. Mild tics generally do not require treatment if they do not affect a child's life at school, in social interactions, or at home [15]. However, individuals with mild tics may still face challenges in certain social situations [14]. Psychoeducation serves as an initial intervention, providing patients with information about TS and strategies for coping with it [16]. It also enhances the understanding and fosters positive attitudes among peers, teachers, and family members toward individuals with TS [17-20]. If psychoeducation alone does not meet the care needs of youths with TS, psychological interventions [19] or pharmacological treatments [21] can be considered, provided they are regularly assessed and tailored to the individual needs of the youths [22].

Gamification is one of the psychoeducational interventions developed by health care providers to promote the physical and mental health of youths with chronic conditions. Gamification also transforms the process of imparting health-related knowledge and behaviors into an enjoyable interactive experience [11,12]. Situated learning theory, as a theoretical framework for board game design, incorporates simulation scenarios alongside gamification mechanics and elements to enhance participants' learning motivation [23,24]. Situated learning theory advocates creating a situated problem-solving context for learners, immersing them in learning. By designing scenarios analogous to learners' real-life contexts, it facilitates knowledge and skill acquisition through observation and exploratory actions [23-25]. Board games require players to engage in immersive learning scenarios and activities, fostering the development of problem-solving and social skills through discussions or role-playing tasks with other players [11,23,25]. Previous studies have demonstrated the effectiveness of board games in improving care-related skills, knowledge, adjustment, and physical and mental health in young people with chronic conditions [11,26-29]. Two recent studies have shown gamification in behavioral intervention for reducing tic severity [30], but video game elements increase tic frequency [31]. However, whether board games can improve tic and psychosocial functioning in adolescents with TS remains unclear. Parents' understanding of and attitudes toward TS influence how they educate their children about symptom management [32]. Parental evaluation is especially crucial for adolescents with TS, as it often mirrors the adolescents' concerns about their future [33]. However, as adolescents grow older, they gradually develop their own strategies for coping with tics and may not always adhere to recommendations from parents or teachers, potentially leading to conflict [13]. Psychoeducation interventions can enhance families' and school administrators' understanding of TS, reduce stereotyping [33], and help safeguard the mental health of young people with TS [9,17,22]. Adolescents may also benefit mentally from conventional psychoeducation approaches, including game-based learning [12]. The innovativeness and significance of this study lie in the fact that no previous research has developed a board game–based psychoeducation intervention to improve tic severity and psychosocial functioning in adolescents with TS.

Aims and Hypotheses

The aim of this study was to examine the effects of a board game intervention on tic severity, mental health, social adjustment, and depression in adolescents with TS compared to the control group. We hypothesized that (1) compared with the control group, implementing the board game intervention would significantly reduce tic severity and depression in adolescents with TS; and (2) compared with the preintervention stage, the intervention group would show improvements in mental health and social adjustment significantly exceeding those of the control group.

Methods

Study Design and Procedure

In this single-blind, 2-arm, parallel randomized controlled study, we evaluated the impacts and effectiveness of a board game intervention on tic severity, mental health, social adjustment, and depression in adolescents with TS. In Taiwan, care as usual (CAU) for youths with TS at pediatric outpatient clinics typically includes daily pyridoxine (50 mg) and psychoeducation [34]. Participants were randomly assigned to one of two groups: (1) the intervention group, which received CAU plus a board game intervention, and (2) the control group, which received outpatient CAU only. Changes in participants' primary outcome measures (tic severity and positive mental health) and secondary outcome measures (social adjustment and depression) were measured at baseline (preintervention, T1), 4 weeks after baseline (postintervention, T2), and 12 weeks after baseline (follow-up, T3). Due to the difficulty and challenging nature of blinding the participants and interveners, only the author

XSL•FO
RenderX

responsible for data analysis (CYL) was blinded. This study was registered in the ClinicalTrials.gov registry of the US National Library of Medicine (registration number: NCT05566236).

Setting and Participants

From September 2022 to July 2024, participants were recruited from a medical center in northern Taiwan and the pediatric neurology outpatient clinics of its branch hospitals. The inclusion criteria were as follows: (1) patients diagnosed with TS [1] by a pediatrician for over a year and aged 12-18 years; (2) patients able to speak and understand Mandarin; and (3) patients willing to participate in the study with parental approval and signed consent. The exclusion criterion was diagnosis with a severe illness, mental illness, or brain injury.

Sample Size and Randomization

The required sample size was estimated using G*Power 3.1.9.2 statistical software (Heinrich Heine University Düsseldorf), referencing a similar study [34]. With 3 repeated measurements, a Cohen *d* effect size of 0.3**,** a significance level of .05, and a statistical power of 0.80, the required sample size was calculated to be 62. Accounting for a 20 % dropout rate, at least 37 participants were needed in each group.

Randomization and blinding were performed independently by a nonparticipating researcher. Permuted block randomization was conducted using Random Allocation Software 2.0 (Informer Technologies), with 2 groups, a sample size of 92, and a fixed block size of 4. The computer-generated random sequences and group assignments were placed sequentially into tinted envelopes with continuous serial numbers, which were then sealed. After obtaining consent forms from eligible participants, the researcher opened the envelopes in order and assigned participants to either the intervention group or the control group based on the allocation inside the envelope.

Study Program

Intervention Group

The board game designed in this study, called the CO-Tics Board Game, aimed to encourage adolescents with TS to coexist with their tics through an interactive parent-adolescent gaming experience. To enhance engagement, the game features 2 maps—The Forest of the Lost Rabbit and The Mighty Rabbit Slays the Dragon—along with various game mechanics, such as interactive scenario cards and challenges for acquiring knowledge about TS and accumulating energy coins (see Figure 1). Drawing on insights from our published qualitative studies on youths with TS [9,13] and other studies [20,33], the interactive scenario cards were developed to reflect adolescents' lived experiences and potential interactions at home, school, and in public settings. The game promotes a nonjudgmental and inclusive attitude among players, encouraging parent and child participants to learn about symptom management while discussing simulated scenarios, sharing their feelings, and exploring coping strategies. The CO-Tics Board Game has acquired a utility model patent (patent number: M636691, applied on January 21, 2023) in Taiwan.

The game sessions were designed to last 1 hour per week over 4 consecutive weeks, with 2 to 3 players per session. The game set includes 2 maps, a game instruction, and a Tips and Tricks Manual, which contains 2 sections: question and answer about TS and interactive scenario card suggestions. Each player starts with 10 energy coins and rolls the dice to determine the order of play. Players roll the dice and move forward. The number on the dice determines the number of squares they can move, and they can take the cards and energy coins indicated in the spaces or move as instructed. There are 45 squares in the Forest of the Lost Rabbit map. The player's mission is to collect the weapon cards and energy coins needed to complete the Mighty Rabbit Slays the Dragon. There are 6 levels (with 24 interactive scenario cards) in the Mighty Rabbit Slays the Dragon map (Multimedia Appendix 1 provides a more detailed description). First, the player who draws the interactive scenario card is determined by spinning the wheel. After that player responds with their personal thoughts, the other players take turns expressing their opinions, and finally, they can consult the advice in the Tips and Tricks manual for coping strategies for the simulated situation.

The interactive scenario cards address 4 areas: family interactions, school life, social interactions, and coexisting with TS. For example, scenario 1 depicts the following incident:

> *One time at the bookstore, I was constantly shrugging and clearing my throat. I noticed that the owner kept staring at me and eventually said, 'You're acting weird. You should see a doctor if you're sick.' I felt hurt by their comment and left the bookstore immediately. I didn't want to leave the house for a while afterward.*

The Tips and Tricks Manual suggests the following for this scenario: some adolescents with TS have experienced judgmental looks from strangers in places like cinemas, restaurants, or buses and have even been directly asked about their unusual movements or sounds. These situations may escalate into conflicts and highlight a widespread lack of understanding about TS. Parents are encouraged to listen to the adolescent's positive and negative feelings, offering support and validation. To aid players in expressing their emotions about various scenarios, a set of cards was created, featuring 30 types of positive and negative feelings. In addition to the game instructions, we also created a 10-minute video, which can be accessed by scanning a QR code to enhance the participants' understanding about the gameplay of the board game. We also send weekly reminders through SMS, email, or the LINE app to remind the participants to play the board game for a 60-minute session. We also asked adolescents and their parents to report the time spent playing the game and to submit feedback and questions via a Google Form. These insights allow us to understand and assess the status of the participants' board game engagement. In this study, the intervention group received CAU and participated in the CO-Tics Board Game to enhance their understanding of TS and develop interpersonal skills and coping strategies.

XSL•FO
RenderX

Figure 1. CO-Tics Board Game.



Control Group

The control group (CAU) in this study received daily pyridoxine (50 mg, orally) and at least one paper-based and oral psychoeducation session delivered by a physician or nurse in the outpatient clinic, providing youths with TS and their families with coping strategies for daily life issues caused by tics and comorbidities, such as medication use, reducing tic frequency, and enhancing interpersonal interaction skills, as well as offering psychological support and relevant health care resources.

Outcomes and Measures

Sociodemographic Variables

The demographic information collected for adolescents with TS included age, sex, education level, age at TS diagnosis, comorbidities, and medication use. For their parents, demographic information included age, education level, and marital status.

Primary Outcome Measures

Yale Global Tic Severity Scale: The Yale Global Tic Severity Scale (YGTSS) is a widely used instrument for assessing tic severity in adolescents [35]. Items on the YGTSS are scored on a 0-5 Likert scale across 5 dimensions: frequency, intensity, number, complexity, and interference. The YGTSS also includes a total impairment scale that evaluates the patient's functioning in interpersonal, academic, and occupational domains. The total tic score is calculated by summing the total motor tic score (0-25 points), vocal tic score (0-25 points), and total impairment score (0-50 points), with higher scores indicating greater tic severity. The YGTSS has demonstrated high reliability and validity [36]. The YGTSS scores of the participants in both groups were scored by a pediatric neurologist (a member of our research team).

Positive Mental Health Scale: Adolescence is a critical period for mental health development as it is essential for establishing social interactions and emotional habits [37]. Grounded in positive psychology, Chen [38] developed the Positive Mental Health Scale (PMHS) to assess participants' mental health status. The PMHS comprises 5 subscales: self-acceptance (3 items), human relationships (5 items), family harmony (6 items), emotional balance (4 items), and optimism and enterprise (7 items). Adolescents self-report their responses, scoring items on a 5-point Likert scale, with a total score range of 25 to 125 points. Higher scores indicate better mental health status. The Cronbach α for the PMHS subscales ranges from 0.76 to 0.92, with a total scale Cronbach α of 0.92. Confirmatory factor analysis has shown the PMHS to have good construct validity and model fit [38].

Secondary Outcome Measures

Social Adjustment Scale for Adolescents with Tourette Syndrome: The Social Adjustment Scale for Adolescents with Tourette Syndrome (SASATS) was developed by Lee et al [13] in 2020 based on a qualitative study to measure social

XSL•FO
RenderX

adjustment in adolescents. The SASATS is a self-reported measure of an adolescent's level of adjustment during interpersonal and social interactions over the past 2 weeks. The 17-item SASATS consists of 4 subscales: the relationship between self and TS (4 items), academic performance (5 items), family relationship (5 items), and peer interaction (4 items) [39]. Each item is scored on a scale of 1 to 4 points, yielding a total score range of 17 to 68 points, with higher scores indicating better social adjustment. The SASATS has demonstrated strong psychometric properties, with an α coefficient of 0.87 and a test-retest reliability of 0.81 [39]. A previous study reported a Cronbach α of 0.82 for the SASATS [5].

Beck Youth Inventory II - Depression scale: The Beck Youth Inventory II (BYI-II)- Depression scale evaluates adolescents' negative thoughts about themselves, life, and the future, as well as feelings of sadness and physiological symptoms of depression [40]. This study used the Chinese version of the BYI-II to measure the self-reported level of depression in adolescents with TS [41]. The scale consists of 20 items scored on a 4-point Likert scale (never, seldom, sometimes, or always), with a total score range of 0 to 60 points. Higher scores indicate more severe depression. The criterion-related validity of the Chinese version is significantly correlated with the original BYI-II. The Chinese version has excellent reliability and validity, with a Cronbach α of 0.93 and a test-retest reliability of 0.81 [41].

Data Analysis

The data were analyzed using intention-to-treat analysis with SPSS 29.0 software (IBM Corporation). Statistical analysis was conducted at the preparticipation, group assignment, and enrollment stages. Demographic data were analyzed using descriptive statistics, with categorical variables represented as frequency distributions and percentages, and continuous variables represented as means and SDs. The homogeneity of the 2 groups' preintervention demographic variables and psychosocial functioning was examined using chi-square tests for categorical variables and independent samples *t* tests for continuous variables. For statistical inference, generalized estimating equations were used to analyze differences in social adjustment, positive mental health, depression levels, and tic severity between the 2 groups, over time, and for group-time interactions. The level of statistical significance was set at $P<.05$.

Ethical Considerations

Ethical approval for the study was obtained from the Institutional Review Board of the Chang Gung Memorial Hospital (201801998B0). The research team strictly adhered to the ethical principles outlined in the Declaration of Helsinki, such as confidentiality, beneficence, respect, impartiality, and nonmaleficence. The researchers explained the study's purposes and procedures to eligible adolescents and their parents, emphasizing that their rights and interests were protected, that participation was voluntary, and that they could withdraw at any time without providing a reason and without affecting their access to treatment. Participant enrollment commenced only after the adolescents and their parents had signed a consent form.

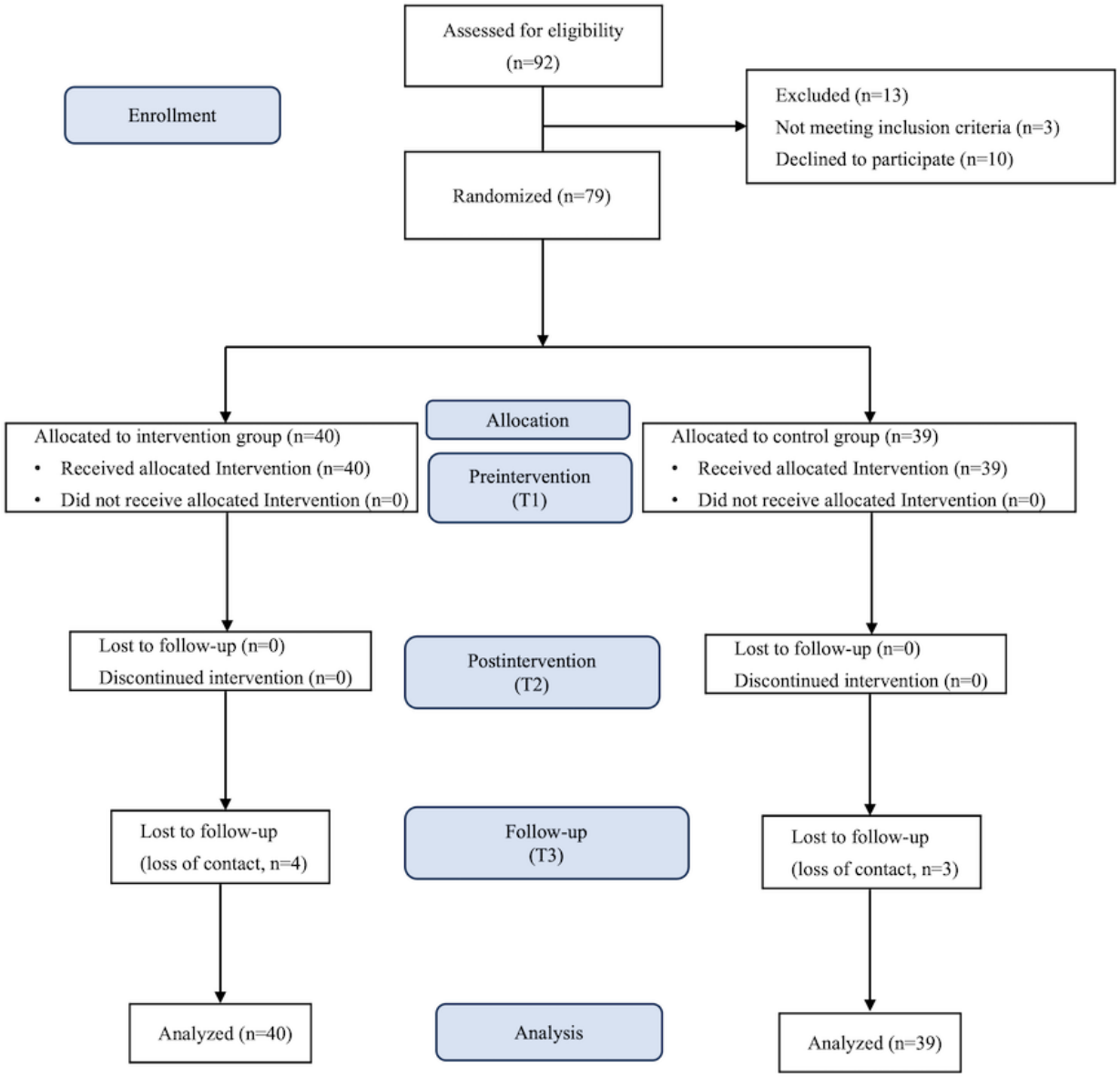
Results

Participant Characteristics

Figure 2 shows the CONSORT (Consolidated Standards of Reporting Trials) flow diagram for the study (Multimedia Appendix 2). Initially, 92 adolescents with TS were screened, and 79 were randomly assigned to either the intervention group (n=40) or the control group (n=39). Seven participants were lost to follow-up, resulting in a dropout rate of 8.9%. A large percentage of the participants were male (n=68, 86%). Approximately 70% (n=56) had at least one comorbidity, primarily attention deficit hyperactivity disorder, and about 32% (n=25) were taking medications, predominantly clonidine. Significant differences were observed between the 2 groups in terms of age ($P=.002$) and education level ($P=.046$), but not for other demographic variables (descriptive statistics for participants' characteristics are shown in Table 1).

XSL•FO
RenderX

Figure 2. CONSORT (Consolidated Standards of Reporting Trials) flow diagram of participants' enrolment and recruitment process.



XSL•FO
RenderX

Table 1. Characteristics of the participating adolescents and their parents.

Characteristics	Intervention group (n=40)	Control group (n=39)	Chi-square (*df*)	*P* value
Adolescents				
Sex			0.78 (1)	.78[a]
Male	34	34		
Female	6	5		
Age (years), mean (SD)	14.33 (1.66)	15.72 (2.11)	–3.265 (77)	.002[b]
Education level			6.161 (2)	.046[c]
Junior high school	31	20		
Senior high school	7	13		
University	2	6		
Age (years) at TS[d] diagnosis, mean (SD)	7.63 (3.746)	6.33 (4.10)	1.462 (77)	.07[b]
Comorbidity			0.002 (1)	.96[a]
No	12	11		
Yes	28	28		
ADHD[e], n (%)	18 (46)	21 (54)		
OCD[f], n (%)	4 (57)	3 (43)		
Other[g], n (%)	10 (59)	7 (41)		
Medication intake			1.284 (1)	.26 [a]
No	25	29		
Yes	15	10		
Clonidine, n (%)	10 (53)	9 (47)		
Aripiprazole, n (%)	4 (100)	0 (0)		
Other[h], n (%)	6 (75)	2 (25)		
Parents				
Paternal age (years), mean (SD)	47.85 (4.52)	48.64 (6.81)	–0.609 (77)	.27[b]
Father's education			2.344 (3)	.50[c]
Junior high school	2	0		
Senior high school	5	7		
University or higher	33	32		
Maternal age (years), mean (SD)	45.35 (4.29)	47.15 (6.66)	–1.434 (77)	.08[b]
Mother's education			1.353 (3)	.72 [c]
Junior high school	1	0		
Senior high school	7	7		
University or higher	32	32		
Parental marital status			1.201 (2)	.55[c]
Married	37	36		
Divorced	2	3		
Separated	1	0		

[a]Chi-square test.

[b]Independent *t* tests.

XSL•FO
RenderX

[c]Fisher exact test.

[d] TS: Tourette syndrome.

[e]ADHD: attention deficit hyperactivity disorder.

[f]OCD: obsessive-compulsive disorder.

[g]Other comorbidities included anxiety and learning disabilities.

[h]Other medications included risperidone and atomoxetine.

Postintervention Outcomes

All participants in the intervention group completed at least one 60-minute board game session per week for 4 consecutive weeks. No adverse reactions were reported among participants throughout the study period. In the present study, the Cronbach α was 0.65 for the YGTSS, 0.94 for the PMHS, 0.88 for the SASATS, and 0.94 for the BYI-II Depression scale. Table 2 present the mean scores, range and CI for the YGTSS, PMHS, SASATS, and BYI-II Depression scale at the preintervention (baseline), postintervention (4 weeks after baseline), and follow-up (12 weeks after baseline). Additionally, we categorized YGTSS scores as mild (<20) or moderate-to-severe (>20). Chi-square and Fisher exact tests showed no significant differences at baseline (*P*=.31), postintervention (*P*=.31), or follow-up (*P*=.78). The difference in SASATS scores between the 2 groups at follow-up was 4.11 points (*t*=2.25; 95% CI 0.44 to 7.34) with an effect size of 0.53, as shown in Table 2. These results indicate that 8 weeks after completing the board game intervention, the intervention group exhibited a significantly higher level of social adjustment compared to the control group. Table 3 presents the generalized estimating equations results for the outcomes of the board game intervention regarding positive mental health, social adjustment, depression, and tic severity after adjusting for adolescents' age and education level. There were no significant differences between the 2 groups' preintervention PMHS, SASATS, BYI-II Depression scale, and YGTSS scores (*P*>.05). Regarding tic severity, no significant differences were observed in YGTSS scores between the 2 groups at either the postintervention (β=–1.5, 95% CI –4 to 1, *P*=.24) or follow-up (β=0.55, 95% CI –2.02 to 3.13, *P*=.68) stages. We further analyzed and found no significant differences in the total tic severity and impairment scores between the 2 groups during the postintervention or follow-up period. The group-time interaction results revealed that, compared to the control group, the intervention group showed significantly greater improvements in PMHS scores at the postintervention (β=5.19, 95% CI 0.36 to 10.02, *P*=.04) and follow-up (β=7.14, 95% CI 2.15 to 12.14, *P*=.005) stages. This suggests that the board game intervention immediately improved positive mental health, with these effects persisting for 8 weeks, indicating time-dependent benefits. The intervention group exhibited significantly more significant improvements in SASATS (β=4.24, 95% CI 1.79 to 6.69, *P*<.001) and BYI-II Depression scale (β=–3.06; 95% CI –6.04 to –0.11, *P*=.04) scores compared to the control group at the follow-up.

We synthesized qualitative information provided via a Google Form by the participants in the intervention group during the board game sessions. Several adolescents reported that the card and map designs were interesting and less rigid and more accessible than standard educational pamphlets or books. Playing the game with a parent not only increased knowledge about TS but also appeared to help parents better understand the disorder's impact on adolescents' daily life and social interactions. Some parents indicated that the game facilitated understanding of their child's thoughts and behaviors; sharing experiences during play enhanced parent-child dialogue and interaction. A few parents further noted that, upon hearing their child's responses to simulated interpersonal scenarios, they felt both concerned about potential misunderstanding by others and surprised at their child's more mature coping strategies than they could expect, prompting their reflection on how to adjust their own interactions with their kids.

XSL•FO
RenderX

Table 2. Comparison of outcome variables between the intervention and control groups at the baseline (T1), postintervention (T2), and follow-up (T3) time points.

Outcome variables	Intervention group (n=40)		Control group (n=39)		Group difference (95% CI)	Effect size
	Mean (SD)	Range	Mean (SD)	Range		
Yale Global Tic Severity Scale						
T1	15.05 (10.99)	0-52	14.77 (13.27)	0-63	0.103 (–5.17 to 5.73)	0.023
T2	14.55 (10.73)	2-50	15.77 (12.69)	0-60	–0.461 (–6.48 to 4.04)	–0.104
T3	15.17 (9.94)	3-49	14.05 (10.99)	3-55	0.453 (–3.78 to 6.01)	0.106
Total tic severity score						
T1	11.05 (6)	0-28	9.36 (6.54)	0-25	1.198 (–1.119 to 4.501)	0.27
T2	11.05 (6.36)	2-30	11.67 (7.67)	0-40	–0.39 (–3.769 to 2.536)	–0.088
T3	12.11 (6.32)	3-39	10.81 (6.93)	3-35	0.837 (–1.797 to 4.397)	0.196
Total impairment score						
T1	4 (6.72)	0-30	5.38 (8.22)	0-40	–0.821 (–4.745 to 1.975)	–0.185
T2	3.5 (5.8)	0-20	4.1 (6.37)	0-20	–0.44 (–3.33 to 2.125)	–0.099
T3	3.06 (5.25)	0-10	3.24 (5.3)	0-20	–0.152 (–2.649 to 2.274)	–0.036
Positive Mental Health Scale						
T1	93.28 (17.91)	53-124	94.79 (15.42)	40-117	–0.40 (–9.02 to 5.98)	–0.091
T2	97.93 (18.58)	52-125	94.26 (15.84)	47-125	0.94 (–4.08 to 11.41)	0.212
T3	95.67 (15.44)	63-125	90.36 (14.82)	44-114	1.488 (–1.81 to 12.42)	0.351
Social Adjustment Scale for Adolescents with Tourette Syndrome						
T1	51.55 (8.54)	25-66	52.36 (6.16)	30-64	–0.48 (–4.15 to 2.53)	–0.108
T2	53.75 (7.52)	39-67	53.33 (6.64)	31-68	0.26 (–2.76 to 3.60)	0.059
T3	55.42 (7.57)	40-68	51.31 (6.88)	29-67	2.25 (0.44 to 7.34)	0.53
Beck Youth Inventory II Depression Scale						
T1	13.50 (9.96)	0-36	12.26 (10.69)	0-52	0.54 (–3.38 to 5.87)	0.120
T2	13.18 (10.29)	0-46	13.74 (10.27)	0-36	–0.25 (–5.18 to 4.04)	–0.055
T3	13.14 (9.99)	0-39	13.94 (10.78)	0-51	–0.33 (–5.69 to 4.08)	–0.078

XSL•FO
RenderX

Table 3. The results of generalized estimating equation on tics severity and psychosocial functions.

Parameters	Estimate (β)	SE	95% CI	Wald chi-square (*df*)	*P* value
Yale Global Tic Severity Scale[a]					
Intercept	8.55	17.17	–25.11 to 42.20	0.25 (1)	.62
Intervention group[b]	1.46	2.72	–3.87 to 6.79	0.29 (1)	.59
Intervention group × T2[c]	–1.5	1.28	–4 to 1	1.39 (1)	.24
Intervention group × T3[c]	0.55	1.31	–2.02 to 3.13	0.18 (1)	.68
Total tic severity score[a]					
Intercept	25.88	8.77	8.69 to 43.06	8.71 (1)	.003
Intervention group[b]	0.26	1.7	–3.08 to 3.59	0.02(1)	.09
Intervention group × T2[c]	–2.31	0.78	–3.83 to 0.79	1.82(1)	.10
Intervention group × T3[c]	–1.04	1.36	–3.71 to 1.62	0.59 (1)	.44
Total impairment score[a]					
Intercept	0.79	8.66	–16.18 to 17.75	0.008 (1)	.93
Intervention group[b]	–0.45	1.49	–3.37 to 2.46	0.093 (1)	.76
Intervention group × T2[c]	0.78	0.35	–0.09 to 1.48	2.88 (1)	.08
Intervention group × T3[c]	1.34	0.77	–0.16 to 2.84	3.072 (1)	.08
Positive mental health scale[a]					
Intercept	139.32	23.21	93.82 to 184.81	36.02 (1)	<.001
Intervention group[b]	–4.16	3.81	–11.64 to 3.32	1.19 (1)	.28
Intervention group × T2[c]	5.19	2.47	0.36 to 10.02	4.44 (1)	.04
Intervention group × T3[c]	7.14	2.55	2.15 to 12.14	7.87 (1)	.005
Social adjustment scale for adolescents with Tourette syndrome[a]					
Intercept	67.28	10.36	46.97 to 87.60	42.15 (1)	<.001
Intervention group[b]	–1.44	1.73	–4.82 to 1.95	0.69 (1)	.41
Intervention group × T2[c]	1.23	1.21	–1.14 to 3.59	1.03 (1)	.31
Intervention group × T3[c]	4.24	1.25	1.79 to 6.69	11.56 (1)	<.001
Beck Youth Inventory II Depression Scale[a]					
Intercept	-6.55	14.83	–35.61 to 22.52	0.19 (1)	.66
Intervention group[b]	2.30	2.41	–2.43 to 7.03	0.91 (1)	.34
Intervention group × T2[c]	–1.81	1.46	–4.67 to –1.04	1.55 (1)	.21
Intervention group × T3[c]	–3.06	1.51	–6.04 to –0.11	4.12 (1)	.04

[a]Adjusted adolescents' age and education.

[b]Control group.

[c]Control group × preintervention.

Discussion

Principal Findings

Partially in line with our expectations, the board game group's baseline to postintervention positive mental health, social adjustment, and depression improved significantly compared to the control group. However, no significant changes were observed in tic severity. According to the European clinical guidelines for TS, psychoeducation involves the precise delivery of information about the symptoms, management, prognosis,

XSL•FO
RenderX

and daily experience of the condition [17]. It is also a frequently used component in designing comprehensive behavioral interventions for tics [34,42].

Comparison With Previous Work

Health professionals and researchers often use psychoeducation tools such as written vignettes, videos, workshops, and training programs to provide information about TS [20]. Unlike previous approaches, this is the first study to design a board game intervention based on the psychosocial experiences of adolescents with TS [9,13]. The board game promotes parent-adolescent interactions while addressing the daily life experiences of adolescents with TS and their families. To avoid imbalanced parent-adolescent power dynamics, both players—the adolescent with TS and their parent—begin the game with equal roles, fostering inclusive and nonjudgmental attitudes during gameplay. Mirror cards were specifically designed to enhance parent-adolescent interaction and communication by encouraging both participants to express their feelings about the scenario cards they drew and collaboratively devise coping strategies.

A systematic review found that the duration of board game play in interventions varies significantly, ranging from 5 minutes to 30 hours depending on the study's objectives and game design [28]. Compared to other studies involving adolescents [11,43], the duration of play in this study was 1 hour per week for 4 consecutive weeks. The 4-week intervention period was chosen because it aligned with the 24 interactive scenario cards designed for the game, allowing for the discussion of 6 different scenarios each week. We found that gamification enhances adolescents' mental health, consistent with previous studies [11,12,28]. In this study, group-time interaction analysis revealed significantly higher PMHS scores in the intervention group at both the postintervention and follow-up. In other words, the board game enhanced positive mental health in adolescents with TS, demonstrating time-dependent effects. Similar to findings in other studies, board games have been shown to reduce negative emotions in youths with cancer [27] and depression [44]. In the current study, group-time interactions showed that, compared to the control group, the intervention group had significantly lower depression inventory scores, with immediate effects observed at the postintervention stage.

Understanding the social adjustment of youths with chronic illnesses enables health care professionals to develop effective strategies that address their psychosocial functioning and needs [45]. In adolescents with TS, the evaluation of social adjustment includes 4 dimensions: academic performance, family relationships, peer interactions, and the relationship between self and TS [39]. Support from friends, peers, and family plays a critical role in promoting healthy social adjustment in adolescents with TS [13]. An Indonesian study found that a board game intervention reduced stigma, raised awareness of mental health problems, and encouraged positive peer involvement in health-seeking behaviors among senior high school students in remote areas [12]. Similarly, a situated family board game fostered interaction and dialogue between youths with chronic illnesses and their parents, facilitated discussions about daily life, and enhanced self-care skills [26]. In line with these findings, group-time interaction analysis revealed that the intervention group's SASATS score at follow-up was significantly higher than that of the control group by 4.11 points, indicating that social adjustment improvements were greater 8 weeks after the board game intervention.

Regarding tic severity, the intervention group's YGTSS score at postintervention was lower by 1.22 point, but no significant group, time, or group-time differences were observed. A possible explanation is because most of the participants had mild tics, and the postintervention follow-up period was only 8 weeks, making it difficult to observe significant long-term changes. Furthermore, in this study, the YGTSS demonstrated a Cronbach α of 0.65, slightly lower than the standard threshold for internal consistency. This may be related to most participants having mild tics and the small sample size; future studies are advised to increase the sample size to potentially improve reliability.

Strengths

This is the first study to develop a serious board game based on the psychosocial experiences of adolescents with TS [9,13], simulating real-life social interaction contexts to meet their care needs. In addition to enhancing knowledge about TS, the board game also includes a wide array of interactive scenario cards for adolescents with TS and their parents to simulate real-life interpersonal interactions, discuss and share their respective coping methods, and foster parent-child communication. Furthermore, to validate the effectiveness of this intervention, we used several reliable and valid tools to measure the variables related to tic severity and psychosocial functioning. Psychoeducation for TS can be delivered without requiring specialist training in psychotherapy [17]. We recommend that health care professionals consider using the board game as a psychoeducation tool in research and practice to improve psychosocial functioning outcomes in adolescents with TS.

Limitations and Future Work

Several limitations of this study must be acknowledged. First, the intervention could not be blinded, as participants were required to play the board game. Additionally, performance bias could not be eliminated because the same researcher was responsible for both group allocation and intervention delivery after opening the envelopes. Second, considering that individuals with mild tics also require psychoeducation [14], no tic severity threshold was set as an inclusion criterion, which may have introduced sampling bias. Moreover, this study did not collect data on whether participants in either group experienced changes in medication use during the follow-up period; therefore, the relationship and potential impact of medication on changes in tic severity remain unclear. Third, because the study design was aimed at a parent-adolescent board game and parental feedback was collected via a Google Form, our ability to comprehensively assess parents' postintervention changes and experiences was limited. Finally, the board game intervention was followed for only 12 weeks, making it impossible to determine its long-term benefits.

Methodological recommendations for future studies are as follows: (1) assign group allocation and intervention delivery to separate personnel to reduce the risk of bias and strengthen

XSL•FO
RenderX

internal validity; (2) treat baseline tic severity as an inclusion or stratification criterion to enable subgroup comparisons, and measure and control medication use as a potential confounder; (3) incorporate parental perspectives via focus group interviews or quantitative surveys of parent-child interaction satisfaction, using parental reports as additional indicators of intervention effectiveness; and (4) extend follow-up to 6-12 months to clarify the longer-term effects of the serious board game on adolescents' psychosocial functioning. We also acknowledge that digital and nondigital board games offer distinct gaming experiences. Future research could develop an interactive digital version of the board game and compare the effects of digital board games, nondigital board games, and CAU on the physical and mental health of adolescents with TS, thereby gaining further insights into the mechanisms and relative effectiveness of different game formats. In addition, the simulated scenarios in this board game were designed specifically for adolescents aged 12-18 years with TS. Future research could adapt the game to the developmental characteristics and life contexts of school-age children, aiming to improve their understanding of TS and reduce potential bullying or stigma among peers.

Conclusions

In conclusion, this is the first study to develop a serious board game simulating social interaction contexts based on the psychosocial experiences of adolescents with TS. Our findings show that combining CAU with the board game effectively enhances adolescents' positive mental health and social adjustment and reduces depression, but does not significantly improve tic severity. Our findings support the use of a serious board game to improve psychosocial functioning in adolescents with TS; however, we did not obtain definitive evidence of clinically meaningful reductions in tic severity. To improve the quality of care for youth with TS, future work should develop and rigorously evaluate a digital version of the board game, determining its feasibility and effectiveness in meeting care needs.

Acknowledgments

We thank all the adolescents and their parents for their participation and support. This work was funded by the National Science and Technology Council (grant NSTC 110-2314-B-227-007-).

Data Availability

Data are available from the corresponding author upon reasonable request. The data are not publicly available due to privacy or ethical restrictions.

Authors' Contributions

MYL was the chief investigator and contributed to the design, execution, and supervision of the study. HSW contributed to the design and data collection. CYL and ZHC were responsible for data analysis. MYL drafted the article. All authors revised the article critically and approved the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Board game summaries.
[PDF File (Adobe PDF File), 475 KB - games_v13i1e76208_app1.pdf]

Multimedia Appendix 2
CONSORT-ehealth checklist.
[PDF File (Adobe PDF File), 1289 KB - games_v13i1e76208_app2.pdf]

References

1. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, Text Revision. Washington, DC: American Psychiatric Association; 2022.
2. Jafari F, Abbasi P, Rahmati M, Hodhodi T, Kazeminia M. Systematic review and meta-analysis of Tourette syndrome rrevalence; 1986 to 2022. Pediatr Neurol 2022 Dec;137:6-16. [doi: 10.1016/j.pediatrneurol.2022.08.010] [Medline: 36182698]
3. Hirschtritt ME, Lee PC, Pauls DL, Dion Y, Grados MA, Illmann C, Tourette Syndrome Association International Consortium for Genetics. Lifetime prevalence, age of risk, and genetic relationships of comorbid psychiatric disorders in Tourette syndrome. JAMA Psychiatry 2015 Apr;72(4):325-333 [FREE Full text] [doi: 10.1001/jamapsychiatry.2014.2650] [Medline: 25671412]
4. Ho CS, Chen YC, Lee MY, Chen CJ, Tsai HJ. Factors associated with Tourette syndrome among adolescents in Taiwan: a cross-sectional correlational study. J Pediatr Nurs 2023;73:e75-e82. [doi: 10.1016/j.pedn.2023.07.012] [Medline: 37544859]

XSL•FO
RenderX

5. Lee MY, Wang HS, Lee TY. Psychosocial stress, self-esteem, and social adjustment: a moderated mediation analysis in Taiwanese adolescents with Tourette syndrome. J Pediatr Nurs 2022;62:e84-e90. [doi: 10.1016/j.pedn.2021.07.013] [Medline: 34330565]
6. Lin YJ, Lai MC, Gau SS. Youths with ADHD with and without tic disorders: comorbid psychopathology, executive function and social adjustment. Res Dev Disabil 2012;33(3):951-963. [doi: 10.1016/j.ridd.2012.01.001] [Medline: 22285732]
7. Stokes A, Bawden H, Backman J, Dooley J, Camfield P. Peer problems in Tourette's disorder. Pediatrics 1991;87(6):936-942. [doi: 10.1542/peds.87.6.936]
8. Johnco C, McGuire JF, McBride NM, Murphy TK, Lewin AB, Storch EA. Suicidal ideation in youth with tic disorders. J Affect Disord 2016 Aug;200:204-211 [FREE Full text] [doi: 10.1016/j.jad.2016.04.027] [Medline: 27136419]
9. Lee MY, Wang HS, Chen CJ. Psychosocial experiences in youth with Tourette syndrome: a systematic review and meta-synthesis. Eur Child Adolesc Psychiatry 2024 Nov;33(11):3787-3802. [doi: 10.1007/s00787-023-02339-w] [Medline: 38129352]
10. Pring K, Malli M, Hardy BW, Rapp SR, Storch EA, Mink JW, et al. Reframing stigma in Tourette syndrome: an updated scoping review. Eur Child Adolesc Psychiatry 2025 Jan;34(1):19-39. [doi: 10.1007/s00787-023-02332-3] [Medline: 38159134]
11. Gkintoni E, Vantaraki F, Skoulidi C, Anastassopoulos P, Vantarakis A. Promoting physical and mental health among children and adolescents via gamification-a conceptual systematic review. Behav Sci (Basel) 2024 Jan 29;14(2):102 [FREE Full text] [doi: 10.3390/bs14020102] [Medline: 38392455]
12. Respati T, Feriandi Y, Frederico R, Nugroho E, Pardosi JF, Withall A. Efficacy of a mental health game-board intervention for adolescents in remote areas: reducing stigma and encouraging peer engagement. TOPHJ 2024 Jun 06;17(1):e18749445310785. [doi: 10.2174/0118749445310785240603045859]
13. Lee MY, Wang HS, Chen CJ, Lee MH. Social adjustment experiences of adolescents with Tourette syndrome. J Clin Nurs 2019 Jan;28(1-2):279-288. [doi: 10.1111/jocn.14564] [Medline: 29893437]
14. Wadman R, Tischler V, Jackson GM. 'Everybody just thinks I'm weird': a qualitative exploration of the psychosocial experiences of adolescents with Tourette syndrome. Child Care Health Dev 2013 Nov;39(6):880-886. [doi: 10.1111/cch.12033] [Medline: 23363392]
15. Hartmann A, Ansquer S, Brefel-Courbon C, Burbaud P, Castrioto A, Czernecki V, et al. French guidelines for the diagnosis and management of Tourette syndrome. Rev Neurol (Paris) 2024 Oct;180(8):818-827. [doi: 10.1016/j.neurol.2024.04.005] [Medline: 38760282]
16. Set KK, Warner JN. Tourette syndrome in children: an update. Curr Probl Pediatr Adolesc Health Care 2021 Jul;51(7):101032. [doi: 10.1016/j.cppeds.2021.101032] [Medline: 34305006]
17. Andrén P, Jakubovski E, Murphy TL, Woitecki K, Tarnok Z, Zimmerman-Brenner S, et al. European clinical guidelines for Tourette syndrome and other tic disorders-version 2.0. Part II: psychological interventions. Eur Child Adolesc Psychiatry 2022 Mar;31(3):403-423 [FREE Full text] [doi: 10.1007/s00787-021-01845-z] [Medline: 34313861]
18. Lee MY. Living with tics: nursing care of pediatric tourette syndrome. Biomed J 2022 Apr;45(2):280-285 [FREE Full text] [doi: 10.1016/j.bj.2021.10.011] [Medline: 34710638]
19. Müller-Vahl KR, Szejko N, Verdellen C, Roessner V, Hoekstra PJ, Hartmann A, et al. European clinical guidelines for Tourette syndrome and other tic disorders: summary statement. Eur Child Adolesc Psychiatry 2022 Mar;31(3):377-382 [FREE Full text] [doi: 10.1007/s00787-021-01832-4] [Medline: 34244849]
20. Nussey C, Pistrang N, Murphy T. How does psychoeducation help? A review of the effects of providing information about Tourette syndrome and attention-deficit/hyperactivity disorder. Child Care Health Dev 2013 Sep;39(5):617-627. [doi: 10.1111/cch.12039] [Medline: 23461278]
21. Roessner V, Eichele H, Stern JS, Skov L, Rizzo R, Debes NM, et al. European clinical guidelines for Tourette syndrome and other tic disorders-version 2.0. Part III: pharmacological treatment. Eur Child Adolesc Psychiatry 2022 Mar;31(3):425-441 [FREE Full text] [doi: 10.1007/s00787-021-01899-z] [Medline: 34757514]
22. Pringsheim T, Okun MS, Müller-Vahl K, Martino D, Jankovic J, Cavanna AE, et al. Practice guideline recommendations summary: treatment of tics in people with Tourette syndrome and chronic tic disorders. Neurology 2019 May 07;92(19):896-906. [doi: 10.1212/wnl.0000000000007466]
23. Hou HT, Li MC. Evaluating multiple aspects of a digital educational problem-solving-based adventure game. Computers in Human Behavior 2014 Jan;30:29-38. [doi: 10.1016/j.chb.2013.07.052]
24. Zainuddin Z, Chu SKW, Shujahat M, Perera CJ. The impact of gamification on learning and instruction: a systematic review of empirical evidence. Educational Research Review 2020 Jun;30:100326. [doi: 10.1016/j.edurev.2020.100326]
25. Wu WH, Hsiao HS, Wu PL, Lin CH, Huang SH. Investigating the learning‐theory foundations of game‐based learning: a meta‐analysis. Comput Assist Learn 2011 Aug 15;28(3):265-279. [doi: 10.1111/j.1365-2729.2011.00437.x]
26. Kennedy A, Semple L, Alderson K, Bouskill V, Karasevich J, Riske B, et al. Don't push your luck! Educational family board (not bored) game for school-age children living with chronic conditions. J Pediatr Nurs 2017;35:57-64. [doi: 10.1016/j.pedn.2017.02.032] [Medline: 28728770]

XSL•FO
RenderX

27. Bettini A, Amore E, Vagnoli L, Maffei F, Martin R. Acceptability and feasibility of a therapeutic board game for children and adolescents with cancer: the Italian version of Shop Talk. Support Care Cancer 2019 Dec;27(12):4479-4485. [doi: 10.1007/s00520-019-04755-8] [Medline: 30911914]
28. Gauthier A, Kato PM, Bul KC, Dunwell I, Walker-Clarke A, Lameras P. Board games for health: a systematic literature review and meta-analysis. Games Health J 2019 Apr;8(2):85-100. [doi: 10.1089/g4h.2018.0017] [Medline: 30256159]
29. Noda S, Shirotsuki K, Nakao M. The effectiveness of intervention with board games: a systematic review. Biopsychosoc Med 2019;13(1):22 [FREE Full text] [doi: 10.1186/s13030-019-0164-1] [Medline: 31641371]
30. Rotstein MS, Zimmerman-Brenner S, Davidovitch S, Ben-Haim Y, Koryto Y, Sion R, et al. Gamified closed-loop intervention enhances tic suppression in children: a randomized trial. Mov Disord 2024 Aug;39(8):1310-1322. [doi: 10.1002/mds.29875] [Medline: 38881244]
31. Raz G, Davidovitch S, Halevi M, Zuckerman M, Ben-Haim Y, Koryto Y, et al. Impact of movie and video game elements on tic manifestation in children. Eur J Neurol 2024 Feb;31(2):e16120. [doi: 10.1111/ene.16120] [Medline: 37946628]
32. Evans J, Seri S, Cavanna AE. The effects of Gilles de la Tourette syndrome and other chronic tic disorders on quality of life across the lifespan: a systematic review. Eur Child Adolesc Psychiatry 2016 Sep;25(9):939-948 [FREE Full text] [doi: 10.1007/s00787-016-0823-8] [Medline: 26880181]
33. Atkinson-Clement C, Duflot M, Lastennet E, Patsalides L, Wasserman E, Sartoris T, et al. How does Tourette syndrome impact adolescents' daily living? A text mining study. Eur Child Adolesc Psychiatry 2023 Dec;32(12):2623-2635 [FREE Full text] [doi: 10.1007/s00787-022-02116-1] [Medline: 36460852]
34. Chen CW, Wang HS, Chang HJ, Hsueh CW. Effectiveness of a modified comprehensive behavioral intervention for tics for children and adolescents with tourette's syndrome: a randomized controlled trial. J Adv Nurs 2020 Mar;76(3):903-915. [doi: 10.1111/jan.14279] [Medline: 31782167]
35. Leckman JF, Riddle MA, Hardin MT, Ort SI, Swartz KL, Stevenson J, et al. The Yale Global Tic Severity Scale: initial testing of a clinician-rated scale of tic severity. J Am Acad Child Adolesc Psychiatry 1989 Jul;28(4):566-573 [FREE Full text] [doi: 10.1097/00004583-198907000-00015] [Medline: 2768151]
36. Storch EA, Murphy TK, Geffken GR, Sajid M, Allen P, Roberti JW, et al. Reliability and validity of the Yale Global Tic Severity Scale. Psychol Assess 2005 Dec;17(4):486-491. [doi: 10.1037/1040-3590.17.4.486] [Medline: 16393016]
37. Mental health of adolescents. World Health Organization. URL: https://www.who.int/zh/news-room/fact-sheets/detail/adolescent-mental-health [accessed 2024-10-10]
38. Chen LC. The development of positive mental health scale for high school students. J Educ Res Dev 2014;10(4):1-22. [doi: 10.3966/181665042014121004001]
39. Lee MY, Wang HS, Chen CJ. Development and validation of the Social Adjustment Scale for adolescents with Tourette syndrome in Taiwan. J Pediatr Nurs 2020;51:e13-e20 [FREE Full text] [doi: 10.1016/j.pedn.2019.05.023] [Medline: 31196591]
40. Beck JS, Beck AT, Jolly JB, Steer RA. Beck Youth Inventories - Second Edition. San Antonio, TX: Psychological Corporation; 2005.
41. Cho SL, Hung LY, Su CL, Chen HC. A research of the Chinese version Beck Youth Inventories. Psychol Test 2009;56(4):639-669. [doi: 10.7108/pt.200912.0639]
42. Rachamim L, Zimmerman-Brenner S, Rachamim O, Mualem H, Zingboim N, Rotstein M. Internet-based guided self-help comprehensive behavioral intervention for tics (ICBIT) for youth with tic disorders: a feasibility and effectiveness study with 6 month-follow-up. Eur Child Adolesc Psychiatry 2022 Feb;31(2):275-287 [FREE Full text] [doi: 10.1007/s00787-020-01686-2] [Medline: 33231786]
43. Maclachlan M, Chimombo M, Mpemba N. AIDS education for youth through active learning: a school-based approach from Malawi. Int J Educ Dev 1997 Jan;17(1):41-50. [doi: 10.1016/s0738-0593(96)00013-2]
44. Wiener L, Battles H, Mamalian C, Zadeh S. ShopTalk: a pilot study of the feasibility and utility of a therapeutic board game for youth living with cancer. Support Care Cancer 2011 Jul;19(7):1049-1054 [FREE Full text] [doi: 10.1007/s00520-011-1130-z] [Medline: 21451945]
45. Spirito A, DeLawyer DD, Stark LJ. Peer relations and social adjustment of chronically ill children and adolescents. Clin Psychol Rev 1991 Jan;11(5):539-564. [doi: 10.1016/0272-7358(91)90003-d]

Abbreviations

BYI-II: Beck Youth Inventory II
CAU: care as usual
CONSORT: Consolidated Standards of Reporting Trials
PMHS: Positive Mental Health Scale
SASATS: Social Adjustment Scale for Adolescents with Tourette Syndrome
TS: Tourette syndrome
YGTSS: Yale Global Tic Severity Scale

XSL•FO
RenderX

Edited by A Coristine; submitted 18.04.25; peer-reviewed by A Gazis, CJ Chen, LC Rodríguez Timaná, AHP Heijerman; comments to author 07.08.25; revised version received 11.08.25; accepted 29.08.25; published 09.09.25.

Please cite as:
Lee MY, Wang HS, Lien CY, Chen ZH
Effects of a Board Game on Tic Management and Psychosocial Functioning in Adolescents With Tourette Syndrome: Randomized Controlled Trial
JMIR Serious Games 2025;13:e76208
URL: https://games.jmir.org/2025/1/e76208
doi:10.2196/76208
PMID:

©Mei Yin Lee, Huei Shyong Wang, Chung Yueh Lien, Zhi Hong Chen. Originally published in JMIR Serious Games (https://games.jmir.org), 09.09.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Virtual Reality–Based Assessment of Attention-Deficit/Hyperactivity Disorder and Comorbid Symptoms in Children: Framework Development and Standardization Study

Harim Jeong[1], PhD; Minjoo Kang[1], MEd; Kennet Sorenson[2], BS; Jacob Moore[1], MD; Robert James Blair[3], PhD; Ellen Leibenluft[4], MD; Jeffrey H Newcorn[5], MD; Beth Krone[5], PhD; Singi Jeong[6], MS; Donghee Kim[7], MS; Soonjo Hwang[1], MD

[1]Department of Psychiatry, University of Nebraska Medical Center, Omaha, NE, United States

[2]University of Nebraska Medical Center, Omaha, NE, United States

[3]Mental Health Services in the Capital Region of Denmark, Brondby, Denmark

[4]National Institute of Mental Health, Bethesada, MD, United States

[5]Icahn School of Medicine at Mount Sinai, New York, NY, United States

[6]Department of Electrical and Computer Engineering, Sungkyunkwan University, Seoul, Republic of Korea

[7]Department of Computer Science and Engineering, Sungkyunkwan University, Seoul, Republic of Korea

Corresponding Author:
Soonjo Hwang, MD
Department of Psychiatry
University of Nebraska Medical Center
510 S 42nd St.
Omaha, NE, 68131
United States
Phone: 1 4055526002
Email: soonjo.hwang@unmc.edu

Abstract

Background: As virtual reality (VR) technology becomes increasingly prevalent, its potential for collecting objective behavioral data in psychiatric settings has been widely recognized. However, the lack of standardized methodologies limits reproducibility and data integration across studies, particularly in assessing attention-deficit/hyperactivity disorder (ADHD) and associated behaviors, such as irritability and aggression.

Objective: This study examines the use of VR-based movement data to operationalize core ADHD symptoms (hyperactivity and inattention) and comorbid disruptive behaviors (irritability and aggression), aiming to identify reproducible and clinically actionable metrics and evaluate their explanatory power for each symptom domain to assess the overall use of these variables.

Methods: A total of 45 children (mean age 9.06, SD 2.11 years; n=14/45, 31% female) participated in the study and were divided into 2 groups: 28 (62%) diagnosed with ADHD and 17 (38%) controls. Seven VR-derived movement variables were analyzed: average speed, acceleration, total distance, area occupied, distance between the hands and head, frequency of movement, and time spent still. Correlation and stepwise regression analyses identified which variables best predicted ADHD symptoms and comorbid behaviors.

Results: Among the 7 VR-derived variables, average speed (mean r=0.460, SD 0.097) and total distance (mean r=0.442, SD 0.116) showed the broadest associations, each correlating with 8 measures. In contrast, frequency of movement was related only to hyperactivity (r=0.416; P=.004), suggesting strong but narrow predictive value. Stepwise regression identified total distance as the sole and strongest predictor of hyperactivity (R^2=0.411) and, except for participant-reported irritability, yielded significant models for all other measures (mean R^2=0.282, SD 0.064; all P<.05).

Conclusions: This study provides empirical evidence on VR-derived movement variables that can inform the development of standardized methodologies for ADHD and comorbid behavior assessment. The identified metrics and their predictive patterns offer a basis for integrating VR-based measures into future research and clinical applications.

(JMIR Serious Games 2025;13:e69146) doi:10.2196/69146

XSL•FO
RenderX

KEYWORDS

virtual reality; attention-deficit/hyperactivity disorder; ADHD; behavioral assessment; digital health

Introduction

Background

The application of virtual reality (VR) technology in psychiatry holds significant promise for enhancing diagnostic accuracy and complementing traditional assessment methods, particularly in disorders such as attention-deficit/hyperactivity disorder (ADHD) and its comorbid symptoms, including irritability and aggression [1-4]. Traditional psychiatric assessments often rely on subjective self-reports and parent reports, which can be influenced by recall and other biases and can be limited to clinical settings that may not reflect a patient's behavior in real-world environments [5,6]. These challenges are particularly pronounced in youth populations, where self-reports may be inconsistent, parent reports may not fully capture key psychopathologies, and discrepancies between multiple informants are common [7,8]. VR addresses these limitations by enabling controlled, realistic simulations in which behavior can be objectively observed and recorded, offering a detailed view of symptom manifestation in various settings that are more representative of daily life [9].

Specifically, VR's potential is particularly relevant for psychiatric disorders in youth with ADHD and its behavioral symptoms, as it allows for the collection of precise behavioral data—including attention span, impulsivity, and hyperactivity—which is critical for understanding the symptom manifestations in naturalistic settings [10]. Given that these symptoms often vary with context and observer attributes, VR provides a unique advantage by enabling data collection in controlled, dynamic environments that mimic real-life scenarios, thereby reducing the need to rely solely on subjective reporting by patients or caregivers [11]. This positions VR as a promising complementary assessment tool to traditional assessments.

Despite the promise VR holds in ADHD, practical implementation in a clinical setting faces several challenges. Most existing studies have small sample sizes and lack standardized frameworks for defining how specific types of VR-collected data relate to ADHD and comorbid disruptive behavior symptoms. This lack of standardization creates challenges for developing generalized VR assessment tools that can be reliably replicated and validated across diverse populations [12,13]. This gap highlights the need for systematic research that clarifies the roles of individual VR-derived data elements in the assessment of ADHD and comorbid disruptive behavior symptoms, ultimately facilitating VR's integration into clinical practice as an objective, data-driven tool.

Objectives

To address these challenges, this study seeks to define and evaluate specific behavioral data collected through VR that are essential for assessing ADHD and comorbid disruptive behavior symptoms. By establishing a clear mapping between VR data types (eg, attention-related behaviors and impulsivity indicators) and ADHD or disruptive behavior symptom domains, this study aims to create a replicable framework that other researchers and clinicians can use. Our goal is to provide a standardized framework that can facilitate consistent data collection across studies and enable the accumulation of larger sample data over time to build the evidence base required to eventually apply VR-based tools in clinical settings for ADHD and comorbid disruptive behavior symptom assessment.

Therefore, this study is guided by the following research questions (RQs):

1. Which specific types of VR-derived behavioral data are most relevant in assessing ADHD and comorbid disruptive behavior symptoms?
2. How can these VR-derived data elements be combined into a composite measure to enhance the comprehensive detection of ADHD and comorbid disruptive behavior symptoms, thereby contributing to the development of a standardized assessment tool?

Through this research, we aim to lay a foundational framework that can be built upon in future studies, thereby facilitating the clinical adoption of VR-based ADHD and comorbid disruptive behavior symptom assessment tools. By providing a clear and replicable structure for VR data collection and interpretation, this study seeks to support the accumulation of evidence necessary for integrating VRs into clinical settings and disorders, ultimately contributing to more accessible, objective, and accurate assessments in clinical settings.

Methods

Experimental Setting

The study recruited 45 children (mean age 9.06, SD 2.11 years), including 14 (31%) girls. Participants were divided into 2 groups: 28 (62%) children diagnosed with ADHD and 17 (38%) children with no history of ADHD or related psychiatric diagnoses. This sample composition allowed us to capture diverse behavioral patterns and identify ADHD and comorbid disruptive behavior symptom–specific movement characteristics within the VR environment. Rather than dividing the ADHD and control groups in our analyses, we focused on the continuum of movement data across both groups, aligning with the research domain criteria framework, which emphasizes transdiagnostic dimensions of behavior and neurobiology rather than categorical diagnoses. This approach allowed us to examine how specific movement patterns may reflect core symptom dimensions across individuals, regardless of diagnostic labels.

A 5-minute VR session, developed using Meta's Oculus Quest 2 system and Unity (Unity Technologies), was used as the experimental setting. Within this session, children engaged in unstructured interactions, as no specific instructions or goals were provided. Instead, a virtual avatar initiated casual conversation prompts (eg, asking about the participant's day or mood), allowing for spontaneous verbal and nonverbal response. The avatar also responded to physical actions, such as getting closer to the avatar (proximity) or being struck by the

XSL•FO
RenderX

participant, allowing a broader range of naturalistic interaction. The virtual room setting was designed to mimic a real-life environment with familiar objects (eg, desk, chair, and bookshelf) spatially arranged. This setup aimed to evoke naturalistic behaviors, such as exploring the space, interacting with objects in the room, and engaging in casual conversations with the virtual avatar about daily routines and moods. Figure 1 provides an illustration of the VR environment.

Movement data were recorded as 3D coordinates (x, y, and z) at 0.5-second intervals using the head-mounted display and hand controllers. These data reflected real-time behavior, including locomotion, gesture, and object-directed movement, and were continuously monitored throughout the session. By structuring the VR interaction into a natural, task-free conversational format, we aimed to gather unbiased behavioral data reflecting the children's spontaneous movements and social interaction styles to specifically capture movement patterns associated with core ADHD symptoms (eg, hyperactivity and inattention) and comorbid disruptive behavior symptoms (eg, irritability and aggression).

Figure 1. Virtual environment used to assess attention-deficit/hyperactivity disorder and comorbid behaviors in children through naturalistic interaction during a virtual reality session. (A) Participant's first-person view; (B) overview of the virtual room layout.



(A) (B)

Data Processing

The recorded 3D coordinates of participants' head and hand movements were subsequently processed to derive behavioral indicators for analysis. These variables were then categorized into 3 main groups.

The first group of data includes velocity-related features, including the average speed of head and hand movements and acceleration, which is defined as the rate of change of speed over time for the head and hands. These variables were chosen based on evidence suggesting that children with ADHD tend to exhibit faster movements and more frequent acceleration than those without the condition [14,15].

The second group of data was spatial features, which include the total distance moved by the head and hands, the area occupied by a participant during movement, and the distance between the hands and head. These variables were selected based on research indicating that body movement can vary significantly across different psychiatric disorders [16]. Specifically, the distance between the hands and head is a movement pattern associated with certain mental health conditions, such as ADHD [17].

The final group of data was frequency and duration features, which identify how often a participant moves and measure the time the participant spends not moving. These variables were selected because frequent movement or prolonged stillness can indicate manifestations of hyperactivity symptoms in children with ADHD [18].

In total, 7 variables (average speed, acceleration, total distance, area occupied, distance between hands and head, frequency of movement, and time spent still) representing these 3 distinct characteristics were derived from the raw data for analysis. The selected variables and their calculation methods are detailed in Table 1.

XSL•FO
RenderX

Table 1. Definitions and equations for the 7 virtual reality (VR)–derived movement variables used to assess attention-deficit/hyperactivity disorder and comorbid disruptive behaviors in children during a VR session.

Category and variables	Equations	References
Velocity		[14,19-21]
Average speed		
Acceleration		
Spatial		[16,22-24]
Total distance		
Area occupied		
Distance between the hands and head		
Frequency and duration		[25-27]
Frequency of movement		
Time spent still		

[a]The "threshold$_1$" is the minimum distance for detecting movement.

[b]The "threshold$_2$" represents the maximum distance within which a participant is considered stationary.

[c]The "distance" refers to the Euclidean distance calculated between consecutive positions of the head or hands.

External Psychopathology Assessment

First, the ADHD symptom profile was evaluated with the ADHD Rating Scale (ADHD-RS) [28]. The ADHD-RS is an 18-item questionnaire scored on a 0 to 3 scale, with 9 questions related to hyperactivity and 9 related to inattention, completed by parents [29]. The ADHD-RS was used to assess the core symptoms of ADHD—hyperactivity and inattention—by examining their relationship with movement features captured through VR.

To assess participants' disruptive behavior symptoms, we collected relevant data using validated assessment tools. In addition to ADHD core symptoms, the disruptive behavior symptoms evaluated in this study included irritability and reactive and proactive aggression. Irritability was included as it is one of the most common features related to various psychopathologies in youth, including ADHD [30]. Thus, we hypothesized that irritability could be predicted through movement analysis in a virtual environment. Data on irritability were collected from both parents and participants using the Affective Reactivity Index (ARI), a 7-item questionnaire scored on a 0 to 2 scale [31].

Next, we assessed reactive and proactive aggression using the Reactive-Proactive Aggression Questionnaire (RPQ) [32]. Aggressive cognitions are more likely to become aggressive actions in youth with ADHD and impulse control. Thus, aggressive behavior can be a relevant target for prediction through movement data in virtual settings [2]. To leverage VR's capability to capture and analyze these behaviors in a controlled environment, we included the RPQ in our study. The RPQ consists of 23 items, each scored on a 0 to 2 scale, with data collected from both parents and participants. The specific items for each assessment tool used in this study are provided in Multimedia Appendix 1.

Analysis of Psychopathology Descriptive Statistics

To explore the relationship between VR-derived movement data and ADHD or disruptive behavior symptoms, we first calculated descriptive statistics for all psychopathology measures. This step provided an overview of the data distribution and informed the subsequent analyses, and these scores were used as dependent variables in the analysis.

Analysis to Identify Key VR-Derived Predictors of ADHD and Comorbid Symptoms

We then conducted Pearson correlation analyses to examine associations between VR-derived movement variables and assessment scores. This step helped determine which specific behavioral indicators captured in the VR environment were most relevant to ADHD and comorbid disruptive behavior symptoms.

Analysis to Identify Core ADHD Symptoms Captured by VR-Derived Data

We applied a backward stepwise multiple linear regression (entry criterion: $P<.05$; removal criterion: $P>.10$) to identify the most predictive combination of VR-derived variables for each symptom domain. Multicollinearity among predictors was assessed after model selection using the variance inflation factor (VIF), with VIF<5 considered acceptable.

Ethical Considerations

Participants were recruited from a university hospital and the local community in the central United States. Written informed consent was obtained from both the children and their legal guardians before participation. The study protocol was approved by the institutional review board at the University of Nebraska

XSL•FO
RenderX

Medical Center (580-23). All collected data were deidentified to ensure confidentiality and privacy. Participants received no compensation for their involvement in the study. No identifiable images of participants were collected or included in the manuscript.

Results

Descriptive Statistics of Psychopathology Measures

The ADHD-RS scores had a mean score of 10.44 (SD 7.24) for hyperactivity, 11.8 (SD 7.33) for inattention, and 22.24 (SD 13.23) for the total ADHD-RS score. The ARI scores, collected from both parents and participants, showed mean scores of 4.31 (SD 3.42) for parent-reported data and 4.35 (SD 3.55) for participant-reported data. For reactive and proactive aggression, as measured by the RPQ, the parent-reported reactive aggression had a mean of 9.15 (SD 5.29), proactive aggression had a mean of 2.37 (SD 3.29), and the total RPQ score had a mean of 11.84 (SD 8.44). Participant-reported RPQ scores had mean scores of 7.6 (SD 4.38) for reactive aggression and 2.11 (SD 2.42) for proactive aggression and a total mean score of 10.04 (SD 6.39).

These detailed descriptive statistics offer foundational insights into the distribution of each dependent variable and are further analyzed in the subsequent sections. Full descriptive statistics are presented in Table 2.

Table 2. Descriptive statistics of attention-deficit/hyperactivity disorder (ADHD) symptoms and comorbid disruptive behavior scores (ADHD Rating Scale [ADHD-RS], Affective Reactivity Index [ARI], and Reactive-Proactive Aggression Questionnaire [RPQ]) reported by parents and participants.

Assessment tool and score type	Scores, mean (SD)
ADHD-RS	
Hyperactivity	10.44 (7.24)
Inattention	11.8 (7.33)
Total	22.24 (13.23)
ARI total (parent)	4.31 (3.42)
ARI total (participant)	4.35 (3.55)
RPQ (parent)	
Reactive	9.15 (5.29)
Proactive	2.37 (3.29)
Total	11.84 (8.44)
RPQ (participant)	
Reactive	7.6 (4.38)
Proactive	2.11 (2.42)
Total	10.04 (6.39)

Identifying Key VR-Derived Predictors of Core ADHD Symptoms and Comorbid Disruptive Behavior Symptoms

Several VR-derived movement variables showed significant correlations with ADHD and comorbid symptom scores, indicating specific behavioral patterns associated with these conditions. Among the velocity-related variables, average speed showed strong correlations with 8 assessment scores, including the ADHD-RS hyperactivity (r=0.639; P<.001) and inattention (r=0.500; P<.001) subscales; the parent-reported RPQ total score (r=0.504; P<.001), proactive aggression subscore (r=0.477; P<.001), and reactive aggression subscore (r=0.464; P=.001); and youth-reported RPQ total score (r=0.394; P=.007), proactive subscore (r=0.367; P=.01), and reactive subscore (r=0.332; P=.03).

Regarding the spatial variables, the area occupied was significantly correlated with 5 assessment scores, including the participant-reported RPQ proactive aggression subscore (r=0.362; P=.01) and the ADHD-RS hyperactivity subscale (r=0.357; P=.02), the youth-reported RPQ total score (r=0.327; P=.03) and the parent-reported RPQ reactive aggression subscore (r=0.314; P=.03), and the ADHD-RS inattention subscore (r=0.295; P=.049). Furthermore, the total distance variable was significantly correlated with 8 assessment scores, including the ADHD-RS hyperactivity (r=0.641; P<.001) and inattention (r=0.494; P<.001) subscales; the parent-reported RPQ total scores (r=0.498; P<.001), proactive aggression subscore (r=0.503; P<.001), and reactive aggression subscore (r=0.439; P=.002); the youth-reported RPQ proactive (r=0.323; P=.03) and total (r=0.302; P=.04); and the parent-reported ARI score (r=0.337; P=.02).

Finally, in the category of frequency and duration variables, the frequency of movement showed a significant correlation with the ADHD-RS hyperactivity subscale (r=0.416; P=.004). The time spent still variable was negatively correlated with 6 assessment scores, including the ADHD-RS hyperactivity (r=−0.448; P=.002) and inattention (r=−0.363; P=.01) subscales, the parent-reported RPQ reactive (r=−0.349; P=.02) and total (r=−0.349; P=.02) scores, and the participant-reported RPQ reactive (r=−0.317; P=.03) and total (r=−0.333; P=.03) scores. These correlations are summarized in Table 3.

XSL•FO
RenderX

Table 3. Significant correlations between virtual reality (VR)–derived movement variables and attention-deficit/hyperactivity disorder (ADHD) and comorbid symptom scores (ADHD Rating Scale [ADHD-RS], Affective Reactivity Index [ARI], and Reactive-Proactive Aggression Questionnaire [RPQ]) among children participating in a VR-based behavioral assessment study.

VR data element and assessment score	Pearson correlation coefficient (*r*)	*P* value
Average speed		
ADHD-RS hyperactivity	0.639	<.001
ADHD-RS inattention	0.500	<.001
RPQ parent total	0.504	<.001
RPQ parent proactive	0.477	<.001
RPQ parent reactive	0.464	.001
RPQ participant total	0.394	.007
RPQ participant proactive	0.367	.01
RPQ participant reactive	0.332	.03
Area occupied		
ADHD-RS hyperactivity	0.357	.02
ADHD-RS inattention	0.295	.05
RPQ participant total	0.327	.03
RPQ participant proactive	0.362	.01
RPQ parent reactive	0.314	.04
Total distance		
ADHD-RS hyperactivity	0.641	<.001
ADHD-RS inattention	0.494	<.001
RPQ parent total	0.498	<.001
RPQ parent proactive	0.503	<.001
RPQ parent reactive	0.439	.002
RPQ participant proactive	0.323	.03
RPQ participant total	0.302	.04
ARI parent	0.337	.02
Frequency of movement		
ADHD-RS hyperactivity	0.416	.004
Time spent still		
ADHD-RS hyperactivity	−0.448	.002
ADHD-RS inattention	−0.363	.01
RPQ parent reactive	−0.349	.02
RPQ parent total	−0.349	.02
RPQ participant reactive	−0.317	.03
RPQ participant total	−0.333	.03

Identifying Core ADHD Symptoms and Comorbid Disruptive Behavior Symptoms Captured by VR-Derived Data

Stepwise regression identified specific combinations of VR-derived movement variables that best predicted ADHD and comorbid symptom scores. The most influential predictors varied by symptom profile, highlighting the ability of VR-based behavioral data to differentiate psychological characteristics.

As all candidate variables represent aspects of physical movement, variables such as total distance (VIF=13.05) and average speed (VIF=13.17) exhibited high VIF values during preliminary examination before final model selection. However, due to the nature of the stepwise regression procedure, only one of these correlated predictors was selected in each final model. As assessed after model selection, all variables retained in the final models showed acceptable VIF levels (<5), indicating that multicollinearity was not a concern for interpretation.

XSL•FO
RenderX

For the parent-reported ARI score related to irritability, the final model identified 2 predictors: total distance (β=.1327; P<.001) and frequency of movement (β=−.0807; P=.003), with an R^2 of 0.287. The positive association with total distance indicates that greater movement distance is linked to higher irritability, whereas the negative association with movement frequency suggests that more frequent movements are linked to lower irritability.

For the parent-reported RPQ total score, 2 predictors remained in the final model: average speed (β=25.6122; P<.001) and frequency of movement (β=−.1270; P=.03), with an R^2 of 0.334. The positive association with average speed indicates that faster movements are related to higher total aggression scores, while the negative association with frequency of movement suggests that frequent movements are associated with lower aggression.

In the parent-reported RPQ proactive aggression score, only total distance (β=.0793; P<.001) was retained, with an R^2 of 0.253. This suggests that proactive aggression is primarily captured by the overall distance covered in the VR environment. For the parent-reported RPQ reactive aggression score, 2 variables were significant: average speed (β=15.8999; P<.001) and frequency of movement (β=−.0886; P=.02), with an R^2 of 0.313. The results show that faster movements are associated with higher reactive aggression, while frequent movements correlate negatively with reactive aggression.

In the youth-reported RPQ scores, the total aggression score was best explained by average speed (β=17.8567; P<.001) and frequency of movement (β=−.1119; P=.02), with an R^2 of 0.263. For proactive aggression, the model included average speed (β=6.3336; P=.001) and frequency of movement (β=−.0399; P=.03), with an R^2 of 0.231. The reactive aggression score was also predicted by average speed (β=10.5277; P=.003) and frequency of movement (β=−.0678; P=.04), with an R^2 of 0.194. These results consistently show that average speed and frequency of movement predicted youth-reported aggression scores.

For the ADHD-RS scores, the hyperactivity subscale was significantly explained by total distance (β=.2219; P<.001), with an R^2 of 0.411, indicating that hyperactivity is explained by the overall distance covered in the VR tasks. The inattention subscale was best predicted by average speed (β=14.0459; P<.001), with an R^2 of 0.250, showing that faster movements are linked to higher inattention scores. The detailed results supporting these observations are provided in Table 4.

XSL•FO
RenderX

Table 4. Stepwise regression models identifying virtual reality (VR)–derived movement predictors of attention-deficit/hyperactivity disorder (ADHD) symptoms and comorbid symptom scores (ADHD Rating Scale [ADHD-RS], Affective Reactivity Index [ARI], and Reactive-Proactive Aggression Questionnaire [RPQ]) in children participating in a VR-based behavioral assessment study.

Assessment score and predictor	Coefficient (β)	*P* value	R^2
ADHD-RS hyperactivity			
Total distance	0.2219	<.001	0.411
ADHD-RS inattention			
Average speed	14.0459	<.001	0.250
ARI (parent)			0.287
Total distance	0.1327	<.001	
Frequency of movement	−0.0807	.003	
ARI (participant)	—[a]	—	—
RPQ total (parent)			0.334
Average speed	25.6122	<.001	
Frequency of movement	−0.1270	.03	
RPQ proactive (parent)			
Total distance	0.0793	<.001	0.253
RPQ reactive (parent)			0.313
Average speed	15.8999	<.001	
Frequency of movement	−0.0886	.02	
RPQ total (participant)			0.263
Average speed	17.8567	<.001	
Frequency of movement	−0.1119	.02	
RPQ proactive (participant)			0.231
Average speed	6.3336	.001	
Frequency of movement	−0.0399	.03	
RPQ reactive (participant)			0.194
Average speed	10.5277	.003	
Frequency of movement	−0.0678	.04	

[a]indicates that no variables were statistically significant for this model.

Discussion

Principal Findings

This study explored the potential of VR-collected movement data to predict ADHD and comorbid disruptive behavior symptoms, aiming to bridge the gap between theoretical possibilities and practical clinical applications. Specifically, we sought to identify VR-derived movement data elements that effectively capture symptoms of ADHD or comorbid disruptive behavior and to examine how these elements can contribute to reliable predictive models for behavioral symptoms.

Our findings suggest that VR-derived data can play a significant role in capturing symptoms of ADHD or disruptive behavior, offering insights that may improve the accuracy, objectivity, and accessibility of assessment tools [33]. By establishing foundational evidence on the associations between VR data and behavioral symptoms, this study may contribute to the broader goal of incorporating VR into psychiatric practice to enhance care precision and accessibility for patients with externalizing psychopathologies [34].

For RQ 1, the results indicate that certain VR-derived movement variables, particularly average speed and total distance, are more closely related to ADHD and disruptive behavior symptoms than other VR-derived variables. Average speed demonstrated significant correlations with a variety of symptoms, suggesting it could function as a versatile marker of ADHD and comorbid disruptive behavior symptoms, including inattention and aggression. Meanwhile, total distance was particularly relevant for predicting hyperactivity symptoms, as shown by its strong correlation with the ADHD-RS hyperactivity subscale.

In addressing RQ 2, we explored how VR-derived behavioral features could be combined to predict specific symptom domains using stepwise regression. This approach identified the most explanatory feature combinations for each behavioral construct.

XSL•FO
RenderX

For ADHD-related symptoms, hyperactivity and inattention were best predicted by total distance and average speed, respectively, indicating relatively direct behavioral correlates for each symptom. In contrast, symptoms related to irritability and aggression (eg, ARI and RPQ scores) required the inclusion of additional features, such as frequency of movement, to improve model fit. These findings imply that comorbid disruptive behaviors may involve more complex behavioral dynamics than core ADHD symptoms [35], possibly reflecting more nuanced or multifactorial movement patterns within the VR environment.

Comparison With Prior Work

Although we collected both parent and child symptom ratings, our analyses showed that movement patterns recorded in VR were more strongly associated with parent reports, as reflected in consistently higher R^2 values across models. This may reflect differences in how each informant perceives and interprets the behavior of parents, as externalizing psychopathologies tend to be outwardly observable better by parents (eg, hyperactivity or aggression), as children may be more attuned to internal states such as anxiety or mood [36]. As our study focused on externalizing problems, such as ADHD, irritability, and aggression, it makes sense that the behavioral signals captured in VR aligned more closely with the parent perspectives. The stronger associations with parent-reported symptoms may indicate that the VR-collected movement patterns correspond more with outwardly visible behaviors.

Furthermore, using movement-based data collected in a virtual environment may help reduce biases commonly associated with caregiver-reported assessments [37]. It offers clinicians an opportunity to observe behavioral patterns more directly and objectively. As more research accumulates using standardized, quantitatively derived behavioral features, there is potential to develop generalized benchmarks that support more efficient and timely assessment of ADHD and comorbid symptoms. This could ultimately enhance clinical workflows by complementing traditional tools with scalable, technology-assisted methods.

These findings align with previous observations that ADHD-related behaviors often manifest in increased physical activity [24], which VR technology can effectively capture in a controlled environment. The patterns observed in this study suggest that physical activity within a VR environment may reflect specific symptom dimensions. For instance, while total distance appears to align with overall physical movement associated with hyperactivity, average speed may capture movement intensity and impulsivity, characteristics often linked to symptoms such as aggression and inattention. However, it is important to note that while these associations were observed in our sample, further research is needed to determine whether these movement patterns consistently predict symptoms across diverse samples and settings.

Limitations and Future Directions

Given the novel nature of VR-derived movement data in this study, these findings provide initial insights but also raise questions that warrant further exploration. The results point to potential indicators of symptom-specific behaviors, but additional studies with larger samples are necessary to confirm our findings and to provide further insight into the complex relations between VR variables and the symptom manifestation of various psychiatric conditions. Moreover, while average speed and total distance emerged as strong predictors for certain behaviors, the unexpected associations observed for frequency of movement and irritability highlight the complexity of interpreting VR data in relation to psychological symptoms [38].

This study is among the first to systematically examine the relationship between VR-derived movement data and ADHD or disruptive behavior symptoms in a controlled, virtual environment. These findings suggest that VR may offer a unique, real-time perspective on behavioral symptoms, providing more objective and context-rich data than traditional assessment methods.

First, the relatively small sample size and the scope of behaviors assessed may impact the generalizability of these findings. To reduce the possibility of type II error, we analyzed the entire study group without dividing them into ADHD versus healthy control groups, thereby maximizing statistical power. Nonetheless, replication with larger and more diverse samples will be necessary to confirm these results.

Second, the broad and generalized virtual environment used in this study might not fully capture the nuances specific to individual psychopathologies. While this design enabled observation of diverse behaviors, future studies should focus on developing VR scenarios tailored to specific behavioral challenges to improve the precision of VR-based assessments.

Third, the unexpected associations observed between frequency of movement and irritability highlight the complexity of interpreting VR data in relation to psychological symptoms. These findings underscore the exploratory nature of this study and the need for replication in other contexts to clarify these associations.

Future research should also consider integrating multimodal measures, such as physiological signals or neuroimaging data, to enhance interpretability and improve the predictive accuracy of VR-based assessment tools.

Conclusions

This study responds to the growing interest in VR-based ADHD assessment by providing detailed definitions and processing methods for specific VR-derived behavioral variables, which may help improve reproducibility in future work. Our analyses identified average speed and total distance as consistent indicators across multiple symptom domains and suggested possible combinations of variables that may be particularly relevant for certain behaviors. By sharing both the design of the VR assessment and the associated data processing approach, we aim to contribute a practical reference point for other researchers. We hope that such collaborative efforts will support the continued development of VR-based tools and encourage their gradual integration into clinical assessment practices.

XSL•FO
RenderX

Acknowledgments

This research was supported by Hippo T&C. The funding organization was not involved in the design, conduct, data analysis, or reporting of the study.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

HJ and SH contributed to the conceptualization and design of the study. MK, KS, and JM collected the data. HJ conducted the formal analysis and led the data interpretation. EL, JHN, BK, and SH contributed to the interpretation of the results. HJ and SH drafted the manuscript. JB, EL, JHN, and BK contributed to the critical revision of the manuscript. All authors reviewed the manuscript for important intellectual content and approved the final version of the manuscript.

Conflicts of Interest

BK and JHN have been consultants to HIPPO T&C, the sponsor of this study. JHN consults to several other pharmaceutical and device companies working in the ADHD space.

Multimedia Appendix 1
Item list for questionnaires used in the study.
[DOCX File , 26 KB - games_v13i1e69146_app1.docx]

References

1. Abbas JR, O'Connor A, Ganapathy E, Isba R, Payton A, McGrath B, et al. What is virtual reality? A healthcare-focused systematic review of definitions. Health Policy Technol 2023 Jun;12(2):100741. [doi: 10.1016/j.hlpt.2023.100741]
2. Terbeck S, Case C, Turner J, Spencer V, Bacon A, Howard C, et al. Assessing reactive violence using immersive virtual reality. PLoS One 2022 May 6;17(5):e0268191 [FREE Full text] [doi: 10.1371/journal.pone.0268191] [Medline: 35522625]
3. Eyre O, Langley K, Stringaris A, Leibenluft E, Collishaw S, Thapar A. Irritability in ADHD: associations with depression liability. J Affect Disord 2017 Jun;215:281-287 [FREE Full text] [doi: 10.1016/j.jad.2017.03.050] [Medline: 28363151]
4. Connor DF, Chartier KG, Preen EC, Kaplan RF. Impulsive aggression in attention-deficit/hyperactivity disorder: symptom severity, co-morbidity, and attention-deficit/hyperactivity disorder subtype. J Child Adolesc Psychopharmacol 2010 Apr;20(2):119-126 [FREE Full text] [doi: 10.1089/cap.2009.0076] [Medline: 20415607]
5. Aboraya A, Rankin E, France C, El-Missiry A, John C. The reliability of psychiatric diagnosis revisited: the clinician's guide to improve the reliability of psychiatric diagnosis. Psychiatry (Edgmont) 2006 Jan;3(1):41-50 [FREE Full text] [Medline: 21103149]
6. Stein DJ, Shoptaw SJ, Vigo DV, Lund C, Cuijpers P, Bantjes J, et al. Psychiatric diagnosis and treatment in the 21st century: paradigm shifts versus incremental integration. World Psychiatry 2022 Oct;21(3):393-414 [FREE Full text] [doi: 10.1002/wps.20998] [Medline: 36073709]
7. Caqueo-Urízar A, Urzúa A, Villalonga-Olives E, Atencio-Quevedo D, Irarrázaval M, Flores J, et al. Children's mental health: discrepancy between child self-reporting and parental reporting. Behav Sci (Basel) 2022 Oct 19;12(10):401 [FREE Full text] [doi: 10.3390/bs12100401] [Medline: 36285970]
8. Harris SK. The influences of adolescents self-reporting a mental health problem [thesis]. Johns Hopkins University. 1996. URL: https://www.proquest.com/docview/304308217?pq-origsite=gscholar&fromopenview=true&sourcetype=Dissertations%20&%20Theses [accessed 2025-08-29]
9. Kim S, Kim E. The use of virtual reality in psychiatry: a review. Soa Chongsonyon Chongsin Uihak 2020 Jan 01;31(1):26-32 [FREE Full text] [doi: 10.5765/jkacap.190037] [Medline: 32612410]
10. Romero-Ayuso D, Toledano-González A, Rodríguez-Martínez MD, Arroyo-Castillo P, Triviño-Juárez JM, González P, et al. Effectiveness of virtual reality-based interventions for children and adolescents with ADHD: a systematic review and meta-analysis. Children (Basel) 2021 Jan 21;8(2):70 [FREE Full text] [doi: 10.3390/children8020070] [Medline: 33494272]
11. Taubin D, Berger A, Greenwald D, Greenwald H, Burke C, Gongora DS, et al. A systematic review of virtual reality therapies for substance use disorders: impact on secondary treatment outcomes. Am J Addict 2023 Jan;32(1):13-23 [FREE Full text] [doi: 10.1111/ajad.13342] [Medline: 36128667]
12. Maples-Keller JL, Bunnell BE, Kim SJ, Rothbaum BO. The use of virtual reality technology in the treatment of anxiety and other psychiatric disorders. Harv Rev Psychiatry 2017;25(3):103-113 [FREE Full text] [doi: 10.1097/HRP.0000000000000138] [Medline: 28475502]

XSL•FO
RenderX

13. Zeng N, Pope Z, Lee JE, Gao Z. Virtual reality exercise for anxiety and depression: a preliminary review of current research in an emerging field. J Clin Med 2018 Mar 04;7(3):42 [FREE Full text] [doi: 10.3390/jcm7030042] [Medline: 29510528]
14. Dahan A, Reiner M. Evidence for deficient motor planning in ADHD. Sci Rep 2017 Aug 29;7(1):9631 [FREE Full text] [doi: 10.1038/s41598-017-09984-7] [Medline: 28852003]
15. Kam HJ, Lee K, Cho SM, Shin YM, Park RW. High-resolution actigraphic analysis of ADHD: a wide range of movement variability observation in three school courses - a pilot study. Healthc Inform Res 2011 Mar;17(1):29-37 [FREE Full text] [doi: 10.4258/hir.2011.17.1.29] [Medline: 21818455]
16. Johnson KA, Dáibhis A, Tobin CT, Acheson R, Watchorn A, Mulligan A, et al. Right-sided spatial difficulties in ADHD demonstrated in continuous movement control. Neuropsychologia 2010 Apr;48(5):1255-1264. [doi: 10.1016/j.neuropsychologia.2009.12.026] [Medline: 20043932]
17. Wing AM, Flanagan JR, Haggard P. Hand and Brain: The Neurophysiology and Psychology of Hand Movements. Cambridge, MA: Academic Press; 1996.
18. Jung JK, Suh YT. Effect of psychomotor program for material experience on the coordination of children with ADHD. Res J Pharm Technol 2017 Oct 17;10(7):2395-2399. [doi: 10.5958/0974-360X.2017.00424.3]
19. Carte ET, Nigg JT, Hinshaw SP. Neuropsychological functioning, motor speed, and language processing in boys with and without ADHD. J Abnorm Child Psychol 1996 Aug;24:481-498. [doi: 10.1007/bf01441570]
20. Wallbott HG. Movement quality changes in psychopathological disorders. In: Kirkcaldy B, editor. Normalities and Abnormalities in Human Movement. Basel, Switzerland: Karger Publishers; 1989:128-146.
21. McGinnis EW, McGinnis RS, Muzik M, Hruschak J, Lopez-Duran NL, Perkins NC, et al. Movements indicate threat response phases in children at risk for anxiety. IEEE J Biomed Health Inform 2017 Sep;21(5):1460-1465. [doi: 10.1109/jbhi.2016.2603159]
22. Sempere-Tortosa M, Fernández-Carrasco F, Mora-Lizán F, Rizo-Maestre C. Objective analysis of movement in subjects with ADHD. Multidisciplinary control tool for students in the classroom. Int J Environ Res Public Health 2020 Aug 04;17(15):5620 [FREE Full text] [doi: 10.3390/ijerph17155620] [Medline: 32759840]
23. Kroes M, Kessels AG, Kalff AC, Feron FJ, Vissers YL, Jolles J, et al. Quality of movement as predictor of ADHD: results from a prospective population study in 5 - and 6 - year - old children. Dev Med Child Neurol 2007 Feb 13;44(11):753-760. [doi: 10.1111/j.1469-8749.2002.tb00282.x]
24. Sempere-Tortosa M, Fernández-Carrasco F, Navarro-Soria I, Rizo-Maestre C. Movement patterns in students diagnosed with ADHD, objective measurement in a natural learning environment. Int J Environ Res Public Health 2021 Apr 07;18(8):3870 [FREE Full text] [doi: 10.3390/ijerph18083870] [Medline: 33917074]
25. Lorber MF, Del Vecchio T, Slep AM, Scholer SJ. Normative trends in physically aggressive behavior: age-aggression curves from 6 to 24 months. J Pediatr 2019 Mar;206:197-203.e1. [doi: 10.1016/j.jpeds.2018.10.025] [Medline: 30429080]
26. Sampasa-Kanyinga H, Colman I, Goldfield GS, Janssen I, Wang J, Tremblay MS, et al. 24-hour movement behaviors and internalizing and externalizing behaviors among youth. J Adolesc Health 2021 May;68(5):969-977. [doi: 10.1016/j.jadohealth.2020.09.003] [Medline: 33069581]
27. Drupp IK. An examination of movement qualities and aggression in a child with a disruptive behavior disorder: a qualitative, single case study [thesis]. Drexel University. 2008. URL: https://researchdiscovery.drexel.edu/esploro/outputs/graduate/991021888770004721 [accessed 2025-08-25]
28. Climie EA, Mitchell K. Parent-child relationship and behavior problems in children with ADHD. Int J Dev Disabil 2016 Feb 08;63(1):27-35. [doi: 10.1080/20473869.2015.1112498]
29. DuPaul GJ, Power TJ, Anastopoulos AD, Reid R. ADHD Rating Scale—5 for Children and Adolescents: Checklists, Norms, and Clinical Interpretation. New York City, NY: Guilford Publications; 2016.
30. Leibenluft E, Allen LE, Althoff RR, Brotman MA, Burke JD, Carlson GA, et al. Irritability in youths: a critical integrative review. Am J Psychiatry 2024 Apr 01;181(4):275-290. [doi: 10.1176/appi.ajp.20230256] [Medline: 38419494]
31. Stringaris A, Goodman R, Ferdinando S, Razdan V, Muhrer E, Leibenluft E, et al. The Affective Reactivity Index: a concise irritability scale for clinical and research settings. J Child Psychol Psychiatry 2012 Nov;53(11):1109-1117 [FREE Full text] [doi: 10.1111/j.1469-7610.2012.02561.x] [Medline: 22574736]
32. Smeets KC, Oostermeijer S, Lappenschaar M, Cohn M, van der Meer JM, Popma A, et al. Are proactive and reactive aggression meaningful distinctions in adolescents? A variable- and person-based approach. J Abnorm Child Psychol 2017 Jan;45(1):1-14 [FREE Full text] [doi: 10.1007/s10802-016-0149-5] [Medline: 27113216]
33. Bell IH, Nicholas J, Alvarez-Jimenez M, Thompson A, Valmaggia L. Virtual reality as a clinical tool in mental health research and practice . Dialogues Clin Neurosci 2022 Apr 01;22(2):169-177. [doi: 10.31887/dcns.2020.22.2/lvalmaggia]
34. van Bennekom MJ, de Koning PP, Denys D. Virtual reality objectifies the diagnosis of psychiatric disorders: a literature review. Front Psychiatry 2017 Sep 05;8:163 [FREE Full text] [doi: 10.3389/fpsyt.2017.00163] [Medline: 28928677]
35. Hoeve M, McReynolds LS, Wasserman GA. The influence of adolescent psychiatric disorder on young adult recidivism. Crim Just Behav 2013 Jun 10;40(12):1368-1382. [doi: 10.1177/0093854813488106]
36. Hope TL, Adams C, Reynolds L, Powers D, Perez RA, Kelley ML. Parent vs. self-report: contributions toward diagnosis of adolescent psychopathology. J Psychopathol Behav Assess 1999 Dec;21:349-363. [doi: 10.1023/a:1022124900328]

XSL•FO
RenderX

37. Kroes G, Veerman JW, De Bruyn EE. Bias in parental reports? Eur J Psychol Assess 2003 Sep;19(3):195-203. [doi: 10.1027//1015-5759.19.3.195]
38. Toohey MJ, DiGiuseppe R. Defining and measuring irritability: construct clarification and differentiation. Clin Psychol Rev 2017 Apr;53:93-108. [doi: 10.1016/j.cpr.2017.01.009] [Medline: 28284170]

Abbreviations

ADHD: attention-deficit/hyperactivity disorder
ADHD-RS: Attention-Deficit/Hyperactivity Disorder-Rating Scale
ARI: Affective Reactivity Index
RPQ: Reactive-Proactive Aggression Questionnaire
RQ: research question
VIF: variance inflation factor
VR: virtual reality

Edited by A Coristine; submitted 31.12.24; peer-reviewed by R Singh, E Maslowski, R Garotti; comments to author 16.06.25; revised version received 20.07.25; accepted 19.08.25; published 07.10.25.

Please cite as:
Jeong H, Kang M, Sorenson K, Moore J, Blair RJ, Leibenluft E, Newcorn JH, Krone B, Jeong S, Kim D, Hwang S
Virtual Reality–Based Assessment of Attention-Deficit/Hyperactivity Disorder and Comorbid Symptoms in Children: Framework Development and Standardization Study
JMIR Serious Games 2025;13:e69146
URL: https://games.jmir.org/2025/1/e69146
doi:10.2196/69146
PMID:

©Harim Jeong, Minjoo Kang, Kennet Sorenson, Jacob Moore, Robert James Blair, Ellen Leibenluft, Jeffrey H Newcorn, Beth Krone, Singi Jeong, Donghee Kim, Soonjo Hwang. Originally published in JMIR Serious Games (https://games.jmir.org), 07.10.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Usability Evaluation of a Virtual Reality Multisensory Sham-Feeding Device for Patients Undergoing Fasting Periods for Colorectal Cancer Surgery: Mixed Methods Study

Xingzhu Yuan[1*], PhD; Peiling Ye[1,2*], MSN; Xinyue Liu[1], MSN; Yuehan Hu[3], BSN; Qin Xu[1], BSN; Wenyi Zhao[1], MSN; Ka Li[4], PhD

[1]West China School of Nursing/Department of Medicine and Engineering Interdisciplinary Research Laboratory of Nursing & Materials, West China Hospital, Sichuan University, Chengdu, China

[2]Division of Gastrointestinal Surgery, Department of General Surgery, West China Hospital, Sichuan University, Chengdu, China

[3]Department of Obstetric Nursing, West China Second University Hospital, Sichuan University, Chengdu, China

[4]Department of Medicine and Engineering Interdisciplinary Research Laboratory of Nursing & Materials, West China Hospital, Sichuan University, Chengdu, China

[*]these authors contributed equally

Corresponding Author:
Ka Li, PhD
Department of Medicine and Engineering Interdisciplinary Research Laboratory of Nursing & Materials
West China Hospital
Sichuan University
Houde Building, No. 37 Guoxue Lane
Wuhou District
Chengdu, 610041
China
Phone: 86 18980601488
Email: likalika127127@163.com

Abstract

Background: Colorectal cancer surgery requires perioperative fasting to ensure safety, but this can cause physiological and psychological discomfort, such as impaired intestinal motility, bloating, immune suppression, anxiety, and appetite loss. To address this challenge, we developed the virtual reality (VR) multisensory sham-feeding device (VRMS-SFD), based on Pavlov conditioned reflex mechanism, where vagal stimulation triggered by seeing, smelling, or thinking about food activates cephalic-phase responses, promoting digestive secretion and intestinal motility. The device integrates multisensory stimulation—visual (food presentation), auditory (eating sounds and relaxing music), and olfactory (food-specific scents)—to create a VR dining experience. It features three VR scenes (Chinese restaurant, fruit shop, and dessert shop) with 23 food options, offering immersive interaction through a head-mounted display and synchronized scent release.

Objective: This study aimed to evaluate the usability, acceptability, and safety of the VRMS-SFD in patients with colorectal cancer during postoperative fasting, assessing its potential to alleviate discomfort, stimulate appetite, and enhance emotional well-being.

Methods: A mixed methods design was used. Participants used VRMS-SFD for 20 minutes during each meal time over 3 postoperative days. Quantitative data included the System Usability Scale (SUS) and demographic or clinical variables. Qualitative data were obtained from 15- to 30-minute semistructured interviews, transcribed verbatim, and thematically analyzed using NVivo (version 15.0; QSR International). When no new themes emerged, the sample size was considered sufficient, and data saturation was achieved. Two researchers independently coded transcripts, with discrepancies resolved by a third reviewer. The study was ethically approved (1087) and clinically registered.

Results: A total of 37 patients were included. The mean SUS score was 77.78 (SD 7.90; range 62.5-97.5), indicating high usability. Participants rated ease of use (mean 4.46, SD 0.56), learnability (mean 4.27, SD 0.69), and confidence (mean 4.27, SD 0.61) positively. There was no correlation between SUS score and age (r=0.05). Thematic analysis revealed four themes: (1) immersive and enjoyable experience, with patients describing the device as “novel” and “engaging”; (2) reduced fasting-related discomfort, including less bloating, improved mood, and reported peristalsis; (3) appetite stimulation, with many noting increased

XSL•FO
RenderX

hunger; and (4) improvement suggestions, such as enhancing scent authenticity, simplifying controls, and diversifying music. One participant experienced transient dizziness, resolving within 5 minutes post–device removal, with no other adverse events. Quantitative and qualitative findings converged, confirming robust usability and clinical benefits.

Conclusions: The VRMS-SFD may be a feasible, well-accepted tool that mitigates fasting-related discomfort, enhances appetite, and improves emotional well-being in patients undergoing colorectal cancer surgery. Future iterations should refine scent calibration and the user interface. Larger trials with objective measures (eg, gastrointestinal motility and hormone levels) are needed to validate efficacy and explore applications in other conditions.

Trial Registration: Chinese Clinical Trial Registry (ChiCTR), ChiCTR2100051419; https://www.chictr.org.cn/showprojEN.html?proj=134263

(JMIR Serious Games 2025;13:e75641) doi:10.2196/75641

KEYWORDS

colorectal cancer; fasting; virtual reality; multisensory stimulation; patient comfort; appetite; System Usability Scale; mixed methods research

Introduction

Background

Colorectal cancer is the third most common malignancy and the second leading cause of cancer-related mortality worldwide, with over 1.9 million new cases and 904,000 deaths reported in 2022 [1]. Surgery is the primary treatment for most patients with colorectal cancer [2]. To minimize perioperative risks, fasting protocols are mandatory for patients, typically requiring 6 hours of preoperative fasting and withholding oral intake until the return of bowel function postoperatively [3]. However, prolonged perioperative fasting (48-120 h) may negatively impact the intestinal mechanical and immune barriers, leading to intestinal ischemia, reperfusion injury, and postoperative gastrointestinal dysfunction [4,5]. Postoperative ileus occurs in approximately 10%-27% of patients following major abdominal surgery, resulting in abdominal distension, delayed recovery, and increased hospital stays [6,7]. Strategies such as early enteral nutrition, early ambulation, and probiotics are incorporated in enhanced recovery protocols to promote gut function, but each has notable limitations, including aspiration risk and inconsistent efficacy [8-10].

Sham feeding, a technique based on Pavlov conditioned reflex theory, involves multisensory stimulation without actual ingestion, and has been shown to stimulate gastric acid secretion, enhance intestinal motility, and support gut function [11]. Based on this, Li et al [12] proposes a new concept of “sham feeding”: simulating oral food intake by integrating visual, auditory, and olfactory sensory stimuli to induce cephalic phase secretion in brain neurons, thereby addressing the shortcomings of traditional nutritional infusion, which lacks the sensory stimulation of food’s color, aroma, and taste, making it difficult to trigger cephalic phase secretion. Simple forms such as gum chewing have demonstrated benefits in patients with colorectal cancer [13,14]. Virtual reality (VR) technology provides immersive, digital environments capable of delivering visual, auditory, and olfactory stimuli simultaneously [15,16]. While VR has been applied in clinical settings for anxiety reduction, pain management, and rehabilitation [17,18]. Meanwhile, VR interventions have shown potential in managing appetite-related behaviors. Recent studies have started to explore the effects of VR on appetite and eating behaviors. Huang et al [19] found that immersive VR reminiscence interventions led to sustained improvements in neuropsychiatric symptoms and appetite in people with dementia. Similarly, Sauchelli and Brunstrom [20] reported that VR exergaming enhanced the affective experience of exercise and significantly reduced postexercise energy intake by 12% in inactive adults. These findings suggest that VR-based sensory stimulation may influence food-related cognition and appetite regulation. However, systematic reviews such as Xia et al [21] indicate that evidence regarding VR’s direct effect on appetite remains inconclusive. However, its use for gut function regulation during perioperative fasting remains underexplored, highlighting the need for further targeted research in specific clinical contexts such as perioperative fasting.

To address this gap, we developed an immersive VR multisensory sham-feeding device (VRMS-SFD), which is designed to simulate cephalic phase responses through synchronized visual, auditory, and olfactory stimulation of food in a 360° immersive environment, thereby bridging the gap between traditional sham feeding and modern digital therapeutics.

Study Aims and Objectives

This study aimed to evaluate the usability and user experience of a newly developed VRMS-SFD for patients with colorectal cancer during perioperative fasting. The objectives were to: (1) assess the usability of VRMS-SFD using the System Usability Scale (SUS), (2) explore patients’ subjective experiences with VRMS-SFD through semistructured qualitative interviews, (3) identify perceived benefits and barriers regarding the use of VR-based sham feeding for appetite stimulation and discomfort relief, and (4) gather user feedback to inform future iterations of the device and guide clinical implementation.

Methods

Study Design

The study used a mixed methods approach and used the IDEAS (Integrate, Design, Assess, and Share) framework for creating digital health behavior change interventions [22]. The IDEAS framework recommends that large-scale randomized controlled trials be preceded by smaller-scale assessments to examine preliminary efficacy and usability. In these early evaluations,

XSL•FO
RenderX

usability and satisfaction can be assessed through questionnaires, while interviews provide deeper insights into user experiences [23]. In this study, qualitative and quantitative evaluation of device usability and safety for patients undergoing fasting periods for colorectal cancer surgery using our self-developed VRMS-SFD. The study has been approved by the ethics committee and clinical trial registry, and written informed consent was obtained from all participants before they joined the study. To maintain confidentiality, each participant was assigned a numerical code. No compensation was provided for participation in this study. This study followed the MMR-RHS (Mixed Methods Reporting in Rehabilitation and Health Sciences) checklist, provided in Multimedia Appendix 1 [24].

Participants—Inclusion and Exclusion Criteria

To be eligible, participants had to (1) be inpatients in the Colorectal Cancer Center ward, (2) be aged 18 to 80 years, (3) be patients who have undergone colorectal cancer surgery, (4) be patients who can cooperate with nurses to use VR equipment, and (5) sign informed consent and voluntarily participate in the study. Exclusion criteria were the following: (1) pregnant or lactating women, or patients with psychiatric disorders; (2) patients with serious and uncontrolled organic diseases or infections, such as decompensated heart failure, lung failure, kidney failure, or other life-threatening diseases at any time; (3) patients with intestinal obstruction, gastric or intestinal perforation, chronic enteritis, or other diseases that affect intestinal function greatly; (4) patients with a history of, or at high risk of gastrointestinal ulcer, (5) patients who have used probiotics, yogurt, probiotics powder, other microecological probiotics, metformin, proton pump inhibitors, or berberine within last 1 month, (6) patients with a history of motion sickness, such as Meniere syndrome or otolithiasis; (7) patients who have undergone preoperative radiotherapy and chemotherapy; 8) patients with concomitant use of other experimental drugs; and (9) patients currently participating in other clinical trials.

Recruitment

Potential participants are recruited through posters posted in the department or by the investigators after the patient has been admitted to the hospital. Their potential eligibility was then assessed, and if deemed eligible, they were invited to join the project for baseline data collection, intervention, and final measurement in the Colorectal Cancer Center ward of West China Hospital, Sichuan University.

VRMS-SFD Intervention Procedures

The VRMS-SFD hardware and software were developed collaboratively by the research team at West China Hospital and an affiliated university-based engineering team. No external commercial vendors were involved. All authors contributed to the conceptualization, iterative prototyping, and pilot usability testing of the device. The complete VR system includes a wireless head-mounted display (PICO Neo 3 Pro, Pico Interactive Inc, United States), PICO controllers, an integrated aroma emitter, a custom wireless gateway, and a mobile treatment cart with a high-performance laptop and display monitor (Figure 1).

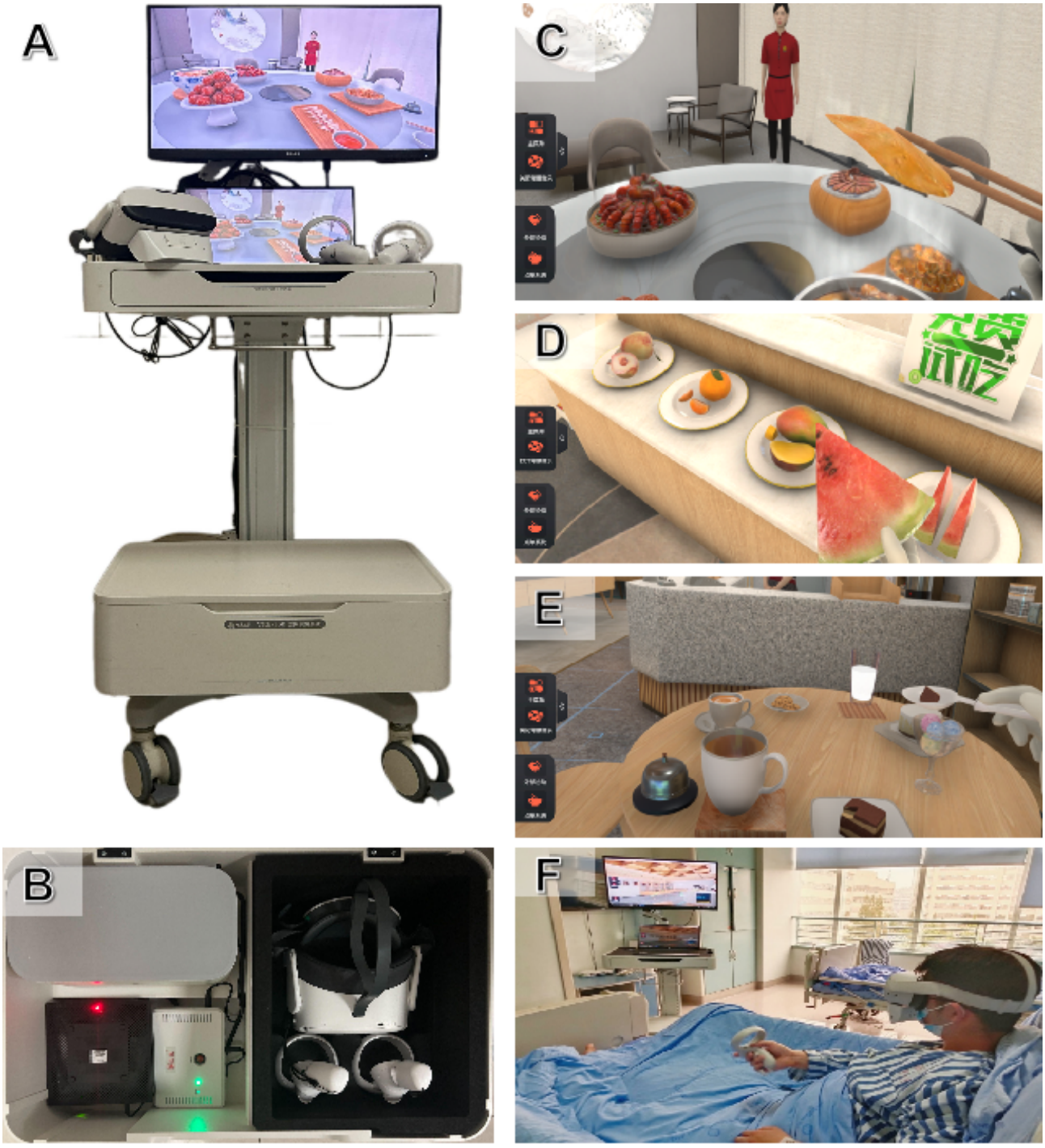
Although the PICO headset is inherently wireless, a custom-designed wireless gateway (TP-Link AX6000) was integrated to ensure real-time synchronization between the VR content and the olfactory stimuli. This gateway enables low-latency (10-20 ms) communication via a 2.4 GHz band within a 0.5-5 m range, linking the VR system and aroma emitter through a stable local area network. When users select and bring a VR food item close to the mouth using the controller, the VR apps transmit a command to the control node via the gateway. The control module translates this command into channel-specific signals, which are amplified and passed to the ultrasonic chip in the aroma emitter, instantly releasing the corresponding scent. This design ensures simultaneous delivery of visual, auditory, and olfactory stimuli, enhancing the immersive experience of VR eating. This real-time alignment, facilitated by the self-developed wireless transmitter and ultrasonic atomization technology, represents a key technical innovation, ensuring immersive multisensory coherence during VR dining.

The device features three VR dining scenes—a Chinese restaurant, a fruit shop, and a dessert shop—offering 23 VR food options such as mapo tofu, maocai, bacon, mango, orange, apple, cappuccino, ice cream, and cookies, each with matching scents. During use, patients hear detailed eating, drinking, and chewing sounds via the headset's built-in speaker, simulating real eating steps. The environment includes light, relaxing background music to promote mental relaxation.

Patients interacted with the VRMS-SFD for 20 minutes during each meal period (7 AM-9 AM, 11 AM-1 PM, and 5 PM-7 PM) from preoperative fasting through postoperative day 3. Further hardware schematics and detailed procedural steps for equipment calibration and patient interaction are provided in Multimedia Appendix 2, and a demonstration video is available in Multimedia Appendix 3.

XSL•FO
RenderX

Figure 1. Hardware setup, immersive VR environment, and patient use of the VRMS-SFD. (A) Photograph showing the integrated hardware system, including a mobile nursing treatment cart, a high-performance laptop (running VR drivers), a 21-inch display monitor, a wireless VR headset (PICO Neo 3, with built-in audio playback), a PICO controller, and an olfactory scent-emitting box attached to the VR headset. (B) Cabinet under the cart with VR headset, controllers, and Wi-Fi 6 router. (C-E) Actual in-VR screenshots from three immersive dining scenes: a fruit shop (C), a Chinese-style restaurant (D), and a dessert shop (E). Participants can select from 23 VR food items (eg, spicy crayfish, bacon, sugar-oil glutinous balls, tiramisu, and milk). When a food item is grasped using the PICO controller and moved toward the mouth, the corresponding scent is emitted in real-time by the scent box. Each food model is rendered with frame-by-frame animation to simulate physical deformation and environmental dynamics (eg, steam and sunlight) depending on the scene. Synchronized audio, including chewing, swallowing, and ambient music, enhances the realism of the simulated eating experience. (F) A postoperative patient with colorectal cancer using the VRMS-SFD. This photograph was taken with patient consent and does not reveal identifiable facial features. It illustrates the real-world application of the device during perioperative fasting. VR: virtual reality; VRMS-SFD: virtual reality multisensory sham-feeding device.



XSL•FO
RenderX

Data Collection and Organization

Overview

The collection and collation of questionnaire data and semistructured interview data were done by two graduate nursing students who had received training. Qualitative and quantitative data were collected from participants following each session after full use of VR. The survey captured basic information about hospitalized patients. The specific content of the patient-facing questionnaire is provided in Multimedia Appendix 4.

At baseline, demographic and clinical information—including gender, age, height, weight, education level, marital status, place of residence, medical insurance coverage, and prior experience with VR—were collected using a structured questionnaire designed for this study. The data were obtained through face-to-face interviews conducted by two trained graduate nursing students before patients engaged with the VR device. Clinical data were extracted from patient charts by researchers using a standardized data collection form. Data were recorded on paper and later digitized and managed using Excel (Microsoft Corp). All responses were documented confidentially and stored securely for subsequent analysis.

Quantitative Measurement

Patients were required to provide their subjective assessment of their experience, in which the SUS was adopted to examine the usability and learnability of the device; it has a high reliability coefficient of 0.85 [25]. The SUS consists of ten 5-point Likert scale questions, from 1=strongly disagree to 5=strongly agree. Both positively and negatively framed items are included. For positively framed items, the corresponding score of each item should be –1; for the negatively framed items, the score is supposed to be subtracted by 5. Finally, scores are summed and multiplied by 2.5; the total usability score ranges from 0 to 100. A score of <50 was considered unacceptable, 50-70 was considered marginal, and >70 was considered acceptable (>85=excellent) [26].

Qualitative Interview

The interview questions were defined based on the research question and topics the study coordinators wanted to address, and their formulation followed basic guidelines for open-ended interview questions, such as omitting yes or no answers or leading questions [27]. Using a semistructured interview, the patient was interviewed for 15-30 minutes after the device had been used. Interviews were conducted in-person with audio recording, ensuring comprehensive capture of participant feedback. Audio-recorded interviews were transcribed verbatim by a professional transcription service using iflyrecClient.exe, with a 100% accuracy check performed by XY and PY to ensure fidelity to the original recordings. Transcripts were anonymized and stored securely in a password-protected database to protect participant confidentiality. The interview outline is detailed in Multimedia Appendix 4.

We used reflexive thematic analysis as described by Braun and Clarke [28], which is underpinned by a constructionist epistemology that recognizes themes as researcher-generated rather than emerging from data. Our analysis focused on shared meaning themes rather than topic summaries, seeking to understand the deeper experiences and perceptions of patients using the VR medical system. The analysis aimed to capture both semantic (explicit) and latent (implicit, interpretive) meanings within the data.

Adverse Events

Adverse reactions, characterization, severity, duration, and resolution will be recorded in detail for each use.

Statistical Analyses

Quantitative Analysis

The total usability score was calculated using the original equation by Brooke [25]:



which results in a score between 0 (min) and 100 (max). For easier visual comparability between the individual SUS item ratings (Qi), we inverted the statements of negatively phrased items (Q2, Q4, Q6, Q8, and Q10) to make all statements positive and adapted the scoring accordingly [29]. The mean and SD of the question and the total usability score were calculated by SPSS Statistics (version 29.0; IBM Corp).

Qualitative Interview

Thematic analysis was conducted following Braun and Clarke's [30] six-phase framework: (1) data familiarization, (2) initial code generation, (3) theme searching, (4) theme review, (5) theme definition and naming, and (6) report production [30]. Two researchers independently coded the transcripts using NVivo (version 15.0; QSR International), with regular meetings to discuss coding decisions and resolve discrepancies through consensus [29]. When no new themes emerged, the sample size was considered sufficient, and data saturation was achieved. Following Hennink and Kaiser's [31] guidance, we used a code frequency counts strategy to assess data saturation. After each interview, transcripts were analyzed to identify new codes and themes. Data saturation was systematically assessed by tracking the emergence of new themes across sequential interviews.

To ensure the rigor of the qualitative findings, we followed Lincoln and Guba's [32] criteria for trustworthiness: (1) credibility was enhanced through member checking, whereby key themes were reviewed by five participants for validation; (2) dependability was addressed by maintaining an audit trail documenting coding decisions and analytical processes; (3) confirmability was ensured through regular peer debriefings among the research team to minimize individual biases; (4) transferability was supported by providing rich, thick descriptions of participants and contexts; (5) reflexivity was maintained throughout, with the researchers engaging in ongoing self-reflection to consider how their backgrounds and preconceptions might influence the interpretation of data.

Mixed Methods Integration and Analysis

The steps for data integration were as follows: quantitative analysis, qualitative analysis, identification of similar and divergent findings, and confirmation, extension, or identification

XSL•FO
RenderX

of discordant results [33]. Confirmation occurs when findings from both types of data reinforce each other. Extension is observed when divergent findings between the two datasets broaden insights into usability by addressing distinct or complementary aspects of the user experience. Discrepancy arises when survey and interview results are inconsistent, contradictory, or in disagreement with each other [23,34].

Ethical Considerations

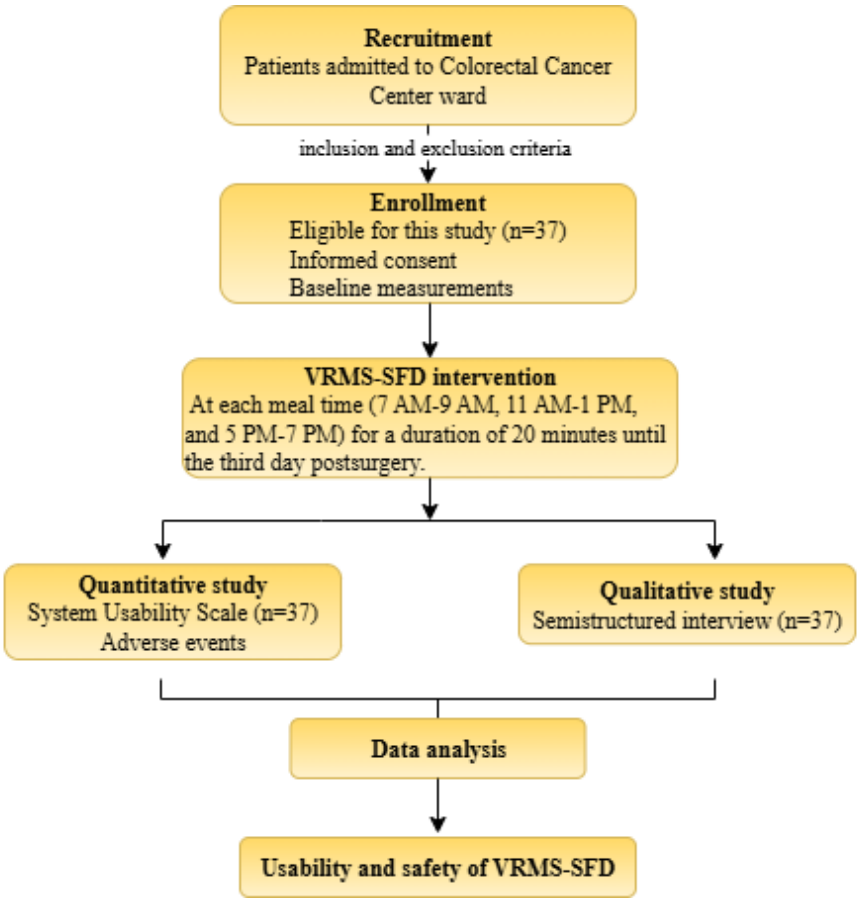
This study was classified as human participants research and was reviewed and approved by the Ethics Committee on Biomedical Research, West China Hospital of Sichuan University (1087). It was also registered as a clinical trial with the Chinese Clinical Trial Registry (ChiCTR2100051419). Written informed consent was obtained from all 37 participants prior to their enrollment, with consent forms explaining the study's purpose, procedures, potential risks (eg, mild dizziness), and benefits. Participants were informed of their right to withdraw at any time without consequences. To protect privacy and confidentiality, each participant was assigned a unique numerical code, and all data (eg, SUS scores and interview transcripts) were deidentified and stored securely on password-protected files accessible only to the research team. Compensation was provided in the form of exemption from postoperative outpatient visit fees and 3-5 free postoperative nursing consultation services to support participants' recovery. No images in the manuscript or supplementary materials (eg, Figure 1E) allow identification of individual participants, as facial features are obscured or not visible, ensuring compliance with privacy standards.

Results

Characteristics of the Participants

A total of 37 participants were recruited for this study, and all completed the use of the VRMS-SFD. The flowchart for the participant inclusion process is shown in Figure 2. In the data collection stage, after 3 days of use, 37 questionnaires were collected, for a recovery rate of 100%. The mean age of the participants was 56.57 (SD 8.05) years; most of the participants were male (21/37, 57%), and all of them were married. The detailed characteristics of the participants are shown in Table 1 and can also be found in Multimedia Appendix 5.

Figure 2. Flow diagram of a mixed methods study evaluating VRMS-SFD for patients undergoing colorectal cancer surgery at West China Hospital, Sichuan University, Chengdu, China, from 2022 to 2023, involving 37 postoperative patients with colorectal cancer, to assess the usability of a VRMS-SFD during perioperative fasting, including participant recruitment, intervention, and data collection processes. VRMS-SFD: virtual reality multisensory sham-feeding device.



XSL•FO
RenderX

Table 1. Characteristics of participants evaluating a VRMS-SFD[a] for patients undergoing colorectal cancer surgery[b].

Sample characteristics	Values (n=37)
Sex, n (%)	
Male	21 (57)
Female	16 (43)
Age (years), mean (SD)	56.57 (8.05)
Height (cm), mean (SD)	162.62 (9.34)
Weight (kg), mean (SD)	60.44 (10.46)
BMI (kg/m^2), mean (SD)	22.82 (2.96)
Nutritional status [c], n (%)	
1	15 (41)
2	8 (22)
3	8 (22)
4	4 (11)
5	1 (3)
6	1 (3)
Education level, n (%)	
Primary school or under	9 (24)
Middle school	12 (32)
High school	6 (16)
University or above	10 (27)
Marital status, n (%)	
Married	37 (100)
Single	0 (0)
Divorced or widowed or else	0 (0)
Place of residence, n (%)	
City	19 (51)
Suburb	18 (49)
Medical insurance	
Yes	37 (100)
No	0 (0)
Have used VR [d] before	
Yes	0 (0)
No	37 (100)

[a]VRMS-SFD: virtual reality multisensory sham-feeding device.

[b]None of the recruited patients had any previous experience with VR.

[c]The NRS2002 (Nutritional Risk Screening) scale was used to assess the nutritional status of participants, where lower scores signified lower nutritional risk and higher scores indicated higher nutritional risk.

[d]VR: virtual reality.

Quantitative Results

The usability of the VR feeding simulation device was assessed using the SUS, a 10-item questionnaire with scores ranging from 0 to 100. Across the 37 participants, the device achieved a mean SUS score of 77.78 (SD 7.90), with individual scores ranging from 62.5 to 97.5, indicating that the usability evaluation of this device is good, but there is still a need to improve it. The detailed scores for each item of the SUS are shown in Table 2 and can also be found in Multimedia Appendix 6. There was no significant correlation between participant age and usability score (Figure 3; r=0.05). Participants also rated their overall

XSL•FO
RenderX

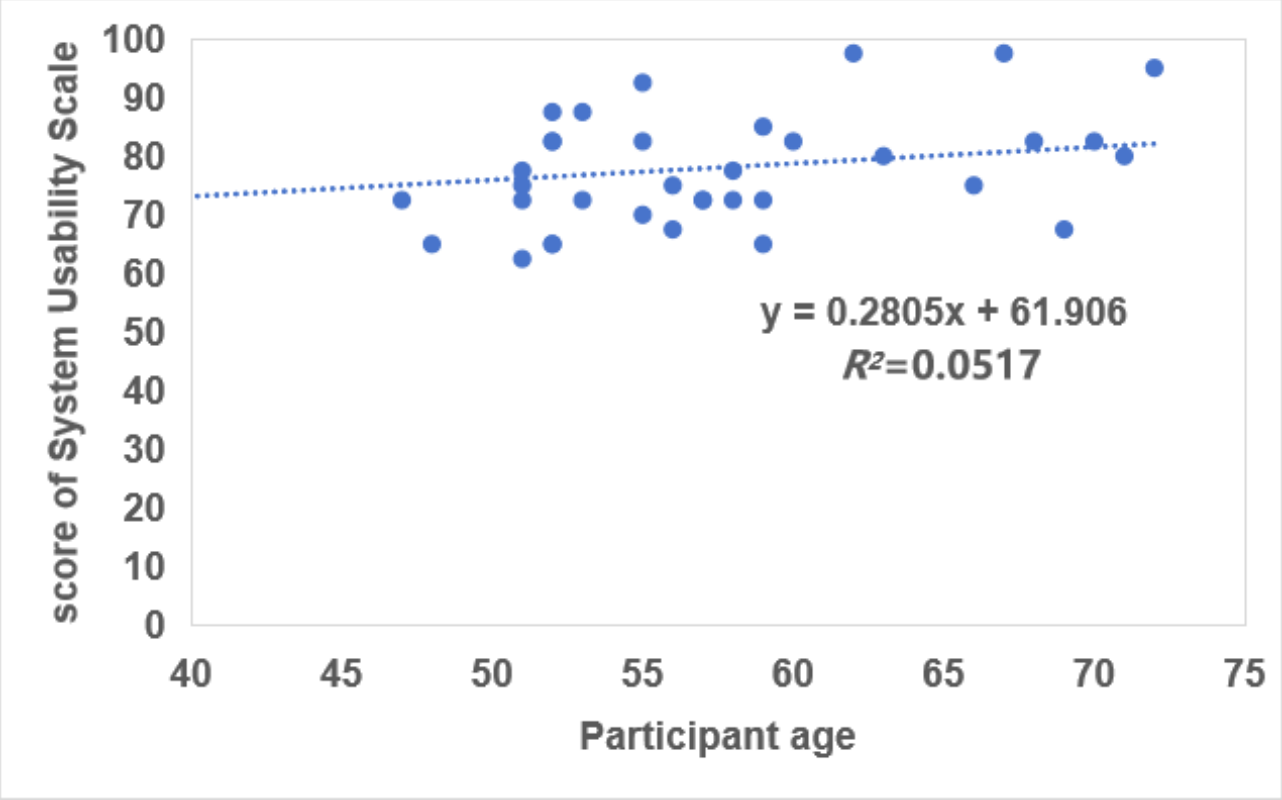
satisfaction with new technology using our device with a mean of 8.38 (SD 1.93) out of 10, indicating relative comfort navigating digital tools independently.

Table 2. System Usability Scale scores for a virtual reality multisensory sham-feeding device for patients undergoing colorectal cancer surgery.

Item No	Item	Score, mean (SD)[a]
Q1	I think that l would like to use this system frequently	4.11 (0.77)
Q2	I found the system unnecessarily complex	2.08 (0.86)
Q3	I thought the system was easy to use	4.46 (0.56)
Q4	I think that l would need the support of a technical person to be able to use this system	2.32 (0.78)
Q5	I found the various functions in this system were well integrated	4.05 (0.85)
Q6	I thought there was too much inconsistency in this system	1.81 (0.66)
Q7	I would imagine that most people would learn to use this system very quickly	4.27 (0.69)
Q8	I found the system very cumbersome to use	1.89 (0.77)
Q9	I felt very confident using the system	4.27 (0.61)
Q10	I needed to learn a lot of things before I could get going with this system	1.95 (0.70)

[a]Total score: mean 77.78 (SD 7.90).

Figure 3. Correlation between age and System Usability Scale scores evaluating a virtual reality multisensory sham-feeding device for patients undergoing colorectal cancer surgery (n=37).



Qualitative Results

Qualitative data were collected through 15- to 30-minute semistructured interviews with 37 postoperative patients with colorectal cancer, to assess the usability and user experience of VRMS-SFD. Interviews were audio-recorded, transcribed verbatim in Mandarin. Two researchers independently coded the transcripts, achieving an intercoder agreement of 92%, and discrepancies were resolved through discussion with a third researcher. Data saturation was reached after 32 interviews, as no new themes emerged, but all 37 interviews were included to ensure comprehensive representation. Reflexivity was maintained by the research team documenting personal biases in reflective journals to minimize influence on analysis.

XSL•FO
RenderX

Confirmability was ensured by maintaining an audit trail of coding decisions and raw data excerpts, linked to the identified themes (eg, immersive experience and reduced discomfort). Dependability was supported by providing detailed methodological descriptions. Credibility was enhanced through member checking, where five participants reviewed preliminary themes to verify accuracy, with minor clarifications incorporated. Transferability was addressed by providing thick descriptions of the study context, participant demographics, and clinical setting, enabling applicability to similar perioperative fasting contexts.

Thematic analysis yielded 4 major themes (overall perception of the device, improvement of perioperative fasting adverse effects, impact on appetite, and suggestions for follow-up improvement of VR equipment), 14 subthemes, and 203 nodes. Multimedia Appendix 7 presents the number of participants whose interviews included each subtheme. For instance, “Stimulates Appetite” was mentioned by 30 of 37 (81%) participants, while “Immersive Experience” was described by 13 of 37 (35%). Themes such as “Relieves Discomfort from Fasting” (20/37) and “Improves Mood” (9/37) showed considerable relevance to perceived benefit. Subthemes with low mention counts (eg, “Scent Too Strong” or “Add More Scenes”) reflect individualized user feedback and inform design iteration. Multimedia Appendix 7 provides the complete qualitative analysis, including themes, subthemes, and supporting participant quotes.

Adverse Events

One of the 37 participants reported mild dizziness. This participant was a first-time user of VR equipment and experienced no associated symptoms such as nausea or vomiting. Upon the onset of dizziness, the VR device was promptly removed by the researchers, and the dizziness resolved within 5 minutes after device removal. Follow-up evaluations were conducted at 30 minutes, 24 hours, and 3 days postincident, with the participant reporting no further discomfort.

Mixed Methods Results

The themes that emerged from the qualitative results were compared and merged with the quantitative results in terms of the overall perception of the device, the improvement of perioperative fasting adverse effects, the impact on appetite, and suggestions for the follow-up improvement of our device. Quantitative data showed that the overall usability of our device was good, especially scores on item 3 (I think the system was easy to use), item 6 (I thought there was too much inconsistency in this system), but some aspects still need further improvement. Specific usability issues (overall experience, its effectiveness in reducing discomfort during preoperative fasting, its impact on appetite, and suggestions) of the VRMS-SFD were deeply explored through qualitative interviews. The overall experience of the device was excellent, immersive, and enjoyable. As for discomfort during perioperative fasting, our device relieved the discomfort of being unable to eat, improved perioperative mood, and promoted intestinal peristalsis. It also stimulated appetite. The optimization suggestions included calibration of scents, simplifying operations, and improving background music. Quantitative results and qualitative results were generally consistent, and no discordant results were observed. The comparison and merging of the 2 datasets resulted in confirmed and expanded findings.

Discussion

Principal Findings

This mixed methods study evaluated the usability, acceptability, and safety of the VRMS-SFD in 37 postoperative patients with colorectal cancer during perioperative fasting. The mean usability score of the device was 77.78 (SD 7.90), above the suggested cut-off value of 70, indicating considered acceptable usability and was graded as B+ [25,26]. According to our usability assessment, item 3 (mean 4.46, SD 0.56) and item 10 (mean 1.95, SD 0.70) indicated easy learnability of our equipment, whereas item 9 (mean 4.21, SD 0.67), item 1 (mean 4.11, SD 0.67), and item 5 (mean 4.05, SD 0.85) demonstrated confidence and general usability among participants [35,36]. This is also supported by comments in the qualitative interviews, where most feedback was positive.

Qualitative analysis produced four principal themes: overall perception of the device, improvement of perioperative fasting adverse effects, impact on appetite, and suggestions for follow-up improvement of VR equipment. Thematic coding achieved an intercoder agreement of 92%; data saturation was reached after 32 interviews, and all 37 interviews were analyzed to ensure comprehensive representation. Participants commonly described the device as “novel” and “enjoyable,” reporting perceived benefits such as improved mood, reduced bloating or discomfort, reports of increased peristalsis or gas passage, and enhanced appetite. Participants also provided actionable feedback, including improving scent authenticity, simplifying controls, and expanding food selections.

Safety findings were favorable: one participant experienced transient dizziness that resolved within 5 minutes after device removal and postural adjustment; no other adverse events were observed. Overall, the quantitative and qualitative results converged to indicate the feasibility, acceptability, and preliminary patient-perceived benefits of the VRMS-SFD in the studied perioperative context.

Comparison to Prior Work

The majority of participants felt that the device offered an immersive experience, describing it as both “novel” and “enjoyable.” This could be attributed to the difference in both form and content between the VRMS-SFD and conventional interventions (eg, medication and injections). By integrating a series of technologies, VR technology provides a multisensory and 3D environments that enable users to become fully immersed in a simulated world, and its immersive property makes the patients actively interact with the vivid VR environment, which demands more attention [37]. Similar to this study, the VR app developed by Langlet et al [38] for treating eating behavior in patients with eating disorders (ED) received an average score of 73.4 (SD 9.2; range 55-90) and was rated as “good” overall. However, due to being in the early stage of development, the app’s usability was evaluated by clinicians, physicians (who have daily interactions with patients

XSL•FO
RenderX

with ED), and IT personnel, rather than by patients with ED themselves.

In addition, the usability of our device does not appear to be affected by age. Previous studies have shown that younger users are more receptive to VR environments [39,40]; however, in this study, no significant correlation between participant age and usability score (r=0.05) was observed. The consistent ease of use of our device across different age groups indicated good universality.

With regard to its clinical effect, some participants indicated that the VRMS-SFD can ease discomfort during perioperative fasting. On the one hand, some participants reported that our device helped improve their mood. As a distraction tool, VR has proven effective in clinical settings, with multiple studies demonstrating its ability to significantly alleviate pain and reduce anxiety in perioperative medical contexts [41,42]. A number of participants highlighted that our device could free them from the frustration of being unable to eat. A likely explanation could be that the multiple stimuli during device use activated the orbitofrontal cortex, which led to an enhanced hedonic response and increased pleasantness [43]. On the other hand, patients also perceived it to have a positive effect in promoting the recovery of gastrointestinal function, such as relieving bloating, promoting intestinal peristalsis, and facilitating gas expulsion. This could be explained from the perspective of neurogastroenterology. The gut-brain axis is the bidirectional communication system between the gastrointestinal tract and the central nervous system [44]. Gastrointestinal motility is regulated by the enteric nervous system and extrinsic factors, including the brain, the autonomic nervous system, the gut-associated immune system, and the gut microbiome [45]. These factors work together to modulate gastrointestinal function and maintain homeostasis. Modulation of intrinsic enteric neuron activity via top-down signals or microbe-associated molecules ultimately influences gastrointestinal physiology [46]. In response to stimuli like the smell and sight of food, hormones that prepare for food consumption—including insulin, ghrelin, gastrin, glucagon-like peptide-1, and cholecystokinin—are released [47]. This is the so-called cephalic response, which is one of the most common forms of conditioning. Some of these pancreatic and gastric secretions exert a stimulatory effect on the gastrointestinal motility [48]. As a result, many participants shared that the device boosted their appetite to some extent. Appetite loss is a pervasive and quality-of-life–altering problem after major abdominal surgery [49]; 55% patients reported decreased appetite after colorectal surgery [50]. Postoperative loss of appetite may lead to weight loss and impaired intestinal absorption, which can subsequently result in malnutrition, adversely affect patients' survival, quality of life, and increase the risk of complications [51,52]. The abdominal discomfort of patients was relieved to some degree. This could be the primary factor contributing to the recognition of the VRMS-SFD among participants.

Several studies have shown that VR-enhanced cognitive behavioral therapy can regulate the appetite and eating frequency of patients with ED [53], highlighting the potential of VR in addressing such problems. Liem et al [54] proposed that food choices were influenced by the sensory input from the environment. Ferrer-Garcia et al [55] conducted VR-based cue exposure therapy that included a library of 30 most craved VR foods among participants and 4 real-life VR environments, aiming to diminish food-related anxiety and cravings, thereby reducing binge eating. Ludovica exposed patients with anorexia nervosa to a VR kitchen environment to reduce anxiety around eating and food avoidance. Previous studies also suggested that the involvement of multiple senses can enhance the feeling of presence and associated food desire [54,56]. However, studies similar to the aforementioned ones have typically been limited to visual cues, with only occasional use of sound or music in the VR environment. In contrast, this study used synchronized visual, auditory, and olfactory modalities associated with food, creating an immersive VR dining experience for participants that effectively aroused their appetite.

Methodologically, our mixed methods design is consistent with contemporary digital-health usability research. Lo et al [57] suggested that the use of a mixed method design could contribute to a better understanding of the intervention. The National Institutes of Health Office of Behavioral and Social Science Research also recommends the application of mixed methods in the evaluation of new instruments or methods, since it offers a comprehensive understanding by combining qualitative and quantitative data. In this way, it fosters innovation, supports the development and refinement of tools, and enhances the scientific rigor and practical applicability of findings [58]. This mixed methods usability study provided evidence on the application of the VRMS-SFD among patients with colorectal cancer who are undergoing surgery. The integration of quantitative metrics and qualitative feedback offers an understanding of both the device's technical performance and its acceptance within the clinical context. These findings reveal several key aspects regarding the feasibility and potential therapeutic value of VR–based feeding intervention in postoperative care.

Limitations

First, this study evaluated short-term usability as a single phase within an iterative design process. While this allowed for focused feedback on initial device performance, it also limits conclusions regarding long-term clinical outcomes. We attempted to mitigate this by integrating both quantitative usability scores and qualitative feedback, ensuring a comprehensive assessment of early feasibility. Nevertheless, longitudinal studies are needed to confirm the durability of effects and assess whether identified usability issues can be resolved through subsequent design iterations.

Second, the study sample was relatively small and drawn exclusively from a single medical institution, which may introduce selection bias and limit generalizability. To reduce this risk, we recruited consecutive eligible patients within the defined study period and ensured demographic diversity within the cohort. However, expanding future trials to include larger and multicenter populations will be essential to enhance the external validity and representativeness of findings.

Third, this study relied heavily on self-reported data, which are susceptible to subjective bias. While triangulation through mixed methods (quantitative SUS scores and qualitative thematic

XSL•FO
RenderX

analysis) improved credibility, the absence of objective physiological markers restricts the interpretation of the biological impact of the intervention. Future studies should incorporate measures such as gastrointestinal motility monitoring, abdominal circumference, and serum biomarkers of the gut-brain axis (eg, ghrelin, leptin, and cholecystokinin), which would provide more robust evidence and strengthen the confirmability and dependability of results.

Fourth, one participant reported experiencing dizziness during the VR session. Since individuals with prior symptoms of dizziness, vestibular disease, or cybersickness were excluded beforehand, the occurrence of dizziness in this case could be attributed to factors such as residual effect of anesthesia, metabolic disorder due to perioperative fasting, adverse effects of postoperative analgesics, or reduced cerebral blood flow caused by abrupt position changes after prolonged immobility postsurgery [59-61]. Besides, mild anxiety might also have played a contributing role [62]. To mitigate such adverse events in the future, sufficient recovery from anesthesia, gradual postural adjustment, and careful monitoring should be prioritized.

Future Directions

Participants in this study actively engaged in suggesting ways to improve the VRMS-SFD, highlighting both technical and experiential aspects. They identified issues such as the limited realism of certain food models and proposed practical enhancements, including more precise calibration of olfactory stimuli and adjustments to the background music to create a more immersive and enjoyable environment. These insights provide a valuable foundation for iterative design and patient-centered refinement.

Looking ahead, future development should prioritize incorporating patient feedback into systematic device upgrades, expanding the diversity of VR dining scenarios, and integrating adaptive features that tailor sensory input to individual preferences or clinical conditions. Beyond technical improvements, advancing this work will require sustained multidisciplinary collaboration among clinicians, engineers, psychologists, and designers. Such cooperation will be essential to optimize usability, enhance ecological validity, and ultimately develop the next generation of multisensory VR sham-feeding interventions with broader clinical applicability.

Conclusions

This mixed methods study evaluated the usability, acceptability, and safety of the VRMS-SFD among patients with colorectal cancer during perioperative fasting. Quantitative findings demonstrated a mean SUS score of 77.78 (SD 7.90), exceeding the threshold for acceptable usability, with consistently high ratings across age groups. Qualitative interviews confirmed these results, revealing that patients perceived the device as novel, engaging, and effective in alleviating fasting-related discomfort and stimulating appetite. Safety outcomes were favorable, with only one transient and self-limited adverse event reported. Taken together, these findings indicate that the VRMS-SFD may be a feasible and well-accepted tool for supporting patient comfort and appetite regulation during perioperative fasting, thereby addressing a critical unmet need in colorectal cancer care.

Acknowledgments

The authors would like to thank the engineering team members for their support in device design and development. This research is supported by the National Natural Science Foundation of China (grant U22A20334) and Key R&D Projects of the Department of Science and Technology of Sichuan Province (grant 2024YFFK0034).

Data Availability

The datasets generated or analyzed during this study are not publicly available due to privacy restrictions and institutional policies on patient data sharing, but are available from the corresponding author on reasonable request.

Authors' Contributions

Design and conceptualization: XY, KL
Participant recruitment: XY, PY
Clinical intervention: XY, PY, QX, WZ
Patient assessment: XY, PY, QX, YH, XL, WZ
Data collection: XL, YH
Writing—original draft: XY, PY
Writing—review and editing: XY, PY, QX, XL
Supervision: KL
All authors reviewed and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

XSL•FO
RenderX

MMR-RHS-checklist.
[PDF File (Adobe PDF File), 726 KB - games_v13i1e75641_app1.pdf]

Multimedia Appendix 2
Intervention process.
[DOCX File , 1534 KB - games_v13i1e75641_app2.docx]

Multimedia Appendix 3
VRMS-SFD patient demonstration video.
[MP4 File (MP4 Video), 3150 KB - games_v13i1e75641_app3.mp4]

Multimedia Appendix 4
Patient questionnaires.
[DOCX File , 16 KB - games_v13i1e75641_app4.docx]

Multimedia Appendix 5
Characteristics of participants evaluating a virtual reality multisensory sham-feeding device for colorectal cancer surgery patients (n=37). None of the recruited patients had any previous experience with virtual reality.
[DOCX File , 13 KB - games_v13i1e75641_app5.docx]

Multimedia Appendix 6
System Usability Scale (SUS) scores for a virtual reality multisensory sham-feeding device of colorectal cancer surgery patients.
[DOCX File , 13 KB - games_v13i1e75641_app6.docx]

Multimedia Appendix 7
Results of the interview content analysis.
[DOCX File , 32 KB - games_v13i1e75641_app7.docx]

References

1. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2024;74(3):229-263 [FREE Full text] [doi: 10.3322/caac.21834] [Medline: 38572751]
2. Vogel JD, Felder SI, Bhama AR, Hawkins AT, Langenfeld SJ, Shaffer VO, et al. The American Society of Colon and Rectal Surgeons clinical practice guidelines for the management of colon cancer. Dis Colon Rectum 2022;65(2):148-177. [doi: 10.1097/DCR.0000000000002323] [Medline: 34775402]
3. Joshi GP, Abdelmalak BB, Weigel WA, Harbell MW, Kuo CI, Soriano SG, et al. 2023 American Society of anesthesiologists practice guidelines for preoperative fasting: carbohydrate-containing clear liquids with or without protein, chewing gum, and pediatric fasting duration-a modular update of the 2017 American society of anesthesiologists practice guidelines for preoperative fasting. Anesthesiology 2023;138(2):132-151. [doi: 10.1097/ALN.0000000000004381] [Medline: 36629465]
4. Hirota K. Hypoxia-dependent signaling in perioperative and critical care medicine. J Anesth 2021;35(5):741-756 [FREE Full text] [doi: 10.1007/s00540-021-02940-w] [Medline: 34003375]
5. Nagai M, Noguchi R, Takahashi D, Morikawa T, Koshida K, Komiyama S, et al. Fasting-refeeding impacts immune cell dynamics and mucosal immune responses. Cell 2019;178(5):1072-1087.e14. [doi: 10.1016/j.cell.2019.07.047] [Medline: 31442401]
6. Sommer NP, Schneider R, Wehner S, Kalff JC, Vilz TO. State-of-the-art colorectal disease: postoperative ileus. Int J Colorectal Dis 2021;36(9):2017-2025 [FREE Full text] [doi: 10.1007/s00384-021-03939-1] [Medline: 33977334]
7. Bragg D, El-Sharkawy AM, Psaltis E, Maxwell-Armstrong CA, Lobo DN. Postoperative ileus: recent developments in pathophysiology and management. Clin Nutr 2015;34(3):367-376. [doi: 10.1016/j.clnu.2015.01.016] [Medline: 25819420]
8. Hill A, Heyland DK, Reyes LAO, Laaf E, Wendt S, Elke G, et al. Combination of enteral and parenteral nutrition in the acute phase of critical illness: an updated systematic review and meta-analysis. JPEN J Parenter Enteral Nutr 2022;46(2):395-410. [doi: 10.1002/jpen.2125] [Medline: 33899951]
9. Willner A, Teske C, Hackert T, Welsch T. Effects of early postoperative mobilization following gastrointestinal surgery: systematic review and meta-analysis. BJS Open 2023;7(5):zrad102 [FREE Full text] [doi: 10.1093/bjsopen/zrad102] [Medline: 37846641]
10. Tang G, Huang W, Tao J, Wei Z. Prophylactic effects of probiotics or synbiotics on postoperative ileus after gastrointestinal cancer surgery: a meta-analysis of randomized controlled trials. PLoS One 2022;17(3):e0264759 [FREE Full text] [doi: 10.1371/journal.pone.0264759] [Medline: 35231076]

XSL•FO
RenderX

11. Feldman M, Richardson CT. Role of thought, sight, smell, and taste of food in the cephalic phase of gastric acid secretion in humans. Gastroenterology 1986;90(2):428-433. [doi: 10.1016/0016-5085(86)90943-1] [Medline: 3940915]
12. Li K, Yuan X, Hu Y, Zhang W, Chen Y, Hong R, et al. [An integrated audio-visual-olfactory virtual reality false feeding device: research, development, and design]. Sichuan Da Xue Xue Bao Yi Xue Ban 2023;54(4):792-797 [FREE Full text] [doi: 10.12182/20230760204] [Medline: 37545076]
13. Sammut R, Trapani J, Deguara J, Ravasi V. The effect of gum chewing on postoperative ileus in open colorectal surgery patients: a review. J Perioper Pract 2021;31(4):132-139. [doi: 10.1177/1750458920917015] [Medline: 32301386]
14. Liu Q, Jiang H, Xu D, Jin J. Effect of gum chewing on ameliorating ileus following colorectal surgery: a meta-analysis of 18 randomized controlled trials. Int J Surg 2017;47:107-115. [doi: 10.1016/j.ijsu.2017.07.107] [Medline: 28867465]
15. Jallad ST, Işık B. The effectiveness of virtual reality simulation as learning strategy in the acquisition of medical skills in nursing education: a systematic review. Ir J Med Sci 2022;191(3):1407-1426. [doi: 10.1007/s11845-021-02695-z] [Medline: 34227032]
16. Zhong D, Chen L, Feng Y, Song R, Huang L, Liu J, et al. Effects of virtual reality cognitive training in individuals with mild cognitive impairment: a systematic review and meta-analysis. Int J Geriatr Psychiatry 2021;36(12):1829-1847. [doi: 10.1002/gps.5603] [Medline: 34318524]
17. Turgut A, İlçe AO, Öztürk H. The effect of immersive virtual reality application on anxiety, pain, and parental satisfaction in the perioperative process of children: a randomized controlled trial. Pain Manag Nurs 2024;25(6):584-590. [doi: 10.1016/j.pmn.2024.06.002] [Medline: 38955550]
18. Kiper P, Szczudlik A, Agostini M, Opara J, Nowobilski R, Ventura L, et al. Virtual reality for upper limb rehabilitation in subacute and chronic stroke: a randomized controlled trial. Arch Phys Med Rehabil 2018;99(5):834-842.e4. [doi: 10.1016/j.apmr.2018.01.023] [Medline: 29453980]
19. Huang LC, Chien CF, Yang YH. Exploring the immediate and long-term effects of immersive virtual reality on behavioral and psychological symptoms of dementia and caregiver burden: longitudinal observational study. JMIR Serious Games 2025;13:e73044 [FREE Full text] [doi: 10.2196/73044] [Medline: 40669069]
20. Sauchelli S, Brunstrom JM. Virtual reality exergaming improves affect during physical activity and reduces subsequent food consumption in inactive adults. Appetite 2022;175:106058 [FREE Full text] [doi: 10.1016/j.appet.2022.106058] [Medline: 35460807]
21. Xia W, Ding J, Yan Y, Chen F, Yan M, Xu X. Effectiveness of virtual reality technology in symptom management of patients at the end of life: a systematic review and meta-analysis. J Am Med Dir Assoc 2024;25(10):105086. [doi: 10.1016/j.jamda.2024.105086] [Medline: 38880120]
22. Mummah SA, Robinson TN, King AC, Gardner CD, Sutton S. IDEAS (Integrate, Design, Assess, and Share): a framework and toolkit of strategies for the development of more effective digital interventions to change health behavior. J Med Internet Res 2016;18(12):e317 [FREE Full text] [doi: 10.2196/jmir.5927] [Medline: 27986647]
23. Leng M, Sun Y, Li C, Han S, Wang Z. Usability evaluation of a knowledge graph-based dementia care intelligent recommender system: mixed methods study. J Med Internet Res 2023;25:e45788 [FREE Full text] [doi: 10.2196/45788] [Medline: 37751241]
24. Tovin MM, Wormley ME. Systematic development of standards for mixed methods reporting in rehabilitation health sciences research. Phys Ther 2023;103(11):pzad084. [doi: 10.1093/ptj/pzad084] [Medline: 37672215]
25. Brooke J. SUS—a quick and dirty usability scale. In: Jordan PW, Thomas B, Weerdmeester BA, McClelland IL, editors. Usability Evaluation in Industry. UK: Taylor & Francis; 1996:189-194.
26. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: adding an adjective rating scale. J Usability Stud 2009;4(3):114-123 [FREE Full text]
27. Wilson C. Interview Techniques for UX Practitioners: A User-Centered Design Method. San Francisco, California: Morgan Kaufmann; 2013.
28. Braun V, Clarke V. Is thematic analysis used well in health psychology? A critical review of published research, with recommendations for quality practice and reporting. Health Psychol Rev 2023;17(4):695-718. [doi: 10.1080/17437199.2022.2161594] [Medline: 36656762]
29. Dittli J, Meyer JT, Gantenbein J, Bützer T, Ranzani R, Linke A, et al. Mixed methods usability evaluation of an assistive wearable robotic hand orthosis for people with spinal cord injury. J Neuroeng Rehabil 2023;20(1):162 [FREE Full text] [doi: 10.1186/s12984-023-01284-8] [Medline: 38041135]
30. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2008;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
31. Hennink M, Kaiser BN. Sample sizes for saturation in qualitative research: a systematic review of empirical tests. Soc Sci Med 2022;292:114523 [FREE Full text] [doi: 10.1016/j.socscimed.2021.114523] [Medline: 34785096]
32. Lincoln YS, Guba EG. But is it rigorous? Trustworthiness and authenticity in naturalistic evaluation. New Dir Program Eval 1986;1986(30):73-84 [FREE Full text] [doi: 10.1002/ev.1427]
33. Creswell JW, Clark VLP. Designing and Conducting Mixed Methods Research. Washington DC, London: SAGE Publications; 2017.

XSL•FO
RenderX

34. Fetters MD, Curry LA, Creswell JW. Achieving integration in mixed methods designs-principles and practices. Health Serv Res 2013;48(6 Pt 2):2134-2156. [doi: 10.1111/1475-6773.12117] [Medline: 24279835]
35. Borsci S, Federici S, Lauriola M. On the dimensionality of the System Usability Scale: a test of alternative measurement models. Cogn Process 2009;10(3):193-197. [doi: 10.1007/s10339-009-0268-9] [Medline: 19565283]
36. Lewis JR, Sauro JF. The factor structure of the System Usability Scale. 2009 Presented at: Human Centered Design, First International Conference; July 19-24, 2009; San Diego, California. [doi: 10.1007/978-3-642-02806-9_12]
37. Li L, Yu F, Shi D, Shi J, Tian Z, Yang J, et al. Application of virtual reality technology in clinical medicine. Am J Transl Res 2017;9(9):3867-3880 [FREE Full text] [Medline: 28979666]
38. Langlet BS, Odegi D, Zandian M, Nolstam J, Södersten P, Bergh C. Virtual reality app for treating eating behavior in eating disorders: development and usability study. JMIR Serious Games 2021;9(2):e24998 [FREE Full text] [doi: 10.2196/24998] [Medline: 33847593]
39. Lorenz M, Brade J, Klimant P, Heyde C, Hammer N. Age and gender effects on presence, user experience and usability in virtual environments-first insights. PLoS One 2023;18(3):e0283565 [FREE Full text] [doi: 10.1371/journal.pone.0283565] [Medline: 36972245]
40. Coxon M, Kelly N, Page S. Individual differences in virtual reality: are spatial presence and spatial ability linked? Virtual Reality 2016;20(4):203-212. [doi: 10.1007/s10055-016-0292-x]
41. Chiu PL, Li H, Yap KY, Lam KC, Yip PR, Wong CL. Virtual reality-based intervention to reduce preoperative anxiety in adults undergoing elective surgery: a randomized clinical trial. JAMA Netw Open 2023;6(10):e2340588 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.40588] [Medline: 37906193]
42. El Mathari S, Shehadeh S, Zwaan WP, Boulidam N, Kuitert L, Twisk JWR, et al. The effect of virtual reality on postoperative anxiety and pain in patients following cardiac surgery: a randomized controlled trial. Eur J Cardiothorac Surg 2024;66(6):ezae440. [doi: 10.1093/ejcts/ezae440] [Medline: 39673743]
43. Spence C. Multisensory flavor perception. Cell 2015;161(1):24-35 [FREE Full text] [doi: 10.1016/j.cell.2015.03.007] [Medline: 25815982]
44. Mayer EA, Knight R, Mazmanian SK, Cryan JF, Tillisch K. Gut microbes and the brain: paradigm shift in neuroscience. J Neurosci 2014;34(46):15490-15496 [FREE Full text] [doi: 10.1523/JNEUROSCI.3299-14.2014] [Medline: 25392516]
45. Margolis KG, Cryan JF, Mayer EA. The microbiota-gut-brain axis: from motility to mood. Gastroenterology 2021;160(5):1486-1501. [doi: 10.1053/j.gastro.2020.10.066] [Medline: 33493503]
46. Agirman G, Hsiao EY. SnapShot: the microbiota-gut-brain axis. Cell 2021;184(9):2524-2524.e1 [FREE Full text] [doi: 10.1016/j.cell.2021.03.022] [Medline: 33930299]
47. Skvortsova A, Veldhuijzen DS, Kloosterman IE, Pacheco-López G, Evers AW. Food anticipatory hormonal responses: a systematic review of animal and human studies. Neurosci Biobehav Rev 2021;126:447-464. [doi: 10.1016/j.neubiorev.2021.03.030] [Medline: 33812978]
48. Mori H, Verbeure W, Schol J, Carbone F, Tack J. Gastrointestinal hormones and regulation of gastric emptying. Curr Opin Endocrinol Diabetes Obes 2022;29(2):191-199. [doi: 10.1097/MED.0000000000000707] [Medline: 35081068]
49. Wagner M, Probst P, Haselbeck-Köbler M, Brandenburg JM, Kalkum E, Störzinger D, et al. The problem of appetite loss after major abdominal surgery: a systematic review. Ann Surg 2022;276(2):256-269 [FREE Full text] [doi: 10.1097/SLA.0000000000005379] [Medline: 35129465]
50. Wennström B, Johansson A, Kalabic S, E-Son Loft AL, Skullman S, Bergh I. Patient experience of health and care when undergoing colorectal surgery within the ERAS program. Perioper Med (Lond) 2020;9:15 [FREE Full text] [doi: 10.1186/s13741-020-00144-6] [Medline: 32467753]
51. Nguyen TH, Ta NT, Dang AK, Nguyen TT, Dam VAT, Latkin CA, et al. A longitudinal assessment of appetite loss and nutritional care among postoperative patients in Vietnam. Front Nutr 2023;10:1008636 [FREE Full text] [doi: 10.3389/fnut.2023.1008636] [Medline: 37032772]
52. Martin L, Lagergren P. Risk factors for weight loss among patients surviving 5 years after esophageal cancer surgery. Ann Surg Oncol 2015;22(2):610-616. [doi: 10.1245/s10434-014-3973-2] [Medline: 25120247]
53. Ferrer-García M, Gutiérrez-Maldonado J, Pla-Sanjuanelo J, Vilalta-Abella F, Riva G, Clerici M, et al. A randomised controlled comparison of second-level treatment approaches for treatment-resistant adults with bulimia nervosa and binge eating disorder: assessing the benefits of virtual reality cue exposure therapy. Eur Eat Disord Rev 2017;25(6):479-490. [doi: 10.1002/erv.2538] [Medline: 28804985]
54. Liem DG, Mawas M, Keast RS. Evoked sensory stimulation of the eating environment, impacts feeling of presence and food desires in an online environment. Food Res Int 2023;167:112645. [doi: 10.1016/j.foodres.2023.112645] [Medline: 37087236]
55. Ferrer-Garcia M, Pla-Sanjuanelo J, Dakanalis A, Vilalta-Abella F, Riva G, Fernandez-Aranda F, et al. A randomized trial of virtual reality-based cue exposure second-level therapy and cognitive behavior second-level therapy for bulimia nervosa and binge-eating disorder: outcome at six-month followup. Cyberpsychol Behav Soc Netw 2019;22(1):60-68. [doi: 10.1089/cyber.2017.0675] [Medline: 30059240]

XSL•FO
RenderX

56. Marucci M, Di Flumeri G, Borghini G, Sciaraffa N, Scandola M, Pavone EF, et al. The impact of multisensory integration and perceptual load in virtual reality settings on performance, workload and presence. Sci Rep 2021;11(1):4831 [FREE Full text] [doi: 10.1038/s41598-021-84196-8] [Medline: 33649348]
57. Lo HHM, Ng M, Fong PYH, Lai HHK, Wang B, Wong SY, et al. Examining the feasibility, acceptability, and preliminary efficacy of an immersive virtual reality-assisted lower limb strength training for knee osteoarthritis: mixed methods pilot randomized controlled trial. JMIR Serious Games 2024;12:e52563 [FREE Full text] [doi: 10.2196/52563] [Medline: 39331525]
58. Creswell JW, Klassen AC, Plano CVL, Smith KC. Best Practices for Mixed Methods Research in the Health Sciences. Bethesda (Maryland): National Institutes of Health; 2013:541-545.
59. Bennett J, McDonald T, Lieblich S, Piecuch J. Perioperative rehydration in ambulatory anesthesia for dentoalveolar surgery. Oral Surg Oral Med Oral Pathol Oral Radiol Endod 1999;88(3):279-284. [doi: 10.1016/s1079-2104(99)70028-4] [Medline: 10503854]
60. Seok Y, Suh EE, Yu S, Park J, Park H, Lee E. Effectiveness of integrated education to reduce postoperative nausea, vomiting, and dizziness after abdominal surgery under general anesthesia. Int J Environ Res Public Health 2021;18(11):6124 [FREE Full text] [doi: 10.3390/ijerph18116124] [Medline: 34204087]
61. Li Y, Dou Z, Yang L, Wang Q, Ni J, Ma J. Oxycodone versus other opioid analgesics after laparoscopic surgery: a meta-analysis. Eur J Med Res 2021;26(1):4 [FREE Full text] [doi: 10.1186/s40001-020-00463-w] [Medline: 33422129]
62. Huang J, Lin J, Xiong Y, Wang Z, Zhu Y, Ye H, et al. Risk factors associated with postoperative discomfort after ambulatory strabismus surgery under general anesthesia. J Pain Res 2020;13:947-953 [FREE Full text] [doi: 10.2147/JPR.S249493] [Medline: 32440200]

Abbreviations

ED: eating disorder
IDEAS: Integrate, Design, Assess, and Share
MMR-RHS: Mixed Methods Reporting in Rehabilitation and Health Sciences
SUS: System Usability Scale
VR: virtual reality
VRMS-SFD: virtual reality multisensory sham-feeding device

Edited by A Coristine; submitted 08.04.25; peer-reviewed by T Vagg, A Gazis; comments to author 10.07.25; revised version received 21.08.25; accepted 16.09.25; published 08.10.25.

Please cite as:
Yuan X, Ye P, Liu X, Hu Y, Xu Q, Zhao W, Li K
Usability Evaluation of a Virtual Reality Multisensory Sham-Feeding Device for Patients Undergoing Fasting Periods for Colorectal Cancer Surgery: Mixed Methods Study
JMIR Serious Games 2025;13:e75641
URL: https://games.jmir.org/2025/1/e75641
doi:10.2196/75641
PMID:

©Xingzhu Yuan, Peiling Ye, Xinyue Liu, Yuehan Hu, Qin Xu, Wenyi Zhao, Ka Li. Originally published in JMIR Serious Games (https://games.jmir.org), 08.10.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Performance and Perceptions of Health Care Professionals Using an Immersive Virtual Reality Tool for Home Care Training: Observational Feasibility and Acceptability Study

José Mira[1,2], PhD; Javier Rios[3], RN; Eva Gil-Hernández[2], BSc; Nida Abed[2], RN; Vanesa Ribeiro-Neves[4], PhD; Clara Pérez-Esteve[2], BSc; Mercedes Guilabert-Mora[1], PhD; Almudena Arroyo[5], PhD; Maria Purificacion Ballester[6], PhD

[1]Universitat de Miguel Hernández d'Elx, Elche, Spain

[2]Fundación para el Fomento de la Investigación Sanitaria y Biomédica de la Comunitat Valenciana, Alicante, Spain

[3]Hospital San Juan de Dios del Aljarafe, Bormujos, Spain

[4]Escola Paulista de Enfermagem-Unifesp, Sao Paulo, Brazil

[5]Fundación San Juan de Dios. Escuela Universitaria de Enfermería y Fisioterapia San Juan de Dios, Universidad Pontificia Comillas, Bormujos, Spain

[6]Pharmacology Department. Universidad Católica de San Antonio de Murcia, Murcia, Spain

Corresponding Author:
José Mira, PhD
Universitat de Miguel Hernández d'Elx
Universidad
Elche, 03202
Spain
Phone: 34 966658984
Email: jose.mira@umh.es

Abstract

Background: Informal caregivers play a crucial role in home care and many lack formal training, potentially compromising patient safety. Immersive virtual reality (VR) offers an innovative approach to training by simulating real-life caregiving scenarios in a risk-free environment. Prior to implementation, the environments and the technique's feasibility and acceptability must be assessed by the professionals who will use it to train caregivers, establishing a performance benchmark based on experienced health care professionals.

Objective: This study aims to test feasibility and develop exploratory benchmarks and acceptability of immersive VR training for home caregiving tasks, using experienced professionals to establish a reference standard for execution quality.

Methods: This observational study was conducted in health care centers in Andalusia, the Valencian Community, and Madrid (Spain). A structured process was followed, including the identification of key home care tasks, the development of best practice guidelines, creation of immersive VR training materials, and the design of a performance evaluation rubric. Health care professionals (n=75) were recruited using a convenience sampling approach. They performed caregiving tasks in VR, and their performance was recorded and assessed using a standardized rubric, which included 205 predefined errors. Participants also completed a posttraining survey evaluating usability, comprehension, and perceived applicability to real-world caregiving.

Results: A total of 75 professionals participated, completing 257 caregiving simulations in a fully immersive VR environment. A total of 417 errors were identified (417/3142, 13.3% of the maximum number of predefined errors), with a mean average of 5.6 (SD 6.8) errors per participant. The most frequent errors occurred in medication management, insulin administration, diaper changing, broncho aspiration prevention, blood pressure monitoring, and hand hygiene. The perceived usefulness of VR training was rated 8.1 out of 10 points (SD 1.9), with 98.7% (74/75) of the participants stating that the time spent in the simulation was worthwhile and 85.3% (64/75) agreeing that the tasks were appropriately represented.

Conclusions: Immersive VR training for informal caregivers is a feasible and well-accepted approach, demonstrating high perceived usefulness among health care professionals. The study establishes a preliminary benchmark for home caregiving task execution, providing a basis for future research evaluating informal caregivers' performance and targeted training interventions to enhance patient safety. Further studies are needed to explore the long-term impact of VR training on caregiver competence and home care quality.

XSL•FO
RenderX

(JMIR Serious Games 2025;13:e75104) doi:10.2196/75104

KEYWORDS

patient safety; virtual reality; augmented reality; health care providers; informal caregivers

Introduction

Background

Caring for others is a core component of the human existence. Caring is as old as humanity itself; it is both a feeling and a need. This universal dynamic transcends cultures and contexts, forming the foundation of human connections and communities [1].

By 2030, the global population aged 60 years or older is projected to grow to 1.4 billion, and by 2050, this figure is expected to nearly double itself, reaching 2.1 billion [2]. Approximately 80% of these individuals will have 1 or more chronic conditions [3], and between 20% and 30% will experience some degree of disability [4].

Although the increase in life expectancy in recent years represents an unprecedented success for humanity, it also poses a significant challenge [5,6]: how can welfare states remain sustainable while ensuring health and social care for the entire population? The response in many countries has been to develop home care plans where informal caregivers play a central role, particularly for individuals living in remote or rural areas [7]. For example, on September 7, 2022, the European Care Strategy for Caregivers and Care Receivers was approved [8] to ensure quality, affordable, and accessible care services across European countries. This strategy aims to improve conditions for both care recipients and caregivers, whether professional or informal (such as family members), with a particular focus on enhancing home care support. Spain has also embraced this new approach, aligning with the growing care economy [9,10], through the *Estrategia Estatal para un Nuevo Modelo de Calidad de Cuidados en la Comunidad, Proceso de Desinstitucionalización* [11] [State Strategy for a New Model of Quality Care in the Community and the Deinstitutionalization Process] (2024-2030). Based on a person-centered care approach, these strategies advocate for a transformation in the care model for dependent individuals, promoting deinstitutionalization policies and enabling people to remain in their homes for as long as they choose [12].

Informal Caregiver Portrait

These caregivers are typically unpaid or engaged in precarious roles, predominantly occupied by women [13]. They provide care within the context of an existing relationship, such as a family member, friend, or neighbor, or are hired through social assistance programs or personal resources, often without formal health care qualifications. Care recipients frequently face various challenges, including disabilities, mental illnesses, chronic noncommunicable diseases (including rare conditions), terminal illnesses, substance use disorders, or age-related frailty. In Spain, nearly 82% of dependent individuals are cared for by family members, 90% of whom are women [14]. Most female caregivers are middle-aged, have limited formal education, receive no salary for their caregiving efforts, and are typically spouses, daughters, or daughters-in-law of the care recipients [5].

Today, caregiving at home has become increasingly demanding and complex, with expectations rising as populations age and caregiving needs for both older adult individuals and children grow [15]. This escalating burden not only impacts caregivers—resulting in emotional, physical, and financial strain—but also affects the health and well-being of care recipients. Evidence indicates that unmet caregiver needs and excessive burdens can directly influence recipient mortality rates [16], underscoring the critical importance of supporting caregivers in their role.

Patient Safety at Home

Ensuring patient safety in home care is a critical yet underexplored aspect of caregiving, particularly as more individuals choose to age or receive treatment at home. While home environments offer comfort and familiarity, they lack the structured safeguards of health care institutions, increasing the risk of unintentional errors that can compromise patient safety and adversely affect health outcomes. Home care patients face various risks of adverse events, including infections, adverse drug reactions, pressure ulcers, fall-related injuries, poisoning, and choking. The rate of medication errors when care is provided by an informal caregiver at home ranges from 2% to 33% of total medications, depending on the complexity of the dosage and the route of administration [17]. Recent estimates suggest that 2% of these medication errors have serious consequences for patients [18]. The risk increases when the patient has low health literacy, receives inadequate caregiving instructions, is cared for by multiple caregivers, or requires complex care. These factors are common causes of errors [8,9].

Errors in home care often led to additional health care resource use, such as emergency consultations or hospital readmissions, placing additional strain on public health care systems [19]. These preventable issues underscore the need for enhanced caregiver support and training to reduce the associated economic and social burden [18]. However, informal caregivers usually do not receive adequate or standardized training before assuming their caregiving responsibilities. At best, they may receive brief and incomplete instructions, which are rarely tailored to their specific caregiving situation or followed up to ensure comprehension and effective application.

The RealityCare Project [20] was developed to create accessible, cost-effective, and efficient training materials and procedures aimed at equipping informal caregivers with the necessary skills to enhance patient safety at home. Unlike traditional passive learning methods, this project leverages immersive and interactive technologies, such as virtual reality (VR), allowing caregivers to acquire essential competencies through active learning. By tailoring training to the specific needs of care recipients, RealityCare seeks to improve care quality, enhance patient safety, and strengthen caregiver preparedness in a highly

XSL•FO
RenderX

personalized manner. This initiative draws on a growing body of literature documenting the use of VR technologies in training health care professionals [21-24]. By integrating these innovative training solutions into informal caregiving, RealityCare aims to bridge the knowledge gap that often leads to errors in home care.

To implement this training effectively, it is essential to develop simulated environments where informal caregivers can practice and acquire skills by replicating real-world caregiving scenarios. As a first step in the development of these training environments, the situations in which informal caregivers are most exposed to risks should be identified. Since caregivers often have difficulties recognizing such risks and acknowledging everyday errors, the input and expertise of health care professionals are particularly valuable. Observing how professionals perform these tasks not only provides a useful benchmark to define the typical margin of error but also highlights which tasks are more complex to execute and therefore require targeted training. By focusing on these critical situations, the design of training scenarios can be better aligned with actual caregiving challenges, theoretically enhancing their effectiveness in preparing caregivers for safe and high-quality home care.

The objective of this study was to develop and evaluate the appropriateness, feasibility, use, and acceptability of VR-based training scenarios and procedures for informal caregivers, drawing on the expertise of health care professionals in tasks where informal caregivers are essential partners in supporting the care of chronically ill and dependent patients. These training environments aim to provide caregivers with a structured, immersive, and interactive learning experience, allowing them to develop the necessary skills to ensure safe and high-quality home care.

Methods

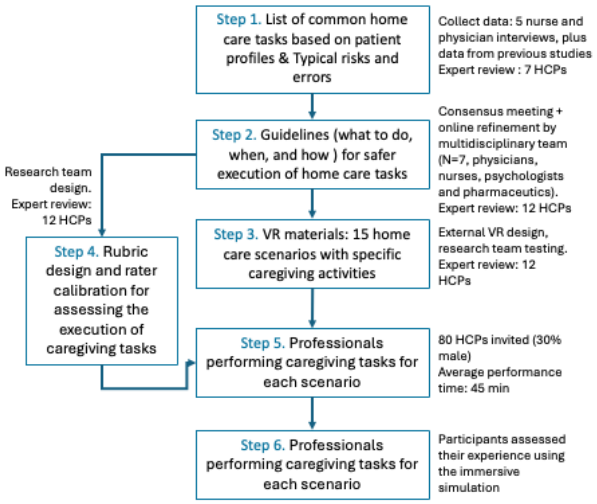
Study Design

This observational study followed a structured process to develop and iteratively refine fully immersive VR instructional materials for informal caregivers, conceived from the perspective of health care professionals who will use them in practice.

In a first phase (steps 1 and 2), a list of common home care tasks was compiled based on the most prevalent patient profiles, ensuring that the selected tasks accurately reflected real-world caregiving challenges. This phase also included the development of guidelines for the safe execution of home care tasks, integrating best practices to minimize errors and enhance patient safety. To support caregiver training, instructional materials were designed, allowing informal caregivers to recreate home care tasks in a virtual environment according to the specific needs of care recipients. For example, the scenarios included administering subcutaneous medication, preparing insulin, or safely mobilizing a patient. These materials provide an interactive and engaging learning experience, enabling caregivers to practice essential skills in a risk-free, controlled environment.

In the second phase (steps 3-6), a rubric for evaluating the execution of home care tasks was then developed, allowing for a structured and consistent assessment of performance and ensuring adherence to best practices. Finally, performance data from a sample of experienced health care professionals were collected within virtual scenarios to establish a reference standard (preliminary benchmark) for home care execution (Figure 1). This benchmark serves as a foundation for future studies assessing informal caregivers' performance, helping to identify key areas for training improvement.

Figure 1. Study flowchart. HCPs: health care providers; VR: virtual reality.



The main intervention in this study was the performance of caregiving tasks in a fully immersive VR environment. No confounders or effect modifiers were formally controlled in this exploratory study aimed at establishing a preliminary benchmark.

The study was conducted between November 27, 2023, and December 11, 2024, in health care centers located in Andalusia, the Valencian Community, and Madrid.

XSL•FO
RenderX

Ethical Considerations

The study adhered to ethical principles for research involving human participants, as established in the Declaration of Helsinki. Participation was entirely voluntary and was granted only after obtaining informed consent. The study protocol was approved by the research ethics committee of Hospital de Sant Joan (22/080, February 1, 2023). To ensure privacy and confidentiality, all data were anonymized prior to analysis, and no identifiable information was retained. During the VR sessions, the recordings captured only the participants' virtual hands, since their faces were fully covered by head-mounted displays (Meta headsets). Participants received a compensation of US $86 for their time.

Definitions

Informal caregivers are those individuals who provide care at home to patients with chronic illnesses, disabilities, or other long-term health or care needs, without having formal qualifications or academic training as health care professionals. These caregivers can be family members or close friends offering support within a personal relationship, but they can also include individuals who are not personally related to the care recipients. This kind of care can be either paid or unpaid. In turn, caregivers who provide professional medical care at home, such as visiting nurses, physiotherapists, social workers, psychologists, or other institutional workers performing medical procedures at home, as well as those offering formal home care assistance services [19], were excluded. *Dependent persons* (recipients) are people with multiple pathologies (a chronic process of more than 2 years of evolution, such as cardiometabolic diseases, respiratory disease, chronic pain, dementia, or reduced mobility) and polypharmacy (using 5 or more drugs daily). *Right care* refers to assisting in the basic and instrumental activities of daily living for persons who require them to promote their well-being meanwhile being safe and respectful of the person's dignity. *Caregiving error* refers to an avoidable and incorrect action on the part of the caregiver that may or could cause harm to the recipient or provide no benefit to their well-being.

Setting

The study was conducted in public and publicly contracted health care centers within the Spanish national health system, located in Alicante, Seville, Madrid, and Granada in Spain. The participating centers comprised primary care, home care, and rehabilitation services that regularly interact with informal caregivers, ensuring ecological validity for the evaluation.

Description of the Study Process

This study followed a structured 6-step process.

Step 1: Identification of Common Home Care Tasks

A list of essential caregiving tasks was compiled based on different patient profiles, usually seen in outpatient and primary care consultations, particularly those with chronic conditions or functional dependence. The focus was placed on frequent care activities required in home settings to ensure that the selected tasks accurately reflected real-world caregiving demands. Input was gathered from 5 nurses and physicians working in home care units and primary care services with home assistance, as well as from previous studies [25,26]. This list was then reviewed by 7 health care professionals, who prioritized typical home care scenarios where informal caregivers require specific training to ensure safe care delivery.

Step 2: Development of Safe Execution Guidelines

Best practice guidelines were established by a multidisciplinary team (n=7) composed of physicians, nurses, psychologists, and pharmacists with the goal of minimizing errors and enhancing patient safety. The initial content was defined during an in-person consensus meeting and further refined through subsequent web-based exchanges, which allowed the team to specify step-by-step instructions for each task. These guidelines were then reviewed and refined as needed by a panel of 12 nursing and medical professionals with extensive experience in home care for patients with diverse chronic conditions (expert panel on home care).

Step 3: Creation of Immersive and Virtual Reality Training Materials

Fully immersive and VR-based instructional content. The first step was the creation of a virtual home environment, complete with different rooms to simulate realistic caregiving settings. In addition, virtual patients were designed, representing both male and female individuals, to reflect diverse care recipients. Once the virtual environment was established, the best practice care guidelines were integrated into the VR system, transforming them into interactive elements within the virtual world. These guidelines were embedded as real-time prompts, decision-making tasks, and scenario-based interventions that caregivers could engage with during training. This approach allows informal caregivers to apply and execute caregiving tasks dynamically, receiving feedback and guidance as they perform each intervention in a risk-free, controlled setting. These scenarios were tested to verify their functionality by 5 members of the research team and by the expert panel on home care, which collaborated in the design of the project's training materials. This procedure also allowed us to ensure the face validity of both the scenarios and the caregiving tasks embedded within them through an iterative qualitative consensus process, whereby every issue raised by the experts was discussed and resolved before the materials were considered finalized. During this process, adjustments were made to address difficulties encountered in navigating the VR environment and to enhance the realism of the scenarios. These refinements ensured that the training materials closely reflected real-world challenges and maximized their relevance and use for subsequent caregiver training. As a result of this development process, a total of 15 scenarios were available for this study. Participants in this step were randomly assigned to the caregiving scenarios they had to perform in a fully immersive environment, serving both as a functionality check and as a means to evaluate performance. Multimedia Appendix 1 includes the scenarios.

Step 4: Design of an Evaluation Rubric

Standardized rubrics for each scenario were created to assess the execution of home care tasks, enabling a structured evaluation of caregiver performance. The rubrics were designed

XSL•FO
RenderX

by the research team and reviewed by the expert panel on home care, on the basis of errors reported by experienced professionals previously, focusing on those considered most characteristic and clinically relevant, as well as on the most serious risks for patient safety. These errors were listed, hierarchized, and reformulated into clear statements to facilitate consistent understanding and application by evaluators. A total of 205 predefined errors could be coded across the 15 scenarios using the rubrics, representing the range of potential deviations from safe practice that evaluators were instructed to identify. To ensure comprehensiveness and relevance, the rubrics were reviewed by 4 external professionals, not affiliated with the research team, all of whom had substantial experience in patient care and clinical training. It was then tested in practical applications with several cases to confirm that it met the requirements for adequate care delivery in alignment with the study objectives. Three nurses trained in the use of rubrics subsequently applied it to evaluate the performance of participants in this study. Before data collection, nurse reviewers independently coded the same set of randomly selected video recordings; discrepancies were then discussed in a consensus meeting to refine decision rules and exemplars for each error type. Interrater agreement was quantified using Cohen κ. Two calibration cycles were conducted to ensure high consistency in ratings. Multimedia Appendix 2 includes the rubrics used.

Step 5: Assessment of Professional Performance in Virtual Scenarios

Experienced health care professionals were enrolled in simulated virtual environment to establish a reference standard (preliminary benchmark) for execution quality. The underlying premise was that the performance of informal caregivers should be assessed against this benchmark, with the expectation that their execution would approximate the standards demonstrated by experienced professionals. All VR sessions were conducted under the supervision of trained personnel to ensure participants' safety and to prevent accidents. Before starting, participants received a standardized introduction to the VR equipment, including instructions on the use of the head-mounted displays and controllers, and were given a few minutes of practice to familiarize themselves with the environment, taking advantage of the rapid learning curve of the system. From the pool of available scenarios, participants were randomly assigned up to 3, which they were asked to complete by performing caregiving tasks for a virtual patient or by solving problem-based situations, such as the "House of Horrors," which simulated common challenges in home medication management.

The list of potential errors predefined in the rubric was not disclosed to the participants. As experienced health care professionals, they were already familiar with the correct procedures for each task and were instructed to perform them accordingly within the VR environment. Errors were subsequently identified and coded based on the participants' performance. An exception was the "House of Horrors" scenario, in which the task explicitly required participants to detect and correct common mistakes typically encountered in home medication management.

The average session lasted approximately 45 minutes, which included an initial period for familiarization with the VR equipment, the virtual environment, and the execution of the designated caregiving tasks. The sessions were recorded by the VR devices. Using standardized rubrics, a primary reviewer assessed each participant's video recordings; when uncertainty arose, a second reviewer cross-checked the case, and any remaining discrepancies were resolved by consensus. This process ensured that all errors were consistently identified and documented.

Step 6. Assessment of the VR Approach by the Participants

Finally, participants evaluated their experience using the immersive simulation through 8 direct questions. The assessment focused on ease of participation, the perceived value of the time spent in the simulation, the attractiveness of the virtual environment, and their ability to adapt to the tool. In addition, participants assessed whether the instructional elements, such as signs and audio, helped them understand the objectives of each scenario, whether the tasks were appropriately represented, and whether the training improved their execution of caregiving tasks. Finally, they evaluated whether the training provided knowledge and skills applicable to their professional practice. The survey was specifically designed by the research team, based on standardized criteria for this type of evaluative instrument [27,28].

Technical Features of the VR Environment

Participants primarily interacted with the virtual environment using their own hand movements, which were tracked by the head-mounted display to simulate realistic home care situations. Although handheld controllers were available, their use was limited; they were mainly used in specific scenarios, such as the "House of Horrors," where participants had to point to and select visual responses displayed on the screen. Interaction was restricted to visual and auditory feedback, without additional haptic stimulation.

Real-time prompts were provided during the performance of tasks. Participants also received initial instructions and were expected to carry out caregiving activities autonomously. As an example, in the hand hygiene scenario, participants were asked to open the tap, apply soap, and wash their hands while following the World Health Organization's Five Moments for Hand Hygiene. On the screen, they received step-by-step visual instructions, and in some cases, a short video demonstration illustrated the correct procedure before they practiced it. This ensured that all participants followed the same standardized sequence across sessions.

The scenarios were designed to allow repeated practice: participants could perform the same sequence as many times as needed until they felt confident that the task had been completed correctly. There was no fixed time limit, but all participants followed identical sequences in each scenario, which ensured comparability and standardization. Sessions were video recorded, but only the care task attempt marked by the participant as valid served as the reference for rubric-based error coding.

XSL•FO
RenderX

Participants Using Virtual Scenarios

Eighty experienced health care professionals were invited to be involved in this study. This sample size was determined pragmatically, reflecting the exploratory nature of the study, consistent with recommendations for pilot or feasibility studies where sample sizes are often based on practical considerations [29]. Participants were recruited from 10 participating health care centers through an open invitation to staff involved in home visits and in conducting health literacy sessions for informal caregivers. Inclusion criteria required that at least 30% of the sample were male, in line with the gender distribution typically observed in the health care sector. Eligible participants were physicians or nurses with a minimum of 7 years of professional experience and all working in primary care, home care, or units where caregivers are trained to care for patients at home (eg, rehabilitation), with regular contact with patients with chronic conditions. Basic demographic data, including sex, age, years of professional experience, and workplace setting, were collected to ensure that participants met the inclusion criteria and that the sample was balanced in terms of gender distribution. Participants were allocated to scenarios using simple random assignment. Their participation involved using VR devices to perform caregiving tasks for virtual patients, and they were compensated for their time. Multimedia Appendix 3 shows the VR-based environment used.

Equipment

The VR simulations were run using Meta head–mounted displays. Particularly, Meta Quest 2 headset was used. VR sessions used stand-alone Meta Quest 2 head-mounted displays with inside-out 6DoF tracking and native hand-tracking; controllers were used only when required by a scenario. Headsets were cleaned between uses and fitted with disposable hygiene covers, and all sites used an identical setup to ensure comparability.

Outcomes

The primary outcome was the number and type of caregiving errors committed by participants during VR scenarios, coded using the standardized rubric. Secondary outcomes included participants perceived usefulness, acceptability, and applicability of VR training, assessed through a postsimulation survey.

Statistical Analysis

Interrater reliability among reviewers was assessed using Cohen κ to ensure consistency in performance scoring before applying the rubric in the actual study context. The number of errors identified using the rubric was summed, resulting in an error count per participant. Errors were then categorized based on the type of care provided. The rubric evaluations were complete, with no missing data. Descriptive statistics, including totals, means, SDs, percentages, and proportions per scenario, were calculated to summarize participants' performance across different caregiving tasks. Frequency distributions were used to categorize the number of errors made by the participants. For selected questions on perceived impact, odds ratios (OR) with 95% CIs were calculated based on chi-square 2×2 tables comparing "yes" and "no" responses. The Exact Binomial Test was used to assess the perceived usefulness of the VR training. For these proportions, 95% CIs were calculated using the exact binomial method (Yes rate [95% CI]). These analyses were exploratory, aimed at identifying patterns and trends.

In addition, two complementary analyses were conducted: (1) the 15 caregiving scenarios were grouped into 3 thematic categories for comparison of error rates using chi-square tests and pairwise comparisons with Bonferroni adjustment: administration and medication management (House of Horrors; Subcutaneous Drug Administration, including Insulin, Glucagon, Heparin, Morphine, and Derivatives), daily care (Diaper Change; Daily Hygiene and Correct Selection of Elements in a Hygiene Process; Patient Transfers to Armchair, Bathroom, or Shower from Bed or Wheelchair; Hand Hygiene; Patient with Orthosis; and Care for the Caregiver), and patient monitoring and prevention (Prevention of Bronchial Aspiration in Patients with Dysphagia; Patient with Heart Failure: Weight Changes, Diet Adjustments; Blood Pressure Monitoring; and Prevention of Pressure Ulcers), and (2) professionals were divided into 2 groups according to the number of errors committed (≤3 vs >3) to compare the overall assessment of the VR training received using the Wilcoxon rank sum test with continuity correction. Statistical significance was considered with a *P* value threshold of .05. All statistical analyses were performed using R software (version 4.4.1; R Core Team) [30]. This study was conducted following the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) [31] guidelines to ensure transparency and quality in reporting observational research.

Results

Overview

A total of 75 health care professionals participated in this study (participation rate of 93.7%), collectively performing 257 VR-based caregiving simulations. Meanwhile, 19 out of 75 (25.3%) were males and 56 out of 75 (74.7%) were females. The participants had a mean age of 37.6 (SD 11.5) years. All participants were employed in units with responsibilities that included regular contact with informal caregivers as part of their professional practice. Those who declined to participate did so because they were unable to allocate the necessary time to complete the VR exercise. Interrater reliability among nurse reviewers was high (Cohen κ between 0.88 and 1.00), supporting consistent rubric-based error coding.

A total of 417 errors were recorded, representing 13.3% (417/3142) of the maximum number of predefined errors that could be coded across 257 evaluated scenarios. This corresponds to a mean average of 5.6 errors (SD 6.8) per professional across all tasks. The distribution of errors per scenario is detailed in Table 1.

XSL•FO
RenderX

Table 1. Distribution of errors.

Scenario	Number of predefined error types per scenario	Participants	Total observed errors (n), with mean (SD) per participant	Error rate within scenario (% of predefined errors)	Proportion of total errors across all scenarios (% of 417)
Administration and medication management	54	84	151 (1.8, 2.5)	3.3	36.2
Medication errors at home (House of Horrors)	8	41	65 (1.6, 1.4)	19.8	15.6
Insulin	12	30	73 (2.3, 3.6)	20.3	17.5
Glucagon	12	6	4 (0.7, 1.6)	5.6	1.0
Heparin	10	4	6 (1.5, 1.3)	15.0	1.4
Morphine and derivatives	12	3	3 (1.0, 1.7)	8.3	0.7
Daily care	116	126	226 (1.5, 2.8)	1.5	54.2
Hand hygiene	8	61	60 (1.0, 1.9)	12.3	14.4
Diaper change	19	31	154 (3.9, 4.1)	26.1	36.9
Daily hygiene and correct selection of elements in a hygiene process	35	13	9 (0.7, 1.4)	2.0	2.2
Patient with orthosis	9	8	0 (0.0, 0.0)	0.0	0.0
Care for the caregiver	17	7	0 (0.0, 0.0)	0.0	0.0
Patient transfers to armchair, bathroom, or shower from bed or wheelchair	28	6	3 (0.5, 0.8)	1.8	0.7
Patient monitoring and prevention	35	47	40 (0.9, 1.0)	2.4	9.6
Prevention of bronchial aspiration in patients with dysphagia	7	15	11 (0.7, 0.5)	10.5	2.6
Blood pressure monitoring	10	26	25 (1.0, 1.3)	9.6	6.0
Patient with heart failure: weight changes, and diet adjustments	11	2	1 (0.5, 0.7)	4.5	0.2
Prevention of pressure ulcers	7	4	3 (0.8, 1.0)	10.7	0.7
Total	205	257	417 (1.6, 2.5)	13.3	100.0

The highest number of errors, with more than 10 occurrences, was observed in several key caregiving tasks. In medication management at home, frequent mistakes included failing to check medication expiration dates, improper storage, and not ensuring a child-friendly first aid kit. In insulin administration, errors involved inadequate hand hygiene, both in terms of duration and technique. During diaper changing, professionals commonly failed to maintain proper hand hygiene, with errors such as insufficient duration and incorrect maneuvers, rolling the diaper toward the groin instead of away from it, cleaning from the side rather than front to back, and allowing continuous moisture in a specific area. In the prevention of broncho aspiration in patients with dysphagia, one of the main mistakes was positioning the patient facing the food instead of correctly aligning their posture. In blood pressure monitoring, errors were related to incorrect recording and interpretation of average readings between both arms. Finally, in hand hygiene, several professionals neglected essential steps such as rubbing between the fingers in a C-shape, cleaning fingernails by rubbing them against the palm, and washing the thumb properly.

The lowest number of errors recorded was 0, achieved by 18 out of 75 (24.0%) participants, while 57 out of 75 (76.0%) made at least 1 error. At the other extreme, a single professional recorded a maximum of 30 errors. Overall, the distribution showed a higher concentration of participants with few errors, as most made between 0 and 4 mistakes, accounting for 61.3% (46/75) of the total participants. This indicates that most of the participants committed only a small number of errors (Table 2).

XSL•FO
RenderX

Table 2. Frequency of errors committed by participants grouped by ranges.

Errors, n	Participants, n	Percentage	Cumulative percentage
0-4	46	61.3	61.3
5-9	17	22.7	84.0
10-14	6	8.0	92.0
15-19	2	2.7	94.7
20-24	3	4.0	98.7
25-30	1	1.3	100.0

Of the 57 professionals who made at least 1 error, 36 out of 57 (63.2%) felt that the simulation improved their execution of the trained task (OR 1.71, 95% CI 1.00-2.94; *P*=.046). In contrast, only 19 out of 57 (33.3%) indicated that the training received applied to their job as well (OR 0.50, 95% CI 0.29-0.87; *P*=.012; Table 3).

Table 3. Perceived impact of the simulation on task execution and job applicability for professionals who made errors.

Participants who made at least 1 error (N=57)	Yes, n (%)	No, n (%)	Chi-square (*df*)	*P* value[a]	OR (95% CI)
Did the simulation improve my execution of the trained task?	36 (63.2)	21 (36.8)	3.95 (1)	.046	1.71 (1.00-2.94)
Did I receive training that I can apply to my job?	19 (33.3)	38 (66.7)	6.34 (1)	.01	0.50 (0.29-0.87)

[a]Chi-square test.

The perceived usefulness of this approach was rated with a mean average score of 8.1 out of 10 points (SD 1.9). Table 4 shows the results of the evaluated aspects.

Table 4. The perceived usefulness of the virtual reality training.

Evaluated aspect	Yes, n (%)	No, n (%)	Yes rate (95% CI)	*P* value[a]
Was my participation in the study easy?	70 (93.3)	5 (6.7)	84.3%-98.2%	<.001
Was the time dedicated to the simulation worthwhile?	74 (98.7)	1 (1.3)	92.8%-99.9%	<.001
Did I find the simulation attractive?	75 (100)	0 (0.0)	95.2%-100%	N/A[b]
Was I able to adapt to the tool?	65 (86.7)	10 (13.3)	77.9%-92.2%	<.001
Did I understand the objective of each situation thanks to the signs and audio?	59 (78.7)	16 (21.3)	67.5%-87.5%	<.001
Did I see the tasks appropriately represented in the simulation?	64 (85.3)	11 (14.7)	75.6%-92.0%	<.001
Did the simulation improve my execution of the trained task?	46 (61.3)	29 (38.7)	49.9%-72.6%	.06
Did I receive training that I can apply to my job?	53 (70.7)	22 (29.3)	59.2%-80.9%	.001

[a]Exact Binomial Test.

[b]N/A: not applicable.

Comparison of Error Rates by Scenario Type

When the scenarios were grouped into 3 thematic categories (administration and management of medication, daily care, and patient monitoring or prevention), significant differences were observed in the error rate relative to the total possible errors (χ^2_2=24.01; *P*<.001). After applying pairwise comparisons, the medication management category showed a higher proportion of errors compared with both daily care (*P*<.001) and patient monitoring or prevention (*P*<.001), while no significant differences were found between the latter 2 (*P*=.60; Table 5).

Table 5. Error rate by type of scenario (chi-square *P*<.001).

Scenario type	Possible errors	Observed errors	Error rate, %
Administration and medication management	836	151	18.1
Daily care	1883	226	12.0
Patient monitoring and prevention	415	40	9.6
Total	3134	417	13.3

XSL•FO
RenderX

Perceived Usefulness and Error Burden

We compared overall usefulness ratings between professionals with ≤3 errors and those with >3 errors (Wilcoxon rank sum test). No statistically significant difference was observed (W=830; P=.09). However, the lower-error group tended to rate usefulness higher (mean 8.53, SD 1.97; median 9) than the higher-error group (mean 7.84, SD 1.72; median 8; Table 6). The distribution of perceived usefulness scores according to the number of errors committed is illustrated in the boxplot (Figure 2).

Figure 2. Perceived usefulness of the tool according to number of errors committed.



Table 6. Overall usefulness ratings by error burden (P=.09 determined using the Wilcoxon rank sum test with continuity correction).

Professional group	N	Mean (SD)	Median (IQR)
≤3 errors	45	8.53 (1.97)	9 (8-10)
>3 errors	30	7.84 (1.72)	8 (7-9)

Discussion

Principal Findings

This study documents the design, development, and appraisal of immersive educational materials aimed at enhancing the patient safety competencies of informal caregivers in home settings, which health care professionals could use to reduce patient safety risks when care is provided by nonprofessional caregivers (eg, family members). These materials address the patient safety challenges associated with the complex clinical profiles and care needs of care recipients at home, aligning with public policies promoting active aging at home and the growing care economy [32,33].

This approach enabled us to establish preliminary performance standards, providing a reference for evaluating informal caregivers in future studies. It also informed the development of training materials through realistic, evidence-based scenarios.

The materials developed constitute a preliminary library of VR resources. These professional-derived benchmarks and scenarios may inform the future design and evaluation of training programs for informal caregivers; direct testing with informal caregivers is required. There was also a trend whereby participants who perceived that they performed better on the task rated the VR training more highly. Although these findings are relevant, replication in larger and more diverse samples is required before they can be generalized.

These results are consistent with previous studies suggesting some benefits of VR in caregiver education, particularly in improving confidence and competence in real-world caregiving scenarios [34,35]. In line with this, simulation-based training programs have been shown to strengthen caregivers' confidence and perceived competence in carrying out essential care activities [36]. Participants also tend to regard these programs as practical and beneficial, which reinforces their overall acceptability [37].

Key Findings and Implications

This study highlights that human error is unavoidable, even among trained professionals performing basic caregiving tasks. Recurring mistakes, especially around the WHO Five Moments for Hand Hygiene [38], reflect persistently suboptimal compliance that can improve with training [39-41].

The study findings indicate a high level of acceptability of the training program among health care professionals, reinforcing the feasibility of implementing VR-based learning tools in caregiver education. However, the scenarios designed for informal caregivers did not always present the expected personal challenges, which may have limited the activation of deeper learning processes. This makes one of the key lessons learned from this study the necessity of incorporating progressive difficulty levels or adaptive challenges to enhance engagement and knowledge retention.

XSL•FO
RenderX

The RealityCare approach focuses on improving patient safety under the care of informal caregivers, addressing both care recipient outcomes and caregiver well-being [42]. In this sense, it is also important to recognize that not all errors have the same implications for patient safety. For instance, medication errors carry a higher risk of severe adverse outcomes than errors in tasks such as diaper changing, which primarily affect patient comfort and dignity. Addressing both dimensions is essential for designing meaningful VR training scenarios that reflect real-world caregiving priorities.

In parallel with the impact that errors may have on patients, these events also increase the emotional burden on caregivers in ways that have received little attention to date. This burden stems from the sense of responsibility associated with making a mistake that directly affects the person under their care—particularly when, as is the case in the majority of situations, a close family relationship is involved. It highlights the need for structured training to reduce stress and mitigate the risk of caregivers experiencing the "double victim" phenomenon [43], where the emotional burden of making mistakes in care impacts both the caregiver and the patient. This aligns with previous studies emphasizing that empowering caregivers with structured training can reduce emotional distress and enhance caregiving quality [44,45].

Another major area of concern was medication management in home care, where frequent errors were observed in tasks such as checking expiration dates, proper storage, and safe administration techniques [46]. Previous studies have shown that medication errors in home settings occur more frequently than assumed, often due to low health literacy, lack of training, and inconsistent caregiving instructions [18,47]. This reinforces the need for targeted training interventions, including problem-solving VR scenarios, where caregivers are required to navigate medication safety challenges to improve their decision-making skills.

The Role of VR in Caregiver Education

New technologies, such as VR, may be reshaping caregiver education by shifting from passive to active learning approaches [48,49]. Studies on VR training for health care professionals suggest that interactive and immersive simulations may enhance skill acquisition and confidence levels [50]. This study is consistent with recent research indicating that training programs integrating disease-specific information, caregiving techniques, and emotional support strategies enable caregivers to deliver higher-quality care without compromising their well-being [51-53].

Practical Implications

VR–based simulation is increasingly applied across multiple sectors [54], particularly in health care [55,56]. Although the experience with non–health care professional profiles (including informal caregivers) is still limited, the potential of this approach to enhance caregiving skills is considerable and applicable across a wide range of contexts [57-60]. These data suggest that VR's impact varies by adaptive scenario design, structured debriefing, progressive difficulty, real-time (including haptic) feedback, appropriate session frequency, curricular integration, cost-effective implementation, and user acceptance [55,56,61-63]. Future work should test whether VR training improves performance on familiar caregiving tasks and addresses navigation or realism issues through the same optimizations to enhance effectiveness and acceptability.

Limitations and Challenges

Several limitations should be considered when interpreting these findings. First, the generalizability of the findings may be constrained by the recruitment strategy, since health care professionals with relevant responsibilities were invited to participate through a nonrandom, convenience approach. Second, training duration varied according to participants' learning curves, which may have influenced overall performance. In addition, prior familiarity with digital devices was not controlled, potentially affecting the ease of use of VR technologies among participants. Third, several measures were taken to minimize potential bias. Selection bias was reduced by recruiting physicians and nurses from multiple centers with a balanced profile of sex and experience. Evaluation bias was addressed through a standardized rubric, independent external reviewers, and interrater reliability checks. All participants received brief familiarization with the VR equipment to limit bias related to digital literacy. Nevertheless, residual bias cannot be completely ruled out and should be considered when interpreting results. Another limitation relates to the random assignment of caregiving scenarios, which may have resulted in participants encountering tasks outside their primary area of expertise, leading to higher error rates in unfamiliar situations. Moreover, feedback on the tool itself was collected using yes/no questions rather than Likert-type scales, which may have limited the granularity and nuance of participants' responses.

In addition, some errors may reflect limitations of the VR environment itself. The immersive environment used may not have fully replicated real-world handling. This factor could have affected user interaction and performance, introducing a source of error that is independent of the participants' actual caregiving competence.

Future studies should stratify by prior caregiving experience, compare informal caregivers against the professional benchmark established here, and include longitudinal designs to assess real-world outcomes and patient safety events. In addition, personalized VR—using adaptive difficulty, real-time (including haptic) feedback, and artificial intelligence–supported performance tracking—should be evaluated to enhance learning effectiveness and engagement.

The choice to include health care professionals in this first phase was deliberate, as their involvement was essential to validate the materials, ensure that the scenarios were feasible, and provide a preliminary benchmark against which informal caregivers' performance can later be evaluated. Moreover, professionals are the ones who will ultimately organize and deliver training; without their endorsement and willingness to implement the program, it would not reach informal caregivers. Nevertheless, future studies must directly involve informal caregivers, who are the ultimate target of the intervention and whose perceptions of usefulness, acceptability, and engagement are crucial for its effectiveness. This study therefore represents

XSL•FO
RenderX

the necessary first stage in a sequential process, with subsequent research focusing on informal caregivers as the direct trainees.

Finally, these results should not be directly transferred to the training of informal caregivers without adaptation. Unlike professionals, informal caregivers will require detailed instructions to compensate for their lack of prior training and experience, which may influence the applicability and effectiveness of VR-based training.

Conclusions

This study indicates that immersive VR training is feasible and acceptable among health care professionals, representing an initial step toward implementation and opening the way for future testing with informal caregivers. The use of VR-based simulations to establish a preliminary benchmark for execution quality allows for a systematic assessment of caregiving performance, using experienced health care professionals as a benchmark. By defining realistic, evidence-based expectations for home caregiving tasks, these findings lay the groundwork for future research evaluating the performance of informal caregivers.

VR-based caregiver training has the potential to bridge the knowledge gap in home care, offering safe, structured, and engaging learning experiences. However, further research is required to validate its long-term impact on caregiving competence, patient safety, and informal caregiver well-being.

Acknowledgments

The collaborative group of participants supported the development of materials.

Data Availability

The data from this study will be shared upon receiving a reasonable request.

Funding

This research was funded by MICIN (Ministry of Science and Innovation)/AEI (State Research Agency)/DOI 10.13039/501100011033 (Digital Object Identifier) (TED 2021-130383B-I00), and by the European Union “NextGenerationEU”/PRTR (Recovery, Transformation and Resilience Plan). During the execution of this study and composition of this manuscript, EG received funding through a Predoctoral Fellowship for Research Training in Health from ISCIII, supported by the European Union NextGenerationEU and the Recovery, Transformation, and Resilience Plan from the Spanish Government (reference FI22/00277).

Authors' Contributions

JM, EG-H, and MPB contributed to conceptualization; JM, EG-H, AA, and MPB contributed to methodology; EG-H, MPB, VR-N, NA, and JR participated in investigation; CP-E contributed to data curation; JM, AA, and MPB participated in supervision; JM contributed to writing original draft preparation; EG-H, MPB, VR-N, NA, and JR contributed to writing, review, and editing; and JM contributed to funding acquisition. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Caregiving scenarios.
[DOCX File , 16 KB - games_v13i1e75104_app1.docx]

Multimedia Appendix 2
Example of rubric.
[DOCX File , 27 KB - games_v13i1e75104_app2.docx]

Multimedia Appendix 3
Virtual reality environment (examples).
[MOV File , 35778 KB - games_v13i1e75104_app3.mov]

References

1. Vos EE, Hilderink HBM, de Bruin SR, van der Beek AJ, Proper KI. The working informal caregiver model: a mixed methods approach to explore future informal caregiving by working caregivers. Sustainability 2022;14(6):3519. [doi: 10.3390/su14063519]
2. World population ageing 2017. World Population Ageing. 2017. URL: https://www.un.org/en/development/desa/population/publications/pdf/ageing/WPA2017_Report.pdf [accessed 2025-10-24]

XSL•FO
RenderX

3. Watson KB, Wiltz JL, Nhim K, Kaufmann RB, Thomas CW, Greenlund KJ. Trends in multiple chronic conditions among us adults, by life stage, behavioral risk factor surveillance system, 2013-2023. Prev Chronic Dis 2025;22:E15 [FREE Full text] [doi: 10.5888/pcd22.240539] [Medline: 40245168]
4. Chowdhury SR, Chandra Das D, Sunna TC, Beyene J, Hossain A. Global and regional prevalence of multimorbidity in the adult population in community settings: a systematic review and meta-analysis. EClinicalMedicine 2023;57:101860 [FREE Full text] [doi: 10.1016/j.eclinm.2023.101860] [Medline: 36864977]
5. Scott AJ. The economics of longevity—an introduction. J Econ Ageing 2023;24:100439. [doi: 10.1016/j.jeoa.2022.100439]
6. Global health estimates: life expectancy and healthy life expectancy. World Health Organization. 2025. URL: https://www.who.int/data/gho/data/themes/mortality-and-global-health-estimates/ghe-life-expectancy-and-healthy-life-expectancy?utm_source=chatgpt.com [accessed 2025-08-23]
7. Martinez-Lacoba R, Pardo-Garcia I, Escribano-Sotos F. Aging, dependence, and long-term care: a systematic review of employment creation. Inquiry 2021;58:469580211062426 [FREE Full text] [doi: 10.1177/00469580211062426] [Medline: 34913376]
8. European care strategy for caregivers and care receivers. European Commission. 2022. URL: https://employment-social-affairs.ec.europa.eu/news/european-care-strategy-caregivers-and-care-receivers-2022-09-07_en [accessed 2025-08-23]
9. Peña-Longobardo LM, Oliva-Moreno J. The economic value of non-professional care: a Europe-wide analysis. Int J Health Policy Manag 2022;11(10):2272-2286 [FREE Full text] [doi: 10.34172/ijhpm.2021.149] [Medline: 34814681]
10. Morrow E, Lynch M. The economics of home support services in Ireland: exploring complex issues of healthcare sustainability and aging populations. Front Public Health 2025;13:1602617 [FREE Full text] [doi: 10.3389/fpubh.2025.1602617] [Medline: 40529697]
11. State Strategy for a New Model of Care in the Community: A Process of Deinstitutionalization (2024–2030). Ministry of Social Rights, Consumer Affairs, and Agenda 2030. Madrid: Ministry of Social Rights, Consumer Affairs, and Agenda 2030; 2024. URL: https://www.observatoriodelainfancia.es/oia/esp/documentos_ficha.aspx?id=8647
12. Barslund M, De Wispelaere F, Lenaerts K, Schepers W, Fries-Tersch E. Policies for long-term carers. In: Publication for the committee on Employment and Social Affairs. Luxembourg: Policy Department for Economic, Scientific and Quality of Life Policies, European Parliament; 2021.
13. Zarzycki M, Morrison V, Bei E, Seddon D. Cultural and societal motivations for being informal caregivers: a qualitative systematic review and meta-synthesis. Health Psychol Rev 2023;17(2):247-276 [FREE Full text] [doi: 10.1080/17437199.2022.2032259] [Medline: 35081864]
14. Alcañiz-Garrán M, García-Sanjuán S, Ramos-Pichardo JD, Sanjuán-Quiles A, Montejano-Lozoya R. The experiences of older individuals providing care to older dependents: a phenomenological study in Spain. PLoS One 2021;16(8):e0255600 [FREE Full text] [doi: 10.1371/journal.pone.0255600] [Medline: 34351991]
15. McKenna O, Fakolade A, Cardwell K, Langlois N, Jiang K, Pilutti LA. Towards conceptual convergence: a systematic review of psychological resilience in family caregivers of persons living with chronic neurological conditions. Health Expect 2022;25(1):4-37 [FREE Full text] [doi: 10.1111/hex.13374] [Medline: 34676951]
16. Schulz R, Beach SR, Czaja SJ, Martire LM, Monin JK. Family caregiving for older adults. Annu Rev Psychol 2020;71:635-659 [FREE Full text] [doi: 10.1146/annurev-psych-010419-050754] [Medline: 31905111]
17. Parand A, Garfield S, Vincent C, Franklin BD. Carers' medication administration errors in the domiciliary setting: a systematic review. PLoS One 2016;11(12):e0167204 [FREE Full text] [doi: 10.1371/journal.pone.0167204] [Medline: 27907072]
18. Gil-Hernández E, Ballester P, Guilabert M, Sánchez-García A, García-Torres D, Astier-Peña MP, Cobos-Vargas, et al. Enhancing safe medication use in home care: insights from informal caregivers. Front Med (Lausanne) 2024;11:1494771 [FREE Full text] [doi: 10.3389/fmed.2024.1494771] [Medline: 39564498]
19. Abed N, Knezevic B, Sousa P, Tella S, Srulovici E, Strametz R, et al. Defining informal caregivers by their characteristics safety roles and training needs in Europe. Sci Rep 2025;15(1):24375 [FREE Full text] [doi: 10.1038/s41598-025-08540-y] [Medline: 40628820]
20. Mira JJ, Ballester P, Gil-Hernández E, Sambrano Valeriano L, Álvarez Gómez E, Olier Garate C, et al. Safe care and medication intake provided by caregivers at home: reality care study protocol. Healthcare (Basel) 2023;11(15):2190 [FREE Full text] [doi: 10.3390/healthcare11152190] [Medline: 37570430]
21. Kokorelias KM, Chiu M, Paul S, Zhu L, Choudhury N, Craven C, et al. Use of virtual reality and augmented reality technologies to support resilience and skill-building in caregivers of persons with dementia: a scoping review. Cureus 2024;16(7):e64082. [doi: 10.7759/cureus.64082] [Medline: 39114214]
22. Albright L, Ko W, Buvanesh M, Haraldsson H, Polubriaginof F, Kuperman GJ, et al. Opportunities and challenges for augmented reality in family caregiving: qualitative video elicitation study. JMIR Form Res 2024;8:e56916 [FREE Full text] [doi: 10.2196/56916] [Medline: 38814705]
23. Gerup J, Soerensen CB, Dieckmann P. Augmented reality and mixed reality for healthcare education beyond surgery: an integrative review. Int J Med Educ 2020;11:1-18 [FREE Full text] [doi: 10.5116/ijme.5e01.eb1a] [Medline: 31955150]

XSL•FO
RenderX

24. Gerup J, Soerensen CB, Dieckmann P. Augmented reality and mixed reality for healthcare education beyond surgery: an integrative review. Int J Med Educ 2020;11:1-18 [FREE Full text] [doi: 10.5116/ijme.5e01.eb1a] [Medline: 31955150]
25. Pérez-Jover V, Mira J, Carratala-Munuera C, Gil-Guillen V, Basora J, López-Pineda A, et al. Inappropriate use of medication by elderly, polymedicated, or multipathological patients with chronic diseases. Int J Environ Res Public Health 2018;15(2):310 [FREE Full text] [doi: 10.3390/ijerph15020310] [Medline: 29439425]
26. Mira JJ. How to reduce medication errors in patients over the age of 65? Expert Opin Pharmacother 2021;22(18):2417-2421. [doi: 10.1080/14656566.2021.1947241] [Medline: 34187264]
27. Hyzy M, Bond R, Mulvenna M, Bai L, Dix A, Leigh S, et al. System usability scale benchmarking for digital health apps: meta-analysis. JMIR Mhealth Uhealth 2022;10(8):e37290 [FREE Full text] [doi: 10.2196/37290] [Medline: 35980732]
28. Lee AT, Ramasamy RK, Subbarao A. Understanding psychosocial barriers to healthcare technology adoption: a review of tam technology acceptance model and unified theory of acceptance and use of technology and utaut frameworks. Healthcare (Basel) 2025;13(3):250 [FREE Full text] [doi: 10.3390/healthcare13030250] [Medline: 39942440]
29. Teresi JA, Yu X, Stewart AL, Hays RD. Guidelines for designing and evaluating feasibility pilot studies. Med Care 2022;60(1):95-103 [FREE Full text] [doi: 10.1097/MLR.0000000000001664] [Medline: 34812790]
30. R: a language and environment for statistical computing. R Core Team. 2024. URL: https://www.r-project.org/ [accessed 2025-03-27]
31. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. PLoS Med 2007;4(10):e296 [FREE Full text] [doi: 10.1371/journal.pmed.0040296] [Medline: 17941714]
32. Ayoubi-Mahani S, Eghbali-Babadi M, Farajzadegan Z, Keshvari M, Farokhzadian J. Active aging needs from the perspectives of older adults and geriatric experts: a qualitative study. Front Public Health 2023;11:1121761 [FREE Full text] [doi: 10.3389/fpubh.2023.1121761] [Medline: 37397742]
33. Shahrestanaki SK, Rafii F, Najafi Ghezeljeh T, Farahani MA, Majdabadi Kohne ZA. Patient safety in home health care: a grounded theory study. BMC Health Serv Res 2023;23(1):467 [FREE Full text] [doi: 10.1186/s12913-023-09458-9] [Medline: 37165357]
34. Huang Y, Ho KHM, Christensen M, Wong DW, Wang S, Su JJ, et al. Virtual reality-based simulation intervention for enhancing the empathy of informal caregivers of people with dementia: a mixed-methods systematic review. Int J Ment Health Nurs 2024;33(2):241-258. [doi: 10.1111/inm.13240] [Medline: 37817470]
35. Siew AL, Chew EYH, Chan E, Siow EKC. Mastering the art of caregiving: instructional approaches to teaching healthcare-related procedural skills to informal caregivers-an integrative review. J Adv Nurs 2025. [doi: 10.1111/jan.16944] [Medline: 40159713]
36. Wang J, Li Q, Cui J, Tu S, Deng Z, Yang R, et al. Effectiveness of virtual reality on the caregiving competence and empathy of caregivers for elderly with chronic diseases: a systematic review and meta-analysis. J Nurs Manag 2023;2023:5449955. [doi: 10.1155/2023/5449955] [Medline: 40225657]
37. Hur Y, Hickman RL. Use and impact of simulation in family caregiver education: a systematic review. West J Nurs Res 2024;46(2):143-151. [doi: 10.1177/01939459231218956] [Medline: 38124438]
38. WHO guidelines on hand hygiene in health care: first global patient safety challenge clean care is safer care. World Health Organization. 2009. URL: https://www.who.int/publications/i/item/9789241597906 [accessed 2023-08-25]
39. Armstrong-Novak J, Juan HY, Cooper K, Bailey P. Healthcare personnel hand hygiene compliance: are we there yet? Curr Infect Dis Rep 2023;25:1-7 [FREE Full text] [doi: 10.1007/s11908-023-00806-8] [Medline: 37361491]
40. Musu M, Lai A, Mereu N, Galletta M, Campagna M, Tidore M, et al. Assessing hand hygiene compliance among healthcare workers in six intensive care units. J Prev Med Hyg 2017;58(3):E231-E237 [FREE Full text] [Medline: 29123370]
41. Ojanperä H, Kanste OI, Syrjala H. Hand-hygiene compliance by hospital staff and incidence of health-care-associated infections, Finland. Bull World Health Organ 2020;98(7):475-483 [FREE Full text] [doi: 10.2471/BLT.19.247494] [Medline: 32742033]
42. Gemito L, Alves E, Moreira J, Marques MF, Caldeira E, Ferreira R, et al. Programmes addressed to informal caregivers' needs: a systematic literature review. Geriatrics (Basel) 2024;9(3):71 [FREE Full text] [doi: 10.3390/geriatrics9030071] [Medline: 38920427]
43. Strametz R. Second and double victims—achievements and challenges in ensuring psychological safety of caregivers. J Healthc Qual Res 2023;38(6):327-328 [FREE Full text] [doi: 10.1016/j.jhqr.2023.05.001] [Medline: 37236907]
44. Wu X, Liang Y, Zheng B, Wang H, Ning M, Zheng H, et al. Care stress in caregivers of disabled stroke patients: a cross-sectional survey. Ann Palliat Med 2020;9(4):2211-2220 [FREE Full text] [doi: 10.21037/apm-20-1369] [Medline: 32692235]
45. Riffin C, Van Ness PH, Wolff JL, Fried T. Multifactorial examination of caregiver burden in a national sample of family and unpaid caregivers. J Am Geriatr Soc 2019;67(2):277-283 [FREE Full text] [doi: 10.1111/jgs.15664] [Medline: 30452088]
46. Alsaeed D, Jamieson E, Gul MO, Smith FJ. Challenges to optimal medicines use in people living with dementia and their caregivers: a literature review. Int J Pharm 2016;512(2):396-404. [doi: 10.1016/j.ijpharm.2015.12.050] [Medline: 26721728]
47. Look KA, Stone JA. Medication management activities performed by informal caregivers of older adults. Res Social Adm Pharm 2018;14(5):418-426 [FREE Full text] [doi: 10.1016/j.sapharm.2017.05.005] [Medline: 28528023]

XSL•FO
RenderX

48. Herur-Raman A, Almeida ND, Greenleaf W, Williams D, Karshenas A, Sherman JH. Next-generation simulation—integrating extended reality technology into medical education. Front. Virtual Real 2021;2:115. [doi: 10.3389/frvir.2021.693399]
49. Borges LADC, Almeida RGDS, Barboza ES, Arruda GOD. Simulation training of caregivers at hospital discharge of patients with chronic diseases: an integrative review. Rev Bras Enferm 2023;76(6):e20230043 [FREE Full text] [doi: 10.1590/0034-7167-2023-0043] [Medline: 38055488]
50. Lindeman DA, Kim KK, Gladstone C, Apesoa-Varano EC. Technology and caregiving: emerging interventions and directions for research. Gerontologist 2020;60(Suppl 1):S41-S49. [doi: 10.1093/geront/gnz178] [Medline: 32057082]
51. Calvó-Perxas L, Vilalta-Franch J, Litwin H, Turró-Garriga O, Mira P, Garre-Olmo J. What seems to matter in public policy and the health of informal caregivers? A cross-sectional study in 12 European countries. PLoS One 2018;13(3):e0194232 [FREE Full text] [doi: 10.1371/journal.pone.0194232] [Medline: 29518147]
52. Uhm KE, Jung H, Woo MW, Kwon HE, Oh-Park M, Lee BR, et al. Influence of preparedness on caregiver burden, depression, and quality of life in caregivers of people with disabilities. Front Public Health 2023;11:1153588 [FREE Full text] [doi: 10.3389/fpubh.2023.1153588] [Medline: 37564425]
53. Bhattacharyya KK, Liu Y, Gothe NP, Fauth EB. Mind-body practice and family caregivers' subjective well-being: findings from the midlife in the United States (MIDUS) study. Gerontol Geriatr Med 2023;9:23337214231185912 [FREE Full text] [doi: 10.1177/23337214231185912] [Medline: 37435003]
54. Xie B, Liu H, Alghofaili R, Zhang Y, Jiang Y, Lobo FD, et al. A review on virtual reality skill training applications. Front Virtual Real 2021;2:645153. [doi: 10.3389/frvir.2021.645153]
55. Nassar AK, Al-Manaseer F, Knowlton LM, Tuma F. Virtual reality (VR) as a simulation modality for technical skills acquisition. Ann Med Surg (Lond) 2021;71:102945 [FREE Full text] [doi: 10.1016/j.amsu.2021.102945] [Medline: 34840738]
56. Foronda CL, Gonzalez L, Meese MM, Slamon N, Baluyot M, Lee J, et al. A comparison of virtual reality to traditional simulation in health professions education: a systematic review. Simul Healthc 2024;19(1S):S90-S97. [doi: 10.1097/SIH.0000000000000745] [Medline: 37651101]
57. Clay CJ, Budde JR, Hoang AQ, Gushchyan A. An evaluation of the effectiveness of immersive virtual reality training in non-specialized medical procedures for caregivers and students: a brief literature review. Front Virtual Real 2024;5. [doi: 10.3389/frvir.2024.1402093]
58. Abayasiri RAM, Padilha Lanari Bo A, Dick TJM, Baghaei N. Influence of virtual reality illusions on balance performance and immersive user experience in young adults: a within-subject experimental study. JMIR Serious Games 2025;13:e70376 [FREE Full text] [doi: 10.2196/70376] [Medline: 40576982]
59. Tan J, Karim M, Tamanna R, Kim S, Patel B. Comparing learning outcomes of virtual reality (VR) simulators using haptic feedback versus box trainer (BT) in laparoscopic training: a systematic review and meta-analysis. Cureus 2025;17(2):e78910. [doi: 10.7759/cureus.78910] [Medline: 40091919]
60. Mergen M, Graf N, Meyerheim M. Reviewing the current state of virtual reality integration in medical education—a scoping review. BMC Med Educ 2024;24(1):788 [FREE Full text] [doi: 10.1186/s12909-024-05777-5] [Medline: 39044186]
61. Tusher HM, Mallam S, Nazir S. A systematic review of virtual reality features for skill training. Tech Know Learn 2024;29(2):843-878. [doi: 10.1007/s10758-023-09713-2]
62. Wang C, Liu M. Digital interventions for cognitive dysfunction in patients with stroke: systematic review and meta-analysis. J Med Internet Res 2025;27:e73687 [FREE Full text] [doi: 10.2196/73687] [Medline: 40705404]
63. Lu W, Shi M, Liu L, Wang S, Deng W, Ma Y, et al. Effect of virtual reality-based therapies on lower limb functional recovery in stroke survivors: systematic review and meta-analysis. J Med Internet Res 2025;27:e72364 [FREE Full text] [doi: 10.2196/72364] [Medline: 40737641]

Abbreviations

OR: odds ratio
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology
VR: virtual reality

XSL•FO
RenderX

Edited by A Coristine; submitted 28.03.25; peer-reviewed by B Mockeviciene, T Mühling, MDP Beristain-Colorado; comments to author 22.08.25; revised version received 21.10.25; accepted 22.10.25; published 20.11.25.

Please cite as:
Mira J, Rios J, Gil-Hernández E, Abed N, Ribeiro-Neves V, Pérez-Esteve C, Guilabert-Mora M, Arroyo A, Ballester MP
Performance and Perceptions of Health Care Professionals Using an Immersive Virtual Reality Tool for Home Care Training: Observational Feasibility and Acceptability Study
JMIR Serious Games 2025;13:e75104
URL: https://games.jmir.org/2025/1/e75104
doi:10.2196/75104
PMID:

©José Mira, Javier Rios, Eva Gil-Hernández, Nida Abed, Vanesa Ribeiro-Neves, Clara Pérez-Esteve, Mercedes Guilabert-Mora, Almudena Arroyo, Maria Purificacion Ballester. Originally published in JMIR Serious Games (https://games.jmir.org), 20.11.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Efficacy of a Mobile Serious Game (SwaziYolo) for Increasing HIV Risk Perception: Randomized Controlled Trial

Bhekumusa Wellington Lukhele[1], MPH, PhD; Mac Delay[1], MS; Fortunate Shabalala[2], MPH, PhD; Mfundi Motsa[3], MPH, PhD; Alexander Kay[4], MD; Christina El Saaidi[5], BDS, MPH, CSM; Masako Kihara[6], PhD; Masahiro Kihara[6], MD, PhD; Katia J Bruxvoort[1], MPH, PhD

[1]University of Alabama at Birmingham, Birmingham, AL, United States

[2]University of Eswatini, Mbabane, Swaziland

[3]Ministry of Health, Mbabane, Swaziland

[4]Baylor College of Medicine, Houston, TX, United States

[5]City of San Antonio Metropolitan Health District, San Antonio, TX, United States

[6]Kyoto University, Kyoto, Japan

Corresponding Author:
Bhekumusa Wellington Lukhele, MPH, PhD
University of Alabama at Birmingham
1720 2nd Ave South Birmingham
Birmingham, AL 35294-0022
United States
Phone: 1 205975 ext 1048
Email: bhekie.l@gmail.com

Abstract

Background: Eswatini has one of the highest HIV prevalence rates worldwide (24.8% among people aged ≥15 years), with unprotected heterosexual transmission accounting for more than 90% of new HIV infections in the country. Low HIV risk perception is known to influence risk behavior. Mobile phone technology is growing rapidly, offering opportunities for technology-driven interventions to improve HIV risk perception and prevention.

Objective: We aimed to design and test a serious game to increase HIV risk perception and intention to engage in protective HIV behaviors among young people in Eswatini.

Methods: Our team developed SwaziYolo, a smartphone-based, interactive, educational story game that places the player in the role of a young adult looking for love in Eswatini's capital city. We conducted the Serious Games HIV Prevention Trial (SGPrev-Trial), a 4-week, 2-arm, unblinded, 1:1 randomized controlled trial of SwaziYolo among people aged between 18 and 29 years in Eswatini. The primary outcome was HIV risk perception using a 10-item and subset 8-item Perceived Risk of HIV Infection Scale (PRHS). We used modified intention-to-treat and per-protocol difference-in-difference (DID) estimation to compare the change between groups in the mean PHRS scores before and after intervention.

Results: Of the 380 participants in the study, 122 (64.2%) in the control group and 119 (62.6%) in the intervention group completed the follow-up, and 95 (79.8%) played the game. In the modified intention-to-treat analyses, no significant differences between groups were observed for the 8-item PRHS (DID: mean 1.1, SD 0.72; *P*=.13) and the 10-item PRHS (DID: mean 1.3, SD 0.80; *P*=.12) scores. In the per-protocol analyses, HIV risk perception increased significantly among participants who played the game (8-item DID: mean 1.6, SD 0.74; *P*=.04 and 10-item DID: mean 1.8, SD 0.83; *P*=.03). Nearly all (94/95, 98.9%) participants strongly agreed or agreed that they would recommend SwaziYolo to their peers.

Conclusions: SwaziYolo was acceptable and increased HIV risk perception among young people in Eswatini who self-selected to play the game. More research is needed to improve and evaluate the SwaziYolo intervention.

Trial Registration: UMIN Clinical Trial Registry UMIN000021781; https://center6.umin.ac.jp/cgi-open-bin/ctr_e/ctr_view.cgi?recptno=R000025103

(JMIR Serious Games 2025;13:e70333) doi:10.2196/70333

XSL•FO
RenderX

KEYWORDS

HIV prevention; youth; digital health; gamification; mobile health; mHealth; Swaziland; Eswatini

Introduction

The HIV epidemic remains a major public health concern worldwide. In 2023, there were approximately 1.3 million new infections, 3 times more than the estimate for 2025 [1]. The countries with the highest HIV prevalence are in southern Africa, including Eswatini. A 2021 national survey in Eswatini found that the prevalence of HIV among people aged 15 years or older was 24.8% [2]. Unprotected heterosexual transmission accounts for more than 90% of all new infections in Eswatini [3], and multiple concurrent sexual partnerships, low HIV testing, and low HIV risk perception are the key drivers of transmission in the country [3-5]. Antiretroviral treatment aids in prevention. Pre-exposure prophylaxis (PrEP), postexposure prophylaxis, and medical male circumcision have helped reduce HIV incidence. However, there is an urgent need for innovative interventions to prevent HIV infection, particularly among young people [1].

Serious games, commonly referred to as gaming for education, motivation, or behavioral change rather than mere entertainment [6-8], are promising strategies for behavior change interventions. Several studies adopting serious games or gamified interventions have observed effective behavioral change outcomes. For example, a randomized controlled trial (RCT) called PlayForward implemented in the United States found improvements in attitudes toward sexual health and an increase in sexual health knowledge [9]. Other RCTs conducted in Kenya, Tanzania, and South Africa found that serious games were associated with significant improvements in self-efficacy, sexual health knowledge and literacy, sexual risk communication, significant changes in behavioral intention to reduce and avoid HIV risk, and higher cognitive affirmative attitudes [10-13].

Low HIV risk perception is associated with an increased risk of HIV infection [14-16], which supports the need for behavior change interventions to focus on increasing HIV risk perception. A large longitudinal study found that increased HIV risk perception was associated with higher odds of increasing protective behavior, such as condom use [17]. Most behavioral theories include the concept of risk perception, which is believed to be necessary for behavior change [18-20]. Previous studies in Eswatini indicate a low-risk perception, particularly among young adults [21].

Despite the potential of serious games in both high-income and low- or middle-income countries to impact perceptions and behaviors, potentially leading to improvements in health outcomes, few studies have examined serious games for HIV prevention, and none have examined serious games for HIV prevention in Eswatini. To address this gap, we designed and tested a serious game to increase HIV risk perception and intention to engage in protective HIV behaviors among young people in Eswatini.

Methods

Study Design and Participants

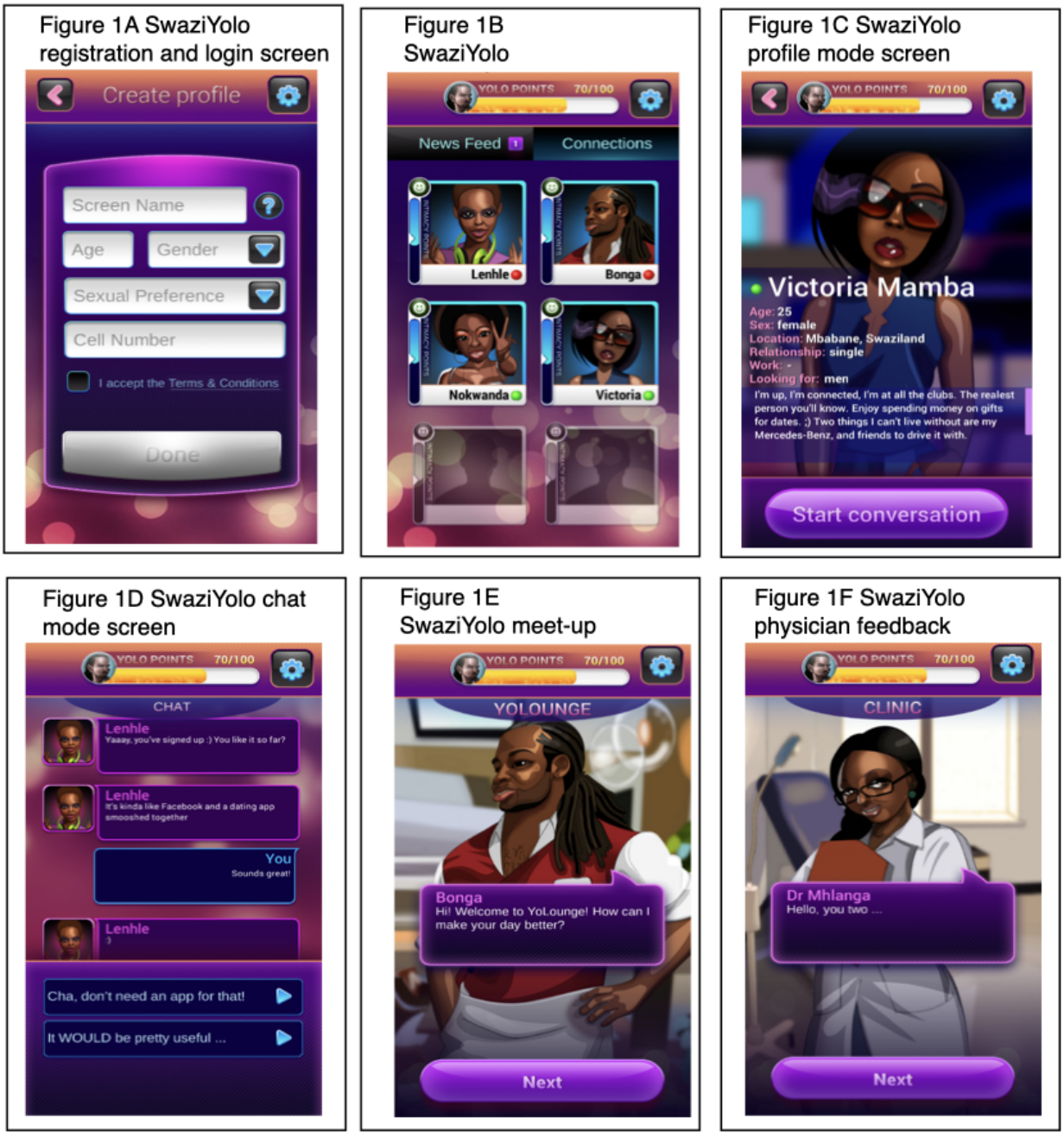
The Eswatini (formerly Swaziland) Serious Games HIV Prevention Trial (SGPrev-Trial) was a 4-week, 2-arm, unblinded, 1:1 RCT of SwaziYolo, a serious games mobile health intervention that we developed for young people in Eswatini. Details of the intervention mapping steps, game design, game screenshots, theoretical framework adopted, study design, ethical considerations, and HIV risk perception scales are detailed in our published trial protocol [6]. Briefly, the trial was powered to detect a difference in the change in HIV risk perception scores before and after the intervention, between the intervention and control groups at 4 weeks. Assuming a moderate effect size of 0.477 based on the study by Chu et al [22], with an α of .05 (2-tailed) and a β of .20, the required total sample size was 146 per group. To account for an anticipated 30% loss to follow-up [23], we increased the sample size by 30%, resulting in 190 participants per group for a final sample size of 380 participants.

Intervention Description

On the basis of our formative research, SwaziYolo was an interactive, educational story game for mobile phones that placed the player in the role of a young adult looking for love in Eswatini's capital city. Players were tasked with making important choices regarding relationships and sexual health, as previously described [6]. SwaziYolo incorporates elements of serious games, such as immersion, role-playing, and a scripted, predetermined storyline that varies based on players' previous selections [6]. The first part of the game was set in an imaginary social network called SwaziYolo, resembling common social media apps. Here, players registered (registration and login screen), viewed profile pictures (potential love interest connections screen) and profiles of potential love interests (profile mode screen), and had web-based chats (chat mode screen) with various characters. The second part of the game took place in various made-up locations around the capital city of Eswatini, such as nightclubs and cafes (meet up at a club screen), where players would go on dates (meetups). In both parts of the game, players were required to make choices to advance conversations or storylines with a friend or love interest (chat mode screen). These choices influenced the opinions and behavior of other characters, as well as the characters' health and safety. At the end of the game, feedback was provided based on the player's choices (feedback from a physician at a clinic screen) [6]. Figure 1 shows selected screenshots of SwaziYolo, including the registration and login screen, the potential love interests' connections screen, the profile mode screen, chat mode screen, meet up screen, and feedback screen.

XSL•FO
RenderX

Figure 1. SwaziYolo screenshots.



Recruitment, Enrollment, and Randomization

Participants were recruited from April 28, 2017, to July 06, 2017. The game was downloadable from the project website and Google Play Store. We also distributed the SwaziYolo app through social media, through popular mobile phone shops, and at universities in major cities around Eswatini [6]. Once the game was downloaded, potential participants were directed to the study information page on our website, and eligibility screening was conducted through a self-administered questionnaire. The eligibility criteria were as follows: aged between 18 and 29 years, owned a smartphone running an Android-based operating system, had WhatsApp, lived in Eswatini at the time of the study, and could adequately grant informed consent. Potential participants who responded yes to all eligibility screening questions (detailed further in the published protocol [6]) were directed to the informed consent page.

We generated 380 5-digit random numbers using Microsoft Excel "=RAND ()" function to serve as one-time passcodes (OTPs). After generating the OTPs, we checked for and removed duplicates. Once participants provided written informed consent, they were each randomized by a computer-generated algorithm that assigned an OTP indicating group assignment. Participants in the intervention group received OTPs that immediately unlocked the SwaziYolo intervention after completing the baseline survey. In contrast, participants in the control group

XSL•FO
RenderX

(waitlist group) received OTPs that would only unlock the intervention after completing a follow-up survey after 4 weeks from the trial registration date.

Data Collection

Participants completed a brief web-based self-administered questionnaire at baseline (before intervention) and then 4 weeks after completion of the baseline survey (after intervention), during which those randomized to the intervention group were able to play the game. The questionnaire included a self-administered, structured, web-based questionnaire created based on a review of the literature and our formative work. For example, questions relating to sociodemographic characteristics were adopted from the 2007 Swaziland Demographic Health Survey, and those related to sexual behaviors and intention to change behaviors were developed from our previous work in Eswatini [21,24].

Perceived Risk of HIV Infection Scale Outcomes

The primary outcome of interest, HIV risk perception, was measured through the validated Perceived Risk of HIV Infection Scale (PRHS) [20]. We used a 10-item index and an 8-item index (a subset of the 10-item index), as shown in Table 1. Although the 10-item scale has been found to perform as well as the 8-item scale [20], we used both scales to capture additional dimensions of risk perception that may be present in our population. Each Likert-type question was scored (see Table 1 for details) such that the final score was the sum of all 8 or 10 items. We also assessed secondary outcomes, including the number of sexual partners in the last 30 days, condom use in the last 30 days, HIV testing in the last 30 days or intention to test for HIV, intention to know the partner's HIV status, intention to reduce multiple sexual partnerships, and intention to use a condom in the next sexual encounter. Details of the administered questionnaire are shown in Multimedia Appendix 1.

Table 1. The 8-item and 10-item Perceived Risk of HIV Infection Scale [20].

	Item content	Response options
Items on both the 8-item and 10-item Perceived Risk of HIV Infection Scales		
Item 1	What is your gut feeling about how likely you are to get infected with HIV?	Extremely unlikely (0), very unlikely (1), somewhat likely (2), very likely (3), or extremely likely (4)
Item 2	I worry about getting infected with HIV	None of the time (0), rarely or some of the time (1), a moderate amount of time (2), a lot of the time (3), or all the time (4)
Item 3	I am sure I will NOT get HIV infected	Strongly disagree (5), disagree (4), somewhat disagree (3), somewhat agree (2), agree (1), or strongly agree (0)
Item 4	I feel vulnerable to HIV infection	Strongly disagree (0), disagree (1), somewhat disagree (2), somewhat agree (3), agree (4), or strongly agree (5)
Item 5	There is a chance, no matter how small, that I could get HIV	Strongly disagree (0), disagree (1), somewhat disagree (2), somewhat agree (3), agree (4), or strongly agree (5)
Item 6	I think my chances of getting infected with HIV are	Zero (0), almost zero (1), small (2), moderate (3), large (4), or very large (5)
Item 7	Picturing myself getting HIV is something I find	Very hard to do (0), hard to do (1), easy to do (2), or very easy to do (3)
Item 8	Getting HIV is something I have	Never thought about (0), rarely thought about (1), thought about some of the time (2), or thought about often (3)
Items on the 10-item Perceived Risk of HIV Infection Scale only		
Item 9	Getting HIV is something I am	Not concerned about (0), a little concerned about (1), moderately concerned about (3), concerned about a lot (4), or extremely concerned about (5)
Item 10	I feel I am unlikely to get infected with HIV	Strongly disagree (5), disagree (4), somewhat disagree (3), somewhat agree (2), agree (1), or strongly agree (0)

Statistical Methods

All data analyses were conducted using SAS 9.4. We described the baseline characteristics of individuals in each group and calculated the mean 8-item and 10-item scores and the proportions with each secondary outcome. For the 8-item and 10-item outcomes, we then conducted difference-in-difference (DID) estimation for the change in prescores and postscores comparing the intervention and control groups using mixed effects linear regression. The primary analysis was based on a modified intention-to-treat (mITT) approach due to postrandomization exclusion of participants lost to follow-up, those with missing data, and those who reported a known HIV positive status at baseline and were unlikely to benefit from the intervention. We also conducted per-protocol (PP) analyses with those who played the game in the intervention group, and subgroup analyses using the same DID approach among participants with multiple sexual partners, those without multiple sexual partners, those who reported condomless sex in the past 30 days, and those who did not report condomless sex in the past 30 days.

To compare pre-post differences in secondary outcomes between the intervention and control groups, we used general logistic

XSL•FO
RenderX

regression models with the postresponse (yes or no or yes, no, or not sure) as the outcome, the intervention versus control group as the predictor, and the preresponse as a confounder.

Ethical Considerations

This study was reviewed and approved by the Swaziland Scientific and Ethical Committee (MH/599C/FWA00015267/IRB0009688) and registered at the UMIN Clinical Trial Registry (UMIN000021781). The protocol [6] was written in accordance with the CONSORT-EHEALTH (Consolidated Standards for Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) checklist [25]. Informed consent was obtained from all participants online before enrollment in the study. Privacy and confidentiality were maintained through the deidentification of data and storage in encrypted servers. Participants' phone numbers were deleted from the data. A lottery-based incentive with a 1:100 chance of receiving US $20 was given to all participants after baseline data collection. Identification of individual participants was not possible in the manuscript or the Multimedia Appendices.

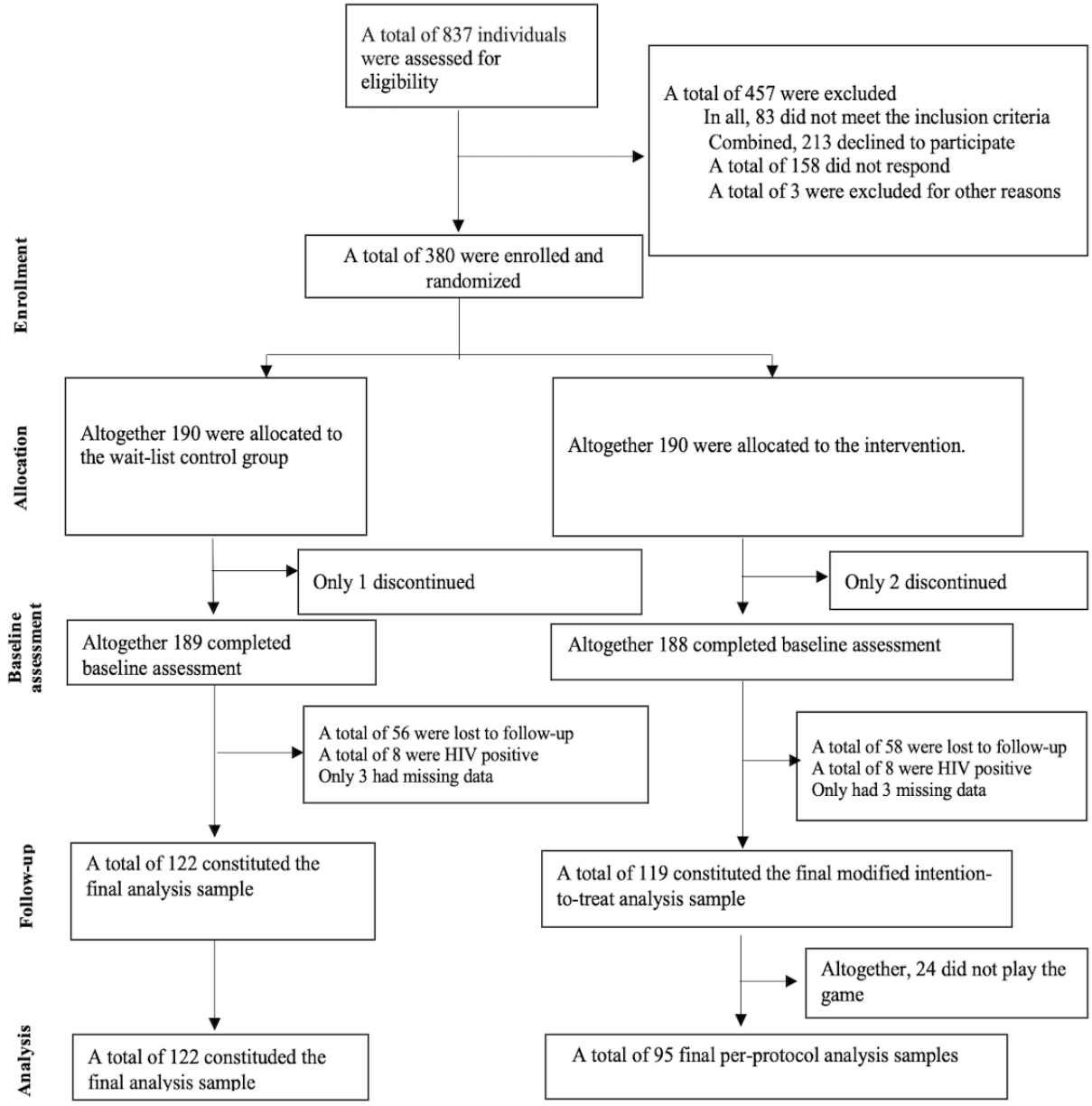
Results

Overview

The last participant completed the 4-week follow-up period on August 10, 2017. In total, 837 potential study participants were assessed for eligibility, and 457 (54.5%) declined to participate or did not meet the inclusion criteria (Figure 2). A total of 380 participants were enrolled in the study and randomized; however, 114 (29.2%) participants were lost to follow-up and were not included in the study, 3 (0.8%) participants discontinued, 6 (1.6%) participants had missing data, and 16 (4.2%) participants reported being HIV positive, leaving 241 (control group: n=122, 64.2%; intervention group: n=119, 62.6%) participants for the mITT analysis. Furthermore, 24 (20.2%) participants in the intervention group did not play the game, leaving 217 (57.1%) participants for the PP analysis (control group: n=122, 100%; intervention group: n=95, 79.8%; Figure 2).

XSL•FO
RenderX

Figure 2. CONSORT (Consolidated Standards of Reporting Trials) diagram for the Serious Games HIV Prevention Trial.



As shown in Table 2, the characteristics of participants in the control and intervention groups were generally similar. The median age of the participants was 22 (IQR 20-24) years in both groups, and 54.9% (67/122) in the control group and 55.5% (66/119) in the intervention group were men. More than 90% (226/241) of the participants in both groups were never married, and 66.4% (81/122) in the control arm and 68.9% (82/119) in the intervention arm were currently in tertiary education. The most common recruitment platform was Facebook, accounting for 43.6% (105/241) of participants, followed by recruitment from the most popular mall in the capital city (40/241, 16.6% of participants). Participants who played the game were asked if they would recommend SwaziYolo to their peers; 98.9% (94/95) either agreed (32/95, 33.7%) or strongly agreed (62/95, 65.3%). Compared to participants who played the game, those who did not play the game were more likely to be female (14/24, 58.3% vs 10/24, 41.1%), recruited from Facebook (15/24, 62.5% vs 35/95, 36.9%), and less likely to have a tertiary education (12/24, 50% vs 70/95, 73.7%; Multimedia Appendix 2).

XSL•FO
RenderX

Table 2. Baseline demographics of the control and intervention groups.

Demographics	Control group (n=122)	Intervention group (n=119)
Sex, n (%)		
Female	55 (45.1)	53 (44.5)
Male	67 (54.9)	66 (55.5)
Age (y), median (IQR)	22 (20-24)	22 (20-24)
Marital status, n (%)		
Single (never married and not living with a partner)	113 (92.6)	113 (95)
Married	2 (1.6)	0 (0)
Living with a partner	7 (5.7)	5 (4.2)
Separated (currently not living together but not divorced)	0 (0)	1 (0.6)
Level of education, n (%)		
None	2 (1.6)	3 (2.5)
Secondary level	3 (2.5)	3 (2.5)
High school level	36 (29.5)	31 (26.1)
Tertiary level	81 (66.4)	82 (68.9)
Employment status, n (%)		
Employed	20 (16.4)	19 (16)
Not employed	27 (22.1)	29 (24.4)
Student	74 (60.7)	63 (52.9)
Self-employed	1 (0.1)	8 (6.7)
Level of monthly income, n (%) (SZL 1=US $0.0718)		
<SZL 249	36 (29.5)	43 (36.1)
SZL 250-1749	74 (60.7)	63 (52.9)
SZL 1750-3000	12 (9.8)	13 (10.9)
≥SZL 3000	0 (0)	0 (0)
How did you hear about the game, n (%)		
Facebook	55 (45.1)	50 (42)
Limkokwing	13 (10.7)	16 (13.4)
Plaza Mbabane	21 (17.2)	19 (16)
Swaziland Christian University	14 (11.5)	14 (11.8)
Other	19 (15.6)	20 (16.8)

Main Outcome: Risk Perception

In the mITT analysis, for both the 8-item and the 10-item PRHS, we observed no significant difference between the control and intervention groups in the change of the mean score for the 8-item PRHS (DID: mean 1.1, SD 0.74; *P*=.13) and the 10-item PRHS (DID: mean 1.3, SD 0.80; *P*=.12; Figure 3; Multimedia Appendix 3). However, PP analyses suggested an increase in risk perception among those who played the game in the intervention group in both PRHS indices (8-item DID mean 1.6, SD 0.74; *P*=.04 and 10-item DID mean 1.8, SD 0.83; *P*=.03; Figure 3; Multimedia Appendix 3).

Table 3 reports the results of the analyses of the secondary outcomes of protective behaviors. The odds ratios for all secondary outcomes were not statistically significant.

The DID mean change in HIV risk perception in the 10-item PRHS was 1.8 (SD 0.89; *P*=.04) among those with multiple sexual partnerships compared to 1.4 (SD 2.22; *P*=.52) among those without multiple sexual partnerships (Figure 4; Multimedia Appendix 4). For both the 8-item and 10-item PRHS, the DID mean change scores were the highest among participants with multiple sexual partnerships.

Among participants reporting condomless sex, the DID mean change in HIV risk perception in the 10-item PRHS was greater (DID: mean of 2.2, SD 0.98; *P*=.02) compared to participants who did not report condomless sex (DID: mean 0.54, SD 1.59; *P*=.73; Figure 5; Multimedia Appendix 5). For both the 8-item and 10-item PRHS, the DID mean change scores were the highest among participants with condomless sex.

XSL•FO
RenderX

Figure 3. Difference-in-difference (DID) analyses of risk perception using the 8-item and 10-item indices for both modified intention-to-treat analysis (mITT) and per-protocol (PP) analyses.

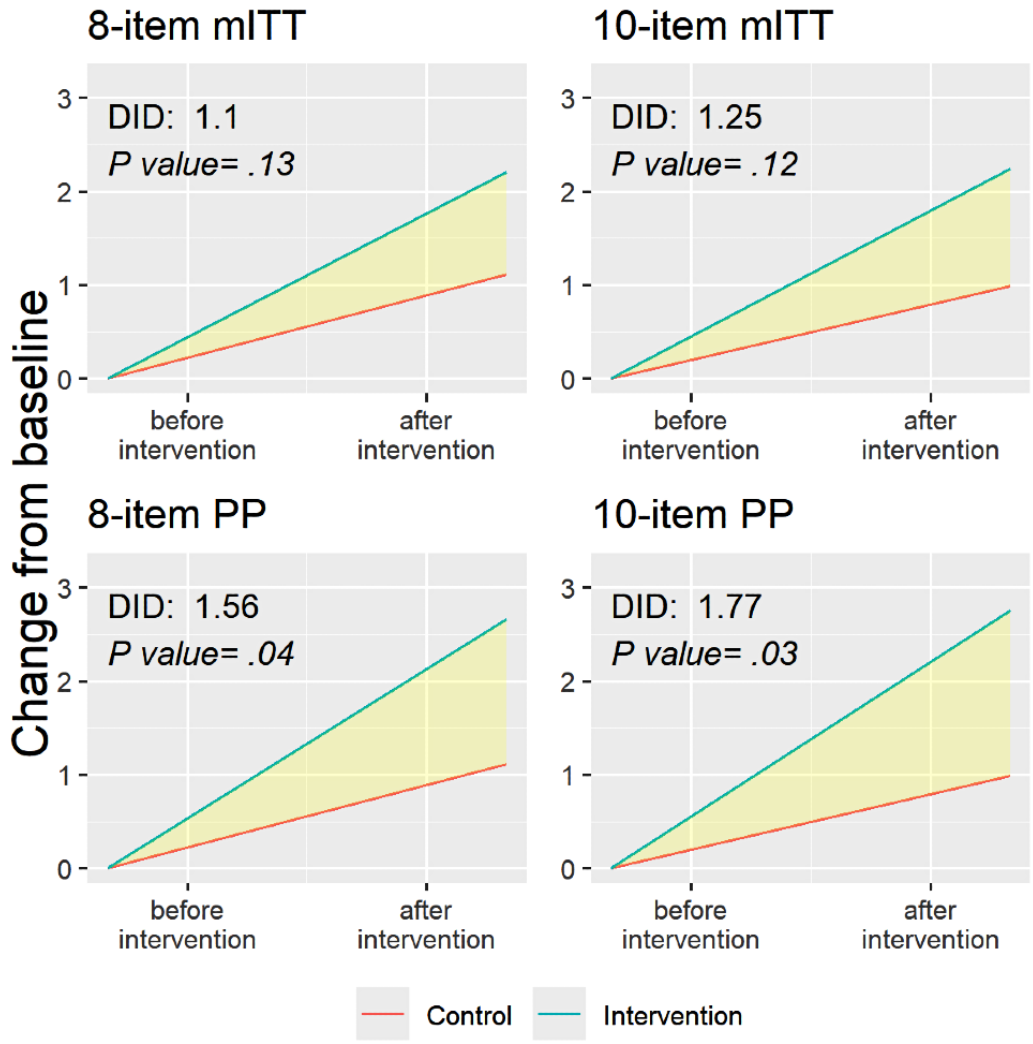


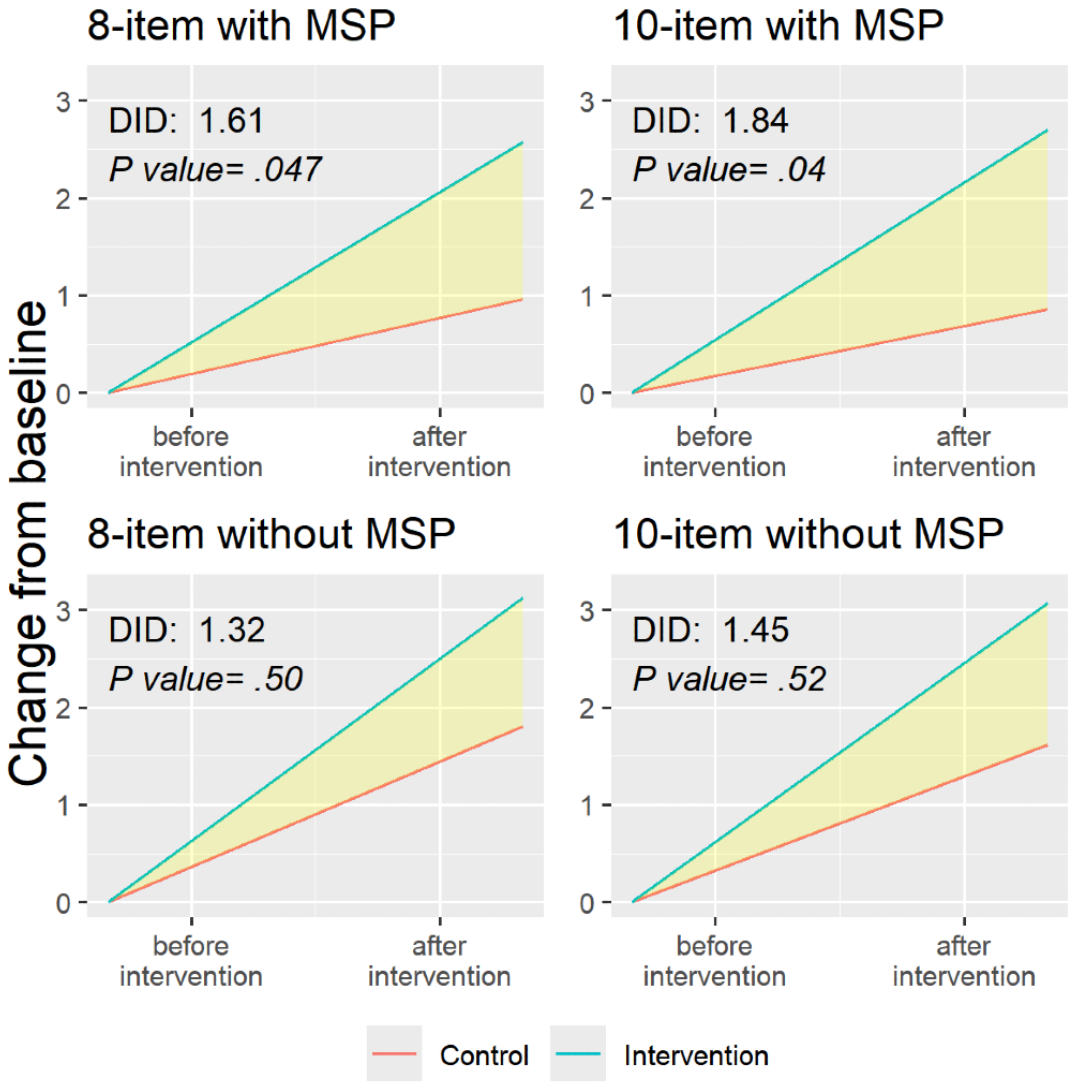
Table 3. Impact of the intervention on secondary outcomes of protective behavior.

Per-protocol secondary outcomes	Control group, n=122		Intervention group, n=95		Odds ratio (95% CI)
	Before intervention, n (%)	After intervention, n (%)	Before intervention, n (%)	After intervention, n (%)	
Number of sexual partners in the last 30 days	__[a]	23 (18.9)	—	19 (20)	1.08 (0.55-2.12)
Sex without a condom in the last 30 days	40 (32.8)	37 (30.3)	24 (25.3)	28 (29.5)	1.03 (0.57-1.88)
Condom use during the last sexual encounter	96 (78.7)	92 (75.4)	78 (82.1)	72 (75.8)	0.91 (0.45-1.85)
Tested for HIV in the last 30 days	22 (18)	19 (15.6)	13 (13.7)	17 (17.9)	1.35 (0.63-2.88)
Intention to test for HIV	10 (8.2)	17 (13.9)	9 (9.5)	14 (14.7)	1.52 (0.09-25.41)
Intention to know the HIV status of partners	109 (89.3)	107 (87.7)	78 (82.1)	83 (87.4)	1.34 (0.54-3.33)
Intention to reduce multiple concurrent partnerships	89 (73)	85 (69.7)	68 (71.6)	72 (75.8)	1.51 (0.77-2.97)
Intention to use a condom in the next sexual encounter	104 (85.2)	97 (79.5)	78 (82.1)	81 (85.3)	1.89 (0.83-4.3)
Steady partner has tested	99 (81.1)	101 (82.8)	77 (81.1)	79 (83.2)	1.04 (0.47-2.26)
Knows partner's HIV status	88 (71.1)	86 (70.5)	64 (67.4)	65 (68.4)	1 (0.52- 1.9)

[a]Not available.

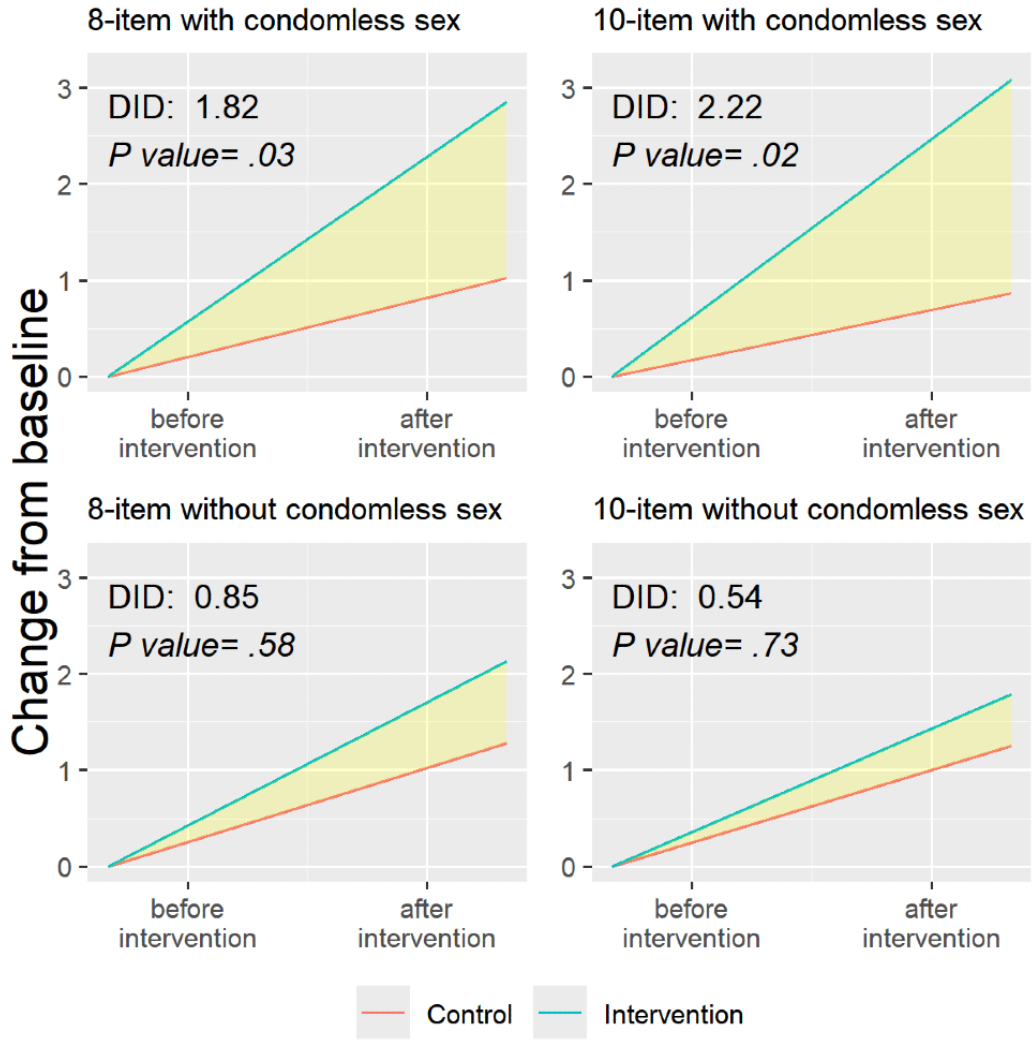
XSL•FO
RenderX

Figure 4. Difference-in-difference (DID) analyses of risk perception among participants with and without multiple sexual partnerships using the 8-item and 10-item indices, per-protocol analysis. MSP: multiple sexual partnerships.



XSL•FO
RenderX

Figure 5. Difference-in-difference (DID) analyses of risk perception among participants with and without condomless sex using the 8-item and 10-item indices, per-protocol analysis.



Discussion

HIV Risk Perception

In primary analyses of this RCT of SwaziYolo—a smartphone-based, interactive, educational story game to address HIV risk perception—the intervention was not associated with a statistically significant change in HIV risk perception among young people in Eswatini. However, the results of the PP analyses suggest that playing the game at least once significantly increased HIV risk perception.

Several factors may explain these results, with implications for the development of the next iteration of SwaziYolo and other serious game interventions. The greater effect observed in the PP analysis and the fact that 20% (24/119) of individuals randomized to the intervention group did not play the game indicate that strategies to increase the uptake of serious game interventions may increase effectiveness. The intervention duration was 30 days, and it is possible that a longer period would have increased efficacy. Challenges associated with uptake may have included carrier network failure, lack of adequate free storage on the phone, and phone resources, as reported by most participants on our online support platform. Furthermore, individuals who did not play the game were more likely to be female, recruited from Facebook, and less likely to have tertiary education. This suggests that different strategies may be needed to engage these populations, such as different game design considerations for young men and women.

HIV Risk Perception and Sexual Behaviors

We found that the impact of the intervention was highest among subgroups reporting multiple sexual partnerships and condomless sex. This finding suggests that serious games, such as SwaziYolo, should be prioritized for people who may be more likely to benefit from interventions focused on HIV risk perceptions. Increasing HIV risk perception can lead to behavior change and increase uptake of HIV prevention interventions [20]. For example, previous studies have found that risk perception is associated with condom use [19,26], PrEP use [15,19,27,28], and HIV testing [26,29,30]. However, the effect size observed in the changes in PP HIV risk perception was modest and may not have been powerful enough to elicit intention to engage in protective behavior. Although the PRHS [20] has been validated to capture multiple dimensions of HIV risk perception, the threshold sufficient to elicit the intention to engage in behavioral change is still subject to further research in our population.

XSL•FO
RenderX

Our study was among the first to use serious games to improve HIV risk perception in a country where 24.8% of the population aged 15 years and older live with HIV [2]. Future studies should examine whether other serious game interventions improve HIV risk perception and whether improved HIV risk perception through serious game interventions is effective in increasing condom use, PrEP uptake, HIV testing, and retention in HIV care among young people in countries such as Eswatini. Studies conducted in similar settings showed that young people were likely to have inaccurate HIV risk perceptions, such that even those at high risk of HIV acquisition had lower HIV risk perceptions [31-33]. Furthermore, the complex intersection of HIV risk perception and HIV stigma remains understudied in high-epidemic settings, such as Eswatini [34,35].

Intention to Change Behavior

Our secondary analysis assessing the impact of SwaziYolo on intention to engage in protective sexual behaviors did not show any statistical difference between groups. This may have been a result of the limited time of exposure to the interventions (4 weeks). Future interventions should assess such effects over longer durations, as the duration of exposure has been highlighted as one of the key factors for sustaining serious game intervention effects [36,37]. For example, in a 24-week follow-up of an HIV medication adherence serious game intervention, the intervention group was 3.75 times more likely to have optimal PrEP dosing compared to the control group that did not receive the intervention [38]. Another plausible reason for the limited impact of SwaziYolo on secondary behavior outcomes could be that HIV risk perception has a limited impact on behavior. However, a substantial body of research has shown that HIV risk perception can be a significant motivator for adopting protective behaviors, especially when combined with other behavioral theory–based constructs, such as self-efficacy, stigma mitigation, and access to HIV prevention tools such as condoms and PrEP [19,27,28,34,39]. It is also possible that, due to broken randomization, participants in the PP analyses had less need to significantly change their behavior. Future research is needed to examine the impact of interventions to improve HIV risk perception on behavioral change and the uptake of protective interventions.

Player Engagement and Acceptability

At least 80% (95/119; Multimedia Appendix 2) of the participants played the game. This engagement rate is higher than that in most serious games [16] and shows the engagement of young people and the potential for future interventions to use serious games. Up to 98% (94/95) of those who played the game said they would recommend it to their peers, indicating that SwaziYolo achieved acceptance among young people in Eswatini. With increased smartphone access, processing power, and advances in artificial intelligence, we hypothesize that future serious games will achieve greater engagement and retention. For example, in 2017, smartphone ownership was 55% in South Africa (a neighboring country to Eswatini sharing many socioeconomic similarities) [34], and in 2024, smartphone ownership increased to 71% [35]. The types of smartphones available in 2024 have considerably more computing power and storage, which will support engagement, immersion, and game aesthetics.

Limitations

First, the engagement data were based on self-reports. The game data used a proprietary architecture, which limited our access to game-generated data and did not allow us to analyze intensity (ie, exposure to different intervention components within the serious game) as generally practiced in the field [40,41]. Future studies should seek full access to data repositories for serious games to allow game paradata analytics. Second, although the characteristics of participants in the mITT analyses were generally balanced, 29.6% (56/189) of the participants in the control group and 30.8% (58/188) of the participants in the intervention group were lost to follow-up. Additionally, 4.2% (8/189) of the participants in the control group and 4.3% (8/188) of the participants in the intervention group reported being HIV positive. However, this was not identified until after the study, and these individuals were not included in the analyses, deviating from standard ITT principles. Consequently, the benefits of randomization in minimizing selection bias might have been diminished in both mITT and PP analyses. Finally, participants in the control (waitlist) group may have experienced the Hawthorne effect, becoming more aware of their behavior and modifying their HIV risk perception due to anticipation of gaining access to SwaziYolo after 4 weeks. This may have increased their risk perception scores and diminished the observed difference between the groups.

Strengths

First, compared to other serious game interventions in sub-Saharan Africa, there was no physical contact with study participants during the intervention period apart from initial recruitment, demonstrating the potential for a fully online serious game intervention. This may have yielded a more realistic estimate of the potential adoption and proximal feasibility of technology-based HIV interventions in low- and middle-income countries. Second, overall, participants found the game highly acceptable, showing potential for future iterations of SwaziYolo and other serious game interventions to promote the uptake of HIV prevention interventions among young people in Eswatini and similar settings.

Conclusions

In a country with high HIV prevalence, SwaziYolo was acceptable and increased HIV risk perception among young people who self-selected to play the game, especially among those who reported multiple sexual partnerships or condomless sex and who were most likely to benefit from interventions to prevent HIV. Although the mITT analysis showed no impact, the findings support the need for future research to optimize and evaluate the intervention.

XSL•FO
RenderX

Acknowledgments

The authors acknowledge the valuable contributions of Patou Musumari, Teeranee Techasrivichien, and S Pilar Suguimoto to the development of the protocol. They also acknowledge Formula D Interactive for the development of the game and Net Storm Technologies for the website development. Furthermore, they thank Mr Norihisa Wada, their game development consultant.

The intervention design, development, and implementation were funded by the Nakayama Foundation for Science, Technology, and Sports (grant). Data analysis was supported by a University of Alabama at Birmingham Faculty Development Grant.

Data Availability

The datasets generated or analyzed during this study are not publicly available due to privacy and ethical restrictions but are available from the corresponding author on reasonable request.

Authors' Contributions

BWL, Masahiro Kihara, and Masako Kihara led study conception, intervention design, methodology, statistical design, programming platform, intervention user interface design, and study trial implementation. BWL and Masahiro Kihara led funding acquisition. BWL led project administration. BWL, KJB, and MD conducted the formal statistical analysis. BWL and KJB interpreted the results and drafted the manuscript. FS, CE, MM, Masahiro Kihara, Masako Kihara, and AK read and edited the drafts of the manuscript and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Serious Games HIV Prevention Trial before and after intervention questionnaire.
[DOCX File , 87 KB - games_v13i1e70333_app1.docx]

Multimedia Appendix 2
Characteristics of participants who did and did not play the game.
[DOCX File , 22 KB - games_v13i1e70333_app2.docx]

Multimedia Appendix 3
Difference-in-difference analysis of risk perception.
[DOCX File , 17 KB - games_v13i1e70333_app3.docx]

Multimedia Appendix 4
Difference-in-difference analysis of risk perception among participants with and without multiple sexual partnerships.
[DOCX File , 35 KB - games_v13i1e70333_app4.docx]

Multimedia Appendix 5
Risk perception among those with and without condomless sex in the past 30 days.
[DOCX File , 35 KB - games_v13i1e70333_app5.docx]

Multimedia Appendix 6
CONSORT-eHEALTH checklist (version 1.6.1).
[PDF File (Adobe PDF File), 1032 KB - games_v13i1e70333_app6.pdf]

References

1. In danger: UNAIDS global AIDS update 2022. Joint United Nations Programme on HIV/ AIDS. 2022. URL: https://www.unaids.org/sites/default/files/media_asset/2022-global-aids-update_en.pdf [accessed 2025-11-17]
2. Eswatini population-based HIV impact assessment 3 2021 (SHIMS3 2021): final report November 2023. Ministry of Health, Eswatini. 2021. URL: https://phia.icap.columbia.edu/wp-content/uploads/2023/12/241123_SHIMS_ENG_RR3_Final-1.pdf [accessed 2024-04-30]
3. The extended national multisectoral HIV and AIDS framework (eNSF) 2014-2018. National Emergence Response Council on HIV and AIDS. URL: http://swaziaidsprogram.org/wp-content/uploads/2016/10/ExtendedNSF.pdf [accessed 2016-03-19]

XSL•FO
RenderX

4. Bicego GT, Nkambule R, Peterson I, Reed J, Donnell D, Ginindza H, et al. Recent patterns in population-based HIV prevalence in Swaziland. PLoS One 2013 Feb;8(10):e77101 [FREE Full text] [doi: 10.1371/journal.pone.0077101] [Medline: 24143205]
5. Nkambule R, Philip NM, Reid G, Mnisi Z, Nuwagaba-Biribonwoha H, Ao TT, et al. HIV incidence, viremia, and the national response in Eswatini: two sequential population-based surveys. PLoS One 2021 Dec 02;16(12):e0260892 [FREE Full text] [doi: 10.1371/journal.pone.0260892] [Medline: 34855890]
6. Lukhele BW, Musumari P, El-Saaidi C, Techasrivichien T, Suguimoto SP, Ono Kihara M, et al. Efficacy of mobile serious games in increasing HIV risk perception in Swaziland: a randomized control trial (SGprev Trial) research protocol. JMIR Res Protoc 2016 Nov 22;5(4):e224 [FREE Full text] [doi: 10.2196/resprot.6543] [Medline: 27876685]
7. Krath J, Schürmann L, von Korflesch HF. Revealing the theoretical basis of gamification: a systematic review and analysis of theory in research on gamification, serious games and game-based learning. Comput Hum Behav 2021 Dec;125:106963 [FREE Full text] [doi: 10.1016/j.chb.2021.106963]
8. Djaouti D, Alvarez J, Jessel JP, Rampnoux O. Origins of serious games. In: Ma M, Oikonomou A, Jain LC, editors. Serious Games and Edutainment Applications. London, UK: Springer; Oct 13, 2011:25-43.
9. Fiellin LE, Kyriakides TC, Hieftje KD, Pendergrass TM, Duncan LR, Dziura JD, et al. The design and implementation of a randomized controlled trial of a risk reduction and human immunodeficiency virus prevention videogame intervention in minority adolescents: PlayForward: Elm City Stories. Clin Trials 2016 Aug;13(4):400-408 [FREE Full text] [doi: 10.1177/1740774516637871] [Medline: 27013483]
10. Winskell K, Sabben G, Ondeng'e K, Odero I, Akelo V, Mudhune V. A smartphone game to prevent HIV among young Kenyans: household dynamics of gameplay in a feasibility study. Health Educ J 2019 Aug 01;78(5):595-606 [FREE Full text] [doi: 10.1177/0017896919832344] [Medline: 34219796]
11. Winskell K, Sabben G, Akelo V, Ondeng'e K, Odero I, Mudhune V. A smartphone game to prevent HIV among young Kenyans: local perceptions of mechanisms of effect. Health Educ Res 2020 Jun 01;35(3):153-164 [FREE Full text] [doi: 10.1093/her/cyaa011] [Medline: 32441759]
12. Haruna H, Hu X, Chu SK, Mellecker RR, Gabriel G, Ndekao PS. Improving sexual health education programs for adolescent students through game-based learning and gamification. Int J Environ Res Public Health 2018 Sep 17;15(9):2027 [FREE Full text] [doi: 10.3390/ijerph15092027] [Medline: 30227642]
13. Datta S, Burns J, Maughan-Brown B, Darling M, Eyal K. Risking it all for love? Resetting beliefs about HIV risk among low-income South African teens. J Econ Behav Organ 2015 Oct;118:184-198 [FREE Full text] [doi: 10.1016/j.jebo.2015.02.020]
14. Corneli AL, McKenna K, Headley J, Ahmed K, Odhiambo J, Skhosana J, FEM-PrEP Study Group. A descriptive analysis of perceptions of HIV risk and worry about acquiring HIV among FEM-PrEP participants who seroconverted in Bondo, Kenya, and Pretoria, South Africa. J Int AIDS Soc 2014 Sep 08;17(3 Suppl 2):19152 [FREE Full text] [doi: 10.7448/IAS.17.3.19152] [Medline: 25224613]
15. Price JT, Rosenberg NE, Vansia D, Phanga T, Bhushan NL, Maseko B, et al. Predictors of HIV, HIV risk perception, and HIV worry among adolescent girls and young women in Lilongwe, Malawi. J Acquir Immune Defic Syndr 2018 Jan 01;77(1):53-63 [FREE Full text] [doi: 10.1097/QAI.0000000000001567] [Medline: 28991885]
16. Prata N, Morris L, Mazive E, Vahidnia F, Stehr M. Relationship between HIV risk perception and condom use: evidence from a population-based survey in Mozambique. Int Fam Plan Perspect 2006 Dec;32(4):192-200 [FREE Full text] [doi: 10.1363/3219206] [Medline: 17237016]
17. Schaefer R, Thomas R, Maswera R, Kadzura N, Nyamukapa C, Gregson S. Relationships between changes in HIV risk perception and condom use in East Zimbabwe 2003-2013: population-based longitudinal analyses. BMC Public Health 2020 May 24;20(1):756 [FREE Full text] [doi: 10.1186/s12889-020-08815-1] [Medline: 32448365]
18. Cao Z, Chen J, Lin B, Zhang C, Zhong X. Factors influencing intention on condom use during sexual intercourse with regular female partners among men who have sex with men in Western China: a structural equation modeling analysis. Sex Transm Dis 2023 Dec 01;50(12):782-788 [FREE Full text] [doi: 10.1097/OLQ.0000000000001876] [Medline: 37944160]
19. Warren EA, Paterson P, Schulz WS, Lees S, Eakle R, Stadler J, et al. Risk perception and the influence on uptake and use of biomedical prevention interventions for HIV in sub-Saharan Africa: a systematic literature review. PLoS One 2018 Jun 14;13(6):e0198680 [FREE Full text] [doi: 10.1371/journal.pone.0198680] [Medline: 29902205]
20. Napper LE, Fisher DG, Reynolds GL. Development of the perceived risk of HIV scale. AIDS Behav 2012 May;16(4):1075-1083 [FREE Full text] [doi: 10.1007/s10461-011-0003-2] [Medline: 21785873]
21. Lukhele BW, Techasrivichien T, Musumari PM, El-Saaidi C, Suguimoto SP, Ono-Kihara M, et al. Multiple sexual partnerships and their correlates among Facebook users in Swaziland: an online cross-sectional study. Afr J AIDS Res 2016 Sep;15(3):203-210. [doi: 10.2989/16085906.2016.1171790] [Medline: 27532641]
22. Chu SK, Kwan AC, Reynolds R, Mellecker RR, Tam F, Lee G, et al. Promoting sex education among teenagers through an interactive game: reasons for success and implications. Games Health J 2015 Jun;4(3):168-174. [doi: 10.1089/g4h.2014.0059] [Medline: 26182060]
23. Arnab S, Dunwell I, Debattista K. Serious Games for Healthcare: Applications and Implications. Hershey, PA: IGI Global; Jul 2013.

XSL•FO
RenderX

24. Lukhele BW, Techasrivichien T, Suguimoto SP, Musumari PM, El-Saaidi C, Haumba S, et al. Structural and behavioral correlates of HIV infection among pregnant women in a country with a highly generalized HIV epidemic: a cross-sectional study with a probability sample of antenatal care facilities in Swaziland. PLoS One 2016 Dec 12;11(12):e0168140 [FREE Full text] [doi: 10.1371/journal.pone.0168140] [Medline: 27942014]
25. Hopewell S, Chan AW, Collins GS, Hróbjartsson A, Moher D, Schulz KF, et al. CONSORT 2025 statement: updated guideline for reporting randomized trials. Nat Med 2025 Jun;31(6):1776-1783. [doi: 10.1038/s41591-025-03635-5] [Medline: 40229553]
26. Ngure K, Thuo N, Ogello V, Kiptinness C, Kamolloh K, Burns BF, et al. Dynamic perceived HIV risk and sexual behaviors among young women enrolled in a PrEP trial in Kenya: a qualitative study. Front Reprod Health 2021 Aug 12;3:637869 [FREE Full text] [doi: 10.3389/frph.2021.637869] [Medline: 36304002]
27. Sewell WC, Patel RR, Blankenship S, Marcus JL, Krakower DS, Chan PA, et al. Associations among HIV risk perception, sexual health efficacy, and intent to use PrEP among women: an application of the risk perception attitude framework. AIDS Educ Prev 2020 Oct;32(5):392-402 [FREE Full text] [doi: 10.1521/aeap.2020.32.5.392] [Medline: 33112674]
28. Hill LM, Maseko B, Chagomerana M, Hosseinipour MC, Bekker LG, Pettifor A, et al. HIV risk, risk perception, and PrEP interest among adolescent girls and young women in Lilongwe, Malawi: operationalizing the PrEP cascade. J Int AIDS Soc 2020 Jun;23(Suppl 3):e25502 [FREE Full text] [doi: 10.1002/jia2.25502] [Medline: 32602649]
29. Lin Y, Li C, Wang L, Jiao K, Ma W. The mediated effect of HIV risk perception in the relationship between peer education and HIV testing uptake among three key populations in China. AIDS Res Ther 2021 Mar 25;18(1):8 [FREE Full text] [doi: 10.1186/s12981-021-00334-2] [Medline: 33766062]
30. Ajayi AI, Mudefi E, Adeniyi OV, Goon DT. Achieving the first of the Joint United Nations Programme on HIV/AIDS (UNAIDS) 90-90-90 targets: understanding the influence of HIV risk perceptions, knowing one's partner's status and discussion of HIV/sexually transmitted infections with a sexual partner on uptake of HIV testing. Int Health 2019 Nov 13;11(6):425-431. [doi: 10.1093/inthealth/ihz056] [Medline: 31365082]
31. Jung MS, Dlamini NS, Cui X, Cha K. Prevalence of HIV testing and associated factors among young adolescents in Eswatini: a secondary data analysis. BMC Pediatr 2022 Nov 14;22(1):659 [FREE Full text] [doi: 10.1186/s12887-022-03698-0] [Medline: 36376807]
32. Onyechi KC, Eseadi C, Okere AU, Otu MS. Effects of Rational-Emotive Health Education Program on HIV risk perceptions among in-school adolescents in Nigeria. Medicine (Baltimore) 2016 Jul;95(29):e3967 [FREE Full text] [doi: 10.1097/MD.0000000000003967] [Medline: 27442633]
33. Tsegay G, Edris M, Meseret S. Assessment of voluntary counseling and testing service utilization and associated factors among Debre Markos University Students, North West Ethiopia: a cross-sectional survey in 2011. BMC Public Health 2013 Mar 19;13:243 [FREE Full text] [doi: 10.1186/1471-2458-13-243] [Medline: 23509888]
34. Machemedze T. Does self-perceived HIV risk mediate the potential association between HIV-related symbolic stigma and sexual behaviour among young adult women in Cape Town, South Africa? BMC Public Health 2023 Jan 28;23(1):188 [FREE Full text] [doi: 10.1186/s12889-022-14862-7] [Medline: 36709267]
35. Embleton L, Logie CH, Ngure K, Nelson L, Kimbo L, Ayuku D, et al. Intersectional stigma and implementation of HIV prevention and treatment services for adolescents living with and at risk for HIV: opportunities for improvement in the HIV continuum in sub-Saharan Africa. AIDS Behav 2023 May;27(Suppl 1):162-184 [FREE Full text] [doi: 10.1007/s10461-022-03793-4] [Medline: 35907143]
36. Jitmun W, Palee P, Choosri N, Surapunt T. The success of serious games and gamified systems in HIV prevention and care: scoping review. JMIR Serious Games 2023 Sep 05;11:e39915 [FREE Full text] [doi: 10.2196/39915] [Medline: 37669098]
37. Smith AU, Khawly GM, Jann J, Zetina AP, Padilla J, Schnall R. A review of serious gaming as an intervention for HIV prevention. Curr HIV/AIDS Rep 2023 Aug;20(4):181-205 [FREE Full text] [doi: 10.1007/s11904-023-00659-9] [Medline: 37213045]
38. Whiteley LB, Olsen EM, Haubrick KK, Odoom E, Tarantino N, Brown LK. A review of interventions to enhance HIV medication adherence. Curr HIV/AIDS Rep 2021 Oct;18(5):443-457. [doi: 10.1007/s11904-021-00568-9] [Medline: 34152554]
39. Tully S, Cojocaru M, Bauch CT. Sexual behavior, risk perception, and HIV transmission can respond to HIV antiviral drugs and vaccines through multiple pathways. Sci Rep 2015 Oct 28;5:15411 [FREE Full text] [doi: 10.1038/srep15411] [Medline: 26507957]
40. Tan JW, Ng KB, Mogali SR. An exploratory digital board game approach to the review and reinforcement of complex medical subjects like anatomical education: cross-sectional and mixed methods study. JMIR Serious Games 2022 Jan 10;10(1):e33282 [FREE Full text] [doi: 10.2196/33282] [Medline: 35006080]
41. Lu W, Griffin J, Sadler TD, Laffey J, Goggins SP. Serious game analytics by design: feature generation and selection using game telemetry and game metrics: toward predictive model construction. J Learn Anal 2023;10(1):168-188. [doi: 10.18608/jla.2023.7681]

XSL•FO
RenderX

Abbreviations

CONSORT-EHEALTH: Consolidated Standards for Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
DID: difference-in-difference
mITT: modified intention-to-treat
OTP: one-time passcode
PP: per-protocol
PrEP: pre-exposure prophylaxis
PRHS: Perceived Risk of HIV Infection Scale
RCT: randomized controlled trial
SGPrev-Trial: Eswatini (formerly Swaziland) Serious Games HIV Prevention Trial

Edited by A Coristine; submitted 22.12.24; peer-reviewed by MP Mabuza, K Comins, X Zhang, G Sabben; comments to author 11.06.25; revised version received 31.10.25; accepted 31.10.25; published 24.11.25.

Please cite as:
Lukhele BW, Delay M, Shabalala F, Motsa M, Kay A, El Saaidi C, Kihara M, Kihara M, Bruxvoort KJ
Efficacy of a Mobile Serious Game (SwaziYolo) for Increasing HIV Risk Perception: Randomized Controlled Trial
JMIR Serious Games 2025;13:e70333
URL: https://games.jmir.org/2025/1/e70333
doi:10.2196/70333
PMID:

©Bhekumusa Wellington Lukhele, Mac Delay, Fortunate Shabalala, Mfundi Motsa, Alexander Kay, Christina El Saaidi, Masako Kihara, Masahiro Kihara, Katia J Bruxvoort. Originally published in JMIR Serious Games (https://games.jmir.org), 24.11.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Super Mario Bros. and Yoshi Games' Affordance of Childlike Wonder and Reduced Burnout Risk in Young Adults: In-Depth Mixed Methods Cross-Sectional Study

Winze Tam[1], BBA, MSc; Congcong Hou[2], MSc, PhD; Andreas Benedikt Eisingerich[1], BSc, MPhil, PhD

[1]Imperial College Business School, Imperial College London, London, United Kingdom

[2]Faculty of Commerce, Kyushu Sangyo University, Fukuoka, Japan

Corresponding Author:
Andreas Benedikt Eisingerich, BSc, MPhil, PhD
Imperial College Business School
Imperial College London
South Kensington Campus
London, SW7 2AZ
United Kingdom
Phone: 44 207589 ext 5111
Email: a.eisingerich@imperial.ac.uk

Abstract

Background: Unrelenting pressure and an "always-on" culture can leave no time for genuine rest among young adults. While playing video games has been noted to afford cognitive escapism and relaxation, critical questions remain about the influence of popular video games, such as Super Mario Bros., and their potential effects on young adults' burnout risk.

Objective: This study examined the extent to which, if at all, popular video games such as Super Mario Bros. and Yoshi could foster childlike wonder. It also investigated the potential of these games to reduce burnout risk among young adults.

Methods: We used a mixed methods approach. First, qualitative data were collected through 41 exploratory, in-depth interviews (women: n=19, 46.3%; men: n=21, 51.2%; prefer not to disclose sex: n=11, 2.4%; mean age 22.51, SD 1.52 years) with university students who had experience playing Super Mario Bros. or Yoshi. Second, quantitative data were collected in a cross-sectional survey (N=336) of players of Super Mario Bros. and Yoshi to examine the games' affordance of childlike wonder, overall happiness in life, and burnout risk.

Results: Insights from in-depth interviews showed that players appreciated the ability of Super Mario Bros. and Yoshi games to instill childlike wonder, enhance happiness in life, and reduce burnout risk. Quantitative analyses showed that the games' affordance of childlike wonder positively affected young adults' happiness (b=0.30, SE=0.04, t=6.80, 95% CI 0.21-0.38; P<.001). In turn, overall happiness significantly reduced the risk of burnout (b=–0.48, SE=0.05, t=–9.55, 95% CI –0.572 to –0.377). Results showed that happiness fully mediated the impact of childlike wonder on burnout, as the direct effect of childlike wonder on burnout risk became insignificant (b=–0.08, SE=0.04, t=–1.88, 95% CI –0.16 to 0.01; P=.06), while the indirect effect of childlike wonder on burnout risk was significant (b=–0.14, bootstrapped SE=0.03, 95% CI –0.20 to –0.09).

Conclusions: The findings showed the significant positive effect of popular video games such as Super Mario Bros. and Yoshi on fostering players' childlike wonder, increasing happiness, and reducing burnout risk. This study was among the first to identify childlike wonder as an emotional pathway through which mainstream video games could enhance well-being and reduce burnout. By moving beyond escapism and nostalgia, it offers a new perspective on how well-designed, globally familiar games can function as accessible, resilience-building digital microenvironments. These findings contributed to research bridging gaming and mental health and have practical implications for game designers, educators, and health professionals interested in promoting mental wellness through everyday play.

(JMIR Serious Games 2025;13:e84219) doi:10.2196/84219

KEYWORDS

childlike wonder; happiness; burnout; video games; qualitative research; survey study

XSL•FO
RenderX

Introduction

Overview of Young Adults and Burnout Risk

Young people today face a unique constellation of interconnected pressures that contribute significantly to anxiety, worry, and burnout [1-3]. Soaring costs for housing, education (leading to significant student loan debt), and basic necessities create significant financial pressure [4,5]. Achieving traditional markers of adulthood (home ownership and a stable career) feels increasingly out of reach for numerous young people. Witnessing vast wealth disparities and feeling the system is rigged can fuel pessimism and a sense of unfairness about future prospects [6]. Furthermore, intense pressure to excel academically from a young age, secure prestigious internships or jobs, and build a seemingly perfect resume can create chronic stress [7]. Success is often narrowly defined and hypercompetitive. Importantly, these cumulative stressors arise during what developmental psychologists describe as emerging adulthood, a period marked by identity exploration and instability [8].

Young adulthood, often referred to as emerging adulthood, is considered a distinct stage of development characterized by identity exploration, instability, self-focus, and shifts in social roles, typically spanning the late teens and 20s [8]. Epidemiological research indicates that many mental disorders first appear during this life stage, although timing varies across diagnostic categories—for example, anxiety disorders tend to surface earlier, while mood disorders can appear later [9,10]. These onset patterns often coincide with major life transitions involving education, work, and growing autonomy, which have been linked to higher psychological distress and lower well-being among young adults [11]. Hence, focusing on this demographic remains critical for understanding today's mental health challenges and for identifying where digital interventions might best support well-being.

In addition, the current "always-on" digital culture and social media expose young people to curated highlights of others' lives, careers, and achievements 24/7, fostering unrealistic comparisons, fear of missing out, and feelings of inadequacy [12-15]. The digital world also provides new avenues for bullying and harassment, impacting mental health significantly. Moreover, smartphones make it difficult to disconnect from work, studies, or social obligations [16,17]. The expectation to be constantly available and responsive erodes downtime and contributes to burnout. Moreover, doomscrolling or the constant exposure to negative news cycles and global crises (climate change, political or social conflict and unrest, or pandemics) can be overwhelming and induce feelings of helplessness [18-20]. Fueled by social media and competitive environments, many young adults internalize unrealistic expectations of perfection in all areas (looks, career, relationships, and hobbies), leading to self-criticism and anxiety about failure [21-23].

Moreover, while digitally connected, many young adults report feeling lonelier than ever [6,12,24]. Traditional community structures (extended family, local community groups, and religious institutions) often play a less central role nowadays. Despite online networks, forming and maintaining deep, meaningful in-person relationships can be challenging in what is often perceived as a fast-paced, transient world [6]. Taken together, these factors can create a state of chronic, unrelenting stress [6,25-27]. Feeling powerless over economic realities, global issues, or demanding work or school environments fuels burnout. Constant pressure and the "always-on" culture leave little time for genuine rest, relaxation, and hobbies that replenish energy. Hence, these factors do not exist in isolation. They feed into each other.

Social media comparisons fuel anxiety and worry about the future sap motivation, while perceived loneliness makes coping harder. It creates a vicious cycle that can be difficult to break. This study aims to make a step toward addressing the mental health crisis facing many young adults today. More specifically, video games, such as open-world games, have been shown in prior work to afford cognitive escapism, relaxation, and enhanced well-being to young people [28]. Video games and gamification techniques have been noted to positively affect the well-being of players in different ways [29-34]. However, critical questions remain about the potential role popular video games can play in young adults' lives.

Background on Super Mario Bros. and Yoshi Games

The Super Mario Bros. franchise created by Nintendo Co, Ltd stands as one of the most globally popular and recognizable in entertainment history [35,36]. Since the groundbreaking original in 1985, Mario games have sold more than 430 million copies worldwide, making it one of the best-selling video game series ever. Its appeal transcends age, culture, and nationality. Super Mario, Luigi, Princess Peach, and the cute dinosaur Yoshi are all cultural icons, instantly recognizable even to nongamers. Super Mario Bros. and Yoshi games can help act as stress buffers and reduce negative affect in several critical ways. For example, engaging with the demanding spatial navigation, timing, and problem-solving tasks in Super Mario Bros. requires focused attention (working memory load). This acts as a distractor, effectively diverting finite cognitive resources away from rumination on real-world worries or negative thoughts, thereby interrupting cycles of pessimism or anxiety. Players become deeply immersed in "being" Mario or Yoshi, navigating the space, temporarily suspending self-referential thoughts and worries about real-life identity or performance.

Furthermore, the core narrative of Mario and Yoshi persistently overcoming obstacles to achieve goals serves as a powerful symbolic representation of perseverance and eventual triumph. Engaging with this narrative, especially through active participation (the player is Mario or Yoshi), can subtly reinforce schemas related to resilience and positive outcomes. The core objective ("Help Prince Florian" or "Rescue Princess Peach") is unambiguous and morally positive, offering a clear purpose without complex emotional baggage or ambiguity. This reduces cognitive load associated with real-world problem interpretation. While challenging in their own ways, Super Mario Bros. and Yoshi games generally lack the prolonged high-stress tension of survival horror or competitive multiplayer games. Their rhythm often involves bursts of focused action followed by brief moments of lower tension (eg, traversing cleared sections), allowing for a more sustainable and restorative form of play.

XSL•FO
RenderX

Thus, Super Mario Bros. and Yoshi games can be a potent tool for temporary stress reduction, mood enhancement, and cognitive restoration, creating a mental "reset" that can foster a more optimistic and resourceful return to daily life. Their specific game design, combining skill demands with low real-world stakes, predictable rules, and positive affect induction, can make these games an effective vehicle for enhanced mental well-being. Super Mario Bros. and Yoshi games, hence, can act as an antidote by creating a virtual space that is the psychological inverse of a burnout-inducing environment.

Enhancing Childlike Wonder and Reducing Burnout Risk

Super Mario Bros. and Yoshi video games can offer pure, unadulterated joy and playfulness. More specifically, Super Mario Bros. games may help rekindle childlike wonder and help players see magic in the everyday. Jumping, running, breaking blocks, and collecting coins; these actions are inherently joyful and tactile. The satisfying "boing" of a jump, the "ching" of a coin, the smash of a brick; all these can trigger simple, primal pleasure. Mushrooms make players big, flowers shoot fire, stars make players invincible, turtles are cute enemies, clouds and trees have eyes! It presents a world operating on its own delightful, nonsensical logic, bypassing adult cynicism.

In addition, the vibrant colors, upbeat music, and generally sunny atmosphere (even underground or in castles has a playful menace) create an optimistic baseline. It is hard to stay cynical when that iconic tune is playing. In essence, Super Mario Bros. games act like a playful lens. It reminds players of the joy of pure discovery, the thrill of simple challenges, the magic of transformation, and the power of imagination that they possessed as children.

Furthermore, the visual and auditory design of Yoshi games is meticulously crafted to feel safe, warm, and inviting. Games such as *Yoshi's Crafted World* and *Yoshi's Woolly World* use a tangible, diorama-like aesthetic. Everything appears to be made of yarn, cardboard, and craft supplies. This evokes a sense of handmade charm and nostalgia, reminiscent of childhood arts and crafts. The world feels nonthreatening and playful. Yoshi games are bursting with joyful, bright colors, while the music is famously cheerful and melodic. Yoshi's vocalizations (the flutter jump and the egg spit) and the sounds of collectibles are satisfying and cute. This creates a consistent visual and audio landscape of positivity.

Burnout is characterized by exhaustion, cynicism (detachment or negativity), and reduced efficacy. Super Mario Bros. and Yoshi video gameplay can help reduce burnout risk in the following ways. First, the attentional demand of Super Mario Bros. and Yoshi gameplay provides a mental break from work or role-related rumination (a core driver of exhaustion). This allows depleted cognitive resources (eg, prefrontal cortex function taxed by decision fatigue) to replenish. Induced flow states and positive affect (dopamine or serotonin release) counter chronic stress responses (eg, lowered cortisol and reduced sympathetic nervous system activation). This provides physiological recovery crucial for combating exhaustion.

Furthermore, Super Mario Bros. and Yoshi games' rewarding mechanics and light-hearted aesthetic actively elevate mood, counteracting the persistent negative affect associated with burnout. That is, the game's core loop reinforces the belief that effort leads to reward. Collecting coins, finding secrets, or beating a level after multiple tries demonstrates a reliable, predictable link between action and positive outcome, rebuilding trust in effort that burnout erodes. Moreover, the games' narrative (perseverance leading to success) and aesthetics (colorful and playful) provide exposure to positive, noncynical representations, subtly challenging pervasive negative worldviews fostered by burnout.

Moreover, Yoshi games are also famously accessible. This removes the stress of failure and allows players to simply enjoy being in the world. The challenge comes from discovery and 100% completion, not from mere survival, making it a choice rather than a demand. While burnout is often fueled by anxiety about the future or rumination on the past, Yoshi games anchor players in the now.

Taken together, Super Mario Bros. and Yoshi games provide psychological restoration. The warm aesthetics, cheerful music, and a lack of intense pressure are not draining. They are replenishing. Playing feels like a break for the brain, similar to the effect of a walk in a pleasant park. The charm and kindness inherent in these games make it hard to be cynical. The games offer joyful sounds and enchanting visual feedback that help rebuild a sense of competence and agency in a low-pressure environment, countering the helplessness that fuels burnout.

Thus, in this study we aim to investigate if popular video games such as Nintendo's classic Super Mario Bros. and Yoshi franchise are able to promote players' childlike wonder—a state of openness, curiosity, and delight in the face of the new or mundane—and whether such experiences enhance overall happiness in life and reduce the risk of burnout among young adults.

Methods

Study 1

Research Design Overview

The findings of our qualitative study were reported in accordance with the Standards for Reporting Qualitative Research guidelines [37]. In addition, we adhered to the Journal Article Reporting Standards-mixed methods article reporting standards and offered information recommended for inclusion in research that reports the collection and integration of qualitative and quantitative data [38]. We conducted semistructured, cross-sectional interviews with open-ended, predetermined questions to explore and consider different personal opinions and views and help generate novel insights from data gathered in a natural setting [39,40]. Specifically, we invited study participants to share their thoughts about Super Mario Bros. and Yoshi games, and how, if at all, such a video game affects their day-to-day life. That is, during the interview, study participants were given an opportunity to reflect on and discuss how Super Mario Bros. and Yoshi games made them feel, what benefits (if any) they got from playing the game, and

XSL•FO
RenderX

how and why they played the game. In case study participants played more than one Super Mario Bros. and Yoshi game, they were invited to reflect on the game they played the most.

Researcher Description

To help minimize any potential bias, 4 trained research assistants (RAs) conducted the interviews. RAs were trained and instructed not to guide the interviewees and their responses in any way. The team of RAs was diverse in sex (2 women and 2 men) and ethnicity (from 3 different continents) to help ensure increased comfort with study participants during interviews.

Study Participants or Data Sources

Participant Recruitment

As part of our qualitative study, we used a convenience sampling method. Potential participants were recruited via posters and by handing out flyers to students on a university campus in July and August 2025.

Participant Selection

The specific eligibility criteria for taking part in the qualitative study were that participants were full-time students and that they had some personal experience playing a Super Mario Bros. or Yoshi game before.

Data Collection

Over the course of 3 weeks, exploratory, in-depth interviews were conducted with 41 participants (women: n=19, 46.3%; men: n=21, 51.2%; preferred not to disclose sex: n=1, 2.4%; mean age 22.51, SD 1.52 years). All interviewees were full-time students (confirmed by their student IDs) and had experience playing a Super Mario Bros. or Yoshi game (screening questions by the RAs included the name of the specific Super Mario Bros. or Yoshi games respondents had played and which console they used to play the game, eg, Wii U, Nintendo Switch, or Nintendo Switch 2). Interviews were conducted in a university cafeteria and lasted between 25 and 40 minutes. All interviewees were reassured of the anonymity of their responses and were informed that their participation would help inform academic research. At the end of the interview, each interviewee was given a chance to ask any questions they may have had.

Statistical Analysis

We examined our qualitative data gathered from the interviews, following recommended and established procedures for qualitative data analyses [40-42]. RAs had a chance to familiarize themselves with the data collected by transcribing the interviews and carefully reading the transcripts. Furthermore, the coding of interview data was conducted in duplicate by the authors, and any potential discrepancy was discussed and resolved. In doing so, we followed the constant comparison technique in prior published work, collected and analyzed interview data in tandem, and continuously compared the newly collected data against the existing insights. As a first step of our analysis, we grouped conceptually linked data and reduced them to a set of meaningful concepts. That is, we examined and analyzed the data to develop insights and a holistic understanding of interviewees' responses and key elements of Super Mario Bros. or Yoshi games that were noted to affect their lives.

In addition, we used descriptive phrases and open coding to develop further insights regarding key drivers and potential process explanations about interviewees' reflections of Super Mario Bros. or Yoshi games and critical elements of gameplay that may influence their overall happiness in life and burnout risk. Moreover, in joint discussions by the team of RAs and authors, we integrated relevant literature in an iterative process, further explored key elements of Super Mario Bros. or Yoshi games that influenced players' lives, and finalized the key elements of our conceptual framework.

Study 2

Research Design

As part of study 2, we conducted a cross-sectional survey using a convenience sampling approach.

Inclusion and Exclusion

Participants were deemed eligible to take part in the study if they were full-time university students at the time of completing the study and if they had personal experience playing a Super Mario Bros. or Yoshi game. We conducted the study in July and August 2025 and followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines to report the observational, cross-sectional studies (Multimedia Appendix 1) [43].

Sampling Procedures

Potential study participants were recruited by handing out flyers and using posters on a university campus in July and August 2025. Trained RAs informed potential study participants that the study would take no longer than 15 minutes and that each participant would receive US $5 (approximately £4) as compensation for taking part in the study.

Sample Size, Power, and Precision

Over the course of 5 weeks, 4 RAs helped with the data collection (student ID check and simple screening questions before the study, ie, which Super Mario Bros. or Yoshi game participants had played before and which console they had used for gameplay, such as Wii U, Nintendo Switch, Nintendo Switch 2, etc) and managed to recruit a total of 350 participants. Because 14 participants did not complete the entire survey, their responses were discarded from the study, and the effective number of responses was 336. In case study participants noted playing more than one Super Mario Bros. or Yoshi game, they were encouraged to consider and think of the game they played the most when answering the study's survey questions. RAs invited participants to complete a brief paper-and-pencil questionnaire about gaming and their daily life activities.

Participant Characteristics

The final sample consisted of 336 full-time university students (women: 19/41, 46.3%; men: 21/41, 51.2%; prefer not to disclose sex: 1/41, 2.4%; mean age 22.51, SD 1.52).

Measures and Covariates

Specifically, participants answered questions about burnout risk that were adapted from validated, previously published scales designed to measure burnout risk [44,45], and included items

XSL•FO
RenderX

such as "I feel emotionally drained by my daily responsibilities (caregiving, work, etc)," "I have become more cynical or detached about activities I used to care about," and "I doubt whether my efforts actually make a difference").

In addition, participants were invited to answer questions to measure their overall happiness in life. Measurement items to capture life happiness were adapted from previously published happiness in life scales [46,47] and consisted of 3 items ("I feel grateful for the good things in my life," "I am satisfied with the overall direction of my life and look forward to what lies ahead," "Overall I would describe myself as a happy person"). Moreover, we captured the extent to which gameplay enhanced participants' childlike wonder by adapting three items from previously published scale items on noticing novelty and joy in discovery, feeling wonder in response to stimuli [48-50] ("Playing a Super Mario Bros./Yoshi game helps me feel delighted by small, ordinary things that others overlook (e.g., shapes in clouds, patterns in pavement)," "Super Mario Bros./Yoshi gameplay encourages me stop to admire or explore things that spark my curiosity, even if they seem unimportant (e.g., a hidden path, an unusual insect)," "Playing a Super Mario Bros./Yoshi game encourages me to pause and to appreciate playful or whimsical details in my environment"). All responses were measured on a 7-point Likert scale (1=strongly disagree, 7=strongly agree). As control, participants indicated their sex (1=female, 2=male, 3=prefer not to disclose).

To ensure the quality of measurement, all scale items were reviewed by the research team for clarity and consistency before survey administration. The questionnaire was distributed in printed form on campus and completed individually by respondents. RAs who facilitated recruitment and survey distribution were not informed of the study's hypotheses and followed standardized instructions to minimize expectancy or response bias. Psychometric properties of the measures were examined using confirmatory factor analyses and reliability testing. All factor loadings were significant (P<.001). Cronbach α, composite reliability (CR), and average variance extracted (AVE) all exceeded widely accepted thresholds ($\alpha \geq .70$, CR ≥ 0.70, and AVE ≥ 0.50), indicating strong internal consistency and convergent validity [51-53]. The Fornell-Larcker criterion further supported discriminant validity across constructs [53].

Statistical Analysis

We followed recommended procedures for conducting confirmatory factor analyses and examined the factor structure and reliability scores for the measurement items in our survey study [51,52]. This was done to examine whether the items significantly loaded on their intended factors and had weak cross-loadings. Furthermore, we examined the AVE and CR scores for the scales used in our study. Moreover, we tested whether the average variance extracted for each construct was greater than the squared correlations between the construct and other variables in our survey [53]. The hypothesized 3-factor model (childlike wonder, happiness, and burnout risk) was tested using confirmatory factor analysis in IBM SPSS AMOS 29.0 with maximum likelihood estimation.

In addition, we conducted a mediation analysis with the PROCESS macro [54] (model 4; bootstrapped samples=5000; 95% CI) to examine the influence of Super Mario Bros. and Yoshi games' affordance of childlike wonder (independent variable) on overall happiness in life (mediator) and, consequently, on players' burnout risk (dependent variable). In order for PROCESS not to treat sex as a continuous variable, sex was coded as a dummy variable (0=female, 1=male). Due to the insufficient sample size (n=1), the "prefer not to disclose" category was excluded from analysis.

Ethical Considerations

Study 1

Ethics approval for the study was received from Imperial College London's Analytics, Marketing, and Operations (AMO) Department. At the time of data collection, the AMO Department followed Imperial College London's ethics protocol for research involving human participants, which includes internal review and written approval by the departmental head of the research group. The procedure ensures compliance with Imperial College Research Ethics Committee guidelines and the principles of the Declaration of Helsinki. Informed consent was obtained from all participants, and participants were informed that they had the right to withdraw from the study at any time point. Study data were anonymized and kept confidential within the research team. More specifically, we ensured that no identification of individual study participants in any images or supplementary material of the manuscript was possible. After the interview was completed, each participant was thanked, debriefed about the study's purpose, and given £15 (approximately US $20) as compensation for their time.

Study 2

For study 2, ethics approval was received from Imperial College London at the Departmental level, Analytics, Marketing, and Operations. The AMO Department follows Imperial College London's ethics protocol for research involving human participants, which includes internal review and written approval by the departmental head of the research group. This procedure ensures compliance with the Imperial College Research Ethics Committee guidelines and the principles of the Declaration of Helsinki. All data collected were anonymized and kept confidential within the research team. Study participants provided informed consent and were informed that they had the right to withdraw from the study at any time and without having to give a particular reason. We ensured that no identification of individual study participants in any supplementary material or in any images of the manuscript was possible. After completing the survey, each participant was thanked, debriefed, received £4 (approximately US $5) as compensation for their time, and was invited to ask any questions they may have had about the study.

Results

Study 1: Overview

The findings of the in-depth interviews demonstrated that Super Mario Bros. and Yoshi games' affordance of childlike wonder played an important role in their impact on players' lives. In Table 1, we list the representative quotes, which indicate the extent to which players viewed Super Mario Bros. and Yoshi

XSL•FO
RenderX

games as fostering childlike wonder and the influence they had on their day-to-day life. In addition, interviewees noted how Super Mario Bros. and Yoshi gameplay affected their overall happiness in life as well as how gameplay helped reduce their risk of burnout (refer to Table 1 for representative quotes).

Participants frequently described the bright colors and optimistic tone of *Super Mario Bros. Wonder* as evoking feelings of calm and childlike joy. Figure 1 illustrates this visual atmosphere, depicting the cheerful landscape that several interviewees referred to when recalling moments of wonder and relaxation.

Interviewees highlighted how the game's design balances challenge with playfulness, creating a sense of ease that contrasts with daily pressures. Figure 2 exemplifies this tone, showing even "enemies" portrayed in a gentle, nonthreatening way—an aspect the interviewees associated with relaxation and stress relief.

Interviewees often described *Yoshi's Crafted World* as a warm, inviting environment that felt comforting and safe. Figure 3 illustrates the handcrafted, diorama-like aesthetic that many associated with nostalgia and emotional restoration.

XSL•FO
RenderX

Table 1. Representative quotes of interviewees in an in-depth qualitative study about the key effects of Super Mario Bros. and Yoshi games and their influence on their lives.

Game key effects	Representative quotes
Childlike wonder	• "Playing Yoshi transports me back in time. It's a bit like time travel. I see myself as this kid again, wondering about the possibilities and beauty of this world. It's a bit like a fresh new start." [Respondent 8] • "Super Mario Bros. Wonder is crazy. I love how the clouds are all animated and full of life. I hardly see the sky these days. Living in a big city it's hard to get a sense of nature. Deep down that certainly must affect people. We did not evolve over thousands of years to run around like rats underground, squashed in hot and cramped, tight spaces like the subway and compete for the tiniest of spaces. Super Mario Bros. Wonder is like a breath of fresh air with trees, clouds, water and fun everywhere. It makes me feel I am part of nature and nature is part of me. I miss that feeling. When I play Super Mario Wonder it is as if I'm a kid again." [Respondent 13] • "Haha, the moment I switch on and the game starts with the iconic Super Mario tune in my head I sense a moment of relief and I am playing the game as I used to when I was a child. Probably that was the happiest time of my life when I was a child. Super Mario gives me a chance to appreciate the world and everything again with the eyes of a child that is not jaded, not tired but full of energy, full of enthusiasm, interest, and marveling at what surrounds us and what each new day may bring." [Respondent 16] • "Playing with friends or just by myself lying on the bed, Super Mario Bros. games make me feel warm and fuzzy and comfortable. A bit like being wrapped in a cozy, warm blanket on a cold winter day when it is snowing outside and as a kid you are dreaming about the snowflakes and wondering how much snow may fall today. Most days I feel exhausted and no energy to appreciate things anymore. Playing Super Mario gives me this new sense of being a kid with wide, interested eyes wanting to learn and explore again." [Respondent 30] • "There is nothing too fancy about Super Mario games. Come to think about it, that's what probably makes the Super Mario Bros. games so special. They hold your hand and let you find that first super mushroom, make that first jump, all in a way that feels like play. It is not a chore and does not feel like a nuisance but it's a game! The colors are so vibrant and fun. The music and tunes are cheerful and put a smile on my face every single time. I play and I cannot help but wonder about… small things. There is this fly on my window. But I don't mind. Maybe the fly will find a way out. I hear the rain outside and I keep looking at the clouds. I think of the clouds in the game and smile. The game makes me feel like a kid in a candy store. You keep looking at all the sweets with amazement, just wondering which one of these incredible candies should be the one to be picked? What a treat!" [Respondent 39] • It is easy to dismiss this but then I feel it is true. Playing Yoshi games makes me feel like… hey, don't take everything so seriously out there. You see, it is easy to get lost in all sorts of worries and anxieties, stress and negativity. I don't see the flowers, the sky, the clouds, the sunset or the sunrise. I have become blind to the beauty of well everything really. But Yoshi video games open my eyes again to that way of looking at the world like a kid in a candy store. Running full speed on top of the clouds? … Hahaha, let's a go!" [Respondent 40]

XSL•FO
RenderX

Game key effects	Representative quotes
Overall happiness in life	• "Look at your phone and everyone tells you to buy more, to chase the big dream and exotic destinations and you will get happy there. Only to keep chasing, keep buying more, travel to no end to get lost, and all the while happiness may be just around the corner. Super Mario games give me that feeling that happiness is truly in the small things in life. It does not take a trip to the Maldives. What are you chasing? Stop chasing, it's right there in front of your eyes. You don't have to spend millions to feel happy. Do you know what's crazy? Playing my Super Mario on the Nintendo Switch makes me think back of the time when I played Super Mario waiting in the dentist's waiting room. And you go, hold on a minute, the dentist room?!? That's the least likely place to feel happiness and you are right. But I still remember, I am sitting there and I completed the game. I completed the game and listened to the credits roll on and that iconic theme… that music… at that moment, I felt happiness. This is the pure magic of playing Super Mario. You can be sitting in a dentist's waiting room and feel happiness. That's truly awesome. Super Mario is simple, pure, happiness." [Respondent 2] • "So I often think about what it means to be happy. There are zillion of books on this and the internet has even more answers. Everyone seems to know what will definitely make you happy. You need a house. You need to run that marathon. You need to look like this and buy that. When was the last time I felt truly happy. I mean truly happy? It's when I played Super Mario Bros. Wonder and I switched on and was there in the moment. When I jumped, I jumped, and when I ran, I ran. That was happiness." [Respondent 17] • "The most precious resources we may have are time, health, knowledge and skill, family and the most previous things in life are free like health. Social media tells me I need to dream big and do big things to achieve big things, everything has to be crazy big. Happiness may be in small things, in the quiet moments, in the time when I am allowed to be myself and just be there and appreciate the moment I have. Super Mario is this simple happiness that made me feel so… happy, there is no other word for it, when I was a kid. Happiness as a kid did not have to be complicated. Happiness was completing a level. Happiness was spending the summer holidays doing nothing special or grand but just enjoying the life one had." [Respondent 25] • "I love level one in my Super Mario Bros. U Deluxe. The grass is green and the clouds and the sky look so cheerful and endearing. It makes me feel I want to walk through the woods, climb trees, balance, and jump off. Or maybe I should lie on the grass and watch the clouds fly by high in the sky waiting for rain to come or a summer thunderstorm or even rolling thunder and snow? Super Mario Bros. U Deluxe reminds me that I am part of nature and it's here for me. Jumping, climbing, jumping, picking up a leaf, a flower. Nature is not just an amazing backdrop in Super Mario, it's integral to the game and how it makes me smile again." [Respondent 26] • "I kept jumping and falling. Jumping and falling all the time. Finally I made the jump and I was over the moon. What a feeling of happiness. I still remember it. That Super Mario theme is pure, catchy video game magic and happiness for the tired soul! Cannot feel not happy when listening to it. It's not a big, dramatic piece. Super Mario games are like that. Happiness in the simplest and clean form." [Respondent 33] • "A while ago I did have the strangest of feelings. I was sitting on my sofa and played Super Mario when I looked outside the window and spotted the moon shining. I haven't seen the moon for ages. I mean ages. And there it was shining and smiling at me. At least I felt it was smiling at me. I even discovered a few stars. Super Mario is a magical treat for the soul. The game does not shout but it gently holds you and lets you explore by yourself. Why would playing a video game make me appreciate my natural surroundings more? It is not obvious to me. Probably on the subconscious level more than anything Super Mario allows me to be part of nature how humans did for centuries, running around, trying to get a lucky break and find a leaf, flower, mushroom that gives them some much needed energy boost, jumping, exploring, avoiding danger, and finally making it back home." [Respondent 34] • "Keep collecting those coins of joy. Power-up with little moments of happiness and never stop looking for hidden blocks  . That's what Super Mario taught me more than anything." [Respondent 37]
Burnout risk reduction	

XSL•FO
RenderX

Game key effects	Representative quotes
	• "Yoshi games help me see the world and my life with softness, playfulness and wonder. The fact that it allows me to appreciate the magic in something as simple and joyful as Super Mario is what brings me joy. Never underestimate the power of play. Life can feel meaningless and empty. Whether it's jumping as Yoshi or noticing the way light hits a glass window, you are actively co-creating a kinder, more magical world just by choosing to see it that way. This may be Yoshi's biggest super power." [Respondent 8] • "Sad how cynical life has become. Do you know when it all started? What is the cause of this? Everyone is cynical and tired of life. It's almost as if humans have given up already. Tragic really when to think about it. When I play my Nintendo Yoshi games I feel less anxious and worry and the crises in my life seem that little bit further away. That moment of inner peace is priceless." [Respondent 11] • "It is hard to feel excited and energized how I used to. Life just keeps happening to you. It's everywhere you look though. Yoshi, Princess Peach, Super Mario and the whole gang brighten my days just keep on running and jumping up that flagpole as high as you can. It gives me a bit more energy and less worry, a bit more strength to keep going another day and not keep thinking about all the things." [Respondent 28] • "Burnout may be the shadow of deep care. If you feel its weight. Breathe. Remember you don't need to carry the wonder. You are the wonder. Mario is our fellow traveler in this journey. Super Mario games allow me to see the magic in small things. It's a true gift to feel life has still something in store and not being tired of it all." [Respondent 33] • "Do you sometimes feel tired of it all? I do. What helps me is to step out and do things that give me space to stop worrying and overthinking things. For example when I play my favorite Yoshi on my Nintendo Switch and get some energy back and motivation back in life. Not everything seems lost anymore when you got Yoshi to cheer you on." [Respondent 32] • "Weird how pointless life can be or feel that way. The news gives me thousand things happening at the same time. So much violence, war, hate, abuse, disconnection, all the time. I need a break from this. It just makes me feel exhausted and I cannot make a difference anyway. I am not going to stop that war or people from acting selfishly. I feel lost and sad. Super Mario gives me back some happiness and moment of when I can enjoy the time so I do not lose hope entirely or willingness to do anything. It's a happiness booster that helps me keep going a bit longer and keep hanging in there. It's nice to feel life still has some happy moments to offer and not just trouble. To level up and keep looking forward to the next adventure." [Respondent 41]

Figure 1. Cheerful landscape in *Super Mario Bros. Wonder* (reproduced with permission from Nintendo).



XSL•FO
RenderX

Figure 2. Super Mario facing sleeping enemies in *Super Mario Bros. Wonder* (reproduced with permission from Nintendo).



Figure 3. The bright and cheerful environment of *Yoshi's Crafted World* (reproduced with permission from Nintendo).



Study 2: Statistical Analysis Results

Detailed survey measurement items and reliability scores are listed in Table 2. Confirmatory factor analyses showed that all measurement items significantly loaded on their intended factors (>0.79; P<.001) with weak cross-loadings (<0.27). In addition, AVE and CR scores exceeded the recommended thresholds. In addition, the Cronbach α scores were generally high and results showed that the AVE for each construct was greater than the squared correlations between the construct and other variables in our survey in support of discriminant validity (Table 2). The model demonstrated good fit to the data (χ^2_{24}=30.6, Tucker-Lewis index=0.99, comparative fit index=0.99, normed fit index=0.98, and root mean square error of approximation=0.03; 90% CI 0.00-0.06; P=.17). Estimation of standardized residuals showed that no values were particularly large (>2.00), indicating no substantial localized areas of misfit.

XSL•FO
RenderX

Table 2. Cross-sectional survey measurement items' factor loadings and reliabilities.

Survey study constructs and measurement items	Factor loadings
Burnout risk[a]	
I feel emotionally drained by my daily responsibilities (caregiving, work, etc)	.84
I have become more cynical or detached about activities I used to care about	.79
I doubt whether my efforts actually make a difference	.87
Happiness in life[b]	
I feel grateful for the good things in my life	.85
I am satisfied with the overall direction of my life and look forward to what lies ahead	.80
Overall I would describe myself as a happy person	.87
Childlike wonder[c]	
Playing a Super Mario Bros/Yoshi game helps me feel delighted by small, ordinary things that others overlook (eg, shapes in clouds, patterns in pavement)	.92
Playing a Super Mario Bros/Yoshi game encourages me to pause and to appreciate playful or whimsical details in my environment	.90
Super Mario Bros/Yoshi gameplay encourages me to stop to admire or explore things that spark my curiosity, even if they seem unimportant (eg, a hidden path, an unusual insect)	.91

[a]α=.83; composite reliability (CR)=0.87; average variance extracted (AVE)=0.70.

[b]α=.86; CR=0.88; AVE=0.71.

[c]α=.92; CR=0.94; AVE=0.83.

Study 2: Mediation Analysis Results

The results of the mediation analysis with PROCESS macro [54] (model 4; bootstrapped samples=5000; 95% CI; sample size=336) showed that Super Mario Bros. and Yoshi games' affordance of childlike wonder (independent variable) had a significant, positive influence on players' overall happiness in life (b=0.30, SE=0.04, *t*=6.8, 95% CI 0.21-0.38; *P*<.001; Table 3). In turn, overall happiness in life significantly reduced players' burnout risk (b=–0.48, SE=0.05, *t*=–9.55, 95% CI –0.57 to –0.38; Table 4). Gender had no significant effect on happiness (b=–0.05, SE=0.12, *t*=–0.42, 95% CI –0.29 to 0.19; *P*=.68; Table 3) and no significant effect on burnout (b=0.02, SE=0.11, *t*=0.14, 95% CI –0.20 to 0.23; *P*=.89; Table 4). Furthermore, the results showed that happiness fully mediated the impact of childlike wonder on burnout, as the direct effect of childlike wonder on burnout risk became insignificant (b=–0.08, SE=0.04, *t*=–1.88, 95% CI –0.16 to 0.01; Table 4), while the indirect effect of childlike wonder on burnout risk was significant (b=–0.14, BootSE=0.03, 95% CI –0.20 to –0.09).

Table 3. Cross-sectional survey mediation regression results: effect of childlike wonder on happiness.

Predictor	b (95% CI)	SE	*t* test	*P* value
DV[a]**: Happiness**[b]				
Constant	2.50 (2.07 to 2.93)	0.22	11.43	<.001
Childlike wonder	0.30 (0.21 to 0.38)	0.04	6.79	<.001
Gender	–0.05 (–0.29 to 0.19)	0.12	–0.42	.68

[a]DV: dependent variable.

[b]R=0.35, R^2=0.12, mean squared error=1.43, $F_{2, 333}$=23.16; *P*<.001.

XSL•FO
RenderX

Table 4. Cross-sectional survey mediation regression results: effect of happiness on burnout.

Predictor	b (95% CI)	SE	*t* test	*P* value
DV[a]: Burnout[b]				
Constant	5.37 (4.91-5.83)	0.23	22.99	<.001
Childlike wonder	–0.08 (–0.16 to 0.00)	0.04	–1.88	.06
Happiness	–0.48 (–0.57 to –0.38)	0.05	–9.56	<.001
Gender	0.02 (–0.20 to 0.23)	0.11	0.14	.89

[a]DV: dependent variable.

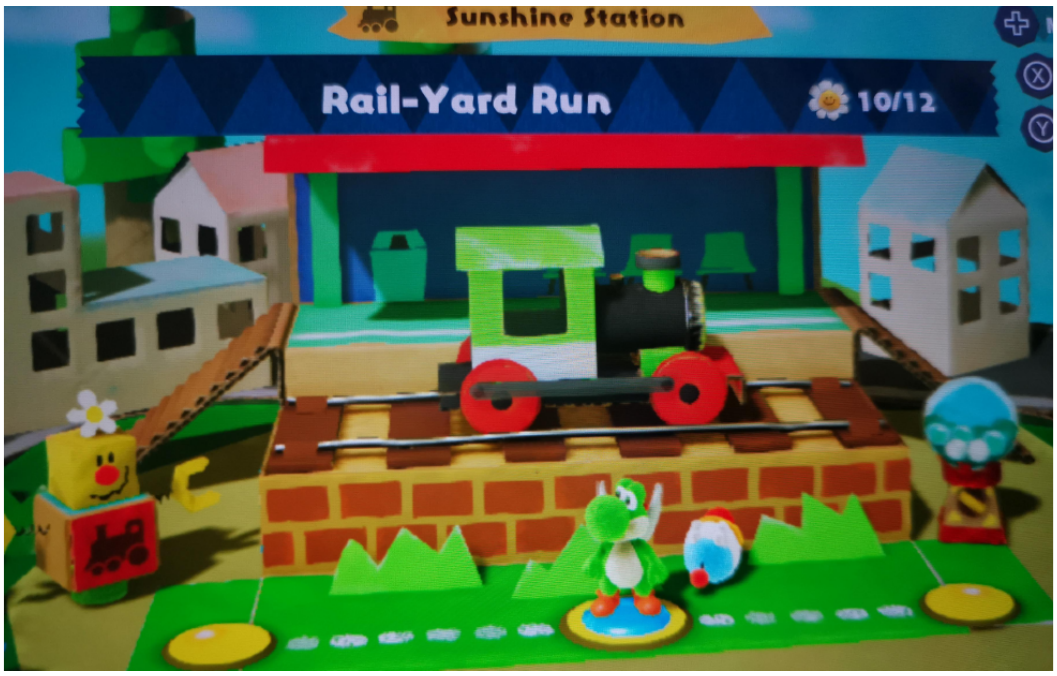
[b]R=0.52, R^2=0.27, mean squared error=1.17, $F_{3, 332}$=40.80; P<.001.

Discussion

Principal Findings

The findings of this study show that players appreciated Super Mario Bros. and Yoshi games' affordance of childlike wonder as well as enhanced feelings of happiness. In addition, the findings of the study indicated the extent to which players noted Super Mario Bros. and Yoshi games' ability to reduce burnout risks. The findings of the quantitative study further showed that Super Mario Bros. and Yoshi games' affordance of childlike wonder was positively associated with people's overall happiness in life. Finally, happiness was negatively associated with burnout risk of study respondents, hence indicating the potential for popular video games such as Super Mario Bros. and Yoshi games to help reduce burnout risk. The colorful, low-pressure worlds of Yoshi games were described by interviewees as a "vacation for the mind." Figure 4 illustrates this atmosphere, conveying the playful and uplifting tone that helps transform ordinary play into a restorative emotional experience.

Figure 4. Joyous scenery in *Yoshi's Crafted World* (reproduced with permission from Nintendo).



Comparison to Prior Work

Prior work has noted the role video games can play in people's lives, including enhanced cognitive escapism and relaxation [28-30]. Beyond escapism, prior research using Self-Determination Theory indicates that games can support basic psychological needs for autonomy, competence, and relatedness, which are associated with intrinsic motivation and need-contingent, short-term improvements in well-being [55]. Players have also been noted to engage with games strategically for mood management and coping-based emotion regulation, selecting particular gameplay challenges or mechanics to reduce stress, elevate mood, or regain a sense of control [56-58]. How popular video games such as Super Mario Bros. and Yoshi may affect young adults' ability to look at the world and their lives in renewed, childlike wonder and how this may affect their overall happiness and reduce risk of burnout, however, remains unanswered. Moreover, nostalgia has been noted to promote well-being by drawing on cherished past experiences to restore positive affect, meaning, and a sense of connection [59], and mainstream, long-running franchises often cue players' nostalgia, especially when they prompt recall of past play experience with familiar series, settings, or characters [60]. However, nostalgia alone may not fully explain the uplifting, playful stance reported by players of whimsical franchises such as Super Mario and Yoshi, pointing to a distinct experiential

XSL•FO
RenderX

state-childlike wonder-that has not yet been systematically examined.

While prior research examined how various gamification techniques may enhance people's feelings of hope or connection with a new digital offering [61-64], critical questions remained about the role highly popular video games, such as Super Mario Bros. and Yoshi, may play in reducing people's burnout risk. Although previous studies shed light on designing offerings that make people's lives easier, put a smile on their face, and possibly inspire them and help form a stronger bond with them [65-70], work has only recently begun to examine how specific content and structural features of games shape emotional outcomes. For example, game aesthetics such as visual style, world design, music, and pacing have been proposed to elicit both hedonic enjoyment and deeper, reflective emotions [71,72]. Likewise, soundtrack music in games has been shown to shape players' emotional responses, for example, by intensifying positive emotions and enhancing horror experiences [73]. In addition, in-game success has been found to predict mood repair and satisfaction of competence and autonomy to mediate effects on enjoyment [74]. Notably, these mechanisms have seldom been linked to broader indicators of well-being, such as sustained happiness or resilience to burnout, particularly among young adults.

Recent reviews suggest that commercial video games show promise for alleviating symptoms of anxiety and depression [2,75]. While these findings are encouraging, the present study takes a different perspective: rather than positioning entertainment games as treatments, it considers whether everyday play can function as a subtle, restorative micro-environment. Thus, the current study contributes to work examining the influence of gaming on mental well-being and, in doing so, further helps bridge gaming and mental health science [76,77] by identifying childlike wonder as a unique emotional pathway through which mainstream games may support happiness and reduce burnout risk.

Limitations

The benefits of video gameplay, including Super Mario Bros. and Yoshi, are likely to be contingent on moderate, voluntary play primarily for enjoyment. Compulsive video gameplay or excessive escapism used solely to avoid real problems can be maladaptive and undermine optimism by neglecting real-world agency. Moreover, playing Super Mario and Yoshi games can help mitigate individual experiences of burnout, but does not address systemic causes (eg, excessive workload, unfair practices). Thus, it may complement, not replace, organizational interventions. Moreover, in this study, we examined self-reported burnout risk. Future research examining the long-term effects of gameplay on actual burnout risk is richly deserving. In addition, we encourage future work to study the potential effects of gameplay on people of varying ages, demographic backgrounds, etc., and examine the influence that such gameplay may have on various aspects of people's overall well-being. Examining such variations could reveal how developmental context and lived experience shape how popular games relate to well-being.

Future Research

The specific design elements of Super Mario Bros., as well as Yoshi games and their focus on skill-based mastery, low violence, predictable rules, achievable goals, and positive aesthetics, may be crucial for fostering the beneficial aspects described in this study. Not all video games offer this combination. Future research that further investigates the potential positive and negative effects of video gameplay on human happiness, resilience, and empathy is richly deserving. Current work examines the extent to which internet use may affect people's anxiety and stress [78,79]; however, there is also work suggesting offerings to put a smile on users' faces and make their lives easier [80-84].

Because additional interviews continued to offer new and nuanced insights, saturation in our qualitative study may not have been achieved, and we invite future research to explore and further investigate the role of popular games in people's lives and their potential impact on the resilience, coping ability, physical and emotional health, as well as overall well-being of people across ages and various lifestyles. As previous research indicates an increasing number of people searching for meaning in their lives and feeling detached from nature as well as fellow human beings [85-88], additional work studying the effects of popular video games on people's empathy, perceived loneliness, biophilia, as well as gratitude and levels of overconsumption is richly deserving.

Nintendo's Super Mario Bros., as well as Yoshi games, can be more than just platformers. They are carefully constructed wonder simulators. They reject the high-stakes, high-stress models of many modern games in favor of an experience that prioritizes safety, curiosity, and joy. For a person whose life may be dominated by deadlines, overwhelming demands, and a perceived loss of control, stepping into Super Mario's or Yoshi's crafted world can be a form of digital therapy. It allows the person to practice being curious again, to experience uncomplicated joy, to succeed in small, measurable ways, and to simply exist in a moment of bright, colorful, and kind-hearted joy.

These games, in essence, are a vacation for the mind, facilitating the very childlike wonder that burnout so effectively extinguishes. Intentional game design can do more than just entertain - it can heal, restore, and inspire. Future research further exploring the various ways gameplay may affect people's outlook and approach to life is richly deserving.

Conclusions

Insights from the qualitative study indicated players' appreciation of Super Mario and Yoshi games' affordance of childlike wonder, happiness in life, and a reduction in burnout risk. In addition to this, the findings of the quantitative analyses showed how Super Mario and Yoshi games' affordance of childlike wonder influenced happiness in life and significantly reduced burnout risk. These video games, hence, can be effective vehicles for enhanced mental well-being. More specifically, the current findings show that Super Mario Bros. and Yoshi games act as an effective antidote by creating a virtual space that is the psychological inverse of a burnout-inducing environment.

XSL•FO
RenderX

In conclusion, the findings for this study demonstrate that popular video games, such as Super Mario Bros. and Yoshi, can enhance childlike wonder, which in turn increases overall happiness and reduces burnout risk. This work is innovative in identifying childlike wonder as a previously underexplored emotional mechanism that links everyday gameplay with improved well-being. It differs from prior studies that have focused mainly on escapism, stress relief, or need satisfaction by showing how mainstream entertainment games can evoke curiosity, playfulness, and joy that counteract cynicism and exhaustion. The study contributes to the growing field connecting gaming and mental health by positioning popular titles as restorative digital spaces rather than merely sources of distraction. Practically, the findings suggest that thoughtfully designed, low-pressure games may serve as simple, real-world tools to rekindle curiosity, uplift mood, and support resilience among young adults facing burnout.

Acknowledgments

The authors thank all the study participants for their time. In addition, the authors would like to thank Nintendo for permitting us to take photos of their Super Mario and Yoshi games for inclusion in the manuscript. The authors used DeepSeek R1 for spelling and grammar checks of the manuscript.

Funding

No external financial support or grants were received from any public, commercial, or not-for-profit entities for the research, authorship, or publication of this article.

Data Availability

The data for the study are available from the authors on reasonable request.

Authors' Contributions

WT was responsible for methodology, data collection, formal analysis, investigation, writing, and reviewing and editing the manuscript. CH was responsible for methodology, data collection, formal analysis, investigation, writing and editing the manuscript, and supervision. ABE was responsible for conceptualization, methodology, data collection, formal analysis, investigation, writing, reviewing, and editing the manuscript, and supervision.

Conflicts of Interest

None declared.

Multimedia Appendix 1
STROBE checklist for cross-sectional studies.
[PDF File (Adobe PDF File), 179 KB - games_v13i1e84219_app1.pdf]

References

1. Hu F, Tai Z, Liu J. The positive effect of video-game play on college students' anxiety and depression symptoms during the COVID-19 pandemic shelter-in-place lockdowns: mixed methods study. JMIR Serious Games 2025;13:e58857 [FREE Full text] [doi: 10.2196/58857] [Medline: 40446291]
2. Kowal M, Conroy E, Ramsbottom N, Smithies T, Toth A, Campbell M. Gaming your mental health: a narrative review on mitigating symptoms of depression and anxiety using commercial video games. JMIR Serious Games 2021;9(2):e26575 [FREE Full text] [doi: 10.2196/26575] [Medline: 34132648]
3. Potts C, Kealy C, McNulty J, Madrid-Cagigal A, Wilson T, Mulvenna M, et al. Digital mental health interventions for young people aged 16-25 years: scoping review. J Med Internet Res 2025;27:e72892 [FREE Full text] [doi: 10.2196/72892] [Medline: 40344661]
4. Zhou L, Parmanto B. Development and validation of a comprehensive well-being scale for people in the university environment (Pitt Wellness Scale) using a crowdsourcing approach: cross-sectional study. J Med Internet Res 2020;22(4):e15075 [FREE Full text] [doi: 10.2196/15075] [Medline: 32347801]
5. Mistry R, Elenbaas L. It's All in the family: parents' economic worries and youth's perceptions of financial stress and educational outcomes. J Youth Adolesc 2021;50(4):724-738 [FREE Full text] [doi: 10.1007/s10964-021-01393-4] [Medline: 33515373]
6. Haidt J. The Anxious Generation: How the Great Rewiring of Childhood Is Causing an Epidemic of Mental Illness. New York, NY: Penguin; 2024.

XSL•FO
RenderX

7. Hartnup B, Dong L, Eisingerich A. How an environment of stress and social risk shapes student engagement with social media as potential digital learning platforms: qualitative study. JMIR Med Educ 2018;4(2):e10069 [FREE Full text] [doi: 10.2196/10069] [Medline: 30006324]
8. Arnett JJ. Emerging adulthood: A theory of development from the late teens through the twenties. American Psychologist 2000;55(5):469-480. [doi: 10.1037/0003-066x.55.5.469]
9. Jones PB. Adult mental health disorders and their age at onset. Br J Psychiatry Suppl 2013;54:s5-s10. [doi: 10.1192/bjp.bp.112.119164] [Medline: 23288502]
10. Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the national comorbidity survey replication. Arch Gen Psychiatry 2005;62(6):593-602. [doi: 10.1001/archpsyc.62.6.593] [Medline: 15939837]
11. Schulenberg JE, Maggs JL. A developmental perspective on alcohol use and heavy drinking during adolescence and the transition to young adulthood. J Stud Alcohol Suppl 2002(14):54-70. [doi: 10.15288/jsas.2002.s14.54] [Medline: 12022730]
12. Anto A, Asif R, Basu A, et al. Exploring the impact of social media on anxiety among university students in the United Kingdom: qualitative study. JMIR Form Res 2023;7:e43037 [FREE Full text] [doi: 10.2196/43037] [Medline: 37327030]
13. Plackett R, Blyth A, Schartau P. The impact of social media use interventions on mental well-being: systematic review. J Med Internet Res 2023;25:e44922 [FREE Full text] [doi: 10.2196/44922] [Medline: 37565693]
14. Karila L, Scher N, Draghi C, et al. Understanding problematic smartphone and social media use among adults in france: cross-sectional survey study. JMIR Ment Health 2025;12:e63431 [FREE Full text] [doi: 10.2196/63431] [Medline: 40047899]
15. Han X, Zhou E, Liu D. Electronic media use and sleep quality: updated systematic review and meta-analysis. J Med Internet Res 2024;26:e48356 [FREE Full text] [doi: 10.2196/48356] [Medline: 38533835]
16. Herriger C, Merlo O, Eisingerich A, Arigayota A. Context-contingent privacy concerns and exploration of the privacy paradox in the age of AI, augmented reality, big data, and the internet of things: systematic review. J Med Internet Res 2025;27:e71951 [FREE Full text] [doi: 10.2196/71951] [Medline: 40367513]
17. Chhabra J, Pilkington V, Benakovic R, Wilson M, La Sala L, Seidler Z. Social media and youth mental health: scoping review of platform and policy recommendations. J Med Internet Res 2025;27:e72061 [FREE Full text] [doi: 10.2196/72061] [Medline: 40540734]
18. Kesner L, Juríčková V, Grygarová D, Horáček J. Impact of media-induced uncertainty on mental health: narrative-based perspective. JMIR Ment Health 2025;12:e68640 [FREE Full text] [doi: 10.2196/68640] [Medline: 40513096]
19. Kermavnar T, Visch V, Desmet P. Games in times of a pandemic: structured overview of COVID-19 serious games. JMIR Serious Games 2023;11:e41766 [FREE Full text] [doi: 10.2196/41766] [Medline: 36634265]
20. Selim R, Basu A, Anto A, Foscht T, Eisingerich A. Effects of large language model-based offerings on the well-being of students: qualitative study. JMIR Form Res 2024;8:e64081 [FREE Full text] [doi: 10.2196/64081] [Medline: 39729617]
21. Anto A, Asif RO, Basu A, et al. Exploring the facilitators and barriers to addressing social media's impact on anxiety within primary care: a qualitative study. BJGP Open 2024;8(2) [FREE Full text] [doi: 10.3399/BJGPO.2023.0190] [Medline: 38272495]
22. Vannucci A, Flannery KM, Ohannessian CM. Social media use and anxiety in emerging adults. J Affect Disord 2017;207:163-166. [doi: 10.1016/j.jad.2016.08.040] [Medline: 27723539]
23. Keles B, McCrae N, Grealish A. A systematic review: the influence of social media on depression, anxiety and psychological distress in adolescents. International Journal of Adolescence and Youth 2019;25(1):79-93 [FREE Full text] [doi: 10.1080/02673843.2019.1590851]
24. Luo L, Yuan J, Xu C, et al. Mental health issues and 24-hour movement guidelines-based intervention strategies for university students with high-risk social network addiction: cross-sectional study using a machine learning approach. J Med Internet Res 2025;27:e72260 [FREE Full text] [doi: 10.2196/72260] [Medline: 40512996]
25. Ergün N, Özkan Z, Griffiths MD. Social media addiction and poor mental health: examining the mediating roles of internet addiction and phubbing. Psychol Rep 2025;128(2):723-743. [doi: 10.1177/00332941231166609] [Medline: 36972903]
26. Li S, Jin X, Song L, Fan T, Shen Y, Zhou J. The association between internet addiction and the risk of suicide attempts in Chinese adolescents aged 11-17 years: prospective cohort study. J Med Internet Res 2025;27:e52083 [FREE Full text] [doi: 10.2196/52083] [Medline: 39752720]
27. Peng J, Liu Y, Wang X, Yi Z, Xu L, Zhang F. Physical and emotional abuse with internet addiction and anxiety as a mediator and physical activity as a moderator. Sci Rep 2025;15(1):2305 [FREE Full text] [doi: 10.1038/s41598-025-85943-x] [Medline: 39824886]
28. Anto A, Basu A, Selim R, Foscht T, Eisingerich A. Open-World Games' affordance of cognitive escapism, relaxation, and mental well-being among postgraduate students: mixed methods study. J Med Internet Res 2024;26:e63760 [FREE Full text] [doi: 10.2196/63760] [Medline: 39689301]
29. Martinez K, Menéndez-Menéndez MI, Bustillo A. Awareness, prevention, detection, and therapy applications for depression and anxiety in serious games for children and adolescents: systematic review. JMIR Serious Games 2021;9(4):e30482 [FREE Full text] [doi: 10.2196/30482] [Medline: 34927589]

XSL•FO
RenderX

30. Zeiler M, Vögl S, Prinz U, Werner N, Wagner G, Karwautz A, et al. Game design, effectiveness, and implementation of serious games promoting aspects of mental health literacy among children and adolescents: systematic review. JMIR Ment Health 2025;12:e67418 [FREE Full text] [doi: 10.2196/67418] [Medline: 40324175]
31. Martínez-Miranda J, Espinosa-Curiel I. Serious games supporting the prevention and treatment of alcohol and drug consumption in youth: scoping review. JMIR Serious Games 2022;10(3):e39086 [FREE Full text] [doi: 10.2196/39086] [Medline: 36006694]
32. Lin Y, Tudor-Sfetea C, Siddiqui S, Sherwani Y, Ahmed M, Eisingerich A. Effective behavioral changes through a digital mhealth app: exploring the impact of hedonic well-being, psychological empowerment and inspiration. JMIR Mhealth Uhealth 2018;6(6):e10024 [FREE Full text] [doi: 10.2196/10024] [Medline: 29907557]
33. Tudor-Sfetea C, Rabee R, Najim M, et al. Evaluation of two mobile health apps in the context of smoking cessation: qualitative study of cognitive behavioral therapy (CBT) versus non-CBT-based digital solutions. JMIR Mhealth Uhealth 2018;6(4):e98 [FREE Full text] [doi: 10.2196/mhealth.9405] [Medline: 29669708]
34. El-Hilly A, Iqbal S, Ahmed M, et al. Game On? Smoking cessation through the gamification of mhealth: a longitudinal qualitative study. JMIR Serious Games 2016;4(2):e18 [FREE Full text] [doi: 10.2196/games.5678] [Medline: 27777216]
35. Schartmann A. Koji Kondo's Super Mario Bros. Soundtrack. London, United Kingdom: Bloomsbury; 2015.
36. Kuittinen P. University of Art and Design Helsinki. Very important game people in the history of computer and video games [dissertation].: University of Art and Design Helsinki; 2006. URL: https://aaltodoc.aalto.fi/server/api/core/bitstreams/97bc6a50-fae6-4420-8284-8b835ad77e02/content [accessed 2025-12-08]
37. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: a synthesis of recommendations. Acad Med 2014;89(9):1245-1251 [FREE Full text] [doi: 10.1097/ACM.0000000000000388] [Medline: 24979285]
38. Mixed methods article reporting standards (MMARS) information recommended for inclusion in manuscripts that report the collection and integration of qualitative and quantitative data. APA Style Journal Article Reporting Standards. URL: https://apastyle.apa.org/jars/mixed-table-1.pdf [accessed 2025-12-03]
39. Glaser BG, Strauss AL. The Discovery of Grounded Theory: Strategies for Qualitative Research. New York, NY: Routledge; 2009.
40. Edwards R, Holland J. What Is Qualitative Interviewing?. London, United Kingdom: Bloomsbury; 2013.
41. Gioia DA, Corley KG, Hamilton AL. Seeking qualitative rigor in inductive research. Organizational Research Methods 2012;16(1):15-31. [doi: 10.1177/1094428112452151]
42. Zeithaml VA, Jaworski BJ, Kohli AK, Tuli KR, Ulaga W, Zaltman G. A theories-in-use approach to building marketing theory. Journal of Marketing 2019;84(1):32-51. [doi: 10.1177/0022242919888477]
43. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, STROBE Initiative. The Strengthening The Reporting of Observational studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. J Clin Epidemiol 2008;61(4):344-349. [doi: 10.1016/j.jclinepi.2007.11.008] [Medline: 18313558]
44. Demerouti E, Mostert K, Bakker AB. Burnout and work engagement: a thorough investigation of the independency of both constructs. J Occup Health Psychol 2010;15(3):209-222. [doi: 10.1037/a0019408] [Medline: 20604629]
45. Leiter MP, Schaufeli WB. Consistency of the burnout construct across occupations. Anxiety, Stress & Coping 1996;9(3):229-243. [doi: 10.1080/10615809608249404]
46. Lyubomirsky S, Lepper HS. A measure of subjective happiness: preliminary reliability and construct validation. Social Indicators Research 1999;46(2):137-155. [doi: 10.1023/a:1006824100041]
47. Diener E, Emmons RA, Larsen RJ, Griffin S. The satisfaction with life scale. J Pers Assess 1985;49(1):71-75. [doi: 10.1207/s15327752jpa4901_13] [Medline: 16367493]
48. Yaden DB, Kaufman SB, Hyde E, Chirico A, Gaggioli A, Zhang JW, et al. The development of the Awe Experience Scale (AWE-S): a multifactorial measure for a complex emotion. The Journal of Positive Psychology 2018;14(4):474-488. [doi: 10.1080/17439760.2018.1484940]
49. Kashdan TB, Gallagher MW, Silvia PJ, et al. The curiosity and exploration inventory-II: development, factor structure, and psychometrics. J Res Pers 2009;43(6):987-998 [FREE Full text] [doi: 10.1016/j.jrp.2009.04.011] [Medline: 20160913]
50. Proyer RT. A new structural model for the study of adult playfulness: assessment and exploration of an understudied individual differences variable. Personality and Individual Differences 2017;108:113-122. [doi: 10.1016/j.paid.2016.12.011]
51. Anderson JC, Gerbing DW. Structural equation modeling in practice: a review and recommended two-step approach. Psychological Bulletin 1988;103(3):411-423. [doi: 10.1037//0033-2909.103.3.411]
52. Bagozzi RP, Yi Y. On the evaluation of structural equation models. JAMS 1988;16(1):74-94. [doi: 10.1007/bf02723327]
53. Fornell C, Larcker DF. Evaluating structural equation models with unobservable variables and measurement error. Journal of Marketing Research 1981;18(1):39. [doi: 10.2307/3151312]
54. Hayes A. Introduction to Mediation, Moderation, and Conditional Process Analysis: a Regression-Based Approach. New York, NY: Guilford Press; 2017.
55. Ryan RM, Rigby CS, Przybylski A. The motivational pull of video games: a self-determination theory approach. Motiv Emot 2006;30(4):344-360. [doi: 10.1007/s11031-006-9051-8]

XSL•FO
RenderX

56. Bowman ND, Tamborini R. “In the Mood to Game”: selective exposure and mood management processes in computer game play. New Media & Society 2013;17(3):375-393. [doi: 10.1177/1461444813504274]
57. Hemenover SH, Bowman ND. Video games, emotion, and emotion regulation: expanding the scope. Annals of the International Communication Association 2018;42(2):125-143. [doi: 10.1080/23808985.2018.1442239]
58. Villani D, Carissoli C, Triberti S, Marchetti A, Gilli G, Riva G. Videogames for emotion regulation: a systematic review. Games Health J 2018;7(2):85-99 [FREE Full text] [doi: 10.1089/g4h.2017.0108] [Medline: 29424555]
59. Routledge C, Wildschut T, Sedikides C, Juhl J. Nostalgia as a resource for psychological health and well‐being. Social & Personality Psych 2013;7(11):808-818. [doi: 10.1111/spc3.12070]
60. Wulf T, Bowman ND, Velez JA, Breuer J. Once upon a game: exploring video game nostalgia and its impact on well-being. Psychology of Popular Media 2020;9(1):83-95. [doi: 10.1037/ppm0000208]
61. Eisingerich AB, Marchand A, Fritze MP, Dong L. Hook vs. hope: how to enhance customer engagement through gamification. International Journal of Research in Marketing 2019;36(2):200-215. [doi: 10.1016/j.ijresmar.2019.02.003]
62. Eisingerich AB, MacInnis DJ, Park CW. A theoretical framework exploring three foundational benefits of brand attachment. In: A Research Agenda for Brand Management in a New Era of Consumerism. Cheltenham, United Kingdom: Edward Elgar Publishing; 2023:161-175.
63. Webb J, Lin Y, Ang A, et al. Feasibility and preliminary outcomes of a mobile intervention combining cognitive behavioral therapy, virtual coaching, and nicotine replacement therapy for nicotine vaping cessation. Telemed Rep 2023;4(1):48-52 [FREE Full text] [doi: 10.1089/tmr.2023.0009] [Medline: 37102136]
64. Bell SJ, Eisingerich AB. Work with me. Harvard Business Review 2007;85(6):32 [FREE Full text]
65. Fritze MP, Eisingerich AB, Benkenstein M. Digital transformation and possession attachment: examining the endowment effect for consumers’ relationships with hedonic and utilitarian digital service technologies. Electron Commer Res 2018;19(2):311-337. [doi: 10.1007/s10660-018-9309-8]
66. Eisingerich AB, Kretschmer T. In e-commerce, more is more. Harvard Business Review 2008;86(3):20-21 [FREE Full text]
67. Duffek B, Eisingerich AB, Merlo O, Lee G. Authenticity in influencer marketing: how can influencers and brands work together to build and maintain influencer authenticity? Journal of Marketing 2025 May 05;89(5):21-46. [doi: 10.1177/00222429251319786]
68. Eisingerich AB, Boehm L. Hospital visitors ask for more shopping outlets. Harvard Business Review 2009;87(5):21 [FREE Full text]
69. Merlo O, Eisingerich A, Shin H, Britton R. Avoiding the pitfalls of customer participation. MIT Sloan Manag Rev 2019;61(1) [FREE Full text]
70. Lin Y, Foscht T, Eisingerich AB. A time for heroes? Conceptualization, development and validation of the brand hero scale. EJM 2023;57(13):1-26. [doi: 10.1108/ejm-09-2021-0700]
71. Possler D, Klimmt C. Towards a model of the entertaining appeal of video game aesthetics. In: Biermann J, Kiefer F, editors. Computerspielforschung: Interdisziplinäre Einblicke in das digitale Spiel und seine kulturelle Bedeutung. Leverkusen, Germany: Verlag Barbara Budrich; 2023:141-158.
72. Daneels R, Bowman ND, Possler D, Mekler ED. The ‘eudaimonic experience’: a scoping review of the concept in digital games research. Media Commun 2021;9(2):178-190. [doi: 10.17645/mac.v9i2.3824]
73. Klimmt C, Possler D, May N, Auge H, Wanjek L, Wolf A. Effects of soundtrack music on the video game experience. Media Psychology 2018;22(5):689-713. [doi: 10.1080/15213269.2018.1507827]
74. Rieger D, Wulf T, Kneer J, Frischlich L, Bente G. The winner takes it all: the effect of in-game success and need satisfaction on mood repair and enjoyment. Comput Hum Behav 2014;39:281-286. [doi: 10.1016/j.chb.2014.07.037]
75. Pine R, Fleming T, McCallum S, Sutcliffe K. The effects of casual videogames on anxiety, depression, stress, and low mood: a systematic review. Games Health J 2020;9(4):255-264. [doi: 10.1089/g4h.2019.0132] [Medline: 32053021]
76. Arigayota A, Duffek B, Hou C, Eisingerich A. Effects of the legend of Zelda: breath of the wild and Studio Ghibli films on young people's sense of exploration, calm, mastery and skill, purpose and meaning, and overall happiness in life: exploratory randomized controlled study. JMIR Serious Games 2025;13:e76522 [FREE Full text] [doi: 10.2196/76522] [Medline: 40750097]
77. Eisingerich A, Arigayota A, Duffek B, Hou C. Response: bridging gaming and engaging with storytelling with mental health science. JMIR Serious Games 2025:e83239. [doi: 10.2196/preprints.83239]
78. Hou C, Foscht T, Duffek B, Arigayota A, Eisingerich A. Mitigating the negative effects of internet browsing on young people's resilience and outlook on life through Classic Grimms' Fairy Tales: exploratory randomized controlled study. JMIR Form Res 2025;9:e76770 [FREE Full text] [doi: 10.2196/76770] [Medline: 40768252]
79. Eisingerich AB, Chun HH, Liu Y, Jia H, Bell SJ. Why recommend a brand face‐to‐face but not on Facebook? How word‐of‐mouth on online social sites differs from traditional word‐of‐mouth. J Consum Psychol 2014;25(1):120-128. [doi: 10.1016/j.jcps.2014.05.004]
80. Park CW, Eisingerich AB, Park JW. Attachment–aversion (AA) model of customer–brand relationships. J Consum Psychol 2013;23(2):229-248. [doi: 10.1016/j.jcps.2013.01.002]

XSL•FO
RenderX

81. Park CW, Macinnis DJ, Priester J, Eisingerich AB, Iacobucci D. Brand attachment and brand attitude strength: conceptual and empirical differentiation of two critical brand equity drivers. Journal of Marketing 2010;74(6):1-17. [doi: 10.1509/jmkg.74.6.1]
82. Eisingerich AB, Bell SJ. Perceived service quality and customer trust. J Serv Res 2008;10(3):256-268 [FREE Full text] [doi: 10.1177/1094670507310769]
83. Merlo O, Eisingerich AB, Gillingwater R, Cao JJ. Exploring the changing role of brand archetypes in customer-brand relationships: Why try to be a hero when your brand can be more? Bus Horiz 2023;66(5):615-629. [doi: 10.1016/j.bushor.2022.11.001]
84. Eisingerich AB, MacInnis DJ, Whan Park C. Do CSR efforts that focus on helping the environment influence brand purchase more than other forms of CSR? J Bus Res 2023;168:114240. [doi: 10.1016/j.jbusres.2023.114240]
85. Frankl VE. Man's Search for Meaning: An Introduction to Logotherapy. Boston, MA: Beacon Press; 1984.
86. Lorenz K. Die Acht Todsuenden Der Zivilisierten Menschheit. Munich, Germany: Piper; 1973.
87. Waggerl KH. Brot. Salzburg-Vienna, Austria: Otto Mueller Verlag; 2005.
88. Frankl VE. The Will to Meaning: Foundations and Applications of Logotherapy. New York, NY: Penguin/Plume; 1988.

Abbreviations

AMOS: Analysis of Moment Structures
BootSE: bootstrapped standard error
MSE: mean squared error
RA: research assistant
SPSS: Statistical Package for the Social Science
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

Edited by S Brini; submitted 16.09.25; peer-reviewed by B Duffek, K Yoshimura; comments to author 11.10.25; accepted 24.11.25; published 19.12.25.

Please cite as:
Tam W, Hou C, Eisingerich AB
Super Mario Bros. and Yoshi Games' Affordance of Childlike Wonder and Reduced Burnout Risk in Young Adults: In-Depth Mixed Methods Cross-Sectional Study
JMIR Serious Games 2025;13:e84219
URL: https://games.jmir.org/2025/1/e84219
doi:10.2196/84219
PMID:

©Winze Tam, Congcong Hou, Andreas Benedikt Eisingerich. Originally published in JMIR Serious Games (https://games.jmir.org), 19.12.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Enhancing Equity in Schoolchildren's Basic Life Support Education in Brazil Through Serious Games: Cohort Study

Uri Adrian Prync Flato[1], PhD, MBA, MD; Adriana Flato[2], DDS; Isabella Bispo Diaz T Martins[3], MD; Giuliana Simoes Nakano[3], MD; Julia Caroline Romao[3], Msc; Manuela Simões Nakano[3], MD; Emilio José Beffa dos Santos[3], MD; Yuuki Daniel Tahara Villas Boas[3], MD; Leonardo Escobar Medeiros[3], MD; Vinicius Gazin Rossignoli[3], CEng; Gabriel Zanatta Cunha[3], MD; Rafael Carreira Batista[3], MD; Pedro Gazotto Rodrigues da Silva[3], MD; Miguel Florentino[3], MD; Amanda Rabelo[4], PhD; Thais Dias Midega[4], MD; Rogerio Passos[4], MD; Helio Guimaraes[4], MD; Karl Kern[5], PhD, MD

[1]Critical Care Unit, Hospital Vila Nova Star-Rede D'Or São Paulo/ São Luiz Itaim-Rede D'Or, Dr. Alceu de Campos Rodrigues, 126, São Paulo, Brazil

[2]Sanare Institute, São Paulo, Brazil

[3]Universidade de Marília, Marília, Brazil

[4]Hospital Israelita Albert Einstein, Jose antonio coelho, 626, São Paulo, Brazil

[5]Sarver Heart Center, Tucson, AZ, United States

Corresponding Author:
Uri Adrian Prync Flato, PhD, MBA, MD
Critical Care Unit, Hospital Vila Nova Star-Rede D'Or São Paulo/ São Luiz Itaim-Rede D'Or, Dr. Alceu de Campos Rodrigues, 126, São Paulo, Brazil

Abstract

Background: Out-of-hospital cardiac arrests (OHCAs) predominantly occur in residential settings, often witnessed by children who could act as first responders. The World Health Organization (WHO) supports the Kids Save Lives (KSL) initiative, recommending basic life support (BLS) training for children aged ≥11 years. However, disparities in BLS education persist globally, particularly in low-resource regions where socioeconomic barriers, such as school type, malnutrition, and limited infrastructure, hinder implementation. Younger children (aged <11 years) face additional challenges due to physical limitations (eg, height, weight, and grip strength), which may compromise their ability to achieve adequate chest compression depth. While gamified learning has shown promise in improving BLS engagement and skill acquisition, its efficacy across diverse socioeconomic groups remains understudied.

Objective: This study aimed to compare cardiopulmonary resuscitation (CPR) performance (compression depth, rate, and recoil) between public and private school students following a game-based BLS intervention and evaluate the feasibility of a serious game (Kids Save Hearts) in improving BLS knowledge and irrespective of socioeconomic background.

Methods: We conducted an observational cohort study with 336 students aged 7 - 17 years from 10 public and 10 private schools in Brazil (April to November 2022). Participants received 40 minutes of video-based CPR training (American Heart Association CPR in Schools) followed by 10 minutes of gamified training using the Children Save Hearts serious game (SG). CPR quality was assessed via quality of cardiopulmonary resuscitation (QCPR) scores (Laerdal QCPR manikin), measuring compression depth (mm), compression rate (per minute), and chest recoil. Anthropometric data (height, weight, and grip strength) and socioeconomic indicators (school type) were collected. Nonparametric tests (Mann-Whitney *U* and chi-square tests) and multivariate regression (SPSS v27.0; IBM Corp) were used to analyze associations between demographics, physical characteristics, and CPR performance.

Results: Older students (11 - 17 years) outperformed younger peers (7 - 10 years) in median compression depth (48 mm vs 37 mm; *P*<.001) and overall QCPR scores (84 vs 42, *P*<.001). Private school students had higher grip strength (24.92 vs 21.48 g/cm²; *P*=.001), but school type did not significantly affect CPR quality. Postintervention SG scores improved universally (*P*<.001), with no age or socioeconomic disparities. Multivariate analysis identified age (*P*<.001), height (*P*<.001), and grip strength (*P*<.001) as independent predictors of high QCPR scores (≥70).

Conclusions: Age and physical development were stronger determinants of CPR quality than socioeconomic factors. The game-based intervention effectively improved BLS knowledge and skills across all participants, demonstrating its potential as an equitable training tool. These findings support the scalability of gamified BLS programs in resource-limited settings.

Trial Registration: Clinicaltrials.gov NCT06565767; https://clinicaltrials.gov/study/NCT06565767

International Registered Report Identifier (IRRID): RR2-10.2196/55333

XSL•FO
RenderX

(JMIR Serious Games 2025;13:e69252) doi:10.2196/69252

KEYWORDS

cardiopulmonary; resuscitation; training; schoolchildren; gamification; basic life support

Introduction

Background

Cardiovascular diseases remain the leading cause of mortality worldwide, accounting for 80% of deaths [1]. Out-of-hospital cardiac arrests (OHCAs) frequently occur at home [2], often in the presence of family members, including children. Over the past 2 decades, OHCA survival rates have improved in North America, Europe, and Asia [3-5], attributed to early recognition, prompt emergency system activation, high-quality chest compressions, and rapid defibrillation. However, achieving similar outcomes globally requires transformative public policies, such as those recommended by the Institute of Medicine in 2015 [6]. Strengthening the chain of survival hinges on layperson training in basic life support (BLS), particularly in high-risk populations [7].

The World Health Organization (WHO) endorses the Kids Save Lives (KSL) initiative [8], advising that children aged 11 years and older receive at least 2 hours of annual BLS and automated external defibrillator training [9]. However, the WHO does not prescribe specific methodologies, leaving room for innovation, particularly through digital tools. Disparities in access to BLS education further complicate implementation, with students in public schools often lacking resources (eg, manikins and trained instructors) compared to their private school counterparts. Socioeconomic factors, including malnutrition and delayed physical development, may also impair cardiopulmonary resuscitation (CPR) performance in underserved communities. Addressing these inequities demands adaptable solutions, such as low-cost manikins, strength-building exercises, and inclusive, technology-driven teaching strategies.

Schoolchildren are ideal first responders and multipliers of BLS knowledge within their communities. Evidence suggests that gamified learning can enhance engagement and long-term retention of CPR skills in children [10]. Serious games, combined with quantitative psychomotor assessments, offer a promising avenue for scalable BLS training [11]. Yet, few studies have evaluated the impact of such interventions across socioeconomic strata.

Study Objective

This study aimed to evaluate disparities in CPR quality (chest compression depth, rate, and chest recoil) between public and private school students following a game-based BLS training intervention (Kids Save Hearts Project), assessing the feasibility of digital tools to bridge gaps in life-saving competency.

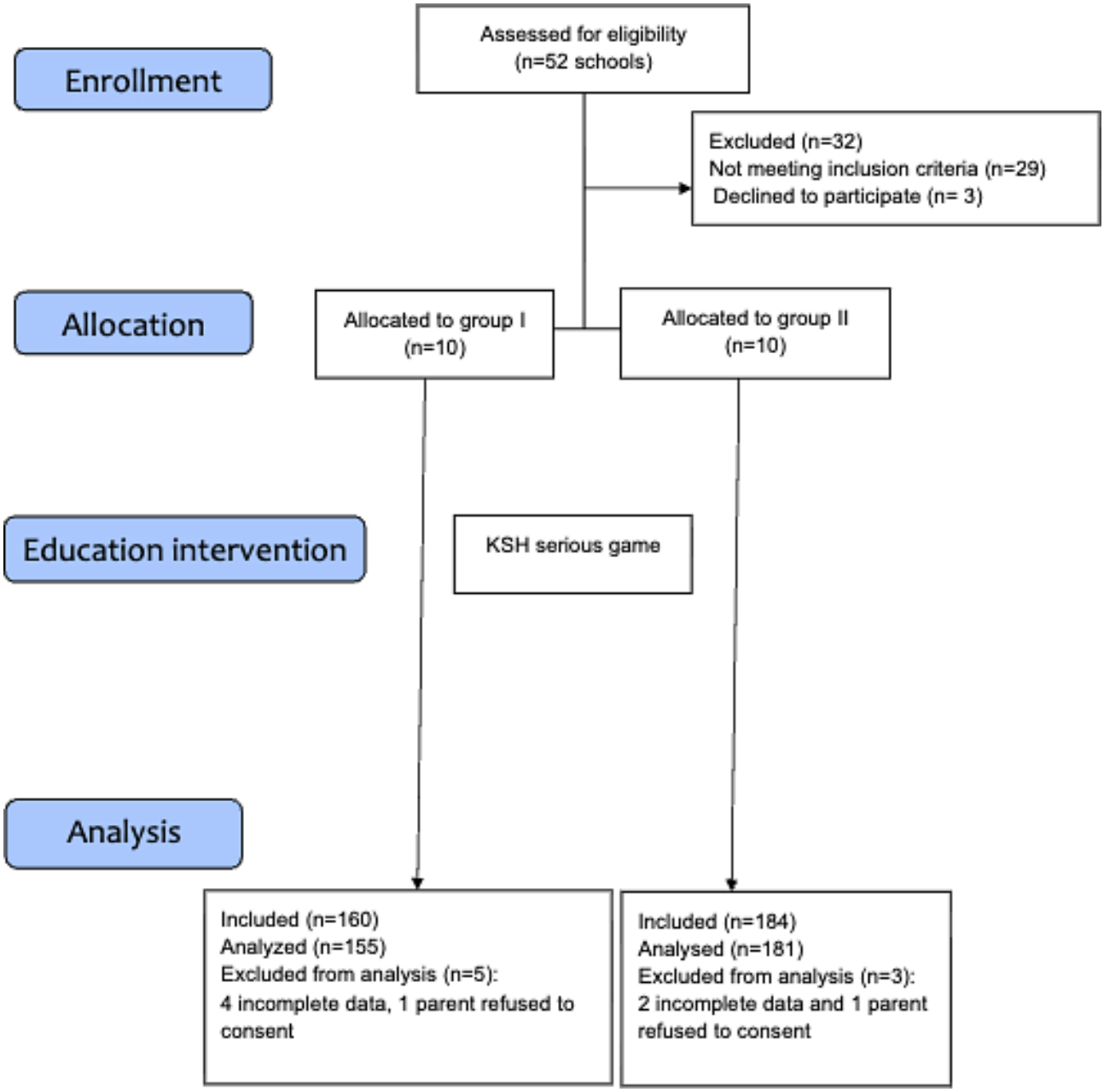
Methods

Overview

This study was designed as an observational cohort study following the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) [12] guidelines (Figure 1). A total of 328 children from elementary and high schools were included, divided into 2 groups based on age: 7-10 years (Group I) and 11-17 years (Group II). Schools, both public and private, were selected through a scientific cooperation agreement with the University of Marilia, located in the western region of São Paulo, Brazil. Selection was conducted sequentially, alternating 1:1 between public and private schools. Classes were enrolled to match the required sample size, considering an average of 20 students per class. The study took place from April to November 2022. Schoolchildren and their legal guardians provided written informed assent and consent to participate. Inclusion criteria required participants to be enrolled in elementary or secondary school, have no prior training in BLS, and provide informed assent and consent. Exclusion criteria included previous CPR training or inability to attend all planned activities (Textbox 1). CPR training was not previously part of the Brazilian educational curriculum in elementary and high schools.

XSL•FO
RenderX

Figure 1. Study design flowchart. KSH: Kids Save Hearts.



Textbox 1. Inclusion and exclusion criteria for participant selection in the cohort study.

Inclusion criteria

- Participants enrolled in elementary or secondary school
- No prior training in basic life support (BLS)
- Provision of informed assent and consent

Exclusion criteria

- Previous cardiopulmonary resuscitation (CPR) training
- Inability to attend all planned sessions

CPR Training

The training was developed following a pattern (Figure 2) that included: welcoming the children, obtaining demographic data (weight, height, bicep circumference, and palmar strength test), qualitative evaluation of the student's expectations in a written way, serious game (SG) play, video-based training, serious game evaluation (SGT1), postcourse written qualitative perception and evaluation for 1 minute with quality of

XSL•FO
RenderX

cardiopulmonary resuscitation (QCPR) manikin. All participants received 40 minutes of video-based training on CPR, practicing while watching based on the American Heart Association (AHA) CPR in Schools Training Kit and 10 minutes of training with continuous chest compressions with audio feedback. The training sessions were conducted with classes of 20 participants during the school period. Mini Little Anne Manikin (Laerdal) and Little Anne QCPR (Laerdal Medical Inc) with real-time feedback software (QCPR training 4.13.3, Laerdal Medical Inc) were used for training. Certified basic life support instructors (AHA) conducted the training based on 2020 resuscitation guidelines [13] with a ratio of 10:1 schoolchild to instructors.

Figure 2. Cardiopulmonary resuscitation training flow: the implementation framework traditional video-based cardiopulmonary resuscitation instruction with supplemental digital game-based learning (KSH serious game) components, with outcomes measured using Laerdal QCPR (quality cardiopulmonary resuscitation) manikins.



QCPR Score

The QCPR score is a standardized metric designed to assess the effectiveness of CPR performance by evaluating key parameters, such as chest compression depth, rate, recoil, and ventilation volume. The QCPR score is calculated as a percentage, with higher scores indicating closer adherence to guideline-recommended CPR standards. By quantifying performance, the QCPR score helps identify areas for improvement, ensuring that high-quality CPR is delivered during cardiac arrest scenarios, which is critical for improving patient outcomes.

Evaluation criteria for QCPR focus on the consistency and accuracy of chest compressions and ventilations. Key metrics include maintaining a compression depth of 5 - 6 cm, a rate of 100 - 120 compressions per minute, complete chest recoil between compressions. Deviations from these targets result in lower QCPR scores, reflecting suboptimal CPR quality. In addition, hands-off time (interruptions in compressions) is factored into the evaluation, as excessive pauses can diminish perfusion efficacy. The QCPR assessment not only serves as a training tool for health care providers but also supports research initiatives aimed at correlating CPR quality with clinical outcomes, reinforcing the importance of adherence to evidence-based resuscitation protocols.

Serious Game Children Save Hearts

The serious game Children Save Hearts was developed on the Smile Game Builder platform (open source) and implemented on Windows 7, 8, 10, and 11 operating systems. The game was developed with the objective of teaching children from 7 to 11 years old to perform BLS care. The script and storytelling are based on the ILCOR (International Liaison Committee on Resuscitation) guidelines.

In the game, the user must follow commands to perform CPR in an avatar. At the beginning of the game, the user must answer questions on how to conduct the service. At the end, the user receives a grade based on the overall performance. The game includes the use of a joystick and simple commands to make the user experience simpler, allowing greater focus on the content. We define a first contact (time) with the SG as (SGTO) and the evaluation phase of SGT1. A usability evaluation was validated and published previously [14], contemplating 5 elements (recognition of unconsciousness, call for help, initiate fast and strong chest compressions, maintain compressions at least 100 to 120 per minute, and automated external defibrillator) with maximum and minimum game score from 0 to 500.

Sample Size

The sample size calculation determined that a minimum of 296 participants was required to detect a 12-mm [15,16] difference in chest compression depth with an expected SD of 10 mm, assuming a significance level of 0.05, power of 90%, and an estimated 20% dropout rate. To account for potential exclusions and ensure the representativeness and robustness of the results, we included a total of 336 participants in the study. This approach minimized the risk of underpowering the study due to unanticipated data losses or incomplete participation.

XSL•FO
RenderX

Statistical Analysis

All statistical analyses were performed using IBM SPSS Statistics (v27.0) with a significance threshold of $P<.05$. Data normality was assessed using the Shapiro-Wilk test, which guided the choice of parametric or nonparametric tests. Continuous variables (compression rate, depth, and QCPR scores) were compared between groups using independent 1-tailed *t* tests (for normally distributed data) and Mann-Whitney *U* tests (for nonnormally distributed data). Categorical variables (school type and gender) were analyzed via chi-square and Fisher exact tests. For correlational analyses, Spearman rank correlation (nonparametric) and Pearson r (parametric) tested associations between anthropometric measures (height, weight, and grip strength) and CPR quality (QCPR scores). Multivariate logistic regression identified predictors of high-quality CPR (QCPR ≥70), adjusting for age, school type, and gender. Pre- and posttraining knowledge gains (serious game scores) were evaluated using Wilcoxon signed-rank tests, while between-group differences (age and school type) were assessed via Kruskal-Wallis tests.

Ethical Considerations

The study protocol was approved by the Ethics Committee of the University of Marília under registration number CAAE: 57160121400005496 [17] and NCT06565767 [18]. Informed consent and informed assent were obtained from all participants before their inclusion in the study. All collected data were deidentified to ensure participant confidentiality. Participants were not compensated for their involvement, and no identifiable personal information is included in this manuscript. Identifiable images published herein were included with written consent.

Results

Study Participants

The study ultimately included 336 participants from 10 public and 10 private schools (Table 1). Among them, Group 1 comprised 155 participants with a median age of 9.2 years, while Group 2 had 181 participants with a median age of 14.4 years. Anthropometric measurements varied among participants, with a notable prevalence of obesity and overweight.

XSL•FO
RenderX

Table . Baseline characteristics and cardiopulmonary resuscitation (CPR) performance metrics in schoolchildren stratified by age group: a cross-sectional study of 336 participants (Group 1: 7 - 10 years, n=155; Group 2: 11 - 17 years, n=181).

Variable	Group 1 7 - 10 years (n=155)	Group 2 11 - 17 years (n=181)	*P* value
Age (years), median (IQR)	9.20 (7 - 10)	14.40 (11 - 17)	<.001[a]
Male sex, n (%)	78 (50.3)	81 (44.8)	.30[b]
Private school, n (%)	53 (34.2)	51 (28.2)	.28[b]
Height (cm), median (IQR)	141.00 (135.50 - 146)	164 (155 - 171)	<.001[a]
Weight (kg), median (IQR)	37.60 (30.90-45.75)	60.40 (49.80 - 69.90)	<.001[a]
BMI			
Underweight, n (%)	6 (4)	10 (5)	.24[b]
Normal weight, n (%)	79 (51)	119 (65)	<.001[b]
Overweight, n (%)	37 (24)	30 (16)	<.001[b]
Obese, n (%)	33 (21)	22 (12)	<.001[b]
Biceps circumference (cm), median (IQR)	23 (20.50-25.50)	27 (24 - 30)	<.001[a]
HGS[c] (g/cm^2), median (IQR)	18.80 (15-28.50)	24.50 (20.25 - 31.70)	<.001[a]
Compression (per minute), median (IQR)	110 (104-123)	107 (104-116)	.01[a]
Chest recoil (%), median (IQR)	98 (96-100)	95 (91-100)	<.001[a]
Compression depth (mm), median (IQR)	37 (31-44)	48 (39-53)	<.001[a]
QCPR[d] score, median (IQR)	42 (30-64)	84 (55-98)	<.001[a]
Serious game KSH[e]			
T0 Score	350 (350-450)	300 (300-450)	.34[a]
T1 Score	450 (400-500)	450 (400-500)	.29[a]

[a]Group comparisons used Mann-Whitney *U* tests.

[b]Chi-square tests.

[c]HGS: handgrip strength.

[d]QCPR: Quality of Cardiopulmonary Resuscitation.

[e]KSH: Kids Save Hearts.

Chest Compression Performance

Compression Rate

There were no significant differences in compression rates between the age groups, with Group 1 performing at a median of 110 compressions per minute compared to Group 2 at 117 compressions per minute (*P*=.01).

Compression Depth

Group 2 demonstrated a significantly greater compression depth compared to Group 1, with a median depth of 48 mm in Group 2 versus 37 mm in Group 1 (*P*<.001).

Quality of Chest Compressions

The overall quality of chest compressions, as measured by the QCPR score, was significantly better in Group 2, with a median score of 84 compared to 42 in Group 1 (*P*<.001).

Knowledge Acquisition

Serious game scores: there was a statistically significant improvement in knowledge acquisition posttraining, with serious game scores showing a marked increase (*P*<.001). However, there were no significant differences in posttraining scores between the age groups.

Comparisons by School Type and Gender

School type (public vs private): the performance of students from public and private schools (Table 2) did not differ significantly overall. However, private school students exhibited

XSL•FO
RenderX

a higher median hand grip strength compared to their public-school counterparts (24.92 vs 21.48 g/cm²; *P*=.001). Gender differences: gender did not have a significant impact on CPR performance (Tables 1 and 2), as there were no notable differences based on gender in the results.

Table . Comparison between the private versus public schools. Analysis of 336 participants (Public School: n=232; Private school: n=104) of urban schools in Marilia, Brazil. Data include sex distribution, age, height, weight, biceps circumference, handgrip strength (HGS), and CPR[a] quality measures (compression rate, depth, recoil percentage, and QCPR[b] scores) assessed using Laerdal LittleAnne QCPR manikins. Continuous variables are reported as median [interquartile range (IQR)]; categorical variables as n (%). Group comparisons used Mann-Whitney *U* tests.

Variable	Public school (n=232)	Private school (n=104)	*P* value
Male sex, n (%)	110 (47.4)	49 (47.1)	>.99[c]
Age (years), median (IQR)	11 (10 - 16)	10 (10 - 14.25)	.29[d]
Height (cm), median (IQR)	153 (141 - 165)	150 [142.75 - 162.25]	.50[d]
Weight (kg), median (IQR)	49.20 (36.45 - 62.02)	48.75 (38.90 - 59.95)	.99[d]
Biceps circumference (cm), median (IQR)	25 (22 - 28)	25.50 (22 - 28.50)	.35[d]
HGS[e] (g/cm^2), median (IQR)	21.48 (16 - 29.52)	24.92 (20.25 - 32)	.001[d]
Compression (minute), median (IQR)	108 (104 - 117)	111 (104 - 122)	.19[d]
Chest recoil (%), median (IQR)	100 (99 - 100)	100 (99 - 100)	.39[d]
Compression depth (mm), median (IQR)	44 (35 - 50.25)	40 (33.75 - 49.25)	.09[d]
QCPR score, median (IQR)	67.50 (40 - 91.25)	47 (34.75 - 90)	.05[d]

[a]CPR: cardiopulmonary resuscitation.

[b]QCPR: quality of cardiopulmonary resuscitation.

[c]For proportions.

[d]For continuous variables and chi-squaretests.

[e]HGS: handgrip strength.

Factors Correlated With High-Quality CPR

QCPR score ≥70: factors associated with achieving a QCPR score of 70 or higher (Table 3), indicative of high-quality CPR, included higher age, greater height, weight, biceps circumference, and hand grip strength, with all factors showing significant associations (*P*<.01).

XSL•FO
RenderX

Table . Characteristics and CPR[a] performance metrics stratified by QCPR[b] score (QCPR <70: n=186; QCPR ≥70: n=150) The table compares demographics (sex, school type), anthropometric measures (age, height, weight, and biceps circumference), handgrip strength (HGS), and CPR quality indicators (compression rate, depth, and chest recoil) between participants achieving adequate (QCPR ≥70) versus inadequate (QCPR <70) CPR performance scores. Continuous variables presented as median (IQR); categorical variables as n (%). Statistical comparisons performed using Mann-Whitney *U* test.

Variable	QCPR score <70, (n=186)	QCPR ≥70, (n=150)	*P* value
Male sex, n (%)	87 (46.8)	72 (48)	.90[c]
Private school, n (%)	64 (34.4)	40 (26.7)	.15[c]
Age (years), median (IQR)	10 (9 - 11)	15 (11 - 16)	<.001[d]
Height (cm), median (IQR)	144 (137 - 152.75)	162 (154 - 170.75)	<.001[d]
Weight (kg), median (IQR)	40.30 (31.85 - 50.27)	60.65 (48.32 - 71.22)	<.001[d]
Biceps circumference (cm), median (IQR)	23 (21 - 26)	27.75 (25 - 30)	<.001[d]
HGS[e] (g/cm^2), median (IQR)	20.20 (15.4-26.9)	25.32 (20.3-33.9)	<.001[d]
Compression (per minute), median (IQR)	109 (104 - 122.7)	108 (104 - 116)	.17[d]
Chest recoil (%), median (IQR)	100 (100-100)	100 (94-100)	<.001[d]
Compression depth (mm), median (IQR)	36 (31-40)	50 (47-55)	<.001[d]

[a]CPR: cardiopulmonary resuscitation.

[b]QCPR: Quality of Cardiopulmonary Resuscitation.

[c]For proportions, with significant differences (*P*<.05) highlighted for key performance metrics including compression depth (*P*<.001) and HGS (*P*<.001).

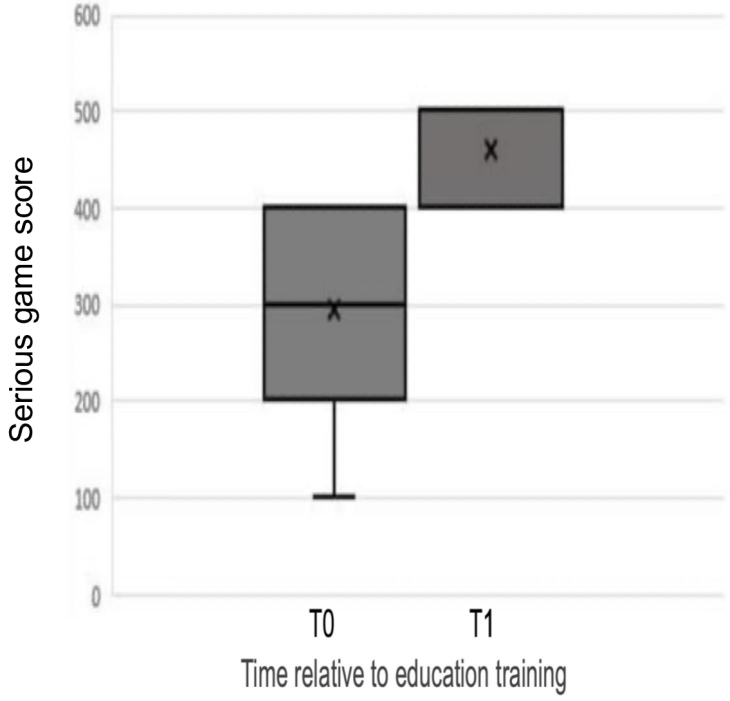
[d]For continuous variables and chi-square test.

[e]HGS: handgrip strength.

Pre- and Posttraining Knowledge Scores

Serious game scores: scores from the serious game (Figure 3) showed a significant increase from pretraining (SGT0) to posttraining (SGT1) for all participants, indicating improved knowledge (*P*<.001).

Figure 3. Box plot analysis of CPR knowledge in a kids save hearts serious game: comparative distributions of serious game score before (T0) and after BLS training (T1). CPR: cardiopulmonary resuscitation.



XSL•FO
RenderX

Discussion

Principal Findings

The importance of BLS training in schools has been increasingly recognized [19]. Our study underscores the importance of early and effective CPR training and adds valuable insights into how age and physical attributes influence the quality of chest compressions among schoolchildren.

Age-Related Performance Differences

Effective bystander CPR is crucial for improving survival rates in cases of OHCA. Research suggests that introducing CPR training to elementary children, starting as early as 4 years old, can significantly impact bystander intervention rates and survival outcomes [16]. Studies indicate that schoolchildren can effectively learn and replicate CPR techniques, with notable improvements in knowledge, attitudes, self-efficacy, and confidence observed after training [20]. The study clearly indicates that older children (aged 11 - 17 years) outperform their younger counterparts (aged 7 - 10 years) in key performance metrics, including compression depth and overall CPR quality.

Recent evidence [21] demonstrates that schoolchildren aged 10 - 14 years can effectively perform chest compressions at rates consistent with AHA and European Resuscitation Council (ERC) guidelines, achieving the recommended 100 - 120 compressions per minute in controlled training environments. However, younger students often face challenges in achieving the recommended chest compression depth due to physical factors such as height, weight, and strength [22]. Despite these challenges, consistent early CPR education fosters confidence and a positive attitude toward performing CPR, which is critical for effective bystander intervention [23]. The current body of evidence demonstrates that schoolchildren across various age groups face significant physiological barriers that render them largely incapable of performing effective chest compressions to a depth of 5 - 6 cm according to established resuscitation guidelines. Multiple studies consistently reveal age-related performance deficits [15] showing that no child in the 9 - 10 years age group could achieve the guideline-recommended compression depth, while only 19% of 11 - to 12-year-olds and 45% of 13 - to 14-years-olds met these standards. A study involving Korean elementary school students found that while younger children struggle with achieving the recommended chest compression depth, consistent training can improve their performance over time [20].

Alternative approaches, such as the foot chest compression method [24] compared to traditional hand placement, failed to demonstrate significant improvement in achieving correct compression depth (47% vs 45%; *P*=.76) among schoolchildren, indicating that the fundamental issue lies in force generation capacity rather than technique modification. In addition, age and physical attributes such as height and weight, which are indirectly related to BMI, play a crucial role in CPR performance [15]. Older children generally perform better, achieving the necessary compression depth and rate more frequently than younger children, regardless of sex and BMI [25].

Gender and Socioeconomic Factor

Gender differences in CPR training reveal nuanced insights [21]. Finke et al [26] described a meta-analysis that gender does not independently affect CPR performance quality; age plays a more significant role, with older schoolchildren performing better regardless of gender. Female high school students often exhibit a higher willingness to perform CPR compared to their male counterparts, both before and after training [27]. This willingness is crucial for the overall effectiveness of CPR interventions in real-life scenarios. Despite this, knowledge and practical application of CPR among high school students, including females, remain moderate, with notable gaps in skills such as the breathing check method. The effectiveness of CPR training and student retention can vary by school type, whether public or private. In Heraklion, Greece, no significant differences in CPR knowledge and skill retention were found between students trained by health care professionals, schoolteachers, or peer students, suggesting that the instructor type, rather than the school environment, plays a crucial role [26]. Similarly, a quasi-experimental study in rural South India showed significant improvements in CPR knowledge among adolescents following a structured training program, indicating that well-designed training can be effective regardless of the school setting [27]. In Brazil, Fernandes et al [28] found that private school students performed better than public school students in assessments of prior knowledge, immediate learning, and delayed learning. This suggests differences in educational quality or resources between the 2 types of schools. Socioeconomic factors also play a significant role, as private school students often come from families with higher socioeconomic status, providing additional educational support and resources [29]. The Brazilian Education Development Index (IDEB) results consistently show superior performance by private school students compared to their public-school counterparts [30]. In addition, the COVID-19 pandemic highlighted disparities, with private schools better equipped to transition to remote learning, maintaining higher performance levels, and potentially impacting long-term learning [31].

Technological Integration and Future Directions

Computer-based learning [32,33] tools can increase engagement among schoolchildren compared to traditional methods, such as internet-based programs [7] and simulations, making BLS training more appealing and memorable. Digital platforms [34] can provide standardized training, ensuring consistent quality and adherence to the latest guidelines, crucial in medical education and community CPR programs [35]. Computer-based BLS training allows students to learn independently, revisiting complex topics as needed, which can benefit younger learners.

Online platforms and applications can make BLS training more accessible, reaching a wider audience, including schools in remote or under-resourced areas, particularly in low-income resource countries with diminishing inequities. Advanced computer programs [36] can offer real-time feedback during BLS training, allowing students to correct techniques immediately and learn more effectively. Virtual and augmented reality can simulate real-life scenarios, providing immersive training experiences without the risks associated with practicing

XSL•FO
RenderX

on live participants. Computer-based training tools [37] can collect student performance data, providing valuable insights for educators to tailor and improve the BLS curriculum.

Implementing technology-driven BLS programs requires significant resources, including hardware, software, and training for educators, which might be a barrier for some institutions. However, using freeware and open-source software can overcome these barriers. The rapidly evolving nature of medical guidelines necessitates regular updates to the training content, which can be resource intensive. While technology can enhance learning, balancing it with human interaction is crucial, primarily to address the emotional and psychological aspects of emergency response training.

Impact of Training Methods

Recently published the 2024 cardiopulmonary resuscitation guidelines [38], recommending that the incorporation of gamified learning be considered as a component of resuscitation training for all BBLS and advanced life support (ALS) courses. This recommendation is made with a weak endorsement due to the current evidence being of very low quality. Nonetheless, gamification has the potential to enhance learner engagement and motivation [39], which may improve the overall effectiveness of these training programs. It can lead to a more informed and prepared generation capable of responding effectively to emergencies.

However, the success of this approach depends on addressing challenges related to resources and content management and maintaining a balance between technological and human elements in education. Continuous research and innovation in this field are essential to maximize the benefits of technology in BLS training for schoolchildren. Implementing novel pedagogical approaches, such as the flipped classroom model [40], has resulted in promising results when evaluating the influence of this method on BLS skills in elementary school children. The study revealed that these SG and student-centered strategies can improve learning outcomes. This method allows students to engage with the material at home and then apply what they have learned in a practical classroom setting, potentially leading to better skill acquisition and retention. Research conducted by Abelairas et al [41] evaluated the effectiveness of very short 4-month rolling refreshers was compared to that of annual retraining. The results indicated that increasing the frequency of training sessions while reducing their duration could enhance effectiveness compared to decreasing the frequency and increasing the duration of sessions. This strategy could facilitate the preservation of a high level of proficiency among pupils, guaranteeing their constant readiness to react in emergencies.

Protocol Deviations

The study protocol was adaptively refined from its original approved form to better align with post-pandemic realities and opportunities for enhanced engagement. All modifications were formally documented and received IRB approval before implementation, ensuring the continued integrity and relevance of the research. In response to overwhelming interest from partner schools, the study population was expanded from two to twenty schools, enrolling 336 students. This adaptation allowed the research to extend its reach to a broader and more developmentally diverse cohort (ages 7 - 17 years), maximizing the public health impact of the training.

Further refinements were made to the intervention and evaluation methodology to improve efficacy and feasibility. Leveraging additional resources, the training was optimized into a concise, 50-minute module comprising a video-based lesson and a gamified training component delivered via a serious game. Critically, the primary outcome measure was transitioned from a written test to direct, quantitative CPR performance metrics, including data from the serious game and QCPR scores. This essential change ensured equitable assessment for all participants, particularly younger children whose literacy was affected by pandemic-related educational disruptions. Consequently, the 6-month follow-up was deemed no longer applicable to the new primary endpoint. The study timeline was also adjusted to accommodate this expanded scope. All changes were guided by a primary commitment to participant welfare and scientific rigor.

Limitations

This study has several methodological limitations that should be acknowledged. First, the single-center design conducted in one geographic region of Brazil may limit the generalizability of findings to other populations with different socioeconomic or educational contexts. The short-term evaluation period also precludes assessment of long-term skill retention, a critical factor for BLS training efficacy. While we controlled for key anthropometric variables, unmeasured confounders (cognitive development and prior gaming experience) may have influenced results. Finally, the study did not evaluate actual willingness to perform CPR in emergencies, a crucial behavioral outcome that may not correlate directly with manikin-based performance metrics. Future multicenter randomized trials with longer follow-up are needed to address these limitations.

Recommendations

Future studies should investigate optimal training intervals through longitudinal designs assessing skill retention at 6 - 12-month intervals. Multicenter randomized controlled trials comparing game-based training versus traditional methods are needed to establish causal efficacy, particularly in low-income countries. Research should also examine real-world translation by measuring actual bystander CPR rates among trained students. Finally, serious game developers should incorporate adaptive algorithms that adjust difficulty based on the user's age and physical capabilities, while maintaining fidelity to evidence-based CPR guidelines. These evidence-based recommendations could significantly advance global implementation of the WHO's Kids Save Lives initiative while addressing critical gaps in resuscitation education.

Conclusion

This study demonstrates that a serious game-based intervention provides an effective and equitable method for delivering BLS training to schoolchildren. Multivariate analysis confirmed that while inherent physiological factors, specifically age, height, and grip strength, were independent predictors of higher CPR

XSL•FO
RenderX

performance scores, the knowledge acquired from the game-based training was universal. Postintervention scenario performance improved significantly across all groups, with no observed disparities attributable to age or socioeconomic status.

Acknowledgments

We would like to thank the University of Marília for the use of their facilities and resources. We also thank the participants for their time and dedication to our study.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

UAPF and HPG performed supervision. UAPF, AdSLF, EJBdS, IBDTM, GSN, JCR, MSN, YDTVB, LEM, GZRdC, RCB, PGRdS, MFA, AR, VR, TDM, and RdHP performed investigation and data curation. UAPF, EJBdS, AR, TDM, and HPG contributed to methodology. UAPF and HPG handled project administration. IBDTM, AR, and TDM managed software. UAPF, EJBdS, AR, TDM, and Kk conducted validation. UAPF, EJBdS, AR, TDM, and HPG contributed to visualization. UAPF, AdSLF, and HPG wrote the original draft. All authors contributed to review and editing of the manuscript.

Conflicts of Interest

None declared.

References

1. Cardiovasc diseases. World Health Organization. 2023. URL: https://www.who.int/health-topics/cardiovascular-diseases#tab=tab_1 [accessed 2025-10-22]
2. McCarthy JJ, Carr B, Sasson C, et al. Out-of-hospital cardiac arrest resuscitation systems of care: a scientific statement from the American Heart Association. Circulation 2018 May 22;137(21):e645-e660. [doi: 10.1161/CIR.0000000000000557] [Medline: 29483084]
3. Yan S, Gan Y, Jiang N, et al. The global survival rate among adult out-of-hospital cardiac arrest patients who received cardiopulmonary resuscitation: a systematic review and meta-analysis. Crit Care 2020 Feb 22;24(1):61. [doi: 10.1186/s13054-020-2773-2] [Medline: 32087741]
4. Xie X, Zheng J, Zheng W, et al. Efforts to improve survival outcomes of out-of-hospital cardiac arrest in China: BASIC-OHCA. Circ Cardiovasc Qual Outcomes 2023 Feb;16(2):e008856. [doi: 10.1161/CIRCOUTCOMES.121.008856] [Medline: 36503279]
5. Jerkeman M, Sultanian P, Lundgren P, et al. Trends in survival after cardiac arrest: a Swedish nationwide study over 30 years. Eur Heart J 2022 Dec 7;43(46):4817-4829. [doi: 10.1093/eurheartj/ehac414] [Medline: 35924401]
6. Board on Health Sciences Policy, Institute of Medicine, Committee on the Treatment of Cardiac Arrest: Current Status and Future Directions. In: Graham R, McCoy MA, Schultz AM, editors. Strategies to Improve Cardiac Arrest Survival: A Time to Act. [doi: 10.17226/21723] [Medline: 26225413]
7. Wyckoff MH, Greif R, Morley PT, et al. 2022 international consensus on cardiopulmonary resuscitation and emergency cardiovascular care science with treatment recommendations: summary from the basic life support; advanced life support; pediatric life support; neonatal life support; education, implementation, and teams; and first aid task forces. Circulation 2022 Dec 20;146(25):e483-e557. [doi: 10.1161/CIR.0000000000001095] [Medline: 36325905]
8. Nakagawa NK, Oliveira KMG, Lockey A, et al. Effectiveness of the 40-minute handmade manikin program to teach hands-on cardiopulmonary resuscitation at school communities. Am J Cardiol 2021 Jan 15;139:126-130. [doi: 10.1016/j.amjcard.2020.09.032] [Medline: 33007308]
9. Schroeder DC, Semeraro F, Greif R, et al. Kids save lives: basic life support education for schoolchildren: a narrative review and scientific statement from the international liaison committee on resuscitation. Circulation 2023 Jun 13;147(24):1854-1868. [doi: 10.1161/CIR.0000000000001128] [Medline: 37194575]
10. Fijačko N, Masterson Creber R, Gosak L, et al. Evaluating quality, usability, evidence-based content, and gamification features in mobile learning apps designed to teach children basic life support: systematic search in app stores and content analysis. JMIR Mhealth Uhealth 2021 Jul 20;9(7):e25437. [doi: 10.2196/25437] [Medline: 34283034]
11. Lim XMA, Liao WA, Wang W, Seah B. The effectiveness of technology-based cardiopulmonary resuscitation training on the skills and knowledge of adolescents: systematic review and meta-analysis. J Med Internet Res 2022 Dec 15;24(12):e36423. [doi: 10.2196/36423] [Medline: 36520524]
12. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. J Clin Epidemiol 2008 Apr;61(4):344-349. [doi: 10.1016/j.jclinepi.2007.11.008] [Medline: 18313558]

XSL•FO
RenderX

13. Panchal AR, Bartos JA, Cabañas JG, et al. Part 3: adult basic and advanced life support: 2020 American Heart Association guidelines for cardiopulmonary resuscitation and emergency cardiovascular care. Circulation 2020 Oct 20;142(16_suppl_2):S366-S468. [doi: 10.1161/CIR.0000000000000916] [Medline: 33081529]
14. Flato UAP, Beffa Dos Santos EJ, Bispo Diaz T Martins I, et al. Interactive serious game to teach basic life support among schoolchildren in Brazil: design and rationale. JMIR Serious Games 2024 Oct 9;12:e55333. [doi: 10.2196/55333] [Medline: 39382546]
15. Jones I, Whitfield R, Colquhoun M, Chamberlain D, Vetter N, Newcombe R. At what age can schoolchildren provide effective chest compressions? An observational study from the Heartstart UK schools training programme. BMJ 2007 Jun 9;334(7605):1201. [doi: 10.1136/bmj.39167.459028.DE] [Medline: 17468118]
16. Abelairas-Gómez C, Rodríguez-Núñez A, Casillas-Cabana M, Romo-Pérez V, Barcala-Furelos R. Schoolchildren as life savers: at what age do they become strong enough? Resuscitation 2014 Jun;85(6):814-819. [doi: 10.1016/j.resuscitation.2014.03.001] [Medline: 24614187]
17. Plataforma Brasil. 2025. URL: https://plataformabrasil.saude.gov.br/login.jsf [accessed 2025-10-22]
18. NCT06565767. Clinical trials Gov. URL: https://clinicaltrials.gov/study/NCT06565767 [accessed 2025-10-29]
19. Schroeder DC, Finke SR, Grübl T, Jänig CW, Böttiger BW. Education of schoolchildren in cardiopulmonary resuscitation - overview of the current literature. Curr Opin Crit Care 2023 Dec 1;29(6):616-620. [doi: 10.1097/MCC.0000000000001111] [Medline: 37861212]
20. Ko JS, Kim SR, Cho BJ. Hands-only cardiopulmonary resuscitation education for elementary school students in Korea: tracking by school grade, physical characteristics, and physical strength. Int J Public Health 2023;68:1606054. [doi: 10.3389/ijph.2023.1606054] [Medline: 38375354]
21. Jarosławski J, Burzyński J, Kryczka K, et al. Impact of anthropometric factors on chest compression depth during CPR provided by children aged 11-14 in a community-wide study. Resusc Plus 2024 Sep;19:100708. [doi: 10.1016/j.resplu.2024.100708] [Medline: 39100391]
22. Fariduddin MN, Mohamed M, Jaafar MJ, Baharin K, Siau CS, Bashah K. Compression-only cardiopulmonary resuscitation and automated external defibrillator course for primary school students: a Malaysian pilot study. Children (Basel) 2022 Dec 27;10(1):58. [doi: 10.3390/children10010058] [Medline: 36670609]
23. Pivač S, Gradišek P, Skela-Savič B. The impact of cardiopulmonary resuscitation (CPR) training on schoolchildren and their CPR knowledge, attitudes toward CPR, and willingness to help others and to perform CPR: mixed methods research design. BMC Public Health 2020 Jun 12;20(1):915. [doi: 10.1186/s12889-020-09072-y] [Medline: 32532235]
24. Kherbeche H, Exer N, Schuhwerk W, Ummenhofer W, Osterwalder J. Chest compression using the foot or hand method: a prospective, randomized, controlled manikin study with school children. Eur J Emerg Med 2017 Aug;24(4):262-267. [doi: 10.1097/MEJ.0000000000000335] [Medline: 26485692]
25. Bánfai B, Pandur A, Pék E, Csonka H, Betlehem J. At what age can children perform effective cardiopulmonary resuscitation? - Effectiveness of cardiopulmonary resuscitation skills among primary school children. Orv Hetil 2017 Jan;158(4):147-152. [doi: 10.1556/650.2017.30631] [Medline: 28116931]
26. Finke SR, Schroeder DC, Ecker H, et al. Gender aspects in cardiopulmonary resuscitation by schoolchildren: a systematic review. Resuscitation 2018 Apr;125:70-78. [doi: 10.1016/j.resuscitation.2018.01.025] [Medline: 29408490]
27. Wingen S, Ecker H, Schroeder DC, Bartholme B, Böttiger BW, Wetsch WA. Addressing the helper's and victim's gender is crucial in schoolchildren resuscitation training-A prospective, educative interventional trial. J Clin Med 2022 Apr 24;11(9):2384. [doi: 10.3390/jcm11092384] [Medline: 35566510]
28. Fernandes JMG, Leite ALDS, Auto BDSD, Lima JEGD, Rivera IR, Mendonça MA. Teaching basic life support to students of public and private high schools. Arq Bras Cardiol 2014 Jun;102(6):593-601. [doi: 10.5935/abc.20140071] [Medline: 25004421]
29. Simielli L. Students' access to qualified teachers is unequal, but how much? Measuring gaps and trends in Brazil. Int J Educ Dev 2023 Oct;102:102837. [doi: 10.1016/j.ijedudev.2023.102837]
30. de Oliveira AM, Santos JLF, Capellini SA. Reading processes of public and private middle school and high school students. Psicol Reflex Crit 2024 Apr 15;37(1):14. [doi: 10.1186/s41155-024-00296-0] [Medline: 38619703]
31. Nazerian P, Mueller C, Soeiro ADM, et al. Diagnostic accuracy of the aortic dissection detection risk score plus D-dimer for acute aortic syndromes: the ADvISED prospective multicenter study. Circulation 2018 Jan 16;137(3):250-258. [doi: 10.1161/CIRCULATIONAHA.117.029457] [Medline: 29030346]
32. Kalz M, Lenssen N, Felzen M, et al. Smartphone apps for cardiopulmonary resuscitation training and real incident support: a mixed-methods evaluation study. J Med Internet Res 2014 Mar 19;16(3):e89. [doi: 10.2196/jmir.2951] [Medline: 24647361]
33. Semeraro F, Monesi A, Gordini G, Del Giudice D, Imbriaco G. Kids Save Lives: a blended learning approach to improve engagement of schoolchildren. Resuscitation 2023 Jan;182:109675. [doi: 10.1016/j.resuscitation.2022.109675] [Medline: 36572375]
34. Otero-Agra M, Barcala-Furelos R, Besada-Saavedra I, Peixoto-Pino L, Martínez-Isasi S, Rodríguez-Núñez A. Let the kids play: gamification as a CPR training methodology in secondary school students. a quasi-experimental manikin simulation study. Emerg Med J 2019 Nov;36(11):653-659. [doi: 10.1136/emermed-2018-208108] [Medline: 31409636]

XSL•FO
RenderX

35. Latsios G, Synetos A, Tsioufis K, Kanakakis J. CPR education in schools: the European “Kids-Save-Lives” initiative and our efforts in Greece. Hellenic J Cardiol 2023;69:59-60. [doi: 10.1016/j.hjc.2022.09.006] [Medline: 36167248]
36. Wilson-Sands C, Brahn P, Graves K. The effect of instructional method on cardiopulmonary resuscitation skill performance: a comparison between instructor-led basic life support and computer-based basic life support with voice-activated manikin. J Nurses Prof Dev 2015;31(5):E1-E7. [doi: 10.1097/NND.0000000000000203] [Medline: 26381346]
37. Borovnik Lesjak V, Šorgo A, Strnad M. Retention of knowledge and skills after a basic life support course for schoolchildren: a prospective study. Inquiry 2022;59:469580221098755. [doi: 10.1177/00469580221098755] [Medline: 35652386]
38. Greif R, Bray JE, Djärv T, et al. 2024 international consensus on cardiopulmonary resuscitation and emergency cardiovascular care science with treatment recommendations: summary from the basic life support; advanced life support; pediatric life support; neonatal life support; education, implementation, and teams; and first aid task forces. Circulation 2024 Dec 10;150(24):e580-e687. [doi: 10.1161/CIR.0000000000001288] [Medline: 39540293]
39. Donoghue A, Sawyer T, Olaussen A, Greif R, Toft L. Gamified learning for resuscitation education: a systematic review. Resusc Plus 2024 Jun;18:100640. [doi: 10.1016/j.resplu.2024.100640] [Medline: 38666256]
40. Cons-Ferreiro M, Mecias-Calvo M, Romo-Perez V, Navarro-Patón R. Learning of basic life support through the flipped classroom in secondary schoolchildren: a quasi-experimental study with 12-Month follow-up. Medicina (Kaunas) 2023 Aug 23;59(9):1526. [doi: 10.3390/medicina59091526] [Medline: 37763645]
41. Abelairas-Gómez C, Martinez-Isasi S, Barcala-Furelos R, et al. Training frequency for educating schoolchildren in basic life support: very brief 4-month rolling-refreshers versus annual retraining-a 2-year prospective longitudinal trial. BMJ Open 2021 Nov 30;11(11):e052478. [doi: 10.1136/bmjopen-2021-052478] [Medline: 34848519]

Abbreviations

AHA: American Heart Association
ALS: advanced life support
BLS: basic life support
CPR: cardiopulmonary resuscitation
ERC: European Resuscitation Council
IDEB: Brazilian Education Development Index
ILCOR : International Liaison Committee on Resuscitation
KSL: Kids Save Lives
OHCA: out-of-hospital cardiac arrest
QCPR: quality cardiopulmonary resuscitation
SG: serious game
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology
WHO: World Health Organization

Edited by A Coristine; submitted 25.11.24; peer-reviewed by D Kitapcioglu, P Aikawa, S Lopez-Garcia; revised version received 30.08.25; accepted 31.08.25; published 10.11.25.

Please cite as:
Flato UAP, Flato A, Martins IBDT, Simoes Nakano G, Romao JC, Nakano MS, dos Santos EJB, Villas Boas YDT, Medeiros LE, Rossignoli VG, Cunha GZ, Batista RC, da Silva PGR, Florentino M, Rabelo A, Midega TD, Passos R, Guimaraes H, Kern K
Enhancing Equity in Schoolchildren's Basic Life Support Education in Brazil Through Serious Games: Cohort Study
JMIR Serious Games 2025;13:e69252
URL: https://games.jmir.org/2025/1/e69252
doi:10.2196/69252

© Uri Adrian Prync Flato, Adriana Flato, Isabella Bispo Diaz T Martins, Giuliana Simoes Nakano, Julia Caroline Romao, Manuela Simões Nakano, Emilio José Beffa dos Santos, Yuuki Daniel Tahara Villas Boas, Leonardo Escobar Medeiros, Vinicius Gazin Rossignoli, Gabriel Zanatta Cunha, Rafael Carreira Batista, Pedro Gazotto Rodrigues da Silva, Miguel Florentino, Amanda Rabelo, Thais Dias Midega, Rogerio Passos, Helio Guimaraes, Karl Kern. Originally published in JMIR Serious Games (https://games.jmir.org), 10.11.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

A Serious Game (Health Unit in Focus) for Enhancing Undergraduate Education on Older Adults' Health: Design and Validation Study

Yuri Gustavo de Sousa Barbalho[1,2*], MSc; Calliandra Maria de Souza Silva[1*], PhD; Carla Sílvia Neves da Nora Fernandes[2,3*], PhD; Raíza Rana de Souza Lima Trombini[1*], MSc; Patrícia Littig Melo[1*], BSc; Aline Farias de Oliveira[1*], BSc; Alayne Larissa Martins Pereira[1*], MSc; Alessandro de Oliveira Silva[4*], PhD; Luciano Ramos de Lima[5*], PhD; Marina Morato Stival[1*], PhD; Diana Lúcia Moura Pinho[1*], PhD; Silvana Schwerz Funghetto[1*], PhD

[1]Faculty of Health Sciences and Technologies, University of Brasília, University Campus - Metropolitan Center, Ceilândia Sul, Brasília, Brazil

[2]School of Health, Polytechnic Institute of Viana do Castelo, Viana do Castelo, Portugal

[3]Aditgames – Association for Innovation, Technologies and Games in Health, Porto, Portugal

[4]Physical Education and Medicine Department, University Center of Brasilia, Brasília, Brazil

[5]Nursing Department, University of Brasília, Brasilia, Brazil

[*]all authors contributed equally

Corresponding Author:
Yuri Gustavo de Sousa Barbalho, MSc
Faculty of Health Sciences and Technologies, University of Brasília, University Campus - Metropolitan Center, Ceilândia Sul, Brasília, Brazil

Abstract

Background: Population aging underlines the critical need to improve health professional training to adequately care for adults aged >60 years. Developing educational resources to support academics and professionals presents a valuable opportunity to enhance understanding of health conditions and improve clinical management. Serious games are designed to develop teaching, training, and learning skills. Their use in the educational setting is warranted, as they integrate digital aspects and gamification to create a playful experience for content acquisition. Deepening this theme in nursing education will improve assistance to the older adult population, leading to more qualified care based on gerontological practices and comprehensive health care for older adults.

Objective: This study aims to develop and validate a serious game on older adult health for undergraduate nursing students.

Methods: This quantitative and descriptive methodological study, conducted between February 2023 and December 2023 at a public university in the Federal District of Brazil, involved the active participation of 27 undergraduate nursing students in their eighth to tenth semesters. The game, Health Unit in Focus (HUF), was developed and validated with their input. It features 75 clinical cases distributed across 3 themes: pharmacology, metabolic syndrome, and semiology. Of the 40 students initially enrolled, 27 completed the study. The app was validated using the System Usability Scale and student feedback, and the results were reported following the Game-Based Intervention Reporting Guidelines (GAMING).

Results: The participants had a mean age of 22.67 (SD 1.44) years, were mostly female (20/27, 74%), and were in their eighth semester (26/27, 96%). The game received an average System Usability Scale score of 85.75 (median 86.57), classified as excellent, as all evaluated items scored >75. Participants considered the game easy to use; accessible; practical; and rich in well-founded, useful content. This high usability score, coupled with the overwhelmingly positive feedback from the students, instills confidence in the game's effectiveness. Furthermore, 100% (27/27) of students agreed that learning through games is effective and expressed interest in incorporating more interactive games into their training. The serious game HUF showed good usability, as its overall score was "excellent," with its highest score in the odd-numbered items that addressed the positive aspects identified in the analysis.

Conclusions: The serious game HUF is not just a valid and reliable tool for training nursing students but also an engaging and interactive approach to learning. Its ability to captivate and involve students in the learning process is a testament to its potential to revolutionize nursing education. It is essential that the development of new methodological resources, such as serious games, be based on scientific evidence to guarantee greater reliability and success in achieving their established objectives.

(JMIR Serious Games 2025;13:e66289) doi:10.2196/66289

XSL•FO
RenderX

KEYWORDS

serious games; nursing education; aged; older adults; design; validation

Introduction

Background

On the global stage, the health concerns of older adults (aged >60 y) have gained importance driven by population aging and the epidemiological transition. This transition reflects a shift in the pattern of mortality and disease, with a growing prevalence of chronic noncommunicable diseases in older age groups [1]. United Nations data show a 3-fold increase in older adults between 1980 and 2022. By 2023, this demographic was expected to reach 994 million, accounting for 12% of the global population. Over the next 3 decades, the older adult population worldwide is projected to more than double, reaching >1.5 billion people by midcentury [2].

Under these circumstances, health practices should prioritize promoting active aging by enhancing quality of life, supporting health maintenance, and preventing disease [3]. In the 2000s, the World Health Organization introduced “active aging” policies that addressed various determinants, including health, social, and economic factors, while also emphasizing the need for adequately trained professionals to support and assist this population [4].

In this sense, developing educational resources to support academics and professionals presents a valuable opportunity to enhance understanding of health conditions and improve clinical management. These resources can take various forms, such as guides, brochures, applications, and games, providing information that can effectively optimize the teaching-learning process in health [5].

Serious games are designed to develop teaching, training, and learning skills, with the first records dating back to the 1970s [6]. These days, they are most typically used in digital games, although they can also be used in physical games. Their use in the educational setting is warranted because they integrate digital aspects and gamification to create a playful content acquisition experience [7]. In recent years, they have been widely applied in health education [8,9].

One study compared the effects of traditional lectures with a serious game on nursing students’ knowledge and skills in assessing patients with burns, finding a significant improvement in educational outcomes among those who used the game [10]. Similarly, another study explored the effects of using a serious game in a home environment, reporting an improvement in theoretical knowledge related to the game’s central theme [11].

Objectives

This study introduces a novel approach to teaching issues related to older adult health conditions commonly encountered in primary health care (PHC). The serious game, designed for any mobile device, offers content tailored to the target population and is adaptable to students’ daily lives. At the time of this research, our game was the only application identified that integrated these specific characteristics, making it a unique and potentially impactful tool in nursing education.

Deepening this theme in nursing education will improve assistance to the older adult population, leading to more qualified care based on gerontological practices and comprehensive health care for older adults.

Once developed and validated, a serious learning game focused on the health of older adults with chronic noncommunicable diseases could be a strategic tool for promoting independent learning among students. Building on this potential for self-directed learning, this study, rooted in methodological development research, aims to create and validate a serious game on older adults’ health for undergraduate nursing students.

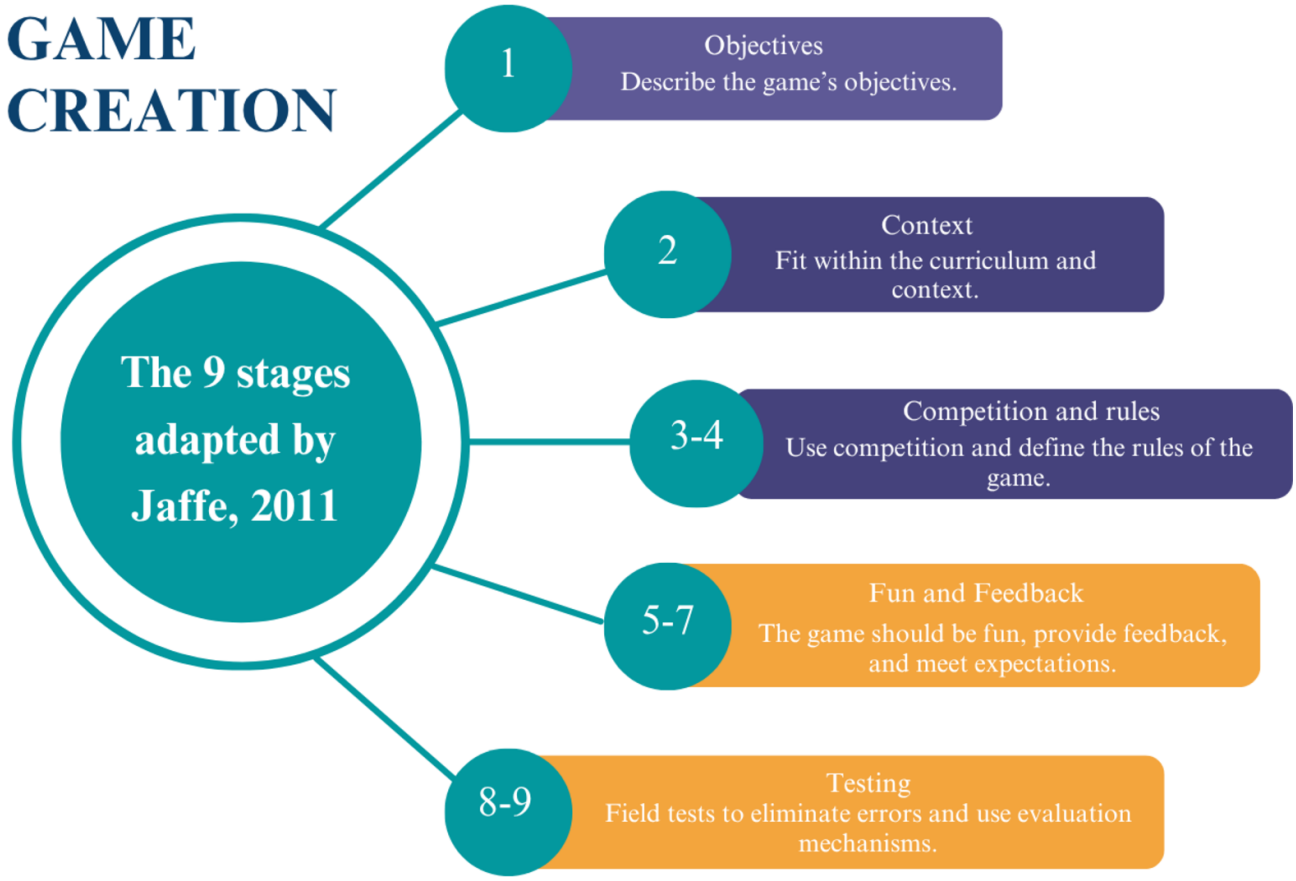
Methods

Study Development and Population

This methodological development study, using a quantitative and descriptive approach, was conducted at a public university in the Federal District of Brazil between February 2023 and December 2023 and was reported according to the recommendations of the Game-Based Intervention Reporting Guidelines (GAMING) [12]. The stages proposed by Jaffe [13] were used for the construction of the serious game. Figure 1 [13] describes the stages of the game’s development, testing, and validation.

XSL•FO
RenderX

Figure 1. Stages of game construction (adapted from Jaffe [13]).



In stage 1, the game's objectives, educational aims, content, and problematizations were established. For stage 2, the context in which students would use the theme was surveyed. In stage 3, the competitive aspects were adjusted to the application's proposal, as it is a single-player game. For the rules, in stage 4, a "help" tab was set up with all the rules and guidelines. Stage 5 aimed to integrate the gamification aspects. Stage 6, immediate feedback, was included immediately after the participants' responses.

In stage 7, the researchers of this study analyzed and carefully studied such variables as students' desire and expectations regarding educational games, especially because they were inserted within the academic context of the population. For evaluation, in step 8, the study conducted a pilot test with 13 undergraduate students who met the minimum knowledge requirements for game use. All participants were enrolled in a discipline that dealt with aging in health. In the last step, the support team adjusted the notes based on the evaluation from the previous step.

The themes included in the serious game were developed based on the needs of nursing undergraduates, focusing on areas of older adults' health where students faced the greatest difficulty or expressed the most interest in deepening their knowledge. This step of assessing students' needs had already been conducted previously. The game incorporates 75 clinical cases, with 25 cases for each of the following themes: pharmacology, metabolic syndrome, and semiology in the context of older adults' health (Multimedia Appendix 1). All clinical cases had been previously validated by expert judges.

The cases were built and validated by experts on older adult health topics. Each case includes contextual information and a true-or-false statement, based on the most prevalent themes in the clinical context of each older adult health topic.

Sample

After the serious game application was fully developed, its usability was evaluated by eighth-semester undergraduate nursing students at a public university. The initial sample consisted of 40 enrolled students, of which n=13 (32.5%) were excluded (10/40, 25% for not attending, 2/40, 5% for withdrawing from the course during collection, and 1/40, 2.5% for dropping out during the validation process), resulting in n=27 final participants. It is assumed that at a 95% CI, at least 70% of the experts will classify the item as appropriate. The calculation would be $n=1.96^2\times0.85\times0.15/0.15^2=22$ students [14]. To detect effect sizes of 0.85 with an α level of .05 and a power of 95%, a sample size of 22 participants was required. Considering the likelihood of participant dropout or ineligibility, we increased our target sample size to 27 participants. Although sample size calculations ensured adequate statistical power, future multicenter studies are recommended to strengthen external validity and account for institutional and regional educational variations.

Data Collection Tools

Before validation, students completed a characterization questionnaire collecting information on age, sex, semester, education, year of entry into their current degree, participation in scientific initiation, prior experience with educational

XSL•FO
RenderX

technologies (such as games), beliefs about the potential to learn through games, and frequency of mobile game use. The scale used for this validation was the System Usability Scale (SUS) [15]. The SUS and Educational Games Evaluation Scale were selected for their demonstrated reliability, ease of interpretation, and extensive validation in educational technology research.

This scale assesses the user's degree of agreement using a Likert scale, ranging from "totally agree" to "totally disagree." The final score is calculated by summing the individual contributions of each item. For even-numbered items, subtract the user's response from 5 (5−user's response); for odd-numbered items, subtract 1 from the user's response (user's response−1). The resulting values are then added and multiplied by 2.5, yielding a "user satisfaction index" ranging from 0 to 100. The score is classified as follows: worst imaginable (0 - 20.5), poor (21 - 38.5), average (39 - 52.5), good (53 - 73.5), excellent (74 - 85.5), and best imaginable (86-100). The system can be further classified based on the following categories: ease of memorization (item 2), ease of comprehension (items 3, 4, 7, and 10), system efficiency (items 5, 6, and 8), user satisfaction (items 1, 4, and 9), and identification of inconsistencies (item 6) [16,17]. This instrument was printed and delivered to participants.

The serious game was evaluated using the Educational Games Evaluation Scale proposed by Savi et al [18] to assess game quality. The scale comprises 3 main components: motivation (10 questions), user experience (16 questions), and learning (3 questions), totaling 29 items. Responses are recorded on a 5-point Likert-type scale.

Statistical Analysis

Data from the answers database were processed and analyzed using SPSS (version 20.0; IBM Corp). Statistical tests were conducted as necessary, with P values <.05 considered statistically significant. The Shapiro-Wilk test was used to assess normality, and the data were found to be normally distributed. Content validity was evaluated based on the questions' relevance, purpose, and clarity of the information. Reliability was calculated by the Cronbach α index, which estimates the internal consistency of the questionnaire items [19].

The Likert-type scales have the following configurations: judgments with responses 1 and 2 characterize total disagreement and disagreement with the statement, respectively. Judgment with response 3 in the answers characterizes indecision; the student neither agrees nor disagrees with the statement. Meanwhile, judgments with responses 4 and 5 characterize agreement with and total agreement with the statement, respectively.

Ethical Considerations

This study was rigorously designed and implemented according to the ethical guidelines established by the Declaration of Helsinki [20], with ethics approval granted by the Research Ethics Committee of the Faculty of Ceilândia at the University of Brasília (protocol 5.237.938). Participants engaged actively by downloading their match data and submitting it securely through the university's educational e-platform. This platform provided a protected digital storage area exclusively accessible to the research team, ensuring participant confidentiality.

Prior to enrollment, all potential participants received clear and detailed explanations of the study's purpose, methods, and procedures, as well as the specific types of data to be collected. Written informed consent was voluntarily obtained from each participant, emphasizing their autonomy and freedom to withdraw from the research at any point without any repercussions or adverse outcomes.

To protect the participants' right to privacy, personal identifiers were safeguarded in secure, password-protected digital files. During data analysis and the reporting of results, only anonymized or pseudonymized information was used. Specifically, each nursing student participant was assigned a unique numeric identifier according to the sequential order of their questionnaire submissions, ensuring complete anonymity throughout the research process.

Participants received no financial incentives for their involvement, aligning with ethical standards promoting voluntary and unbiased participation. Additionally, considerable care was taken to prevent any potential identification through the publication or supplementary materials associated with this study. In instances where the use of potentially identifiable images was unavoidable, explicit written consent was obtained from the concerned individuals. However, no identifiable images were ultimately included in this paper or its supplementary documentation, thereby reinforcing the commitment to participants' confidentiality and ethical research conduct.

Results

Serious Game: Health Unit in Focus

The final product of this study is a serious game app called Health Unit in Focus (HUF) for mobile devices. HUF is a single-player question-and-answer game composed of clinical cases with online and offline functionality. On its home screen are fields for log-in and password if the user is already registered and a field for creating a new account.

Also, on its home screen, the player can access the game's established areas: pharmacy for pharmacology cases, laboratory for metabolic syndrome, and consulting room for semiology. The player can also consult solved cases and other players' achievements in the archive and secretary areas. Furthermore, a help tab provides information on game areas, how to play, emergencies, content review, achievements, and profiles.

Upon entering the game area, participants must respond to a clinical case on their chosen topic. These cases are categorized as easy, medium, or difficult, and participants must complete the initial levels to advance to more challenging ones. Immediate feedback is provided after each response, explaining the correct answer. Emergency scenarios will also appear during gameplay, requiring the participant to respond within 60 seconds. Figure 2 presents some of the tabs from the HUF application.

XSL•FO
RenderX

Figure 2. Tabs of the Health Unit in Focus app.



Stages in the Construction of the Serious Game

The game's scenarios focus on older adults' health issues within the PHC context, combining technological resources with gamification and innovation to create an engaging and effective learning experience [21]. The construction of the serious game involved 9 stages.

Stage 1: Developing and Describing the Game's Specific Objectives

According to Jaffe [13], developing an educational game requires determining the content; problematizing it; and setting clear, context-aligned objectives. Accordingly, the application aimed to enhance knowledge of older adults' health in PHC clinical contexts by engaging students with problem-based content on pharmacology, chronic noncommunicable diseases (metabolic syndrome), and semiology.

The game development explicitly integrated pedagogical theories, such as experiential learning and self-directed learning, to ensure that clinical cases and interactive elements facilitated deeper cognitive engagement and autonomous knowledge construction.

Stage 2: The Context Is Inserted (Contextual Integration)

This serious game model was intended for use by undergraduate nursing students during their final year of study.

Stage 3: Competitive Space

As a single-player game, HUF included a feature that allows users to compare their achievements, introducing an element of competition.

Stage 4: Defining the Game Rules

The game included detailed rule descriptions, gameplay mechanics, game areas, content review, and how to earn new achievements. Immediate feedback and emergency scenarios were purposefully designed to replicate the urgency of clinical decision-making and foster real-time critical thinking skills directly related to clinical practice.

Stage 5: Fun

Gamification aspects commonly seen in games and applications were incorporated to enhance engagement [22].

Stage 6: Immediate Feedback

Upon evaluating a clinical case and selecting a response, the user receives immediate feedback on the success or failure of their choice.

Stage 7: Meeting the Needs of the Participants

The themes in HUF are designed to address the target audience's needs, considering factors such as age, gender, culture, and academic level.

Stages 8 and 9: Testing and Evaluation Mechanisms

This research involves testing the application prototype and assessing its usability with undergraduate nursing students.

During stages 8 and 9, a total of 13 undergraduate nursing students (mean age 22 years; 9/13, 69% female) participated in the study. Their profile closely matched the final target audience. The students assessed the game's quality using a scale developed by Savi et al [18], which was divided into 3 categories: motivation, user experience, and learning. The results indicate that most statements received a satisfactory average in student

XSL•FO
RenderX

evaluations, suggesting that the game can be effectively used as an educational tool based on the evaluated points (Multimedia Appendix 2).

Four items positively highlighted the motivation and user experience categories, with average ratings of 4.6 and 4.8, respectively. These items pertained to the game's relevance, recommendation to friends, interest in playing again, and ease of learning. The learning domain also stood out, with high averages in 2 of its 3 items. These statements focused on the game's contribution to knowledge and its efficiency in learning.

Usability Validation (Sample)

The sample that validated the game's usability closely matches the desired profile of the serious game's end users. It comprised 27 undergraduate nursing students, primarily women, all in the eighth semester of their undergraduate degree.

The need for digital literacy was not a concern for these researchers, as the students already had previous experience of using digital platforms in previous subjects, such as Mentimeter and Kahoot, among others. This academic profile demonstrated previous familiarity with educational technologies, making it easier to adapt to and interact with new digital tools in the teaching-learning context.

Although participants generally possessed adequate digital literacy from prior educational experiences, variability in digital proficiency levels could influence engagement and game effectiveness, which should be considered when generalizing results.

The participants had a mean age of 22.66 (SD 1.44; range 21-27) years. The participants unanimously agreed on the value of incorporating games into academic training and recognized the potential for learning through games (Table 1). The students used the application before they were in contact with their practice scenarios (supervised internship).

Students could download the app through a QR code, which optimized its access. The game takes approximately 35 minutes, so the researchers timed the students for 40 minutes to complete the task.

Table . Characteristics of the usability test participants (N=27).

Characteristics	Participants, n (%)
Sex	
Female	20 (74.1)
Male	7 (25.9)
Semester	
Eighth	26 (96.3)
Tenth	1 (3.7)
Other experiences with ET[a]	
Yes	16 (59.3)
No	11 (40.7)
Do you like games?	
Yes, I like it a lot	18 (66.7)
Yes, I like it a little	9 (33.3)
I do not like it	0 (0)
Do you think it is possible to learn through games?	
Yes	27 (100)
No	0 (0)
Would you like to include interactive games during your undergraduate course?	
Yes	27 (100)
No	0 (0)

[a]ET: educational technology.

Validation

Based on the degree of agreement with the items evaluated using the SUS, which focused on assessing the suitability of the serious game as an educational tool on issues related to older adults' health, the participants found the game easy to use, accessible, useful, and practical with well-founded content. The final score was 85.75, considered excellent, with a median of 86.57. All the other items scored >75, with a Cronbach α of 0.755 (Figure 3).

The items evaluated by SUS have specific usability characteristics with relevant meanings. Their scores have value intervals between 0 and 100 points to help quantify and qualify them, thus demonstrating software quality (Table 2). The score is classified as follows: worst imaginable (0 - 20.5), poor

XSL•FO
RenderX

(21 - 38.5), average (39 - 52.5), good (53 - 73.5), excellent (74 - 85.5), and best imaginable (86-100).

Based on individual participant responses, the serious game's average usability value most frequently fell within the 86 - 100 range on a scale from worst imaginable to best imaginable (Table 3).

Figure 3. Evaluation of Health Unit in Focus application's items using the System Usability Scale.

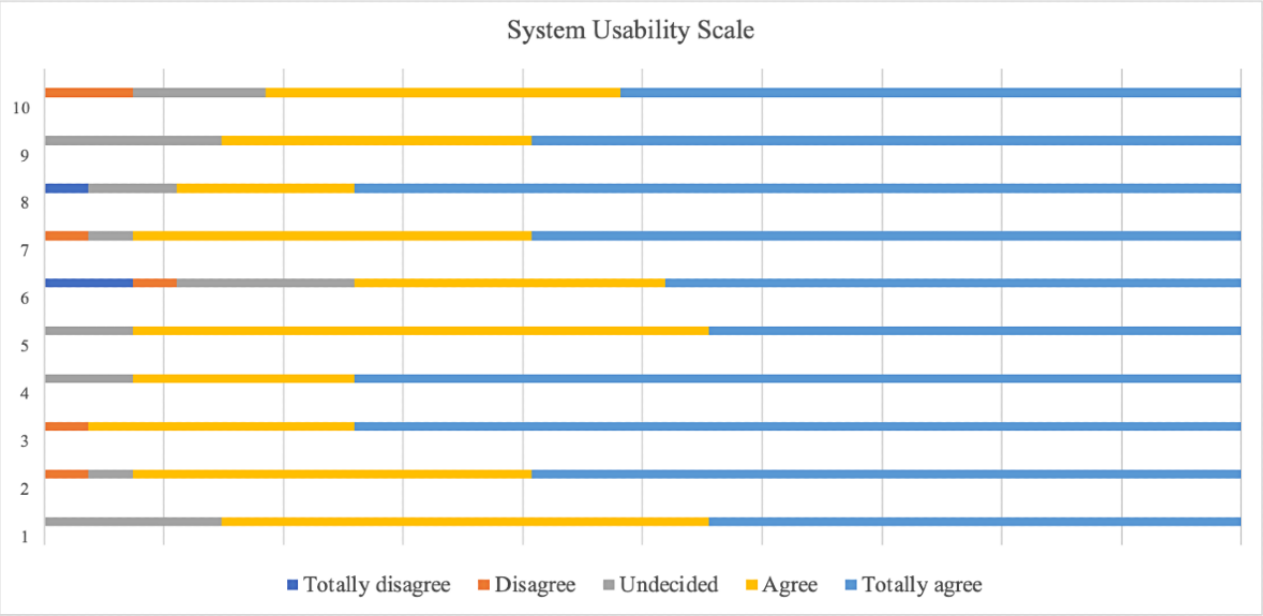


Table . Usability of the "Health Unit in Focus" application as evaluated by the System Usability Scale (N=27).

Usability characteristics	Scores, median	Meaning
Ease of learning	89.34	Easy to understand
Efficiency	84.25	Agility in executing proposed tasks
Inconstancy	75.92	Absence or low number of errors
Ease of memorization	87.03	Easy-to-use system even after a period of inactivity
Satisfaction	86.11	Pleasant design
Total	84.53	Excellent

Table . Categorization and classification of the Health Unit in Focus (HUF) application usability using the System Usability Scale (SUS; N=27).

Categorization scale	Responses, n (%)	SUS classification
0 - 20.5	0 (0)	Worst imaginable
21 - 38.5	0 (0)	Poor
39 - 52.5	1 (3.7)	Average
53 - 73.5	2 (7.4)	Good
74 - 85.5	9 (33.3)	Excellent
86 - 100	15 (55.6)	Best imaginable

Discussion

Principal Findings

The effective use of educational technologies in the teaching-learning process is a hallmark of modern education, particularly in health education. Their impact lies in supporting personal, technical, and theoretical development through an engaging and enjoyable learning experience. In addition to this, these technologies optimize time, work, and even health care. During the training process, they can be used to prepare students for the field because clinical practice already uses several mobile technologies [23].

The development of the HUF serious game intends to facilitate access to everyday content on older adults' health in a playful

XSL•FO
RenderX

and interactive manner, and the target population validated it. The conditions added to this app describe possible clinical cases that appear daily in a nursing professional's routine and are sometimes not covered in traditional classes. In this way, students may be more aware of how to approach these cases.

While results indicate strong usability and satisfaction within this specific academic setting, replication in diverse educational contexts, particularly those with varying degrees of technological infrastructure and student digital literacy, is essential for broader applicability.

It is essential that the development of new methodological resources, such as serious games, be based on scientific evidence to ensure greater reliability and success in their established objectives [24]. Studies show that educational games tend to achieve good results using standardized methods to meet educational demands [25,26]. Therefore, this study sought references previously used in the literature to construct and validate the HUF application [13,15].

Technological resources such as serious games must arouse students' interest, be compatible with their needs, and suit (adapt to) their cognitive style. In this way, this technology can foster knowledge formation [27]. The HUF was developed based on feedback from nursing students, giving prominence to their most challenging subjects. We therefore sought to meet the needs of this audience.

Technologies should be engaging and enjoyable in addition to meeting user expectations. Combining leisure, engagement in the clinical context, and knowledge acquisition are key aspects of gamification and are recognized as essential elements of an enjoyable experience [28]. The proposed serious game incorporates these aspects, offering features such as totems, trophies, and opportunities for advancement and new achievements.

This approach demonstrates how technological resources facilitate the integration of theoretical aspects, practical daily life routines, and student accessibility [9].

Usability evaluation guarantees a system's ease, suitability, and rapid learning in handling while seeking to minimize operational errors. Therefore, usability evaluation seeks greater efficiency, safety, and security in implementing such systems, making the user experience more pleasant [29].

According to Brooke [30], the aspects assessed by the SUS make it a reliable and robust evaluation tool with a high degree of validity. They are essential for a good software result, covering a range of system aspects such as complexity, need for support, interface, and others.

Regarding these aspects, the HUF serious game demonstrated good usability, as the overall score was considered excellent. HUF's highest score was in the odd-numbered items that deal with positive aspects found in the analysis. Other findings present a variation in usability between 78.8 and 88.2 among young students [31-33].

There is a constant growth in the use of serious games as a learning tool in the health sector. These digital games, developed both for educational and entertainment purposes, have been widely used to improve the theoretical and practical knowledge of students and health professionals. Several studies have shown that serious games can improve knowledge retention, clinical skills development, and decision-making, making them a promising pedagogical resource for training more prepared and qualified professionals [34,35].

The SUS is a widely adopted tool for measuring the usability of various software and hardware products. However, its original conception was not specifically aimed at evaluating mobile apps. However, studies have shown its effective use with promising results [36,37].

A study carried out in Brazil evaluated the usability of a digital tool using SUS. The study reported a usability score of 75 [38]. Another study, carried out in Germany, validated a virtual reality serious game for patients. The SUS results showed a final score of 86.9 [39]. These studies are in line with the findings reported in this paper.

The scale used also allows for assessing user satisfaction, which is considered a key component of usability [30]. The HUF demonstrated high user satisfaction, with a score 86 on items 1, 4, and 9, supporting this assessment. Previous studies have shown that the use of serious games is associated with similarly high satisfaction rates [40-42].

The modern scenario involves technologies in the daily lives of the entire population, especially young people and adults. This involvement has brought new needs and challenges to producing relevant and acceptable educational products. In this sense, the researcher must be able to transform knowledge into attractive, coherent, and intriguing products for the user [43]. According to the participants' evaluation, HUF proved to be a pleasant and easy-to-use system with low errors.

Integrating innovative tools and strategies into education and training is not a novel concept. However, combining new dynamic and interactive approaches is critical for enhancing learning [44]. In addition to the educational or training process, technological advancements have resulted in higher-quality nursing care in a variety of contexts [45-47].

Games, especially serious games, are advantageous for undergraduate health courses, positively impacting student motivation and satisfaction. As educators, teachers should be attentive to complementing learning methods with resources such as games as long as they are reliable and validated [48]. Notably, the HUF serious game has reliable internal consistency and has been appropriately validated.

Limitations

This study has several limitations. The application use time was approximately 40 minutes, which could limit its depth if not integrated into the participants' regular academic routine. Immediate access also depended on a stable internet connection or a prior download. The small sample size and the fact that it was conducted with students in a single center are notable limitations. The study did not include before or after testing or comparisons with other instructional formats. As no other studies have addressed this topic to date, the potential influence of the novelty effect on the results remains unexplored. Lastly,

XSL•FO
RenderX

functional technical limitations, such as the lack of communication between synchronous players, may have influenced the outcomes.

Future studies should incorporate longitudinal designs, multicenter participation, and comparative evaluations against traditional and alternative digital education formats to better evaluate long-term effectiveness and transferability of learned skills.

Conclusions

In the current scenario, characterized by rapid technological advancements and easy access to information, the challenge of enhancing the educational process, particularly in the health sector, is significant. Developing technological tools that align with the target audience's needs can enhance the learning process. The HUF serious game was validated with undergraduate students, with promising results for its use and potential replication in similar contexts. Furthermore, HUF features easy access to information, a user-approved interface, and content relevant to older adults' health, making it a valuable educational resource.

Acknowledgments

The authors would like to thank all participants of this study, as well as the Federal District Research Support Foundation, the Coordination for the Improvement of Higher Education Personnel, and the University of Brasília, for supporting this research.

Data Availability

All data generated or analyzed during this study are included in this paper and its supplementary materials.

Authors' Contributions

YGdSB contributed to conceptualization, writing (review and editing), validation, and writing (original draft). CMdSS contributed to funding acquisition, investigation, methodology, and project administration. CSNdNF contributed to formal analysis. RRdSLT, PLM, AFdO, ALMP, and AdOS contributed to funding acquisition, investigation, methodology, and project administration. LRdL contributed to conceptualization and supervision. MMS contributed to formal analysis and supervision. DLMP contributed to conceptualization, supervision, project administration SSF contributed to conceptualization, supervision, and project administration.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Examples of clinical cases.
[DOCX File, 16 KB - games_v13i1e66289_app1.docx]

Multimedia Appendix 2
Responses to the educational games evaluation model.
[DOCX File, 16 KB - games_v13i1e66289_app2.docx]

References

1. Barroso IL, Viana JA, et al. The aging population: aspects of Brazil and the world, under the look of literature. Braz J Dev 2020;6(8):61871-61877. [doi: 10.34117/bjdv6n8-564]
2. World population prospects 2022: summary of results population division. United Nations. URL: https://www.un.org/development/desa/pd/sites/www.un.org.development.desa.pd/files/wpp2022_summary_of_results.pdf [accessed 2025-09-22]
3. Oliveira NGN, Tavares D. Active ageing among elderly community members: structural equation modeling analysis. Rev Bras Enferm 2020;73(suppl 3):e20200110. [doi: 10.1590/0034-7167-2020-0110] [Medline: 33111807]
4. Sousa NDS, Lima MG, Cesar CLG, Barros MDA. Active aging: prevalence and gender and age differences in a population-based study. Cad Saude Publica 2018 Nov 23;34(11):e00173317. [doi: 10.1590/0102-311X00173317] [Medline: 30484561]
5. Gonzaga AR, de Jesus LM, Duque AM. Proposal for an educational guide on active aging and cognitive stimulation for the elderly. Revisbrato 2022;6(4):1308-1327. [doi: 10.47222/2526-3544.rbto53872]
6. Abt CC. Serious Games: Viking Press; 1970.
7. Tsekleves E, Cosmas J, Aggoun A. Benefits, barriers and guideline recommendations for the implementation of serious games in education for stakeholders and policymakers. Brit J Educational Tech 2016 Jan;47(1):164-183. [doi: 10.1111/bjet.12223]

XSL•FO
RenderX

8. da Nova Fernandes CSN, Couto G, Afonso A. An aging simulation game's impact on the attitudes of nursing students. Nurs Pract Today 2019;6(3):142-151. [doi: 10.18502/npt.v6i3.1257]
9. Fernandes CS, Angelo M, Martins MM. Giving voice to caregivers: a game for family caregivers of dependent individuals. Rev Esc Enferm 2018:e03309. [doi: 10.1590/s1980-220x2017013903309] [Medline: 29947701]
10. Nasirzade A, Deldar K, Froutan R, Shakeri MT. Comparison of the effects of burn assessment mission game with feedback lecture on nursing students' knowledge and skills in the burn patients' assessment: a randomized clinical trial. BMC Med Inform Decis Mak 2024 Jun 5;24(1):157. [doi: 10.1186/s12911-024-02558-4] [Medline: 38840136]
11. Fijačko N, Masterson Creber R, Metličar Š, et al. Effects of a serious smartphone game on nursing students' theoretical knowledge and practical skills in adult basic life support: randomized wait list-controlled trial. JMIR Serious Games 2024 Apr 5;12:e56037. [doi: 10.2196/56037] [Medline: 38578690]
12. Warsinsky S, Schmidt-Kraepelin M, Rank S, Thiebes S, Sunyaev A. Conceptual ambiguity surrounding gamification and serious games in health care: literature review and development of Game-Based Intervention Reporting Guidelines (GAMING). J Med Internet Res 2021 Sep 10;23(9):e30390. [doi: 10.2196/30390] [Medline: 34505840]
13. Jaffe L. Games are multidimensional in educational situations. In: Bradshaw M, Lowenstein A, editors. Innovative Teaching Strategies in Nursing and Related Health Professions: Jones and Bartlett Publishers; 2011:175.
14. Lopes MDO, Silva VD, Araujo TD. Methods for establishing the accuracy of clinical indicators in predicting nursing diagnoses. Int J Nurs Knowl 2012 Oct;23(3):134-139. [doi: 10.1111/j.2047-3095.2012.01213.x] [Medline: 23043652]
15. Lourenço DF, Valentim EC, de Moraes Lopes MH. Translation and cross-cultural adaptation of the system usability scale to Brazilian Portuguese. Aquichan 2022 May 13;22(2):1-16. [doi: 10.5294/aqui.2022.22.2.8]
16. Padrini-Andrade L, Balda RDC, Areco KCN, et al. Evaluation of usability of a neonatal health information system according to the user's perception. Rev Paul Pediatr 2019;37(1):90-96. [doi: 10.1590/1984-0462/;2019;37;1;00019] [Medline: 30569950]
17. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: adding an adjective rating scale. J Usability Stud 2009;4:117-124.
18. Savi R, Von WC, Ulbricht V, Vanzin T. Proposta de um Modelo de Avaliação de Jogos Educacionais. Revista Nova Tecnol Educ 2010;8(3):e352.
19. Gliem JA, Calculating GRR. Interpreting, and reporting Cronbach's alpha reliability coefficient for Likert-type scales. In: Proceedings of the 2003 Midwest Research to Practice Conference in Adult, Continuing, and Community Education: University of Wisconsin:82-88.
20. Issue Information-Declaration of Helsinki. J Bone Miner Res 2020 Dec 1;36(12):BMi-BMii. [doi: 10.1002/jbmr.4088]
21. Deguirmendjian SC, Miranda F, Zem-Mascarenhas SH. Serious game desenvolvidos na Saúde: Revisão Integrativa da Literatura. J Health Inform 2016;8(3):110-126 [FREE Full text]
22. Visch V, Kooij K, Hoogendoorn E, Spijkerman R. Validation of games for behavioral change: connecting the playful and serious. Int J Serious Games 2015;2(3):63-75. [doi: 10.17083/ijsg.v2i3.75]
23. Silva ADA, Mascarenhas VHA, Araújo SNM, Machado RDS, Santos A, Andrade E. Mobile technologies in the nursing area. Rev Bras Enferm 2018;71(5):2570-2578. [doi: 10.1590/0034-7167-2017-0513] [Medline: 30304191]
24. Fernandes KT, Lucena M. Uma Experiência na Criação de Game Design de Jogos Digitais Educativos a partir do design thinking. RENOTE 2018;16(1):e396. [doi: 10.22456/1679-1916.85928]
25. Jewell D. Design thinking in higher education. ResearchGate. URL: https://www.researchgate.net/publication/376206883_Design_Thinking_in_Higher_Education [accessed 2025-10-14]
26. Chandler HM. Manual de Produção de Jogos Digitais: Bookman; 2012.
27. Pereira ALM, Leon C, Ribeiro LM, et al. Web-based virtual learning environment for medicine administration in pediatrics and neonatology: content evaluation. JMIR Serious Games 2020 Oct 21;8(4):e18258. [doi: 10.2196/18258] [Medline: 33084579]
28. Viana LS, Oliveira EN, Vasconcelos MIO, Fernandes CAR, Dutra MCX, de AP. Development and validation of an educational game about drug abuse and the risk of suicide. SMAD 2023;19(2):16-25. [doi: 10.11606/issn.1806-6976.smad.2023.188483]
29. de ML, Santana FM, Ramos JVM, et al. Parameterization and usability of the nursing process system. Res Soc Dev 2022;11(3):89. [doi: 10.33448/rsd-v11i3.26088]
30. Brooke J. SUS: a "quick and dirty" usability scale. In: Jordan PW, Thomas B, McClelland IL, Weerdmeester B, editors. Usability Evaluation In Industry: CRC Press; 1996:207-212. [doi: 10.1201/9781498710411-35]
31. da Silva Lima Roque G, Roque de Souza R, Araújo do Nascimento JW, de Campos Filho AS, de Melo Queiroz SR, Ramos Vieira Santos IC. Content validation and usability of a chatbot of guidelines for wound dressing. Int J Med Inform 2021 Jul;151:104473. [doi: 10.1016/j.ijmedinf.2021.104473] [Medline: 33964703]
32. Issom DZ, Hardy-Dessources MD, Romana M, Hartvigsen G, Lovis C. Toward a conversational agent to support the self-management of adults and young adults with sickle cell disease: usability and usefulness study. Front Digit Health 2021;3:600333. [doi: 10.3389/fdgth.2021.600333] [Medline: 34713087]

XSL•FO
RenderX

33. Griffin AC, Khairat S, Bailey SC, Chung AE. A chatbot for hypertension self-management support: user-centered design, development, and usability testing. JAMIA Open 2023 Oct;6(3):ooad073. [doi: 10.1093/jamiaopen/ooad073] [Medline: 37693367]
34. Fang H, Fang C, Che Y, Peng X, Zhang X, Lin D. Reward feedback mechanism in virtual reality serious games in interventions for children with attention deficits: pre- and posttest experimental control group study. JMIR Serious Games 2025 Feb 24;13(1):e67338. [doi: 10.2196/67338] [Medline: 39993290]
35. Palee P, Wongta N, Khwanngern K, Jitmun W, Choosri N. Serious game for teaching undergraduate medical students in cleft lip and palate treatment protocol. Int J Med Inform 2020 Sep;141:104166. [doi: 10.1016/j.ijmedinf.2020.104166] [Medline: 32570197]
36. Amorim AAL, Alvarenga K, Jacob LCB, Araújo ES. Usability evaluation of the Agente Escuta application: translational research. Codas 2023;35(4):e20220149. [doi: 10.1590/2317-1782/20232022149pt] [Medline: 37729328]
37. Hyzy M, Bond R, Mulvenna M, et al. System usability scale benchmarking for digital health apps: meta-analysis. JMIR Mhealth Uhealth 2022 Aug 18;10(8):e37290. [doi: 10.2196/37290] [Medline: 35980732]
38. Flato UAP, Beffa Dos Santos EJ, Bispo Diaz T Martins I, et al. Interactive serious game to teach basic life support among schoolchildren in Brazil: design and rationale. JMIR Serious Games 2024 Oct 9;12(1):e55333. [doi: 10.2196/55333] [Medline: 39382546]
39. Bressler M, Merk J, Gohlke T, et al. A virtual reality serious game for the rehabilitation of hand and finger function: iterative development and suitability study. JMIR Serious Games 2024 Aug 27;12(1):e54193. [doi: 10.2196/54193] [Medline: 39190432]
40. Rickenbacher-Frey S, Adam S, Exadaktylos AK, Müller M, Sauter TC, Birrenbach T. Development and evaluation of a virtual reality training for emergency treatment of shortness of breath based on frameworks for serious games. GMS J Med Educ 2023;40(2):Doc16. [doi: 10.3205/zma001598] [Medline: 37361243]
41. Cavalcanti H, Bushatsky M, Barros M, Melo C, Delgado Filho AJF. Evaluation of the usability of a mobile application in early detection of pediatric cancer. Rev Gaúcha Enferm 2021;42(2):172. [doi: 10.1590/1983-1447.2021.20190384]
42. Costa RDO, Medeiros SD, Coutinho VRD, Mazzo A, Araújo MD. Satisfaction and self-confidence in the learning of nursing students: randomized clinical trial. Esc Anna Nery 2020;24(1). [doi: 10.1590/2177-9465-ean-2019-0094]
43. Moraes VD, Ferraz L. Educational technology on expressing breast milk: development and validation of a Serious Game. Rev Bras Saude Mater Infant 2021;21(3):845-855. [doi: 10.1590/1806-93042021000300007]
44. Galvão ECF, Püschel VAA. Multimedia application in mobile platform for teaching the measurement of central venous pressure. Rev Esc Enferm USP 2012 Oct;46 Spec No:107-115. [doi: 10.1590/s0080-62342012000700016] [Medline: 23250266]
45. Oliveira RMD, Duarte AF, Alves D, Furegato ARF. Development of the TabacoQuest app for computerization of data collection on smoking in psychiatric nursing. Rev Lat Am Enfermagem 2016 Aug 29;24:e2726. [doi: 10.1590/1518-8345.0661.2726] [Medline: 27579930]
46. Araujo JP, Gallo AM, Lima Parada CMG, et al. Mobile applications as a strategy to support parents in the care of newborns: a scoping review. Rev Esc Enferm 2023:e20220470. [doi: 10.1590/1980-220X-REEUSP-2022-0470EN]
47. Pineros N, Tenaillon K, Marin J, et al. Using gamification to improve engagement and learning outcomes in medical microbiology: the case study of “BacteriaGame”. FEMS Microbiol Lett 2023 Jan 17;370:1-6. [doi: 10.1093/femsle/fnad034] [Medline: 37086092]
48. Arruzza E, Chau M. A scoping review of randomised controlled trials to assess the value of gamification in the higher education of health science students. J Med Imaging Radiat Sci 2021 Mar;52(1):137-146. [doi: 10.1016/j.jmir.2020.10.003] [Medline: 33153931]

Abbreviations

GAMING: Game-Based Intervention Reporting Guidelines
HUF: Health Unit in Focus
PHC: primary health care
SUS: System Usability Scale

XSL•FO
RenderX

Edited by A Coristine; submitted 09.09.24; peer-reviewed by A Pina-Oliveira, CI Sartorão Filho, THR da Silva, V Çakar; revised version received 22.04.25; accepted 24.04.25; published 04.11.25.

Please cite as:
Barbalho YGDS, Silva CMDS, Fernandes CSNDN, Trombini RRDSL, Melo PL, Oliveira AFD, Pereira ALM, Silva ADO, Lima LRD, Stival MM, Pinho DLM, Funghetto SS
A Serious Game (Health Unit in Focus) for Enhancing Undergraduate Education on Older Adults' Health: Design and Validation Study
JMIR Serious Games 2025;13:e66289
URL: https://games.jmir.org/2025/1/e66289
doi:10.2196/66289

© Yuri Gustavo de Sousa Barbalho, Calliandra Maria de Souza Silva, Carla Sílvia Neves da Nora Fernandes, Raíza Rana de Souza Lima Trombini, Patrícia Littig Melo, Aline Farias de Oliveira, Alayne Larissa Martins Pereira, Alessandro de Oliveira Silva, Luciano Ramos de Lima, Marina Morato Stival, Diana Lúcia Moura Pinho, Silvana Schwerz Funghetto. Originally published in JMIR Serious Games (https://games.jmir.org), 4.11.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Effectiveness of a Web-Based Virtual Simulation to Train Nursing Students in Suicide Risk Assessment: Randomized Controlled Investigation

Paul Roux[1,2], PhD, MD; Yujiro Okuya[3], PhD; Cristina Morel[4]; Mariane Soulès[5]; Hugo Bottemanne[6,7], MD; Eric Brunet-Gouet[1,2], PhD, MD; Solène Frileux[1,2], MD; Christine Passerieux[1,2], PhD, MD; Nadia Younes[2,8], PhD, MD; Jean Claude Martin[3], PhD

[1]DisAP - MOODS Team, INSERM UMR1018, CESP, Université de Versailles Saint-Quentin-En-Yvelines - Université Paris-Saclay, Le Chesnay, France

[2]Service Universitaire de Psychiatrie d'Adultes et d'Addictologie, Centre Hospitalier de Versailles, 177 rue de Versailles, Le Chesnay, France

[3]Laboratoire Interdisciplinaire des Sciences du Numérique, CNRS, Université Paris-Saclay, Orsay, France

[4]Institut de Formation en Soins Infirmiers de Versailles, Centre Hospitalier de Versailles, Versailles, France

[5]Institut de Formation en Soins Infirmiers de Rambouillet, Centre Hospitalier de Rambouillet, France

[6]Service Hospitalo-Universitaire de Psychiatrie, Hôpital Bicêtre, Assistance Publique - Hôpitaux de Paris, Le Kremlin-Bicêtre, France

[7]MOODS Team, INSERM UMR1018, CESP, Faculté de Médecine Paris-Saclay, Université de Versailles Saint-Quentin-En-Yvelines, Université Paris-Saclay, Villejuif, France

[8]Equipe DevPsy, INSERM UMR1018, Centre de recherche en Épidémiologie et Santé des Populations, Université Paris-Saclay & Université Versailles Saint-Quentin-En-Yvelines, Villejuif, France

Corresponding Author:
Paul Roux, PhD, MD
DisAP - MOODS Team, INSERM UMR1018, CESP, Université de Versailles Saint-Quentin-En-Yvelines - Université Paris-Saclay, Le Chesnay, France

Abstract

Background: Suicide is a leading cause of preventable death worldwide. Nurses play a critical role in suicide prevention; yet, they face significant obstacles. Improving the evaluation and management of patients at risk of suicide requires innovative training techniques that safely and effectively enhance nursing students' skills, knowledge, and confidence. Virtual simulation (VS) based training can be particularly effective because it allows interaction with patients without the risk of causing harm.

Objective: The purpose of this study was to evaluate the pedagogical effectiveness of a novel VS tool featuring a fully automated and emotionally reactive virtual patient by assessing its ability to assist nursing students in learning suicide risk assessment. VS also included an online group debriefing, co-run by a nurse and a medical teacher.

Methods: A randomized controlled investigation was conducted with 68 first-year nursing students recruited from nursing schools offline and online. They were divided into a control group receiving teaching as usual (TAU) and an intervention group receiving TAU plus VS. The intervention was purely web-based and unblinded. Outcomes were self-assessed through questionnaires using Kirkpatrick Training Evaluation Model, which focuses on knowledge, skills, confidence, empathy, and satisfaction among students.

Results: The VS group exhibited significantly higher confidence (3 points of increase after TAU vs 10.6 points of increase after VS, B=7.2; SE 2.5; $t_{111.5}$=2.8; P=.006) and a marginally enhanced ability to respond appropriately to suicidal thoughts (1.6 points of improvement after TAU vs 6.4 points of improvement after VS, B=−4.5; SE 2.5; $t_{119.5}$=−1.8; P=.08) compared with the control group. However, there were no significant differences in knowledge acquisition or the general level of empathy. Satisfaction with VS was high, particularly regarding the authenticity of the virtual patient. Authenticity was perceived as greater when emotional prosody was included with facial emotions.

Conclusions: The use of VS demonstrated promising results in enhancing nursing students' confidence in detecting suicide risk and their skills in counseling individuals experiencing a suicide crisis, suggesting its incorporation into routine teaching methods. Further research is needed to explore its long-term benefits for students and its impact on patient outcomes.

Trial Registration: OSF Registries 10.17605/OSF.IO/7SFA5; https://osf.io/7sfa5

(JMIR Serious Games 2025;13:e69347) doi:10.2196/69347

XSL•FO
RenderX

KEYWORDS

nurse; simulation; virtual patient; suicide; depression; education; confidence; emotion; prosody; facial expression

Introduction

Suicide is among the leading causes of death worldwide, with 703,000 deaths by suicide in 2019, which was more than deaths due to malaria, HIV, breast cancer, war, and homicide [1]. One important axis of suicide prevention is the education of health care professionals in recognizing depression and managing suicidal crises. However, this aspect requires further investigation to assess its possible benefits due to the small number of randomized controlled investigations [2].

Nurses are ideally positioned to intervene and implement preventative measures toward suicide. However, several obstacles in suicide risk evaluation and care by nurses have been identified. First, negative attitudes and a stigma toward patients experiencing suicide crises are high among nurses, although less frequent than among doctors [3]. Second, nurses do not usually use guidelines or suicide assessment instruments when they evaluate suicidality [4,5], which may result from deficits in skill and knowledge [6]. Gaps in the training of nurses have been reported by psychiatric nurses associations in the management of individuals with suicidal ideation [7], and nursing students reported a lack of education and knowledge in suicide prevention [8]. Some qualified nurses also experience not having sufficient knowledge or skills to feel safe in suicide detection [9] and, sometimes, avoid asking questions about suicidality for fear of what to do with the answer [10]. One of the major obstacles in suicide prevention may, thus, result from a lack of confidence. Health sciences students are afraid of talking about suicide with patients [11] in particular, asking direct questions about suicidal thoughts and intentions, as they believe such questions could embarrass or distress patients. To decrease such clinical anxiety, health students are usually taught that, according to the literature, talking about suicide within a screening interview does not increase distress or suicide ideation by patients [12,13] and may even improve their sense of well-being [14]. However, students often doubt that a patient with suicidal intent can feel relief when speaking about suicide until they experience it during a face-to-face interaction. It, therefore, seems necessary for students to experience the relief felt by patients with suicidal ideation when they are interviewed about their suicidal thoughts: this would make it easier for them to put into practice their competence in suicide prevention with real patients. It would also be safer if these first experiences did not occur with real patients.

Usual teaching methods have shown pedagogical effectiveness for suicide prevention in nursing studies. Short lectures and discussions of health care staff experiences and videotaped testimonials by experts on the lived experience of suicidal crises have been shown to be associated with an increase in self-perceived competence [15]. A 3.5-hour suicide prevention training course with video clips, practicing skills, and group discussion was associated with a significant increase in general perceived self-efficacy compared to preintervention [16]. However, some data suggest that simulation may be more efficient than usual teaching methods to improve suicide risk detection and care [17]. The most prominent simulation method in psychiatry involves standardized patients [18], played by professional actors, teachers, or carers. However, it is often difficult and costly to find, train, pay, and debrief standardized patients, thus limiting the spread of this teaching method. By contrast, simulation with fully autonomous virtual patients allows perfect standardization of the environment, a potentially infinite number of learners for one simulation, and remote learning. Randomized controlled studies of the pedagogical efficiency of virtual simulation (VS) targeting suicide risk are scarce. Only one study involving second-year medical students reported that students who benefited from VS inquired more frequently than students who benefited from a video-based intervention in several suicide risk areas during an interaction with a simulated patient [19]. However, no differences were found in communication skills, such as the learner's professional appearance, behavior, and empathy. The virtual patient consisted of an embodied conversational agent portrayed by a static image of her face, without any emotional reaction to the questions asked by the learner. This may have lacked realism, as interpreting subtle nonverbal cues a patient displays is a critical skill for health students. During a psychiatric interview, facial expressions can provide an understanding of the patient's affective processes [20]. Thus, autonomous and emotionally reactive virtual patients are needed to simulate nursing interviews in suicidology teaching properly.

The present study aimed to evaluate the pedagogical effectiveness of a complementary simulation teaching using an autonomous and emotionally reactive virtual patient (Simulation with a Virtual Patient in Psychiatry [SIVIPSY]) compared with traditional teaching alone. We made the hypothesis that this teaching tool would help nursing students to learn how to detect and evaluate a suicidal crisis and behave appropriately. The study also aimed to establish the degree of perceived realism of VS and the impact of the virtual patient's emotional reactions on the realism of SIVIPSY and learner satisfaction.

Methods

Design of the Virtual Patient

SIVIPSY consisted of a VS of a nurse interview with a patient experiencing severe depression with an imminent suicide risk [21]. It was designed by nursing teachers, academic psychiatrists, and computer science researchers specializing in human-computer interactions. In SIVIPSY, learners must dialogue with an emotionally reactive avatar according to an underlying decision tree architecture: they have to select the best question to ask, which is then generated by vocal synthesis. The virtual patient reacts according to the learner's choices by displaying adaptive verbal and facial expressions of emotions.

Relevant Emotions and Their Expression by the Virtual Patient

We implemented the virtual patient using the MARC (Multimodal Affective and Reactive Characters) platform for

XSL•FO
RenderX

animating expressive virtual agents [22]. MARC can combine nonverbal (facial and body animations) and verbal behaviors of 3D characters. The platform features a model that evaluates the current situation according to several criteria and selects facial animation accordingly [23]. This platform offers several characters that have been used in various social-skills training experiments [24,25]. We used a male and female model that suited the patient's and his wife's personae in the scenario.

We selected 8 emotions relevant to a suicidal crisis: sadness, anger, shame, doubt, frustration, discouragement, nervousness, and relief. We designed corresponding facial expressions using the Facial Action Coding System (Carl-Herman Hjortsjö) [26]. For sadness, anger, and shame, we used one of the multiple possibilities of action unit combinations for expressing these basic emotions [27]. For doubt, frustration, discouragement, nervousness, and relief [28], we selected the action units most frequently displayed in the 6 video clips expressing these 5 emotions.

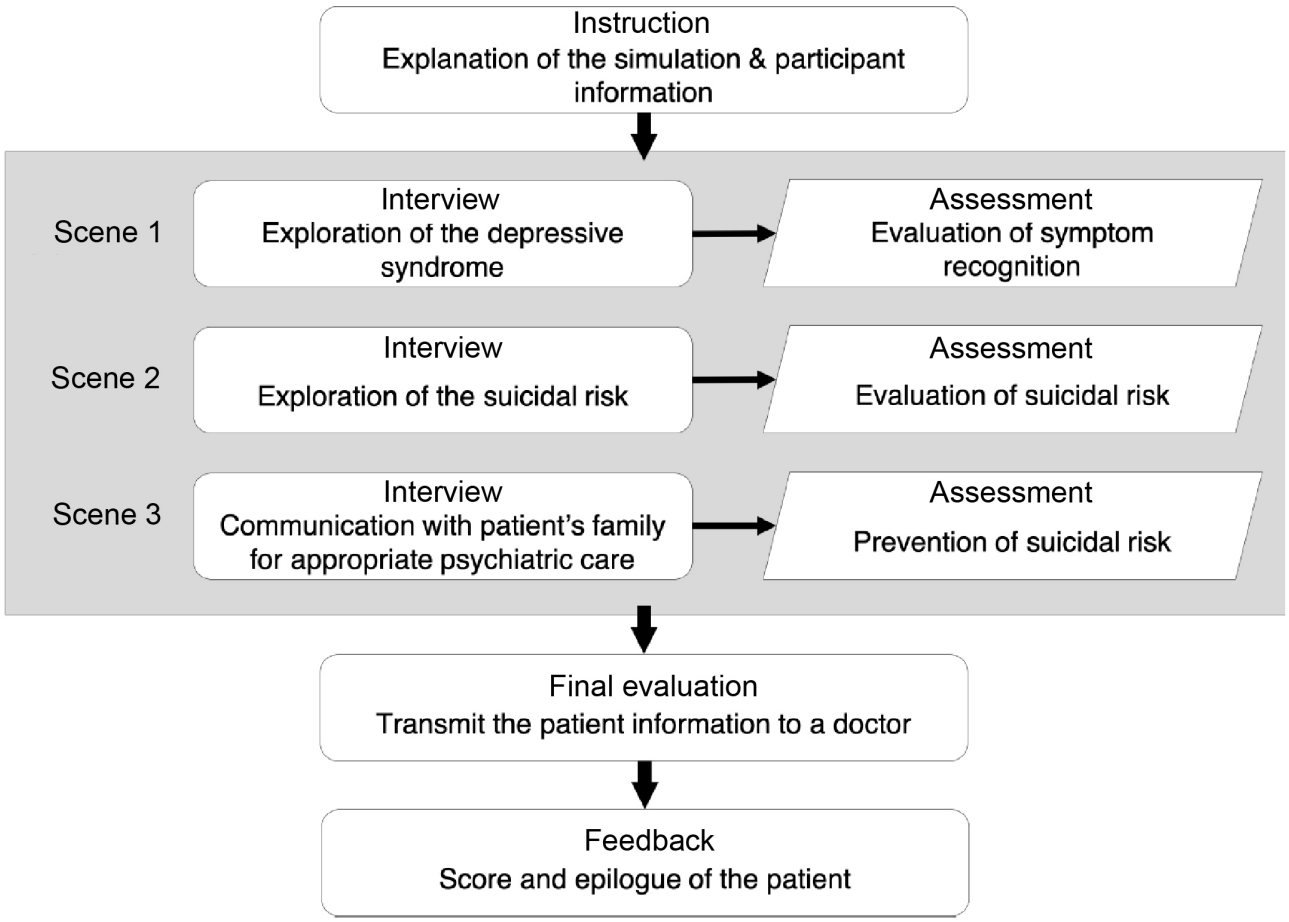
The Acapela speech synthesizer (Acapela Group) [29] was used to generate the audio files corresponding to the spoken utterances of the virtual patient and his wife.

VS Scenario

We used a dialog-based scenario with a decision tree architecture in which the learner has to select the most appropriate question to ask the virtual patient at risk of suicide from a list of choices. Correct choices and distractors were selected based on usual recommendations for a psychiatric interview [20,30]. Thus, users had to pay attention to the virtual patient's key verbal or nonverbal behaviors throughout the scenario to choose the correct question.

At the beginning of the simulation, the participants were given the instructions for the virtual interview and a nursing observation sheet describing the patient's health status. The scenario comprised 3 scenes (see Figure 1), each with 2 phases: the interview with the virtual patient (see Figure 2) and the assessment of observed signs or symptoms (see Figure 3). The nursing observation sheet was accessible on the screen throughout the interview (see Figure 2). Multiple choice questions (MCQ) are also given during the assessment phase. After the 3 scenes, learners had to select an appropriate nursing report describing the patient's condition based on their interviews. The feedback was given to learners at the end of the simulation based on all choices made during the simulation.

Figure 1. Structure of the virtual simulation scenario.

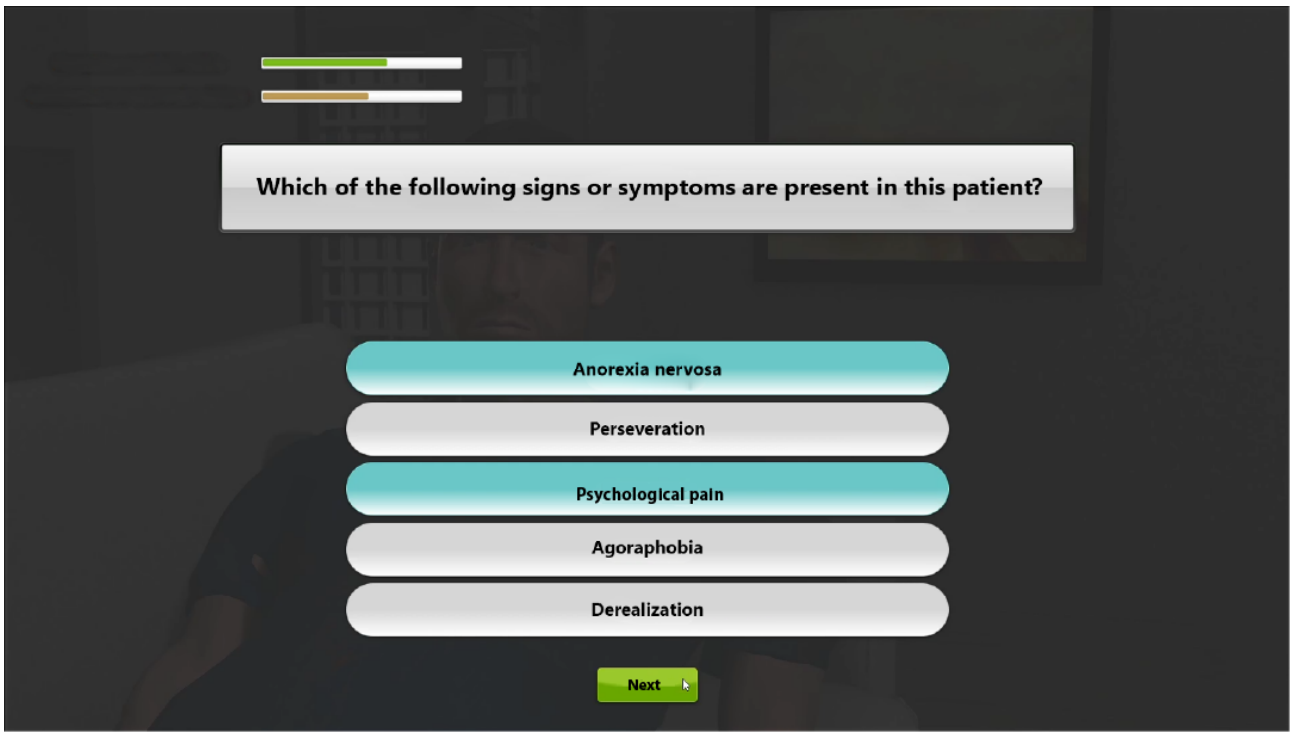


XSL•FO
RenderX

Figure 2. Screen capture of the interview phase.



Figure 3. Screen capture of the assessment phase.

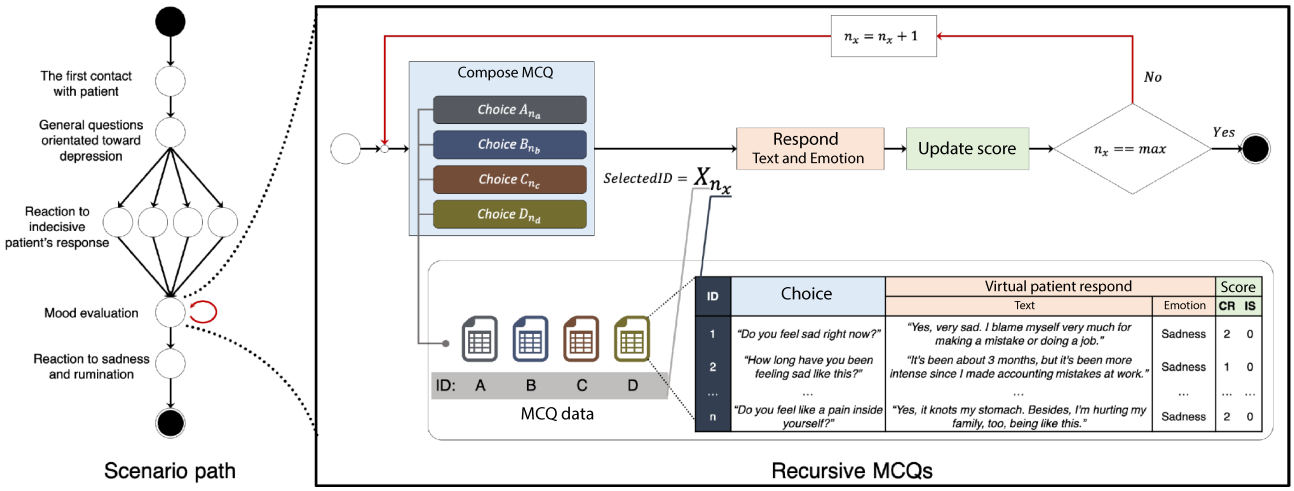


Design of the Interaction With the Virtual Patient

For the interview phase, we adopted a linear-interactive instructional design [31]. This design is widely used in VSs in health care and allows the learner to select available questions to diagnose and care for a patient within a formalized clinical practice model. To give learners a unique yet controlled narrative, the scenario was designed with the string of pearls model in which learners can proceed with a sequence of events based on their choice, but the main path leads to a single endpoint [32,33] (see Figure 4).

XSL•FO
RenderX

Figure 4. Scenario path of the first scene. Learners interview a virtual patient to diagnose the health problem: circles represent events of interviewing with multiple-choice questions format (left). Recursive multiple-choice questions occur when a set of questions needs to be asked to evaluate the patient's health issue (right). CR: clinical reasoning; IS: interpersonal skills; MCQ: multiple choice question.



Multiple Choice Questions

MCQs given to participants during the screening were composed of an appropriate answer and several types of distractors:

- Psychiatric distractors investigated the semiology of obsessive-compulsive disorder, autism, and anorexia nervosa. Although some symptoms of depression might resemble these disorders (rumination and perseveration for obsessive-compulsive disorder, social withdrawal for autism, and loss of appetite for anorexia nervosa), these distractors are inadequate for the simulation of a severe, characterized depressive episode.
- Vague suicidal risk distractors could be selected by participants who wrongly thought that the use of periphrases, allusions, and hints was preferable while exploring suicidal risk to avoid distressing the patient.
- Somatic distractors were questions concerning hypertension, dysphagia, and organic asthenia, which could be plausible for the patient in this scenario. Still, they were too widely related to his actual and urgent psychiatric condition.
- Nonspecific distractors were related to family, job, and education, which are often used in a psychiatric interview but not particularly valuable for this specific situation of a psychiatric emergency.

For some events, a set of MCQs composed of an appropriate answer and 3 distractors was recursively asked of learners (Figure 4, right). Each item in the MCQ had several different questions of the same category. When the learner selected the item, only the selected item was changed in the next composed MCQ, but the other 3 items remained. The order of items was randomized each time. Once one of the items reached the last question, the user could go to the next level. Consequently, if each item had 5 different questions, the user could end the recursive MCQs after 5 to 17 loops. Throughout the simulation, 9 MCQs and 4 recursive MCQs were given during the interview phase, and 11 MCQs were provided for the assessment phase.

Scoring System

Learners' performance was automatically measured according to 3 scores.

The clinical reasoning score evaluated how learners collected and interpreted signs and symptoms to determine the optimal nursing diagnosis and care plan. Clinical Reasoning was assessed during the interview and assessment phases. For example, in the interview phase, asking specific questions about suicide intention ("Have you had any suicidal thoughts in recent days?") gave +2 points; asking vague questions about suicide intention ("Have you had any dark thoughts in recent days?") gave +1 point; and asking a question about an irrelevant psychiatric condition ("Are you afraid of gaining weight?") gave 0 points. The quality of the care plan and nursing clinical handover proposed by the learner had an impact on the patient's outcome. Efficient prevention of the imminent suicidal crisis (eg "Remove shoelaces and belt, and offer uncorded pajamas") or a good handover avoids a suicidal attempt within the psychiatric ward. By contrast, if the participant made a poor care plan (eg "Let the patient rest quietly" or "Lock him in his room") or a poor handover, the patient attempted to commit suicide in his room, without dying, as it is usually not recommended that a patient dies within a simulation session due to the detrimental impact on the learner's stress level and negative emotions [34].

The interpersonal skills score assessed how learners constructed a trustful and empathetic relationship with the patient. Interpersonal skills were evaluated exclusively during the interview phase. Distractors in the MCQ were selected from among untrained students' usual mistakes in interacting with a severely depressed individual, which were collected from medical students in the third year of study during role-playing situation simulations run at the University Versailles Saint-Quentin-En-Yvelines over several years. The most common mistakes were attempts to overstimulate a slowed down and anergic patient, to cheer up a patient with pathological sadness, and nonhelpful beliefs about suicide prevention. For example, "I hear that you have suicidal thoughts and that you are desperate. We're here to help you through this difficult time." gave +1 points whereas "Haven't you thought of all the harm you'll be doing to your children who'll be orphans?" gave −2 points.

XSL•FO
RenderX

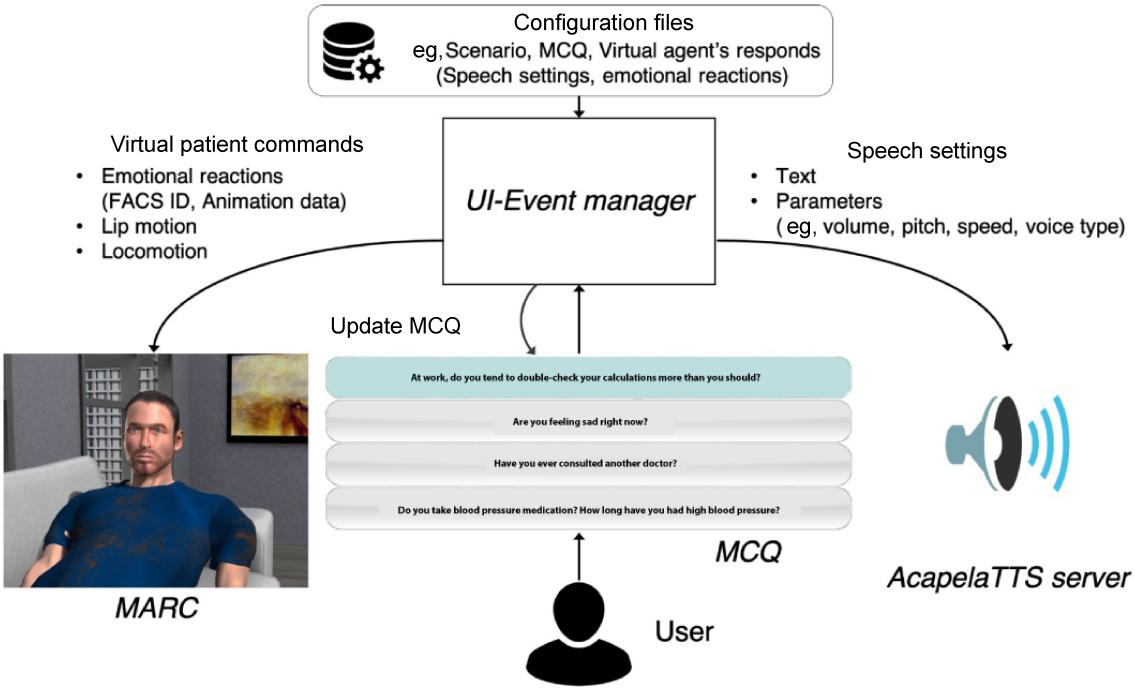
The clinical efficiency score counted the number of questions the learner asked before the interview ended. Clinical efficiency was evaluated during the interview phase. In particular, clinical efficiency increased with the number of distractors users chose. For suicidal risk assessment, it is important to collect as much relevant information from the patient as possible to make a clinical decision in as little time as possible [35].

Clinical reasoning and interpersonal skills were scored at each MCQ and displayed with progressive score bars on the screen (see Figure 2), whereas clinical efficiency was given to the users only at the end of a sequence of recursive MCQs. During the feedback at the end of the simulation (see Figure 1), users received each score ranging from 0 (worst) to 20 (best) scaled from row scores, following the standard metric used in French education.

Software Architecture and Procedure

The SIVIPSY system comprises a UI-Event manager, MARC, and an AcapelaTTS server (Acapela Group; see Figure 5). The UI-Event manager is implemented in Java, updating the graphical user interface and handling each event during the simulation based on user input and a predefined scenario. The content of the scenario is defined in several external configuration files (.xml) in which the experimenter can specify the logical flow of the scenario and the MCQs that include: questions given to learners, the virtual patient's responses, the type of emotional expression and its intensity (Action Unit combinations of each emotion are defined in the other configuration file), settings of spoken utterances (eg speed and pitch), and score settings (clinical reasoning and interpersonal skills). The UI-Event manager communicates with MARC and the AcapelaTTS server via UDP. MARC updates the virtual patient's facial expressions and body movements based on the given requests. The AcapelaTTS server, implemented in C#, synthesizes the voice in real time with specified settings.

Figure 5. System architecture of the virtual patient simulation. MARC: Multimodal Affective and Reactive Character; MCQ: multiple choice questionnaire.



The learners played the simulation on a Windows PC installed at the Versailles Hospital through a remote desktop environment due to the restrictions on movement due to the COVID-19 pandemic. Remote access to the PC was handled by AnyDesk software via an internet connection, offering users remote control of the PC. The rendered resolution was 1.920×1.080 pixels, but remote users might have perceived less resolution as we prioritized the system reaction time over graphics quality in AnyDesk.

Recruitment

Nursing students were recruited from 7 nursing schools affiliated with Versailles Saint-Quentin-En-Yvelines University. During back-to-school meetings involving all first-year students, participants were informed about the study by their nurse training manager and one author of this study (PR). Students also received an email to update them on the study. The inclusion criteria were being in the second semester of the first year of study and being over 18 years of age.

Study Design

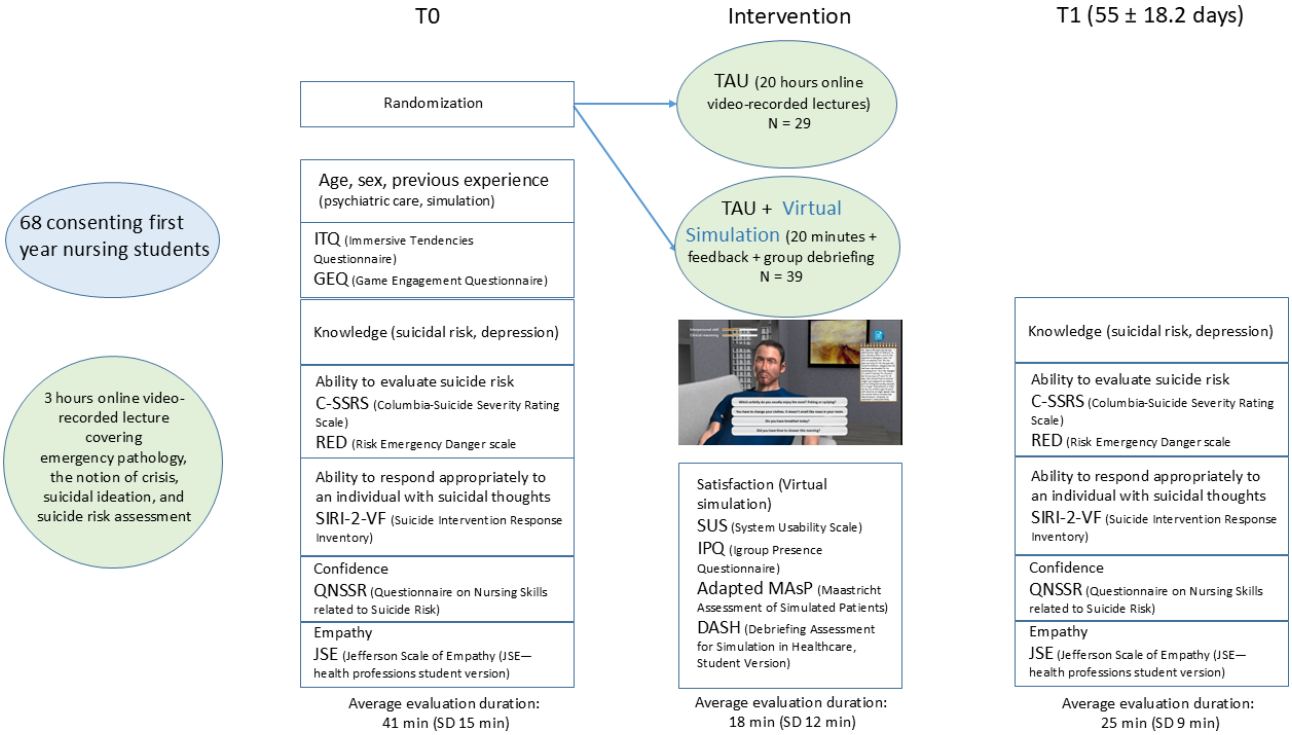
The design of the study is presented in Figure 6, and the CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) checklist is provided in Checklist 1. No important changes were made to the methods after the investigation

XSL•FO
RenderX

commenced. Participants were randomized between 2 arms with a 4:3 allocation ratio favoring VS using the "randomize" function in R (experiment package version: 1.2.1, R version 4.1.3, no random seed, simple randomization without stratification). All participants received the standard teaching program, consisting of 20 hours of online video-recorded lectures on psychiatry and mental health. Before starting the research and being randomized, participants were asked to attend a 3-hour online course covering emergency pathology, the notion of crisis, suicidal ideation, and suicide risk assessment. Randomized participants in the control arm received only the usual teaching program (teaching as usual [TAU] arm). Randomized participants in the intervention arm received TAU plus VS with SIVIPSY. Two versions of SIVIPSY were developed. In the first version, the virtual patient spoke with a neutral prosody, whereas in the second version, he spoke with a sad prosody. The 2 versions were compared with tests to the hypothesis that emotional prosody improved the perceived authenticity of the virtual patient with severe depression. All participants in 2021 were allocated to the neutral prosody version of SIVIPSY, whereas all participants in 2022 were allocated to the sad prosody version of SIVIPSY. The VS lasted approximately 20 minutes. Two types of feedback were provided to learners. The first type was automatic: the program gave 3 scores to the learners (clinical reasoning, interpersonal skills, and clinical efficiency) and revealed what happened to the patient after the interaction with the learner according to the choices that were made. The second type of feedback consisted of a remote group debriefing, co-run by a nurse and a medical teacher, when possible. It followed the rules of debriefing with good judgment [36] and was preceded by an anonymous review of the learner log file to detect performance deficits that may have varied depending on the group. The debriefing groups consisted of 4.1 individuals on average (SD 2.7), resulting in 8 different groups. Learners were grouped based on their availability.

PR generated the random allocation sequence, enrolled participants, and assigned them to interventions. Teachers and learners were not blinded. Participants were prompted by mail to engage in the evaluations, the VS, and the debriefing. The questionnaires for the 3 evaluation times were completed online by participants during several remote sessions led by PR, who addressed participant queries and ensured that all items were submitted. All questionnaires corresponding to one evaluation moment were completed at the same time for each participant. The assessment durations are reported in Figure 6.

Figure 6. Schematic representation of the study design. TAU: teaching as usual.



Sample Baseline Characteristics

Baseline characteristics were recorded at inclusion. Demographic data consisted of age, sex, and native language. We also evaluated self-reported previous experience in psychiatric care (none/[internship/employment]), previous simulation experience (no/yes), and participation in the prerequisite online course (no/yes). Finally, we evaluated the individual's disposition to be particularly receptive to the VS modality through 2 measurements. The first consisted of their immersive tendency, measured using the Immersive Tendencies Questionnaire (ITQ) [37]. Immersion gives individuals the impression that they have left the real world and are now present in the virtual environment. A greater sense of immersion results in higher levels of involvement in a virtual environment, which may increase the pedagogical effectiveness of the VS. The second baseline disposition was the psychological engagement in video game-playing, measured using the Game Engagement Questionnaire (GEQ) [38], which may increase the positive impact of VS on pedagogical outcomes.

XSL•FO
RenderX

Outcomes

SIVIPSY was assessed using Kirkpatrick Training Evaluation Model [39]. This model establishes the impact of a learning intervention according to five levels: (1) reaction effect: satisfaction/dissatisfaction of participants; (2) learning effect: improvement of the participants' knowledge; (3) behavioral effect: changes in attitudes, skills, or learners' confidence; (4) patient results (ie, whether the intervention improves the nurse diagnosis care plan) to approach operational effectiveness and reach patient-reported outcomes; and (5) return on investment. Only the first 3 Kirkpatrick levels were investigated in this study, using validated tools when available.

Knowledge

Knowledge was measured using a 10-item, 5-choice questionnaire on the subject of suicidal risk and depression created for the study.

Skills, Confidence, and Attitude

Two skills were assessed in this study, the ability to evaluate suicide risk and the ability to respond appropriately to an individual with suicidal thoughts.

Ability to Evaluate Suicide Risk

A specific tool was created to measure the ability to assess suicide risk in the current study. We filmed a psychiatric nurse interview of a patient with intermediate suicide risk [40]. Two academic psychiatry teachers portrayed the nurse and patient roles. The first scale used was the Columbia-Suicide Severity Rating Scale (C-SSRS), which has demonstrated good convergent and divergent validity in the international scientific literature [41].

The second scale was the Risk - Emergency - Danger (RED) scale, which is widely used in France in clinical and teaching settings [42] and recommended by the French health authorities [43]. The RED scale follows international recommendations concerning suicide risk assessment [44]. The reference for the evaluation of the suicidal crisis presented by the patient in the video was established by 11 experts, who used the 2 scales to rate the severity of the suicidal crisis (see Multimedia Appendix 1). We scored each item by computing an absolute *z* score using the recruited expert panelist's answers as the reference and summing them for each participant for each scale.

Ability to Respond Appropriately to an Individual With Suicidal Thoughts

The learner's interpersonal skills in managing suicidal crises were assessed using the Suicide Intervention Response Inventory [45], French version 2 (SIRI-2-VF) [46]. The SIRI-2-VF includes 15 statements corresponding to exchanges between patients and caregivers taken from consultation excerpts. Each exchange begins with patients' statements concerning an aspect of their situation. Then, 2 possible answers proposed by 2 different caregivers are presented. The participant is asked to judge the suitability or unsuitability of each proposed response by giving a score ranging from −3 (very inappropriate response) to +3 (very appropriate response) on a 7-point Likert scale. We scored each excerpt by computing the absolute *z* score using published expert panelist answers [46] and summed them for each participant.

Confidence

We measured another behavioral effect of VS: the improvement of learners' confidence. Confidence was measured using the Questionnaire on Nursing Skills related to Suicide Risk (QNSSR), which was adapted from a previous version initially developed for medical students [47]. It included 14 items rated on a visual analog scale ranging from the least possible confidence (coded 0) to the greatest confidence possible (coded 10, see the full scale in Multimedia Appendix 2).

Empathy

Empathy was measured using the Jefferson Scale of Empathy (JSE—health professions student version) [48].

Satisfaction

Satisfaction with the serious game system user-friendliness was assessed using the System Usability Scale (SUS), one of the most widely used questionnaires to measure the perceived ease of use of interactive systems, with an excellent internal consistency (Cronbach α=0.90) [49]. Satisfaction with the immersive experience was assessed using the Igroup Presence Questionnaire (IPQ), which focuses on the subjective sense of being in a virtual environment and exhibited good internal consistency (Cronbach α=0.85 and 0.87) [50].

Satisfaction with the authenticity of the virtual patient was measured using 6 items (see Multimedia Appendix 3 extracted from the Maastricht Assessment of Simulated Patients (MAsP), which evaluates the performance of actors simulating patients in an educational setting with an acceptable internal consistency (Cronbach α=0.73) [51]. These items were chosen to be compatible with the virtual nature of the patient, as no specific scale measuring virtual patient authenticity existed at the time of testing. We also measured satisfaction with the facial expression of emotions displayed by the virtual patient using 4 items (see Multimedia Appendix 3).

Satisfaction with the remote group debriefing was assessed using the Debriefing Assessment for Simulation in Healthcare (DASH) student version [52], which showed good internal consistency with nursing students (Cronbach α=0.82) [53]. This scale explores the climate, debriefing structure, ability to engage in exchange, and strengths and areas for improvement. Finally, a 10-item questionnaire was developed to measure satisfaction with overall simulation training. It was adapted from a previous satisfaction questionnaire about simulation with medical students [47]. This questionnaire explores various aspects of satisfaction, such as the preference for simulation over another pedagogical modality, the perceived realism of the situation, and the importance of being actively involved, etc. It was completed with a question about the difficulty of the simulation and with a global measure of satisfaction on an 11-point scale.

Statistical Analysis

Missing data were estimated using multivariate imputations by chained equations (50 imputations, mice package of R). The fraction of missing information (fmi) is reported in the results. Linear mixed-effects models were run with participants as the

XSL•FO
RenderX

random factors, skills, knowledge, confidence, and empathy as the successive dependent variables, and time (2 modalities: before/after the intervention), arm (2 modalities: VS/TAU), and time-arm interaction as the simultaneous dependent variables. All analyses were run as intention-to-treat to preserve randomization. All learners randomly assigned to one of the arms were analyzed together, regardless of whether they completed the VS. The required sample size was estimated based on a meta-analysis of the educational effectiveness of simulation tools with virtual patients [54], which reports Hedges effect sizes of 0.94 for knowledge improvement. The number of subjects was estimated to be 25 in each arm, with a significance level α=.05 and power β=90%. A linear mixed-effects model was also run to explore the effect of the SIVIPSY version (sad or neutral vocal emotion) on satisfaction with the authenticity of the virtual patient (MaSP adapted).

Ethical Considerations

The study was approved by the university's ethical review board (approval number: CER-Paris-Saclay-2020‐087). Each participant provided written informed consent before inclusion. To ensure that the commitment was entirely voluntary, students were informed that their decision to participate or not would not affect the outcome of their psychiatry exam and, in turn, their progression to the second year. All study data were anonymous, but logistical measures (email confirmation and videocalls) were used to prevent participants from having multiple identities. No compensation was provided to participants for their involvement in the current research, and access to SIVIPSY was free.

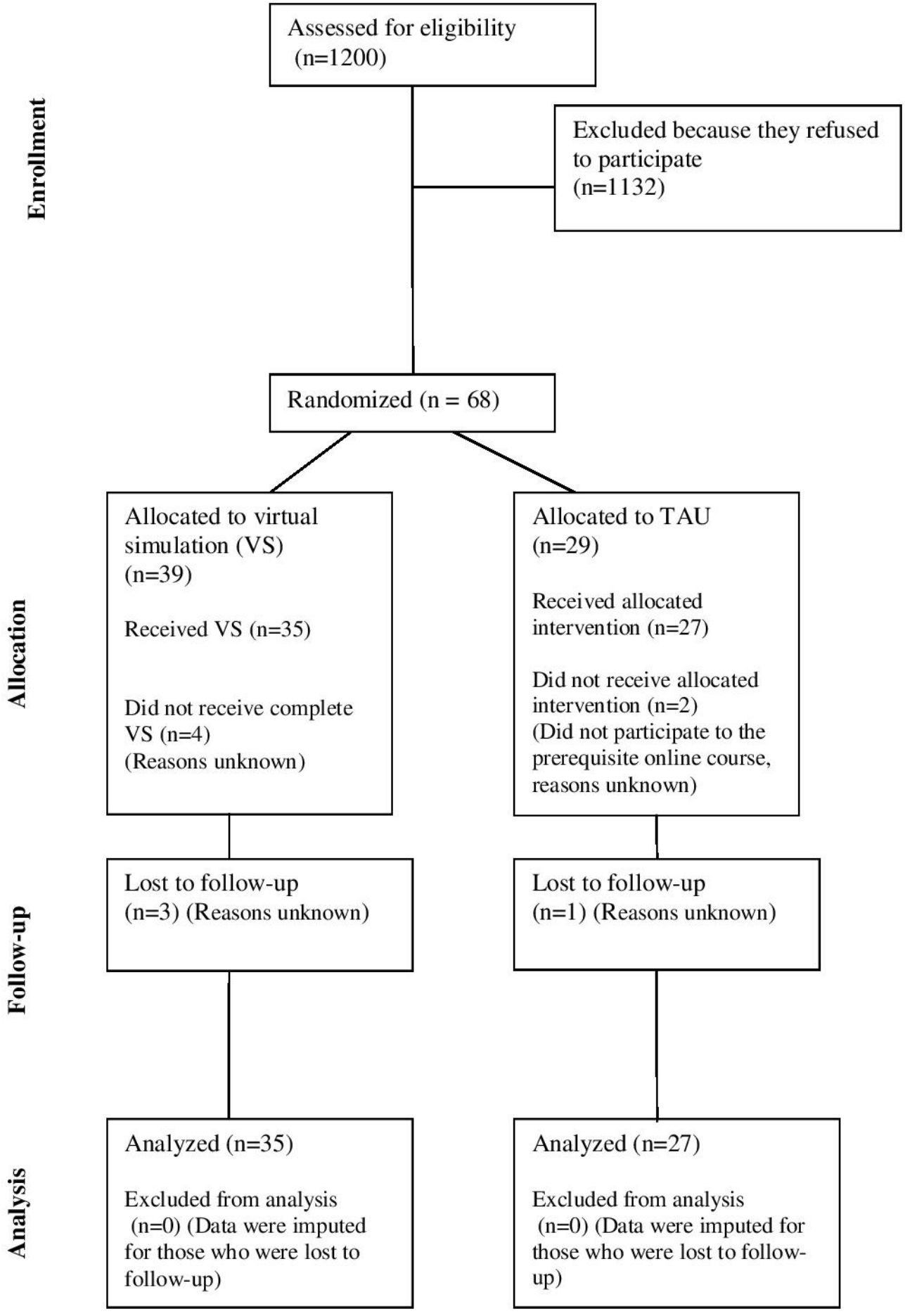
Results

Participants

The recruitment began in February 2021 and was completed in March 2022. Approximately 1200 students were approached, but only 68 consented to participate in the study (response rate of 6%, see the CONSORT (Consolidated Standards of Reporting Trials) flow diagram of study participants in Figure 7). We included 29 students in the TAU arm and 39 in the VS arm. Most students were female, approximately 25 years of age, and few had previous experience in the psychiatric care of patients (see Table 1).

XSL•FO
RenderX

Figure 7. Consolidated Standards of Reporting Trial showing the flow of participants through each stage of the study. TAU: teaching as usual.



XSL•FO
RenderX

Table . Participant characteristics.

Variable	Teaching as usual (N=29)	Virtual simulation (N=39)	Statistic	*P* value
Age (years), mean (SD)	24.6 (7.9)	26.4 (10.5)	$t_{66}=-0.85$	.40
Sex (female), n (%)	27 (93.1)	33 (84.6)	$\chi^2_1=0.48$	.49
Native language (non-French), n (%)	2 (6.9)	4 (10.3)	$\chi^2_1=0$	.96
Previous experience in psychiatric care (as an intern), n (%)	11 (37.9)	7 (17.9)	$\chi^2_2=3.57$	.17
Previous experience in psychiatric care (as an employee), n (%)	3 (10.3)	4 (10.3)		
Previous simulation experience, n (%)	13 (44.8)	16 (41)	$\chi^2_1=0$	.95
Participation in prerequisite online course, n (%)	27 (93)	33 (84.6)	$\chi^2_1=0.48$	.49
Immersion (ITQ)[a], mean (SD)	80.4 (13.3)	75.9 (17.2)	$t_{65.9}=1.2$	.24
Implication (GEQ)[b], mean (SD)	3.1 (1.4)	2.8 (1.5)	$t_{61.7}=0.78$	.44

[a]Immersive Tendencies Questionnaire.

[b]Game Engagement Questionnaire.

By contrast, participants were familiar with simulation, with more than half having previous experience with this pedagogical method (either high fidelity, standardized patient, or versus). Very few participants did not study the online course before the evaluations. The average ITQ score corresponded to a moderate immersive tendency and the average GEQ score showed a slight lack of game engagement.

There were no significant statistical differences between participants in the TAU and VS arms in terms of demographic characteristics, previous psychiatric and simulation experience, participation in the prerequisite course, immersive tendency, or engagement in games (see Table 1). Four participants in the VS arm did not fully complete the intervention as intended; they were retained in the VS arm according to the intention-to-treat analysis. Among the participants in the VS arm, 18 interviewed the virtual patient with neutral prosody and 21 the virtual patient with sad prosody. No significant statistical differences were measured between participants interacting with the virtual patient expressing neutral versus sad prosody, although participants using the sad prosody version were marginally older and marginally more game-engaged (see Multimedia Appendix 4).

Effect of the Simulation on Performance

The descriptive statistics are presented in Table 2 and the test statistics in Multimedia Appendix 4.

Outcomes were measured at an average of 55 (SD 18.2) days after inclusion. There were no missing data for TAU before the intervention, 3% of missing data for TAU after the intervention, no missing data for VS before the intervention, and 8% of missing data for VS after the intervention.

XSL•FO
RenderX

Table . Performance (skills, knowledge, confidence, and empathy) in the 2 arms before and after the simulation.

Variable		Teaching as usual		Virtual simulation		Effect size
	Theoretical range	Before, mean (SD)	After, mean (SD)	Before, mean (SD)	After, mean (SD)	Cohen *d* (95% CI)
C-SSR[a] (sum of absolute *z scores)*	1.1 - 16.7	3.6 (1.6)	3.7 (2.4)	3.6 (1.7)	3 (1.6)	−0.2 (−0.7 to 0.29)
RED[b] (sum of absolute *z* scores)	1.1 - 6.3	3 (1.4)	3.5 (1.8)	2.9 (1.5)	3.2 (1.5)	−0.17 (−0.67 to 0.32)
SIRI[c] (sum of absolute *z* scores)	10.8 - 188	44.8 (15.8)	43.2 (15.9)	46.7 (14.1)	40.3 (11.9)	−0.44 (−0.94 to 0.06)
Knowledge (total)	0 - 50	35.7 (4.8)	35.7 (4.4)	37.8 (3.9)	38.1 (4)	0.03 (−0.46 to 0.53)
QNSSR[d] (total)	0 - 100	56.7 (13.2)	59.7 (15.1)	51 (10.8)	61.6 (12.4)	0.81 (0.29-1.32)
JSE[e] (total)	20 - 140	108.1 (9.5)	107 (9.8)	110.3 (8.7)	110.7 (7.2)	0.06 (−0.43 to 0.56)

[a]C-SSR: Columbia Suicide Severity Rating Scale.

[b]RED: Risk-Emergency-Danger scale.

[c]SIRI: Suicide Intervention Response Inventory.

[d]QNSSR: Questionnaire on Nursing Skills related to Suicide Risk.

[e]JSE: Jefferson Scale of Empathy (Health Professions Student version).

Knowledge

The effect of the time:arm interaction on knowledge was nonsignificant (0 points of increase after TAU versus 0.3 points of increase after VS, B=0.2; SE 0.8; $t_{119.7}$=0.2; *P*=.82; fmi=0.084), suggesting that the intervention was ineffective in improving knowledge of suicidology.

Skills, Attitude, and Confidence

Ability to Evaluate Suicide Risk

The effect of the time, arm interaction on the C-SSRS (0.1 points of worsening after TAU versus 0.3 points of improvement after VS, B=−0.3; SE 0.5; $t_{125.6}$=−0.6; *P*=.58; fmi=0.047) and RED scales (0.5 points of worsening after TAU vs 0.3 points of worsening after VS, B=−0.3; SE 0.5; $t_{124.8}$=−0.5; *P*=.59; fmi=0.052) was nonsignificant, suggesting that the intervention was ineffective in improving the evaluation of suicide risk.

Ability to Respond Appropriately to an Individual With Suicidal Thoughts

The effect of the time, arm interaction on the SIRI was marginally significant (1.6 points of improvement after TAU versus 6.4 points of improvement after VS, B=−4.5; SE 2.5; $t_{119.5}$=−1.8; *P*=.08; fmi=0.085). The time effect was nonsignificant for the TAU (B=−1.6; SE 1.8; $t_{50.9}$=−0.9; *P*=.36; fmi=0.09), whereas it was significant for the VS (B=−6.2; SE 1.8; $t_{64.9}$=−3.5; *P*=.001; fmi=0.136). The performance in response to suicidal thoughts improved after VS, suggesting a possible effectiveness of the intervention on how to interact with a patient experiencing a suicidal crisis.

Confidence

The effect of the time:arm interaction on the QNSSR was significant (3 points of increase after TAU versus 10.6 points of increase after VS, B=7.2; SE 2.5; $t_{111.5}$=2.8; *P*=.006; fmi=0.131). The time effect was marginally significant for the TAU (B=3; SE 1.8; $t_{50.9}$=1.7; *P*=.097; fmi=0.091) and significant for the VS (B=10.2; SE 1.7; $t_{62.8}$=5.8; *P*<.001; fmi=0.159). This suggests the effectiveness of the versus in increasing confidence in one's ability to interact with a patient experiencing a suicidal crisis.

Empathy

The effect of the time:arm interaction on knowledge was nonsignificant (1.1 points of worsening after TAU vs 0.4 points of improvement after VS, B=1.2; SE 2; $t_{122.9}$=0.6; *P*=.54; fmi=0.065), suggesting that the intervention was ineffective in improving empathy.

Satisfaction

The satisfaction results are presented in Table 3.

Internal consistency was excellent for the DASH, good for IPQ and the virtual patient facial emotions quality, acceptable for the virtual patient authenticity, questionable for the satisfaction with overall simulation training, and poor for the SUS. The user-friendliness of SIVIPSY was very high. The simulation induced a moderate sense of presence. The virtual patient was considered relatively authentic, with appropriate natural facial expressions of emotions, neither inhibited nor exaggerated. The debriefing was considered to be very good to excellent, and satisfaction with the overall intervention was between good and very good. The difficulty of the scenario was considered appropriate.

XSL•FO
RenderX

Table . Satisfaction with the virtual simulation.

Variable	Theoretical range	Mean (SD)	Missing (%)	Cronbach α
User-friendliness				
System Usability Scale	0-100	87.1 (9.9)	10.3	0.54
Overall friendliness	1 - 7	5.5 (0.7)	10.3	—
Presence				
Igroup Presence Questionnaire (item average)	1-7	4.2 (1)	10.3	0.84
Satisfaction with the virtual patient				
Authenticity (item average)	1-4	3.1 (0.6)	10.3	0.79
Facial emotions quality (item average)	1 - 4	3 (0.7)	12.8	0.89
Facial emotions quantity	1 - 5	3 (0.9)	12.8	—
Debriefing				
DASH[b] total (item average)	1-7	6.5 (0.5)	12.8	0.96
DASH stage setting (item average)	1 - 7	6.5 (0.5)	12.8	0.83
DASH learning context (item average)	1 - 7	6.6 (0.4)	12.8	0.79
DASH organized debriefing (item average)	1 - 7	6.6 (0.5)	12.8	0.83
DASH reflection on performance (item average)	1 - 7	6.3 (0.6)	12.8	0.81
DASH performance identification (item average)	1 - 7	6.3 (0.8)	12.8	0.91
DASH help to improve (item average)	1 - 7	6.4 (0.7)	12.8	0.81
Satisfaction with the overall simulation training				
Questionnaire (item average)	1-4	3.4 (0.3)	12.8	0.62
Perception of difficulty	1 - 5	2.9 (0.3)	12.8	—
Global rating	0 - 10	8.7 (1)	12.8	—

[a]Not applicable.

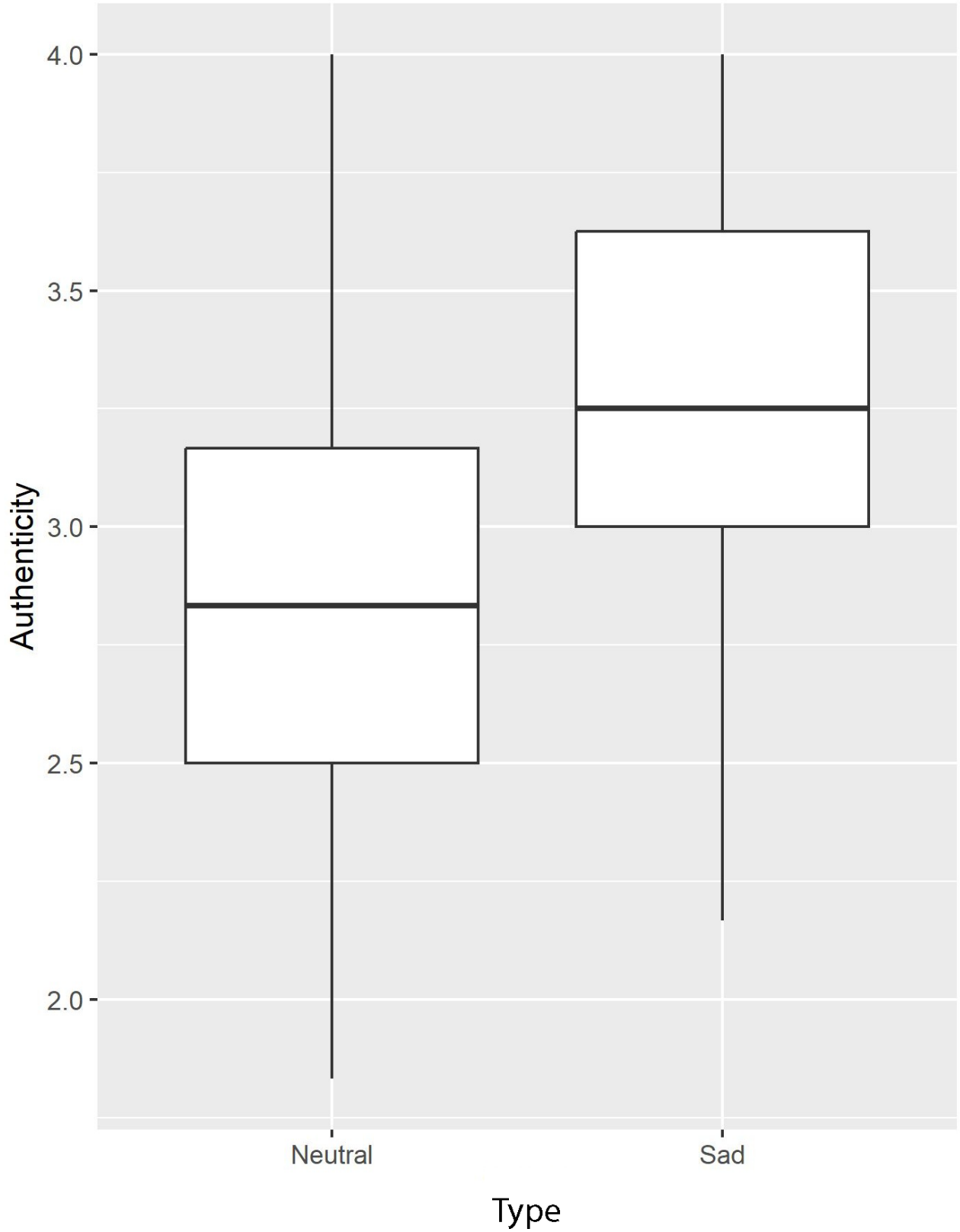
[b]DASH: Debriefing Assessment for Simulation in Healthcare.

Impact of Emotional Prosody on Satisfaction With the Virtual Patient

The effect of Prosody on virtual patient authenticity (MASP adapted) was significant (B=0.4; SE 0.2; t_{71}=2.2; P=.03; fmi=0.065). Learners rated the virtual patient as more authentic when he had a sad prosody than a neutral prosody (see Figure 8). This difference became marginally significant when age and GEQ (which were marginally different between the 2 conditions) were entered as covariates in the model (B=0.4; SE 0.2; $t_{63.7}$=1.9; P=.06; fmi=0.131).

XSL•FO
RenderX

Figure 8. Plot of the virtual patient's authenticity according to his prosody.



Discussion

Overview

This exploratory study aimed to evaluate the effectiveness of VS in training nursing students in suicide risk assessment through a randomized controlled investigation. Pedagogical effectiveness was measured in terms of satisfaction, knowledge, skills, confidence, and empathy. The VS group exhibited higher confidence and a better ability to respond appropriately to suicidal thoughts after VS compared with the control group. Satisfaction with VS was high, particularly regarding the authenticity of the virtual patient. This authenticity was

XSL•FO
RenderX

perceived as greater when emotional prosody was included alongside facial emotions.

Principal Findings

SIVIPSY was more efficient than usual teaching to improve confidence in nursing skills for the assessment and care of an individual experiencing a suicidal crisis. The picture was less clear in terms of the effectiveness of SIVIPSY in improving nursing skills related to suicidal crises. The ability to respond appropriately to an individual with suicidal thoughts significantly increased after the VS, whereas it did not after the usual teaching. However, the difference in improvement after VS and TAU was only marginally significant, suggesting that further studies are needed to confirm the greater effectiveness of VS in improving adequate nursing responses to individuals experiencing a suicidal crisis.

By contrast, SIVIPSY was not associated with an improvement in the ability to assess the severity of a suicidal crisis. SIVIPSY may thus be ineffective in improving the accuracy of suicide risk assessment. Another explanation could be a lack of sensitivity to change in the tool we have developed to measure this construct. SIVIPSY was also not associated with knowledge improvement. Neither VS nor TAU modified nursing students' general level of empathy. The average empathy score in the present sample was comparable to that reported in a US sample [48]. Empathy measured with the JSE may be a relatively stable personality disposition, even in the health care setting [55], which is unlikely to change after a one-off intervention such as a simulation session. Changes in empathy have been observed in medical students, but only over several years, with a progressive decline, especially after having spent more time interacting with patients [56]. The existence of such a decline needs to be clarified concerning nursing students, for which there are conflicting results [57,58].

There was high satisfaction with SIVIPSY user-friendliness and the overall simulation, although these results may lack reliability due to poor and questionable internal consistencies found for these questionnaires. The satisfaction with the group debriefing was very high and very reliable, suggesting that SIVIPSY should be implemented in teaching as usual. A social desirability bias could explain the high level of satisfaction reported in this study, as academic teachers were involved in the debriefing. However, all questionnaires were anonymously completed, which should have limited such bias. The level of satisfaction was reliable and good concerning the sense of presence, the authenticity of the virtual patient, and his emotional reactions. Our results suggest that authenticity was, as expected, improved using emotional prosody, although such improvement could be partially explained by the older age and greater game implication of the participants who used the sad prosody version of the VS. Future virtual patients in pedagogical settings should display vocal expression of emotions, as authenticity is a critical component of interactive multimedia mental health education programs for nurses [59].

Comparison to Previous Work

A previous study reported that nursing students had higher self-confidence after a simulated standardized patient session than those who watched a traditional video-recorded lecture on suicide [27]. Our results suggest that the beneficial impact of simulation involving actors playing simulated patients extends to VS. It is reasonable to assume that changes in attitudes with the improvement of confidence favored by VS may improve suicidal crisis detection, given that attitudes and self-efficacy are associated with the intention to engage in suicide prevention behaviors [60]. Nursing students often feel anxiety and fear before they start their clinical practice in psychiatry [61]. One reason for them to avoid evaluating suicide risk is the anticipation of not knowing what to do in case of suicidal thoughts. Enhancing confidence may have a positive impact on practical behavior (with an increase in suicide risk assessment of actual patients) and lead to improvements in attitude and reduce stigma [3].

Regarding the lack of knowledge improvement after VS, a previous study also reported that knowledge about suicide was comparable after simulation with a standardized patient and after a traditional lecture on suicide [29]. This suggests that the specificity of the pedagogical action of simulation does not appear to lie in improving theoretical nursing knowledge. A recent analysis confirmed this interpretation by reporting a lack of significant differences between simulation, in general, and controls in improving suicide risk assessment and intervention in RCTs, while non-RCT interventions with simulation and pre-post design were associated with knowledge improvement [62].

Strengths and Limitations

The 2 main strengths of this study are the innovative use of an emotionally reactive virtual patient to train students in caring for patients with suicidal ideation and the randomized controlled design of the intervention. The main methodological limitation of this study was the lack of an active comparator (an additional pedagogical intervention in the control arm), which may have inflated the efficiency of VS. As previously reported, attitudes (ways of approaching patients with suicidal ideas, beliefs, self-efficacy, confidence, and sense of preparedness regarding suicidal crises) improved significantly when simulation training was compared to an inactive comparator. By contrast, no significant difference was found concerning these attitudes when simulation training was compared to an active comparator [62]. Other methodological limitations were the lack of preregistration for this exploratory study, with no primary outcome selected, and the lack of a control for multiple comparisons. Despite randomization, groups were unbalanced in terms of the outcomes, with less confidence and a lower ability to respond appropriately to an individual with suicidal thoughts in the intervention group than the control group. Further studies should confirm the present results by stratifying learners based on these 2 outcomes. Finally, the expressiveness of the virtual patient is still limited, particularly postures and gestures.

Future Directions

Further simulation may benefit from increasing online feedback during the interaction with the virtual patient beyond the evolution of bar scores implemented in SIVIPSY. It has been reported that per-interview feedback increased clinicians' empathetic responses toward a patient experiencing a suicide

XSL•FO
RenderX

crisis more than postinterview feedback during training with a virtual patient [63]. The time dedicated to coding the verbal and emotional reactions of the virtual patient was substantial, and further interventions should use a more dynamic approach to emotional expressions based, for example, on the appraisal theory of emotion [64,65] to automate the avatar reactions. This could increase the authenticity of the virtual patient and decrease the burden on the teacher who creates the simulation scenario. Future studies should include an active comparator in the control arm, which may consist of a teaching program without simulation (such as recorded lectures on suicide assessment, delivery of booklets followed by question-and-answer discussions, or role plays) or with simulation using an alternative method (like simulation with standardized patients or expert patients).

Our findings have significant implications for disseminating and implementing SIVIPSY in nursing curricula. The intervention, which improved the management of individuals exhibiting suicidal behaviors, may help reduce unnecessary emergency room visits in the context of a marked deterioration in the mental health of young women in France after the COVID-19 pandemic [66]. The pedagogical effectiveness of VS reported in this study may translate into clinical efficiency, improving patient-reported outcomes, such as reducing suicide attempts and mortality by suicide following the implementation of this simulation training program. One argument favoring this transfer arises from the fact that training primary care doctors and nurses to better screen and treat depression, with supplemental help from psychiatrists, lowered suicide rates, suicide attempts, and suicidal ideation [67]. However, the studies investigating the clinical efficiency of suicide prevention programs often combine education, training, screening, and referral for treatment. The separate effect of each intervention component remains unknown, and further studies should assess the specific clinical efficiency of suicide risk training programs. For instance, long-term patient outcomes should be investigated after the nurses enter clinical practice, such as the number of structured suicide risk screenings conducted in clinical practice, which should increase after the VS training, or the number of suicide attempts, which should decrease. An operational efficiency of the VS would also imply a return on investment by reducing suicidality-related costs.

Conclusions

In conclusion, VS shows encouraging results in improving the confidence and skills of nursing students to behave appropriately with patients in suicidal crises. Further studies are needed to assess the impact of VS on patient outcomes such as the prevention of death by suicide and suicide attempts and the return on training investment.

Acknowledgments

This study was funded by the Agence Régionale de Santé d'Ile-de-France with the grant "Soutien à l'innovation dans les formations des professionnels de santé: simulation en santé" (2019). It was also funded by Versailles Saint-Quentin-En-Yvelines University with the grant "Appel à Projet simplifié interne" (2020). We thank Pr Fabrice Jollant for his valuable suggestions regarding outcomes assessment.

The study was funded by the following grants: Agence Nationale de la Recherche (ANR-17-CE19-0028), Agence Régionale de Santé d'Ile de France (Support to innovation in the training of health professionals: health simulation), and Université de Versailles Saint-Quentin-En-Yvelines (simplified internal call for projects; July 2020).

Data Availability

The datasets analyzed during this study and the code used are attached as Multimedia Appendix 5.

Authors' Contributions

PR contributed to conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, software, resources, supervision, validation, visualization, and writing of the original draft, as well as reviewing and editing the manuscript. YO contributed to conceptualization, formal analysis, investigation, methodology, project administration, software, resources, visualization, and writing of both the original draft and the reviewed manuscript. CM contributed to conceptualization, investigation, project administration, resources, supervision, validation, and manuscript review and editing. MS was involved in conceptualization, investigation, project administration, resources, supervision, validation, and reviewing and editing the manuscript. HB contributed to manuscript review and editing. EBG contributed to conceptualization, validation, and review and editing of the manuscript. SF contributed to reviewing and editing the manuscript. CP was involved in conceptualization, supervision, validation, and review and editing of the manuscript. NY contributed to conceptualization, supervision, validation, and writing of both the original and revised versions of the manuscript. JCM contributed to conceptualization, formal analysis, funding acquisition, methodology, project administration, software, resources, supervision, validation, and writing of both the original draft and the reviewed manuscript.

Conflicts of Interest

None declared.

XSL•FO
RenderX

Multimedia Appendix 1
Establishment of the scoring norms for the Columbia-Suicide Severity Rating Scale and the Risk-Emergency-Danger tool.
[DOCX File, 23 KB - games_v13i1e69347_app1.docx]

Multimedia Appendix 2
Questionnaire on Nursing Skills related to Suicide Risk (QNSSR).
[DOCX File, 28 KB - games_v13i1e69347_app2.docx]

Multimedia Appendix 3
Satisfaction questionnaires.
[DOCX File, 24 KB - games_v13i1e69347_app3.docx]

Multimedia Appendix 4
Test statistics of the effect of the simulation on performance.
[DOCX File, 22 KB - games_v13i1e69347_app4.docx]

Multimedia Appendix 5
Additional material.
[ZIP File, 81 KB - games_v13i1e69347_app5.zip]

Checklist 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File, 9904 KB - games_v13i1e69347_app6.pdf]

References

1. Suicide worldwide in 2019: global health estimates. World Health Organization.: World Health Organization; 2021.
2. Zalsman G, Hawton K, Wasserman D, et al. Suicide prevention strategies revisited: 10-year systematic review. Lancet Psychiatry 2016 Jul;3(7):646-659. [doi: 10.1016/S2215-0366(16)30030-X] [Medline: 27289303]
3. Saunders KEA, Hawton K, Fortune S, Farrell S. Attitudes and knowledge of clinical staff regarding people who self-harm: a systematic review. J Affect Disord 2012 Aug;139(3):205-216. [doi: 10.1016/j.jad.2011.08.024] [Medline: 21925740]
4. Aflague JM, Ferszt GG. Suicide assessment by psychiatric nurses: a phenomenographic study. Issues Ment Health Nurs 2010 Apr;31(4):248-256. [doi: 10.3109/01612840903267612] [Medline: 20218768]
5. Wärdig R, Engström AS, Carlsson A, Wärdig F, Hultsjö S. Saving lives by asking questions: nurses' experiences of suicide risk assessment in telephone counselling in primary health care. Prim Health Care Res Dev 2022 Oct 26;23:e65. [doi: 10.1017/S146342362200055X] [Medline: 36285522]
6. Valente S, Saunders JM. Barriers to suicide risk management in clinical practice: a national survey of oncology nurses. Issues Ment Health Nurs 2004 Sep;25(6):629-648. [doi: 10.1080/01612840490472147] [Medline: 15371147]
7. Puntil C, York J, Limandri B, Greene P, Arauz E, Hobbs D. Competency-based training for PMH nurse generalists: inpatient intervention and prevention of suicide. J Am Psychiatr Nurses Assoc 2013;19(4):205-210. [doi: 10.1177/1078390313496275] [Medline: 23950543]
8. Ferguson M, Reis J, Rabbetts L, et al. The impact of suicide prevention education programmes for nursing students: A systematic review. Int J Mental Health Nurs 2020 Oct;29(5):756-771. [doi: 10.1111/inm.12753]
9. Valente SM. Oncology nurses' knowledge of suicide evaluation and prevention. Cancer Nurs 2010;33(4):290-295. [doi: 10.1097/NCC.0b013e3181cc4f33] [Medline: 20467313]
10. Wärdig RE, Hultsjö S, Lind M, Klavebäck I. Nurses' experiences of suicide prevention in primary health care (PHC) - a qualitative interview study. Issues Ment Health Nurs 2022 Oct;43(10):903-912. [doi: 10.1080/01612840.2022.2089789] [Medline: 35793075]
11. Vedana KGG, Pereira CCM, dos Santos JC, et al. The meaning of suicidal behaviour from the perspective of senior nursing undergraduate students. Int J Mental Health Nurs 2018 Jun;27(3):1149-1161. [doi: 10.1111/inm.12431]
12. Gould MS, Marrocco FA, Kleinman M, et al. Evaluating iatrogenic risk of youth suicide screening programs: a randomized controlled trial. JAMA 2005 Apr 6;293(13):1635-1643. [doi: 10.1001/jama.293.13.1635] [Medline: 15811983]
13. Biddle L, Cooper J, Owen-Smith A, et al. Qualitative interviewing with vulnerable populations: individuals' experiences of participating in suicide and self-harm based research. J Affect Disord 2013 Mar 5;145(3):356-362. [doi: 10.1016/j.jad.2012.08.024] [Medline: 23021191]
14. Dazzi T, Gribble R, Wessely S, Fear NT. Does asking about suicide and related behaviours induce suicidal ideation? What is the evidence? Psychol Med 2014 Dec;44(16):3361-3363. [doi: 10.1017/S0033291714001299]

XSL•FO
RenderX

15. Solin P, Tamminen N, Partonen T. Suicide prevention training: self-perceived competence among primary healthcare professionals. Scand J Prim Health Care 2021 Sep;39(3):332-338. [doi: 10.1080/02813432.2021.1958462] [Medline: 34340646]
16. Kerr S, Martin C, Fleming M. Preventing suicide; nurse education and the occluded issue of gender. Nurse Educ Pract 2018 Sep;32:58-63. [doi: 10.1016/j.nepr.2018.07.004] [Medline: 30031273]
17. Luebbert R, Popkess A. The influence of teaching method on performance of suicide assessment in baccalaureate nursing students. J Am Psychiatr Nurses Assoc 2015;21(2):126-133. [doi: 10.1177/1078390315580096] [Medline: 25979880]
18. Piot MA, Dechartres A, Attoe C, et al. Effectiveness of simulation in psychiatry for nursing students, nurses and nurse practitioners: A systematic review and meta-analysis. J Adv Nurs 2022 Feb;78(2):332-347. [doi: 10.1111/jan.14986] [Medline: 34378236]
19. Foster A, Chaudhary N, Murphy J, Lok B, Waller J, Buckley PF. The use of simulation to teach suicide risk assessment to health profession trainees-rationale, methodology, and a proof of concept demonstration with a virtual patient. Acad Psychiatry 2015 Dec;39(6):620-629. [doi: 10.1007/s40596-014-0185-9] [Medline: 25026950]
20. Shea SC. Psychiatric interviewing e-book: the art of understanding: a practical guide for psychiatrists, psychologists, counselors, social workers, nurses, and other mental health professionals. Elsevier Health Sciences. 2016.
21. Roux P, Okuya Y, Martin JC. Simulation virtuelle: développement du jeu sérieux SIVIPSY pour l'enseignement de la psychiatrie aux étudiants en santé. Prepsy. 2021. URL: https://www.prepsy.fr/s/PROGRAMME_MAPS2021_PREPSY-cck3.pdf [accessed 2025-07-06]
22. Courgeon M, Martin JC, Jacquemin C. Marc: a multimodal affective and reactive character. Presented at: Proceedings of the 1st Workshop on AFFective Interaction in Natural Environments; 2008 p. 1.
23. Courgeon M, Céline C, Martin JC. Modeling facial signs of appraisal during interaction: impact on users' perception and behavior. International Foundation for Autonomous Agents and Multiagent Systems 2014:765-772.
24. Hoque M, Courgeon M, Martin JC, Mutlu B, Picard RW. MACH: my automated conversation coach. : ACM Press Presented at: Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing; Sep 8, 2013 p. 697. [doi: 10.1145/2493432.2493502]
25. Oker A, Prigent E, Courgeon M, et al. How and why affective and reactive virtual agents will bring new insights on social cognitive disorders in schizophrenia? An illustration with a virtual card game paradigm. Front Hum Neurosci 2015;9:133. [doi: 10.3389/fnhum.2015.00133] [Medline: 25870549]
26. Friesen E, Ekman P. Facial Action Coding System: A Technique for the Measurement of Facial Movement: Palo Alto Consulting Psychologists Press; 1978.
27. Cordaro DT, Sun R, Keltner D, Kamble S, Huddar N, McNeil G. Universals and cultural variations in 22 emotional expressions across five cultures. Emotion 2018 Feb;18(1):75-93. [doi: 10.1037/emo0000302] [Medline: 28604039]
28. Golan O, Baron-Cohen S. Systemizing empathy: teaching adults with Asperger syndrome or high-functioning autism to recognize complex emotions using interactive multimedia. Dev Psychopathol 2006;18(2):591-617. [doi: 10.1017/S0954579406060305] [Medline: 16600069]
29. Acapela TTS for windows. Acapela Group. URL: https://www.acapela-group.com/ [accessed 2025-07-06]
30. Nordgaard J, Sass LA, Parnas J. The psychiatric interview: validity, structure, and subjectivity. Eur Arch Psychiatry Clin Neurosci 2013 Jun;263(4):353-364. [doi: 10.1007/s00406-012-0366-z] [Medline: 23001456]
31. Ellaway R, Poulton T, Fors U, McGee JB, Albright S. Building a virtual patient commons. Med Teach 2008 Jan;30(2):170-174. [doi: 10.1080/01421590701874074]
32. Huwendiek S, BA, Zary N, Fischer MR, Ruiz JG, Ellaway R. Towards a typology of virtual patients. Med Teach 2009 Aug;31(8):743-748. [doi: 10.1080/01421590903124708] [Medline: 19811212]
33. Schell J. The Art of Game Design: A Book of Lenses: CRC press; 2008.
34. Khanduja K, Bould MD, Andrews M, et al. Impact of unexpected death in a simulation scenario on skill retention, stress, and emotions: a simulation-based randomized controlled trial. Cureus 2023 May;15(5):e39715. [doi: 10.7759/cureus.39715] [Medline: 37398706]
35. Pham-Dinh C, Laprevote V, Schwan R, et al. Quantifying efficacy of investigation during a simulated psychiatric interview. Encephale 2020 Apr;46(2):96-101. [doi: 10.1016/j.encep.2019.07.010] [Medline: 31522835]
36. Rudolph JW, Simon R, Rivard P, Dufresne RL, Raemer DB. Debriefing with good judgment: combining rigorous feedback with genuine inquiry. Anesthesiol Clin 2007 Jun;25(2):361-376. [doi: 10.1016/j.anclin.2007.03.007] [Medline: 17574196]
37. Witmer BG, Singer MJ. Measuring presence in virtual environments: a presence questionnaire. Presence (Camb) 1998 Jun;7(3):225-240. [doi: 10.1162/105474698565686]
38. Brockmyer JH, Fox CM, Curtiss KA, McBroom E, Burkhart KM, Pidruzny JN. The development of the Game Engagement Questionnaire: A measure of engagement in video game-playing. J Exp Soc Psychol 2009 Jul;45(4):624-634. [doi: 10.1016/j.jesp.2009.02.016]
39. Kirkpatrick D, Kirkpatrick J. Evaluating Training Programs: The Four Levels: Berrett-Koehler Publishers; 2006.
40. Google Drive. URL: https://drive.google.com/file/d/1WcqMhIvjvb4nMXPiF440q21AuCUYKtpq/view?usp=drive_link [accessed 2025-07-30]

XSL•FO
RenderX

41. Posner K, Brown GK, Stanley B, et al. The Columbia-Suicide Severity Rating Scale: initial validity and internal consistency findings from three multisite studies with adolescents and adults. Am J Psychiatry 2011 Dec;168(12):1266-1277. [doi: 10.1176/appi.ajp.2011.10111704] [Medline: 22193671]
42. Viard MC, Grandgenèvre P, Bubrovszky M, et al. Impact de la formation à l'intervention de crise suicidaire sur la confiance et les aptitudes des professionnels hospitaliers des Hauts-de-France. L'Encéphale 2023 Oct;49(5):504-509. [doi: 10.1016/j.encep.2022.05.005]
43. Agence nationale d'accréditation et d'évaluation en santé, Fédération Française de Psychiatrie, Direction Générale de la Santé. La crise suicidaire: reconnaître et prendre en charge. Conférence de consensus Hôpital de la Pitié Salpêtrière - Paris: ANAES. 2000. URL: https://www.has-sante.fr/jcms/c_271964/fr/lacrise-suicidaire-reconnaitre-et-prendre-encharge [accessed 2025-07-30]
44. Weber AN, Michail M, Thompson A, Fiedorowicz JG. Psychiatric emergencies. Medical Clinics of North America 2017 May;101(3):553-571. [doi: 10.1016/j.mcna.2016.12.006]
45. Neimeyer RA, Bonnelle K. The suicide intervention response inventory: a revision and validation. Death Stud 1997;21(1):59-81. [doi: 10.1080/074811897202137] [Medline: 10169714]
46. Kopp-Bigault C, Audouard-Marzin Y, Scouarnec P, Beauchamp G, Séguin M, Walter M. Traduction française et validation d'une version courte du Suicide Intervention Response Inventory-2 (SIRI-2-VF). L'Encéphale 2018 Nov;44(5):435-445. [doi: 10.1016/j.encep.2017.07.006]
47. Younes N, Delaunay AL, Roger M, et al. Evaluating the effectiveness of a single-day simulation-based program in psychiatry for medical students: a controlled study. BMC Med Educ 2021 Jun 16;21(1):348. [doi: 10.1186/s12909-021-02708-6] [Medline: 34134692]
48. Fields SK, Mahan P, Tillman P, Harris J, Maxwell K, Hojat M. Measuring empathy in healthcare profession students using the Jefferson Scale of Physician Empathy: health provider--student version. J Interprof Care 2011 Jul;25(4):287-293. [doi: 10.3109/13561820.2011.566648] [Medline: 21554061]
49. Gronier G, Baudet A. Psychometric evaluation of the F-SUS: creation and validation of the French version of the System Usability Scale. International Journal of Human–Computer Interaction 2021 Oct 2;37(16):1571-1582. [doi: 10.1080/10447318.2021.1898828]
50. Schubert T, Friedmann F, Regenbrecht H. The experience of presence: factor analytic insights. Presence-Teleop Virt 2001 Jun;10(3):266-281. [doi: 10.1162/105474601300343603]
51. Wind LA, Van Dalen J, Muijtjens AMM, Rethans JJ. Assessing simulated patients in an educational setting: the MaSP (Maastricht Assessment of Simulated Patients). Med Educ 2004 Jan;38(1):39-44. [doi: 10.1111/j.1365-2923.2004.01686.x] [Medline: 14962025]
52. Simon R, Raemer D, Rudolph J. Debriefing Assessment for Simulation in Healthcare©-Student Version, Long Form: Cambridge: Center for Medical Simulation; 2010.
53. Dreifuerst KT. Using debriefing for meaningful learning to foster development of clinical reasoning in simulation. J Nurs Educ 2012 Jun;51(6):326-333. [doi: 10.3928/01484834-20120409-02] [Medline: 22495923]
54. Cook DA, Erwin PJ, Triola MM. Computerized virtual patients in health professions education: a systematic review and meta-analysis. Acad Med 2010 Oct;85(10):1589-1602. [doi: 10.1097/ACM.0b013e3181edfe13] [Medline: 20703150]
55. Yu CC, Tan L, LE MK, et al. The development of empathy in the healthcare setting: a qualitative approach. BMC Med Educ 2022 Apr 4;22(1):245. [doi: 10.1186/s12909-022-03312-y] [Medline: 35379249]
56. Neumann M, Edelhäuser F, Tauschel D, et al. Empathy decline and its reasons: a systematic review of studies with medical students and residents. Acad Med 2011 Aug;86(8):996-1009. [doi: 10.1097/ACM.0b013e318221e615] [Medline: 21670661]
57. Ward J, Cody J, Schaal M, Hojat M. The empathy enigma: an empirical study of decline in empathy among undergraduate nursing students. J Prof Nurs 2012;28(1):34-40. [doi: 10.1016/j.profnurs.2011.10.007] [Medline: 22261603]
58. Håkansson Eklund J, Holmström IK, Ollén Lindqvist A, Sundler AJ, Hochwälder J, Marmstål Hammar L. Empathy levels among nursing students: A comparative cross-sectional study. Nurs Open 2019 Jul;6(3):983-989. [doi: 10.1002/nop2.280] [Medline: 31367422]
59. Lamont S, Brunero S. "eSimulation". Part 2: Evaluation of an interactive multimedia mental health education program for generalist nurses. Collegian 2014;21(1):3-9. [doi: 10.1016/j.colegn.2012.11.002] [Medline: 24772985]
60. Kuhlman STW, Walch SE, Bauer KN, Glenn AD. Intention to enact and enactment of gatekeeper behaviors for suicide prevention: an application of the theory of planned behavior. Prev Sci 2017 Aug;18(6):704-715. [doi: 10.1007/s11121-017-0786-0] [Medline: 28444519]
61. Sharif F, Masoumi S. A qualitative study of nursing student experiences of clinical practice. BMC Nurs 2005 Nov 9;4(1):6. [doi: 10.1186/1472-6955-4-6] [Medline: 16280087]
62. Richard O, Jollant F, Billon G, Attoe C, Vodovar D, Piot MA. Simulation training in suicide risk assessment and intervention: a systematic review and meta-analysis. Med Educ Online 2023 Dec;28(1):2199469. [doi: 10.1080/10872981.2023.2199469] [Medline: 37073473]
63. Yao H, de Siqueira AG, Bafna A, et al. A virtual human interaction using scaffolded ping-pong feedback for healthcare learners to practice empathy skills. : ACM Presented at: Proceedings of the 22nd ACM International Conference on Intelligent Virtual Agents Faro Portugal; Sep 6, 2022; Faro Portugal p. 1-8. [doi: 10.1145/3514197.3549621]

XSL•FO
RenderX

64. Scherer KR. The dynamic architecture of emotion: Evidence for the component process model. Cognition and Emotion 2009 Nov;23(7):1307-1351. [doi: 10.1080/02699930902928969]
65. Ortony A, Clore GL, Collins A. The Cognitive Structure of Emotions: cambridge university press; 1988. [doi: 10.1017/CBO9780511571299]
66. Fond G, Pauly V, Brousse Y, et al. Mental health care utilization and prescription rates among children, adolescents, and young adults in France. JAMA Netw Open 2025 Jan 2;8(1):e2452789. [doi: 10.1001/jamanetworkopen.2024.52789] [Medline: 39777444]
67. Mann JJ, Michel CA, Auerbach RP. Improving suicide prevention through evidence-based strategies: a systematic review. Am J Psychiatry 2021 Jul;178(7):611-624. [doi: 10.1176/appi.ajp.2020.20060864] [Medline: 33596680]

Abbreviations

C-SSRS: Columbia-Suicide Severity Rating Scale
CONSORT: Consolidated Standards of Reporting Trials
CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
DASH: Debriefing Assessment for Simulation in Healthcare
fmi: fraction of missing information
GEQ: Game Engagement Questionnaire
IPQ: Igroup Presence Questionnaire
ITQ: Immersive Tendencies Questionnaire
JSE: Jefferson Scale of Empathy
MARC: Multimodal Affective and Reactive Characters
MAsP: Maastricht Assessment of Simulated Patients
MCQ: multiple choice question
QNSSR: Questionnaire on Nursing Skills related to Suicide Risk
RED: Risk - Emergency - Danger
SIRI-2-VF: Suicide Intervention Response Inventory, French version 2
SIVIPSY: Simulation with a Virtual Patient in Psychiatry
SUS: System Usability Scale
TAU: teaching as usual
VS: virtual simulation

Edited by A Coristine; submitted 27.11.24; peer-reviewed by A Foster, MC Stovall; revised version received 07.05.25; accepted 07.05.25; published 01.08.25.

Please cite as:
Roux P, Okuya Y, Morel C, Soulès M, Bottemanne H, Brunet-Gouet E, Frileux S, Passerieux C, Younes N, Martin JC
Effectiveness of a Web-Based Virtual Simulation to Train Nursing Students in Suicide Risk Assessment: Randomized Controlled Investigation
JMIR Serious Games 2025;13:e69347
URL: https://games.jmir.org/2025/1/e69347
doi:10.2196/69347

© Paul Roux, Yujiro Okuya, Cristina Morel, Mariane Soulès, Hugo Bottemanne, Eric Brunet-Gouet, Solène Frileux, Christine Passerieux, Nadia Younes, Jean Claude Martin. Originally published in JMIR Serious Games (https://games.jmir.org), 1.8.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Exploring the Impact of Home-Based Serious Smartphone Resuscitation Gaming on Stress Among Nursing Students Practicing Simulated Adult Basic Life Support: Randomized Waitlist Controlled Trial

Nino Fijačko[1,2]; Benjamin S Abella[3]; Špela Metličar[4]; Leon Kopitar[1,5]; Robert Greif[6]; Gregor Štiglic[1,5,7]; Pavel Skok[2,8]; Matej Strnad[1,2,8,9]

[1]Faculty of Health Sciences, University of Maribor, Žitna 15, Maribor, Slovenia

[2]University Medical Centre Maribor, Maribor, Slovenia

[3]Icahn School of Medicine at Mount Sinai, New York, NY, United States

[4]Medical Dispatch Centre Maribor, University Clinical Centre Ljubljana, Ljubljana, Slovenia

[5]Faculty of Electrical Engineering and Computer Science, University of Maribor, Maribor, Slovenia

[6]Faculty of Medicine, University of Bern, Bern, Switzerland

[7]Usher Institute, University of Edinburgh, Edinburgh, United Kingdom

[8]Faculty of Medicine, University of Maribor, Maribor, Slovenia

[9]Community Healthcare Centre Dr. Adolfa Drolca Maribor, Maribor, Slovenia

Corresponding Author:
Nino Fijačko
Faculty of Health Sciences, University of Maribor, Žitna 15, Maribor, Slovenia

Abstract

Background: Simulation-based training is widely used in resuscitation education, yet limited research exists on how serious smartphone games—especially when used independently at home—impact stress levels during simulated adult basic life support (BLS). Understanding this relationship may offer new approaches to preparing health care students for high-stress clinical situations.

Objective: This study aimed to evaluate the impact of a home-based serious resuscitation game, MOBICPR, on physiological stress markers among nursing students performing simulated adult BLS.

Methods: In this single-center, randomized, waitlist controlled trial, 43 first-year nursing students were assigned to either an intervention group (IG) or a waitlist control group (WL-CG). Stress was measured at baseline and 2-week and 4-week follow-ups using electrodermal activity (EDA), blood volume pulse (BVP), heart rate (HR), and body temperature (BT) collected via the Empatica E4 wearable (Empatica Inc., USA). Each data collection point included 3 phases: mandala coloring before and after simulated adult BLS performance, and the adult BLS scenario itself. The MOBICPR game—a serious mobile game designed per the 2021 European Resuscitation Council adult BLS guidelines—was played at home over 2 weeks by IG (weeks 0 - 2) and WL-CG (weeks 2 - 4). A random forest classifier, trained on the AffectiveRoad dataset, predicted stress levels (none, moderate, and high) based on physiological signals.

Results: Of 124 students invited, 43 participated (22 in IG, 21 in WL-CG; 38/43, 88% female; mean age of 19, SD 0.6 years). EDA, BVP, and BT significantly changed across measurement phases in both groups (P<.05), while HR did not show consistent differences (P>.05). Stress classification showed a significant decrease in stress after simulated adult BLS in the IG at the 2-week follow-up (P=.04), but not in the WL-CG. After 2 weeks of gameplay, 12 of 22 participants in the IG had lower stress levels after performing simulated adult BLS compared to before, suggesting an adaptive stress response. No significant group-level stress reductions were observed over time.

Conclusions: Short-term, home-based gameplay using a serious resuscitation game modestly influenced physiological indicators of stress during simulated adult BLS among nursing students. While overall group stress levels remained stable, individualized responses suggested improved coping for some. Incorporating serious games into curricula could offer learners safe, gamified environments to rehearse stressful clinical scenarios. Future research should explore optimal game frequency and content depth to maximize educational and emotional resilience outcomes.

Trial Registration: ClinicalTrials.gov NCT05784675; https://clinicaltrials.gov/study/NCT05784675

XSL•FO
RenderX

(JMIR Serious Games 2025;13:e67623) doi:10.2196/67623

KEYWORDS

adult basic life support; stress; serious games; gaming; nursing student; simulation; randomized controlled trial; RCT; smartphone

Introduction

Adult basic life support (BLS) poses a substantial challenge for health care workers, as patient survival depends on their prompt and precise intervention [1]. To prepare future health care professionals for such challenges, educators can use various types of simulations [2], extended realities [3], or serious games [4]. According to Bergeron's definition [5], serious games are "interactive computer applications, with or without a significant hardware component, like smartphones, that includes a challenging goal, is fun to play with, incorporates some concept of scoring, and imparts in the user a skill, knowledge, or attitude which can be applied in the real world." Games are called serious when they have an educational purpose [6]. Playing a serious game on resuscitation of a patient in cardiac arrest can induce stress in users [4]. That might be positive, known as eustress, which arises as an individual's response to situations within their manageable limits [7]. Conversely, negative stress, referred to as distress, emerges as an individual's response to situations that surpass their ability to cope [8]. Serious smartphone games, as a component of mobile learning with gamification, have shown a positive impact on enhancing a range of learning outcomes and competencies in nursing as they offer the user the opportunity for repetition, which affects reduced stress levels and self-efficacy, and that within the home environment [9-13]. Beyond adult BLS [14-17], serious smartphone games have also been used to teach various clinical interventions, including advanced life support [18], choking management [19], triage [20], tactical combat casualty care [21], and a range of surgical [22-24] and nursing skills [25,26]. The assessment of stress during real or simulated resuscitation can be measured subjectively via a variety of scales [27]. Objectively through a range of electrophysiological metrics [28,29], including changes in heart rate [30,31], blood pressure [32], or skin temperature [33] alongside biomarkers (eg, cortisol [27,30] or α-amylase in saliva [32]). Over the past decade, new technology has improved our ability to track real-time stress levels using mobile phones and wearable devices [34]. While there are several studies on stress detection in controlled laboratory conditions [35-38], to our knowledge, no studies have examined perceived stress detection following the use of serious smartphone games based on cardiac arrest resuscitation content at home. That triggered us to investigate the effects of home-based serious smartphone gaming on stress among nursing students practicing simulated adult BLS.

Methods

Study Design

This single-center, simulated, randomized, waitlist controlled study was conducted at the Faculty of Health Sciences, University of Maribor (FHS UM) in Slovenia, Europe, from March to May 2023. A randomized, waitlisted, controlled study is a type of scientific experiment where participants are randomly assigned to either receive the intervention being tested immediately or be placed on a waiting list (control group) to receive the intervention after a delay, allowing for comparison of outcomes between the groups.

Participants

Every nursing student registered in the first-degree program at the FHS UM, for the academic year 2022 - 23 was invited to participate in the research project. A verbal presentation explained the research protocol to the students. They could sign up and provide their written consent. The inclusion criteria included nursing students who provided written informed consent, were 18 years or older, were capable of performing simulated adult BLS on a manikin (eg, without injury), and were willing to perform rescue breathing on a manikin. Notably, no exclusion criteria were set.

Study Protocol

Using a computer-generated randomization list (Microsoft 365 Excel Enterprise), study participants were allocated in equal proportions to either the intervention group (IG) or the waitlist control group (WL-CG). The WL-CG participants were placed on a waitlist, receiving the intervention (playing the MOBICPR game at home for 2 weeks) after the IG. The evaluation of stress initially occurred at baseline and was followed up by 2 subsequent measurements: one after 2 weeks and another at week 4. Each of the 3 measurements (baseline and the 2 follow-ups) consisted of four phases: (phase 1) five minutes of coloring mandala at the study site; (phase 2) performing simulated adult BLS with 2 minutes of cardiopulmonary resuscitation (CPR) at the study site; (phase 3) five minutes of coloring mandala at the study site; and (phase 4) playing the MOBICPR game at the study site followed by 2 weeks at home, first for IG, and after 2 weeks for WL-CG. In the first 3 phases, stress was measured by using the Empatica E4 (Empatica Inc., USA) device.

The Empatica E4 device was used to measure stress, as it is a proven and reliable tool for assessing physiological stress indicators [39]. Its efficacy was established by comparing various physiological measures with those recorded by other devices under similar stressors. Notably, the Empatica E4 device demonstrated the highest accuracy in detecting stress compared to its counterparts [40].

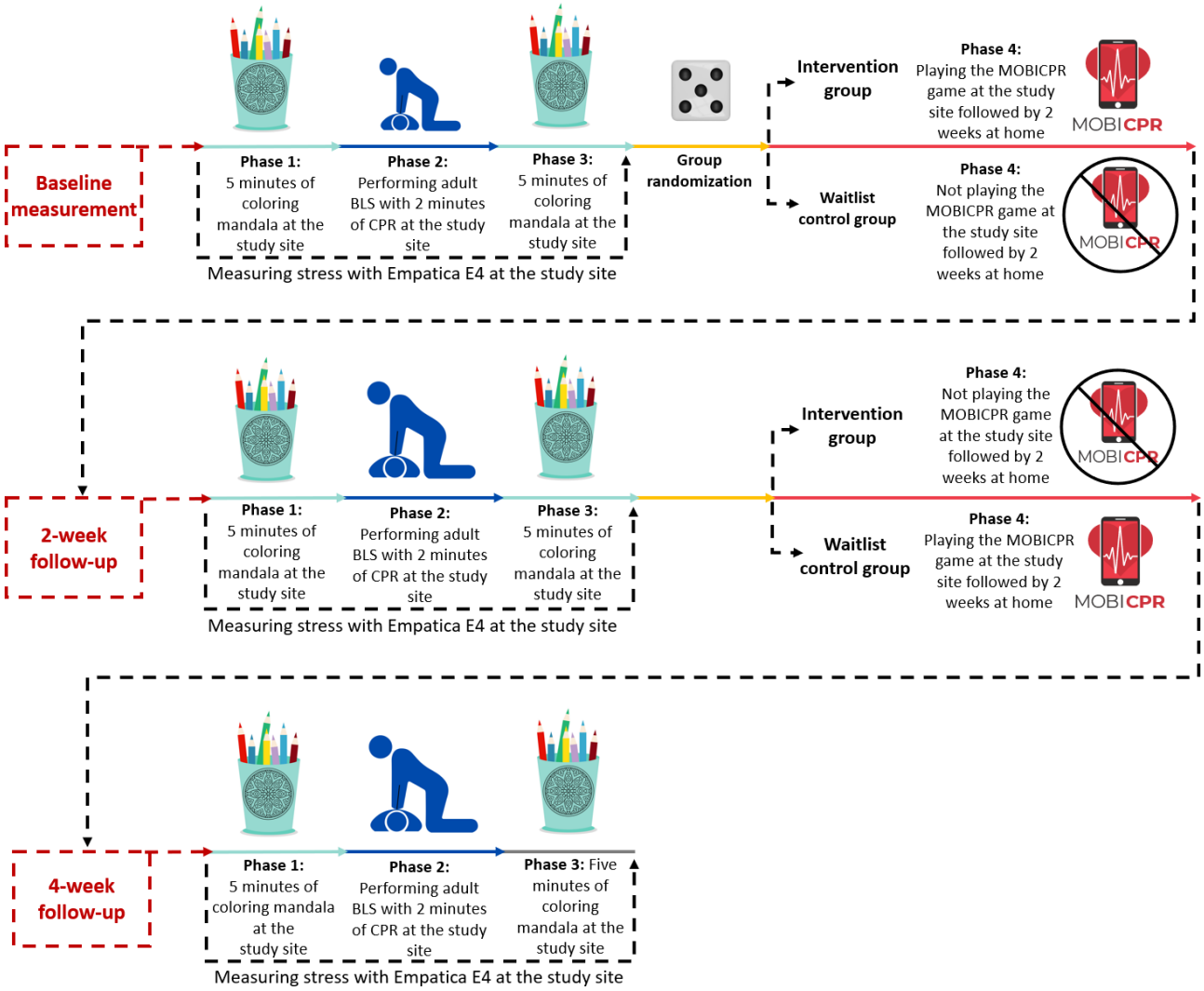
Before commencing the first phase of the measurement, the study investigator took a moment to reemphasize the study's objectives to each nursing student and provided additional information to ensure clarity and understanding. In the first phase, each nursing student in both research groups selected a distinct mandala figure from the *Celestial Mandalas Coloring Book* [41] and marked it with their study number for identification. The inclusion of mandala coloring was primarily designed to divert nursing students' attention away from the realization that they were undergoing stress measurement. This

XSL•FO
RenderX

approach is grounded in the concept that structured coloring activities, particularly those involving predesigned repetitive patterns such as mandalas, are associated with mindfulness practices and contribute to calming the mind and reducing perceived stress levels [42]. Before coloring the mandala figure for 5 minutes, the investigator securely placed the Empatica E4 device on the nondominant hand of the study participant by following Empatica instructions and connected the device to the E4 real-time application (Empatica Inc., USA) installed on a Google Pixel 4 (Google LLC, USA) smartphone to facilitate real-time monitoring and data recording. The coloring activity took place in a separate quiet room without interruptions, where a steady temperature of 21 to 23 °C was maintained. In the second phase, each study participant was given a scenario of out-of-hospital cardiac arrest (OHCA) to read. Then they proceeded to a designated room where they performed simulated adult BLS with 2 minutes of CPR on an adult manikin (Resusci Anne QCPR, Laerdal Medical, Norway) using an automated external defibrillator (Trainer AED, Defibtech LLC, USA). The OHCA scenario was performed in a staged kitchen, which was surrounded by mobile walls in the hospital's simulated room [43,44]. In the third phase, the study participants continued coloring preselected mandalas. After 5 minutes, the investigator removed the Empatica E4 device. After the third phase, nursing students played the MOBICPR game under the investigator's observation to familiarize themselves with the game's mechanics and content correctly. The game's content (eg, the 5-finger mnemonic), which guided the user through OHCA, was developed through the Delphi process [45] and adhered to the 2021 European Resuscitation Council adult BLS guidelines [46]. The MOBICPR game decreases the victim's survival probability with each error made by the player. Upon completion, the MOBICPR game provided a cumulative score as a gamification element, indicating the victim's likelihood of survival. A score above 51% suggests a successful outcome [47]. Playing the MOBICPR game at home has been shown to enhance and maintain theoretical knowledge of adult BLS, although it did not improve practical skills [43]. The average gameplay duration, based on previous studies, was 8 minutes. Subsequently, the IG received a Samsung Galaxy A13 (Samsung Group, South Korea) smartphone, preloaded with the MOBICPR game. Participants were instructed to play the game at home for 2 weeks, as frequently as they desired. After those 2 weeks, nursing students in the IG returned the smartphones. Subsequently, these smartphones were handed over to the study participants in WL-CG to play the MOBICPR game at home for 2 weeks. Then they returned the smartphones too (Figure 1).

XSL•FO
RenderX

Figure 1. Study flowchart illustrating the randomized waitlist controlled trial design conducted at the Faculty of Health Sciences, University of Maribor, Slovenia (March-May 2023), evaluating the impact of home-based MOBICPR gameplay on physiological stress responses among nursing students performing simulated adult BLS. BLS: basic life support; CPR: cardiopulmonary resuscitation.



Outcome

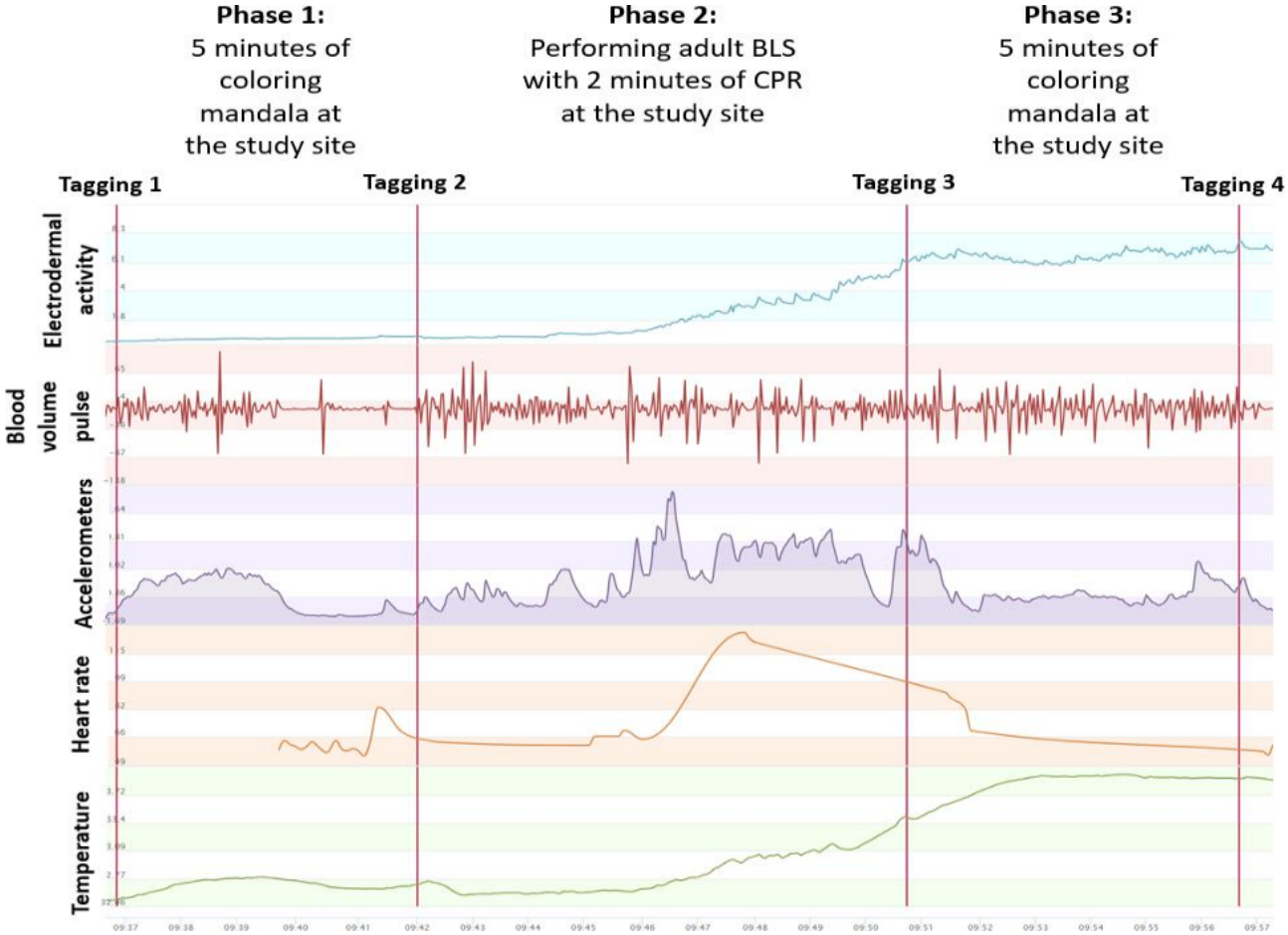
The primary outcome of the study was to assess whether playing the MOBICPR game at home for 2 weeks led to changes in physiological metrics among nursing students in the IG and WL-CG before and after performing simulated adult BLS on a manikin, with stress levels evaluated as a secondary outcome.

Collecting Data, Measuring Stress, and Statistical Analysis

To signify the beginning and the end of each phase for every study participant, the investigator was required to activate a feature on the Empatica E4 device by pressing a button, a process known as "tagging." Each study participant was assigned 4 distinct tags: the first tag signified the start of phase 1, the second marked the conclusion of phase 1 and the commencement of the second phase, the third tag indicated the end of the second phase and the beginning of the third, and the fourth tag represented the conclusion of the third phase. Once the fourth tag was logged, the Empatica E4 device was removed from the nursing student's wrist. Before the Empatica E4 device was turned off, the data were transferred to the Empatica Manager software via the Bluetooth built into the Empatica E4 device. For each nursing student in the study, 3 separate recordings were taken using the Empatica E4 device—at baseline and then after 2 and 4 weeks. These recordings measured five distinct physiological metrics: (1) electrodermal activity (EDA), (2) blood volume pulse (BVP), (3) accelerometer data, (4) heart rate (HR), and (5) body temperature (BT), as shown in Figure 2.

XSL•FO
RenderX

Figure 2. Example of Empatica E4 "tagging" used for stress measurement, showing data collection phases for one nursing student in the randomized waitlist controlled trial on the effects of serious gaming on stress during simulated adult BLS in Slovenia (2023). Phases 1 and 3 represent 5-minute mandala coloring periods before and after simulated adult BLS. BLS: basic life support; CPR: cardiopulmonary resuscitation.



Statistical analyses were carried out over a period extending from May 2023 to September 2024. The research was conducted in the programming language R (R Foundation for Statistical Computing, Austria). We read Empatica E4 device physiological metrics with a library "wearables" [48], while a stacked bar graph was generated using a library ggplot2 [49].

The age of participants was the only measure where we used mean and SD. For physiological metrics, we first assessed normality using the Shapiro-Wilk test. Data with P values ≥.05 were considered normally distributed and analyzed with parametric tests (eg, Student t test), while nonnormal data (P<.05) were analyzed using nonparametric alternatives (eg, Mann-Whitney U test for independent samples or Wilcoxon signed-rank test for paired comparisons).

Primary Outcome

We focused on the measurement of 4 key physiological metrics: EDA, BT, BVP, and HR. Accelerometer data were excluded, as participants wore the Empatica E4 on their nondominant hand, which remained at rest during the mandala coloring activity, consistent with procedures followed in similar studies [35,39]. We analyzed those physiological metrics by comparing the average values along with their specified 95% CIs and relevant statistical tests. To address multiple comparisons in the pairwise analysis of psychological measures, we applied a false discovery rate adjustment to minimize the risk of false positives.

Secondary Outcome

We used physiological metrics (EDA, BT, and HR) to measure stress levels, implementing the pretrained random forest classifier, one of the most widely used machine learning techniques [50], and was developed by Hosseini et al [33]. The classifier was trained on the AffectiveRoad dataset [51], optimized through grid search and cross-validation, and categorized stress into three levels: 0 (no stress), 1 (moderate stress), and 2 (high stress). This classification was based on thresholds from the AffectiveRoad dataset, where S≤0.65 indicated no stress, 0.65≤S≤1.3 indicated moderate stress, and S≥1.3 indicated high stress.

The classifier extracted 48 signal features from a 10-second sliding window of EDA, BT, and HR data, along with contextual features from the preceding 10 windows. While we did not contribute to the classifier's training or the categorization process, we calculated the median stress value for each measurement, leading to occasional intermediate values like 0.5 (between no stress and moderate stress) and 1.5 (between moderate and high stress).

Stress level prediction was determined using the median value. The median was selected because, in a 5-minute session, predicted stress levels could display an imbalanced distribution

XSL•FO
RenderX

(eg, [2, 2, 2, 1, 0]). Using the median was preferred over the mean when predicted values did not adhere to a normal distribution [52]. The Wilcoxon signed-rank test was used to assess the statistical significance of pairwise differences in stress levels at different time points. To address multiple comparisons in the pairwise analysis of psychological measures, we applied a false discovery rate adjustment to minimize the risk of false positives.

We also conducted pairwise comparisons of the MOBICPR game's effect on stress levels during home use, examining positive, negative, and equal ranks to represent how stress levels changed in phase 1 (before performing simulated adult BLS) and phase 3 (after performing simulated adult BLS). A positive rank indicated increased stress (ie, the number of times stress levels increased), negative ranks indicated decreased stress (ie, the number of times stress levels decreased), and equal ranks indicated no change in stress. The Wilcoxon signed-rank test was used to inspect the significant differences in the ranks of stress levels before and after simulated adult BLS.

All tests were 2-tailed, and statistical significance was set at $P<.05$, with P values adjusted where appropriate using the Benjamini-Hochberg false discovery rate procedure.

Ethical Considerations

This study was reviewed and approved by the Slovenian National Medical Ethics Committee (approval number 0120-157/2018) and registered at ClinicalTrials.gov (NCT05784675). Due to the complexity of the registered study protocol, the publication was planned in multiple parts; this study represents one of two main research papers derived from the original design. The study protocol was written in accordance with the CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) checklist [53]. All participants received both verbal and written information about the study and provided written informed consent prior to enrollment. Strict measures were implemented to ensure confidentiality and data protection; all data were anonymized prior to analysis, and no identifying personal information was stored or included in publication materials. Participation was entirely voluntary and not linked to academic performance or evaluation. No images or supplementary materials contain identifiable information about participants, and thus no additional image consent is required. Participants were rewarded for their participation with a free beverage from a vending machine and a printed copy of the Game Changer painting by the street artist Banksy.

Results

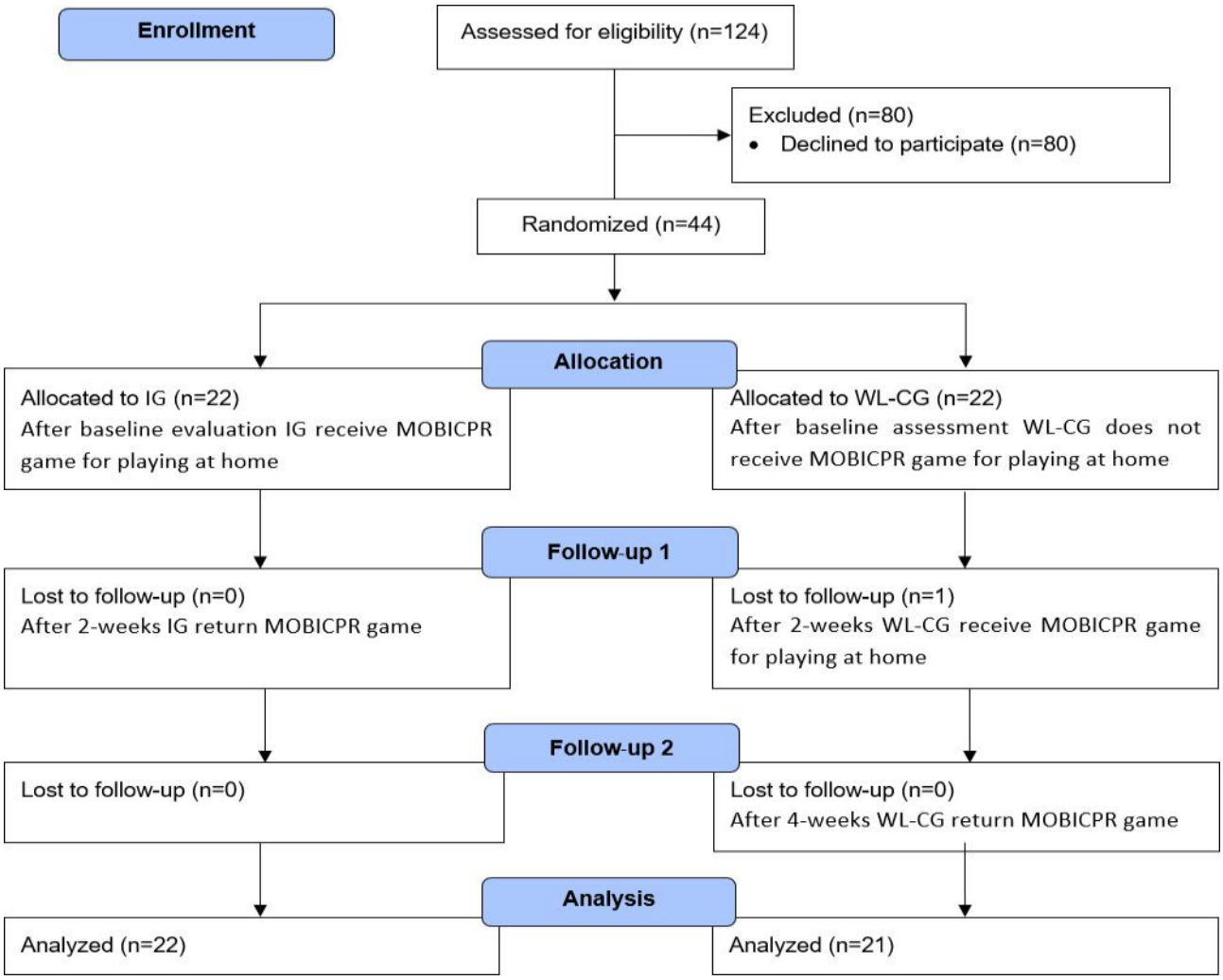
Participant Details

Out of 124 nursing students at the enrollment, 44 (35.5%) participated in the study. At follow-up, 1 nursing student in the WL-CG dropped out. In the end, 43 study participants were included in the final analysis, 22 in IG and 21 in WL-CG (Figure 3).

Most were women (38/43, 88%) with a mean age of 19 (SD 0.6) years. On average, study participants in both groups played the MOBICPR game at home 7 (SD 4) times.

XSL•FO
RenderX

Figure 3. Flow diagram of participant enrollment and group assignment in a randomized waitlist controlled trial with Slovenian nursing students (n=43), evaluating physiological stress responses to serious game-based adult BLS training using Empatica E4 wearables (March-May 2023). BLS: basic life support; IG: intervention group; WL-CG: waitlist control group.



Primary Outcome

First, the average values for the physiological metrics EDA (µS), BVP, HR (beats/minute), and BT (°C) were calculated. BVP represents relative changes in blood flow and is typically expressed without a specific unit [54]. EDA, BVP, and BT show statistically significant changes across measurement phases in both IG and WL-CG, with *P* values below .05 in multiple comparisons. In contrast, HR shows no consistent statistical significance, as most *P* values are above .05, indicating minimal differences between IG and WL-CG in multiple comparisons (Table 1). Overall, there were few statistically significant differences between IG and WL-CG across measurement phases (*P*>.05).

XSL•FO
RenderX

Table . Comparison of physiological metrics (electrodermal activity, blood volume pulse, heart rate, and body temperature) across measurement phases in Slovenian nursing students during a randomized waitlist controlled trial on stress and serious smartphone gaming in adult basic life support training, conducted March-May 2023. All metrics were measured using the Empatica E4 device.

Physiological metrics and measurement (phase)	Intervention group, mean (95% CI)	Waitlist control group, mean (95% CI)	*P* value[a]
Electrodermal activity (µS)			
Baseline			
Baseline (phase 1)	1.69 (0.56 - 2.82)	1.42 (0.47 - 2.36)	.98
Baseline (phase 3)	6.5 (3.3 - 9.69)	7.41 (4.85 - 9.97)	.51
P value[b]	.002	<.001	
2-week follow-up			
2-week follow-up (phase 1)	1.96 (0.45 - 3.46)	1.71 (0.5 - 2.91)	.37
2-week follow-up (phase 3)	6.85 (3.82 - 9.88)	8.54 (5.6 - 11.48)	.34
P value[b]	<.001	<.001	
4-week follow-up			
4-week follow-up (phase 1)	1.89 (-0.06 - 3.85)	0.64 (-0.02 - 1.3)	.84
4-week follow-up (phase 3)	8.01 (3.77 - 12.24)	7.7 (5.28 - 10.13)	.49
P value[b]	<.001	<.001	
Blood volume pulse			
Baseline			
Baseline (phase 1)	30.44 (22.02 - 38.87)	23.07 (16.74 - 29.39)	.29
Baseline (phase 3)	50.62 (37.74 - 63.49)	48.58 (37.21 - 59.94)	>.99
P value[b]	.02	<.001	
2-week follow-up			
2-week follow-up (phase 1)	22.39 (17.05 - 27.73)	23.55 (17.7 - 29.41)	.62
2-week follow-up (phase 3)	44.98 (32.02 - 57.93)	41.27 (32.26 - 50.28)	.97
P value[b]	<.001	<.001	
4-week follow-up			
4-week follow-up (phase 1)	26.7 (20.01 - 33.4)	26.07 (20.62 - 31.53)	.99
4-week follow-up (phase 3)	41.6 (32.76 - 50.43)	45.52 (36.05 - 54.98)	.53
P value[b]	.007	<.001	
Heart rate (beats/minute)			
Baseline			
Baseline (phase 1)	87.53 (81.89 - 93.16)	86.88 (82.03 - 91.73)	.86
Baseline (phase 3)	86.85 (79.41 - 94.29)	88.78 (82.91 - 94.66)	.67
P value[b]	.77	.36	
2-week follow-up			
2-week follow-up (phase 1)	84.05 (78.7 - 89.41)	86.35 (81.59 - 91.11)	.76
2-week follow-up (phase 3)	85.54 (80.49 - 90.59)	92.55 (87.51 - 97.58)	.06
P value[b]	.45	.002	
4-week follow-up			
4-week follow-up (phase 1)	84.59 (79.5 - 89.68)	87.21 (80.72 - 93.69)	.51
4-week follow-up (phase 3)	86.8 (81.7 - 91.91)	92.88 (86.68 - 99.09)	.12

XSL•FO
RenderX

Physiological metrics and measurement (phase)	Intervention group, mean (95% CI)	Waitlist control group, mean (95% CI)	*P* value[a]
P value[b]	.19	.06	
Body temperature (°C)			
Baseline			
Baseline (phase 1)	32.7 (32.27 - 33.13)	31.73 (30.95 - 32.51)	.03
Baseline (phase 3)	33.32 (32.75 - 33.89)	32.73 (31.96 - 33.5)	.20
P value[b]	.005	<.001	
2-week follow-up			
2-week follow-up (phase 1)	32.22 (31.54 - 32.89)	32.71 (32.02 - 33.4)	.30
2-week follow-up (phase 3)	32.94 (32.12 - 33.76)	33.14 (32.34 - 33.94)	>.99
P value[b]	.003	.05	
4-week follow-up			
4-week follow-up (phase 1)	32.52 (31.82 - 33.23)	32.45 (31.38 - 33.52)	.55
4-week follow-up (phase 3)	33.25 (32.43 - 34.06)	33.02 (32.11 - 33.92)	.48
P value[b]	.02	.03	

[a]These *P* values indicate between-group comparisons (intervention group vs waitlist control group) at the same timepoint (baseline or 2-week follow-up or 4-week follow-up).

[b]These *P* values represent within-group comparisons across phases (e.g. Phase 1 vs. Phase 3).

Secondary Outcome

Additionally, stress levels were assessed using the physiological metrics EDA, BT, and HR. In phase 1, the IG showed a significant within-group change between baseline and the 2-week follow-up (*P*=.04). In contrast, no significant changes were observed in the WL-CG. In contrast, no significant changes were found in the WL-CG. Other comparisons between the IG and WL-CG, as well as across follow-up points, did not show significant differences (Table 2).

Table . Average stress level comparisons between pre- and postadult basic life support mandala coloring phases in intervention and control groups across 3 time points (baseline, 2-week, and 4-week follow-ups) in a 2023 study of MOBICPR game use among Slovenian nursing students.

Measurement (phase)	Intervention group (n=22), mean	Waitlist control group (n=21), mean	*P* value[a]
Baseline			
Baseline (phase 1)	0.75	0.74	.96
Baseline (phase 3)	0.71	1.09	.18
P value[b]	.22	.22	
2-week follow-up			
2-week follow-up (phase 1)	0.61	0.57	.69
2-week follow-up (phase 3)	1.11	0.76	.12
P value[b]	.04	.45	
4-week follow-up			
4-week follow-up (phase 1)	0.66	0.45	.26
4-week follow-up (phase 3)	0.68	0.81	.56
P value[b]	.81	.16	

[a]These *P* values represent between-group comparisons at each timepoint.

[b]These *P* values represent within-group phase comparisons.

Table 3 compares the ranks before and after simulated adult BLS training between the IG (n=22) and WL-CG (n=21) at 2 different points: baseline and after 2 weeks of playing the MOBICPR game at home. At baseline, 8 participants in the IG had a negative rank, 7 had a positive rank, and 7 had an equal rank, with a *P* value of 0.77, indicating no significant difference. In the WL-CG, 6 participants had a negative rank, 9 had a positive rank, and 6 had an equal rank, with a *P* value of 0.45,

XSL•FO
RenderX

also showing no significant difference. After 2 weeks of playing the MOBICPR game at home, the IG showed a shift, with 6 participants having a negative rank, 12 having a positive rank, and 4 having an equal rank, with a *P* value of .04, suggesting a significant change. In the WL-CG, 4 participants had a negative rank, 10 had a positive rank, and 7 had an equal rank, with a *P* value of .16, indicating no significant difference.

Table . Pairwise comparison of stress level changes before and after adult basic life support performance, measured before and after 2 weeks of home MOBICPR gameplay in a 2023 randomized trial of Slovenian nursing students using wearable stress data.

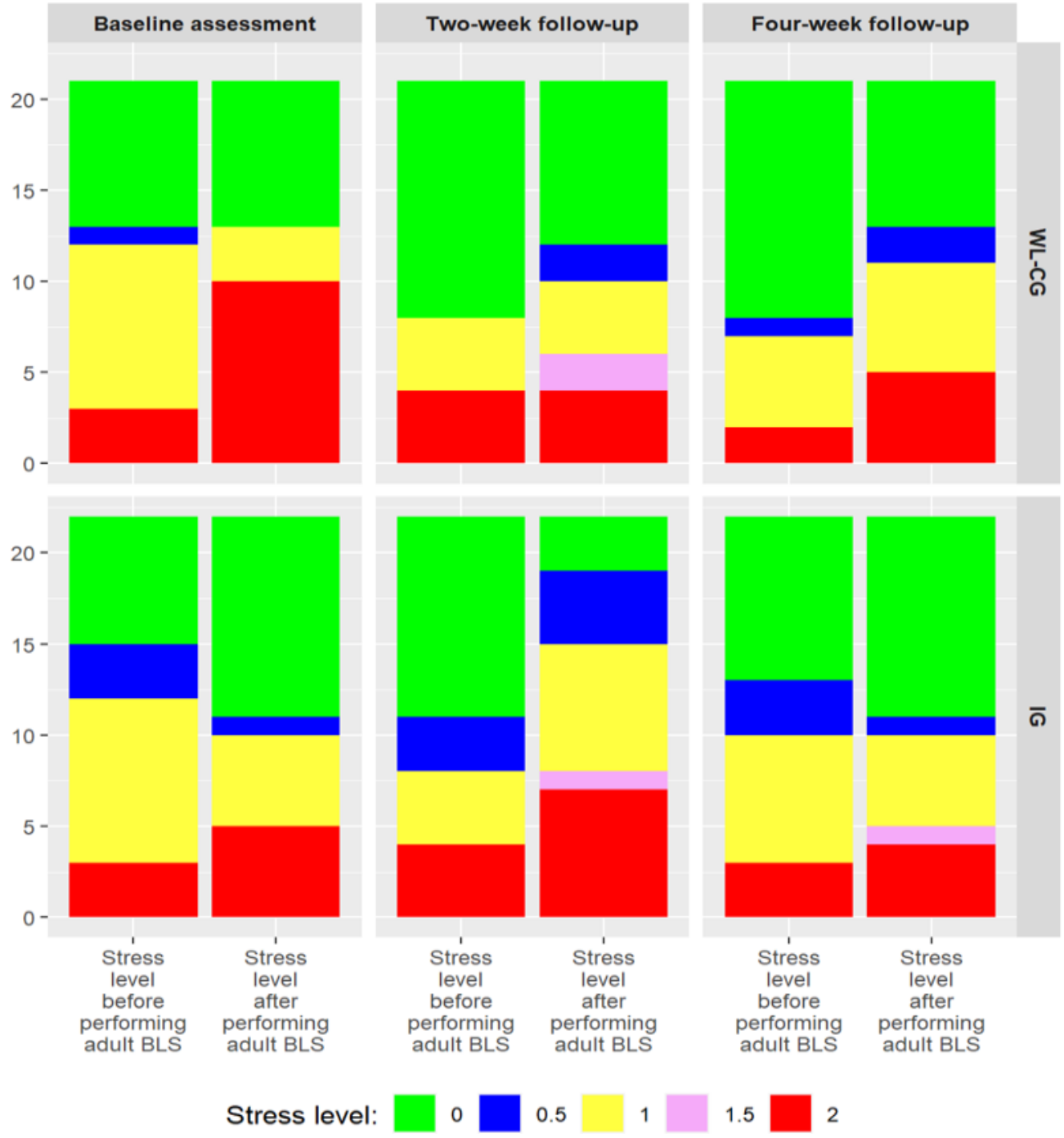
	Intervention group (n=22)		Waitlist control group (n=21)	
Measurement (phase) and rank before and after adult basic life support	Values, n	*P* value	Values, n	*P* value
Baseline		.77		.45
Negative rank	8		6	
Positive rank	7		9	
Equal rank	7		6	
After 2 weeks of playing MOBICPR at home[a]		.04		.16
Negative rank	6		4	
Positive rank	12		10	
Equal rank	4		7	

[a]"After 2 weeks of playing MOBICPR at home" refers to the measurement taken 2 weeks after baseline for the intervention group and 4 weeks after baseline for the waitlist control group, reflecting the point at which each group completed 2 weeks of home gameplay.

At baseline, both groups show a similar distribution of stress levels, with many participants in both groups having stress levels of 0 and 1 before performing simulated adult BLS. There is a noticeable shift in stress levels after performing simulated adult BLS, with more participants reporting higher stress levels (especially stress level 2). During the 2-week follow-up, stress levels before performing simulated adult BLS show a larger distribution of lower stress levels (0 and 0.5) in both groups, but after performing simulated adult BLS, the IG shows a slight shift toward lower stress levels, while the WL-CG still shows a higher proportion of participants experiencing stress level 2. During the 4-week follow-up, stress levels before and after performing simulated adult BLS in the IG appear more balanced, with many participants maintaining lower stress levels (0 and 0.5). In the WL-CG, there is still a shift toward higher stress levels after performing adult BLS, although there is a slight reduction in stress level 2 compared to previous time points (Figure 4).

XSL•FO
RenderX

Figure 4. Stress level distribution before and after simulated adult BLS among nursing students in the intervention and waitlist control groups at baseline, 2-week, and 4-week follow-ups in a 2023 randomized trial using the MOBICPR game and Empatica E4 wearable sensors. BLS: basic life support; IG: intervention group; WL-CG: waitlist control group.



Discussion

Principal Findings

To our knowledge, this study is the first to examine the effects of home-based serious smartphone resuscitation gaming on stress levels, measured using the Empatica E4 device, among nursing students participating in simulated adult BLS. Research using methodologies like ours has mainly focused on assessing cognitive, behavioral, and motivational learning outcomes in educational settings [9-13], while dedicating less attention to the evaluation of stress levels. A 2017 study measured the impact of team leader stress on teams practicing advanced life support in a simulation environment, using the Empatica E4 device and self-reported measures for assessment [55]. A key finding was that EDA levels increased as team performance improved and decreased with poorer performance. This aligns partially with our findings, where EDA, BVP, and BT showed statistically significant changes between groups (IG and WL-CG) and across measurement phases, indicating that these physiological metrics are sensitive to changes over time and may reflect varying levels of stress or engagement. Stress is known to elicit physiological responses, including elevated HR, particularly in simulation environments [56]. Studies conducted in clinical environments with nurses in intensive care using the

XSL•FO
RenderX

Empatica E4 device to measure stress have shown positive correlations between stress and HR, stress and BT, as well as between HR and EDA [33,57,58]. Similarly, elevated HR has also been observed in simulation environments [56]. Our findings regarding HR revealed no consistent statistically significant changes across measurement phases or between groups, which contrasts with prior studies conducted in both clinical and simulation environments. For example, Ahmadi et al [57] and Zhang et al [58] demonstrated positive correlations between HR and stress among intensive care unit nurses during actual patient care, and Harvey et al [56] found similar results in simulated trauma scenarios. These discrepancies may stem from differences in context, measurement duration, and task complexity. While those studies involved longer monitoring periods (up to 12 hours) and real or high-fidelity simulations with greater cognitive load, our study involved short-term simulations (~10 minutes) using manikins and a controlled environment. Furthermore, HR is known to be influenced by various confounding factors such as baseline fitness, breathing patterns, and individual stress perception, which may reduce its sensitivity in short-duration tasks like ours. Thus, while EDA, BVP, and BT showed consistent responses, HR alone may not be a reliable standalone indicator of acute stress in brief, low-to-moderate intensity simulation-based tasks. These contrasting results highlight the importance of multimodal stress measurement in educational research, especially when using wearables in short simulated scenarios. Regarding BT, measured on participants' forearms, our study reported similar values to those observed in a study involving 3 different cognitive load tasks [59], with temperatures ranging between 31 and 33 °C. These results align with the normal regional variation of human skin temperature, which ranges from 29.8 (SD 1.6) °C on the sole to 34.7 (SD 0.6) °C on the neck and 32.6 (SD 0.9) °C on the forearm [60]. Both studies measured temperatures in similar environments; however, they used a digital thermometer, while we used the Empatica E4 device.

The term "stress paradox" refers to the phenomenon where exposure to a task induces stress, which, counterintuitively, can have a positive effect (eustress) on the learning curve by enhancing performance and fostering growth. In our study, this paradox is illustrated by the physiological response patterns observed. EDA, BVP, and BT showed changes across measurement phases, indicating engagement and potential arousal during simulated adult BLS performance. Interestingly, despite participants playing the MOBICPR game at home an average of less than 8 times, stress levels did not change significantly at the group level, as shown in both absolute values and machine learning–derived stress classifications. However, performance in the simulated adult BLS scenario improved over time, particularly in theoretical knowledge retention as reported in our earlier publication [43]. Moreover, we observed a statistically significant shift in stress response within the IG after 2 weeks of gameplay: 12 participants showed reduced stress after performing simulated adult BLS compared to before. This pattern supports the idea that controlled exposure to stress via serious gaming may lead to adaptive physiological responses without overwhelming distress, aligning with prior research that Slovenian nurses often manage stress well during resuscitation due to repeated simulation training [61,62]. Thus, even in the absence of reduced stress at a group level, individual adaptation and performance improvement hint at the beneficial effect of eustress within the framework of the stress paradox. That feeling of eustress benefits learning and leads to "stress-related growth" [63].

Gamification, incorporated into a serious game, allows users to learn and retain content through mobile learning, as it represents a familiar educational environment for the modern user [64]. However, the open market currently provides a limited selection of serious smartphone games specifically tailored for teaching adults [65]. Moreover, several serious smartphone games used in recent studies are no longer available [9-13,66]. By developing the MOBICPR game, we aim to address the current scarcity of accessible serious smartphone games for teaching adult BLS by offering a freely available tool for both mobile [67] and web-based learning [18,68]. The MOBICPR game will be freely available in 2026 following the release of the new European Resuscitation Council adult BLS guidelines. It will be published on the Google Play Store and can also be requested directly from the corresponding author as an Android (.apk) file. Dissemination will include collaboration with academic institutions, promotion at conferences, participation in innovation competitions and award submissions, as well as targeted outreach via social media platforms. Initially, the game will be in English and compatible with Android devices. Language localization and iOS support are planned for future updates, pending available resources. Additionally, our approach incorporates heutagogy, where learners have the autonomy to shape their own learning experience, a method that promises to redefine future educational practices. By integrating gamification features into the MOBICPR game, our goal is for users to engage with it more frequently, thereby enhancing the retention of knowledge over a longer period [43,47]. Additionally, a 2023 study suggests that playing console games for approximately 20 hours over 14 days can enhance endoscopic skills [69]. In contrast, our participants played an average of 7 times during the same period, totaling about 45 minutes. This reduced gameplay duration could potentially impact their level of stress. Although immersive educational methods such as virtual reality hold promise for impacting stress levels, recent research on both adult and pediatric basic and advanced life support presents mixed outcomes, indicating the need for further studies [30,70].

A recent review identified several stress mitigators that could influence stress and decision-making during resuscitation, including cognitive aids such as checklists, stress management training, and meditation [71]. Specifically, meditation involving the use of mandalas, like our approach, may support health care professionals in managing stress. Recent studies suggest that mandalas are linked to mindfulness practices, helping nurses working in critical care to calm the mind and reduce perceived stress levels [42]. Similarly, activities like Sudoku or daydreaming can also have a similar effect on reducing stress [59].

Limitations

Our studies have several limitations. First, although we invited all nursing students enrolled in the 2022/2023 academic year to participate, only 44 chose to enroll, and 1 dropped out,

XSL•FO
RenderX

resulting in a final sample size of 43 participants. While this sample size is comparable to other studies [33,57,58] using the Empatica E4 device for stress assessment, it remains relatively small, which may limit the statistical power and generalizability of our findings. No formal a priori power analysis was conducted due to the exploratory nature of this study and the absence of comparable effect sizes from similar interventions. Future research should include a formal power calculation and a larger, more diverse sample to validate and extend the results presented here. Looking ahead, we could engage a younger audience to determine if the MOBICPR game is effective in teaching them adult BLS content at home prior to participating in actual training sessions. Additionally, the sample size of participants could have been larger if we had the ability to offer incentives, as is done at other universities [72-74]. Despite this, our retention rate remained steady, with no participants dropping out as the study progressed. Second, we have only one Empatica E4 device, which influences the process of collecting data. It happened that event marking was inconsistent, causing event mark times to appear either too close to each other or separated from each other to a large extent. This was resolved by eliminating intermittent tags if the time duration between 2 event marks did not exceed 4.5 minutes. We allowed a 30-second variation since recording the event can hardly be done to a second precisely. This means the first event mark indicated the beginning of the first phase, followed by an event mark recorded at least 4.5 minutes later, indicating the time span of the first phase. Third, due to the limitations of the Empatica E4 device, we were unable to record data during phase 2, where simulated adult BLS was performed, because of hand movements. A more convenient way to accurately measure stress during physical activity may involve the use of electroencephalogram headbands [75] or wireless electrocardiogram monitors [76]. With advancements in artificial intelligence (ie, deep learning), stress detection based on facial expressions has also become increasingly feasible [77]. Future studies could explore these alternative technologies to enhance data collection during physically active scenarios. Fourth, having study participants undertake the same simulated adult BLS scenario may have made them more familiar with it, thereby potentially reducing their stress levels due to repetition and anticipation of what was coming next. Similarly, the limited content and structure of the MOBICPR game may have contributed to decreased cognitive load over time. Although this familiarity reflects the positive effects of repeated practice, it may also have muted the variability in stress responses we aimed to capture. Adding more diverse and randomized scenarios could help reduce this effect, but such improvements were not feasible within the scope of this study due to financial constraints. Fifth, many of the study participants were female, which could introduce gender bias and influence the results.

Conclusions

Our study has shown that playing a serious smartphone game such as MOBICPR game at home can have an impact on stress levels before and after performing adult BLS. The application of serious gaming and its incorporation into the educational curriculum might also be considered more widely. That might give users the opportunity to practice in a safe environment like home, without risk of harm, and empower them for tomorrow's challenges.

Acknowledgments

The authors would like to thank all nursing students who participated in this study. NF, LK, and GS are supported by Slovenian Research Agency grants ARRS P2-0057, ARRS N3-0307, and ARRS BI-US/22-24-138, and NextGenerationEU and MVZI (C3330-22-953012).

Data Availability

The data that support the findings of this study are not publicly available due to privacy and ethical restrictions. Further inquiries can be directed to the corresponding author.

Authors' Contributions

The study presented here was carried out through collaboration among all authors. NF developed a study design and supervised the study. NF, LK, GS, and ŠM drafted the manuscript. NF, GS, LK, and ŠM conducted data collection and analysis. MS, PS, BSA, and RG conducted a comprehensive content review.

Conflicts of Interest

NF is a member of the European Resuscitation Council Basic Life Support Science and Education Committee. RG is European Resuscitation Council Director of Guidelines and International Liaison Committee on Resuscitation, and International Liaison Committee on Resuscitation Task Force chair for Education Implementation and Team. Other authors declare that they have no conflict of interest.

Checklist 1
CONSORT-EHEALTH checklist (v 1.6.1).
[PDF File, 1236 KB - games_v13i1e67623_app1.pdf]

XSL•FO
RenderX

References

1. Vincent A, Semmer NK, Becker C, et al. Does stress influence the performance of cardiopulmonary resuscitation? A narrative review of the literature. J Crit Care 2021 Jun;63:223-230. [doi: 10.1016/j.jcrc.2020.09.020] [Medline: 33046274]
2. Oliveira Silva G, Oliveira FSE, Coelho ASG, et al. Influence of simulation design on stress, anxiety and self-confidence of nursing students: systematic review with meta-analysis. J Clin Nurs 2023 Sep;32(17-18):5668-5692. [doi: 10.1111/jocn.16681] [Medline: 36894868]
3. Snoswell AJ, Snoswell CL. Immersive virtual reality in health care: systematic review of technology and disease states. JMIR Biomed Eng 2019 Sep;4(1):e15025. [doi: 10.2196/15025]
4. Ghoman SK, Schmölzer GM. The RETAIN simulation-based serious game—a review of the literature. Healthcare (Basel) 2020 Mar;8(1):3. [doi: 10.3390/healthcare8010003]
5. Bergeron B. Developing Serious Games: Charles River Media, Inc; 2006.
6. Gorbanev I, Agudelo-Londoño S, González RA, et al. A systematic review of serious games in medical education: quality of evidence and pedagogical strategy. Med Educ Online 2018 Dec;23(1):1438718. [doi: 10.1080/10872981.2018.1438718] [Medline: 29457760]
7. Ahrens D. Serious games – a new perspective on workbased learning. Proc Soc Behav Sci 2015 Aug;204:277-281. [doi: 10.1016/j.sbspro.2015.08.152]
8. Biggs A, Brough P, Drummond S. Lazarus and Folkman's psychological stress and coping theory. In: Cooper CL, Quick JC, editors. The Handbook of Stress and Health: A Guide to Research and Practice, 1st edition: Wiley; 2017:349-364. [doi: 10.1002/9781118993811]
9. de Sena DP, Fabrício DD, da Silva VD, Bodanese LC, Franco AR. Comparative evaluation of video-based on-line course versus serious game for training medical students in cardiopulmonary resuscitation: a randomised trial. PLOS One 2019 Aug;14(4):e0214722. [doi: 10.1371/journal.pone.0214722]
10. Arif A, Arif A, Fasciglione KA, Jafri FN. The iterative development and testing of an interactive mobile application for skill retention of cardiopulmonary resuscitation among high school students: a pilot study. Simul Gaming 2021 Oct;52(5):650-663. [doi: 10.1177/10468781211032692]
11. Aksoy E. Comparing the effects on learning outcomes of tablet-based and virtual reality-based serious gaming modules for basic life support training: randomized trial. JMIR Serious Games 2019 May 1;7(2):e13442. [doi: 10.2196/13442] [Medline: 31042153]
12. Farsi Z, Yazdani M, Butler S, Nezamzadeh M, Mirlashari J. Comparative effectiveness of simulation versus serious game for training nursing students in cardiopulmonary resuscitation: a randomized control trial. Int J Comput Games Technol 2021 Mar 24;2021:6695077. [doi: 10.1155/2021/6695077]
13. Demirtas A, Basak T, Sahin G, Sonkaya M. The serious game and integrated simulator for cardiopulmonary resuscitation training in nursing students. Simul Gaming 2022 Apr;53(2):97-110. [doi: 10.1177/10468781211073162]
14. Flato UAP, Beffa dos Santos EJ, Bispo Diaz T Martins I, et al. Interactive serious game to teach basic life support among schoolchildren in Brazil: design and rationale. JMIR Serious Games 2024 Oct 9;12(1):e55333. [doi: 10.2196/55333] [Medline: 39382546]
15. Latif MH, Ajmal M, Ahmad F, Alam J, Saleem A. LA-VIE: a serious game for cardiopulmonary resuscitation. In: 2017 IEEE 5th International Conference on Serious Games and Applications for Health (SeGAH): IEEE; 2017:1-5. [doi: 10.1109/SeGAH.2017.7939297]
16. Akaltan KF, Önder C, Vural Ç, Orhan K, Akdoğan N, Atakan C. The effect of game-based learning on basic life support skills training for undergraduate dental students. J Dent Educ 2023 Oct;87(10):1458-1468. [doi: 10.1002/jdd.13303] [Medline: 37400112]
17. Cheng P, Huang Y, Yang P, et al. The effects of serious games on cardiopulmonary resuscitation training and education: systematic review with meta-analysis of randomized controlled trials. JMIR Serious Games 2024 Feb 6;12(1):e52990. [doi: 10.2196/52990] [Medline: 38319697]
18. Phungoen P, Promto S, Chanthawatthanarak S, et al. Precourse preparation using a serious smartphone game on advanced life support knowledge and skills: randomized controlled trial. J Med Internet Res 2020 Mar 9;22(3):e16987. [doi: 10.2196/16987] [Medline: 32149711]
19. Boada I, Rodriguez Benitez A, Thió-Henestrosa S, Soler J. A serious game on the first-aid procedure in choking scenarios: design and evaluation study. JMIR Serious Games 2020 Aug 19;8(3):e16655. [doi: 10.2196/16655] [Medline: 32812882]
20. Moon SH, Kim SO. Enhancing triage accuracy in emergency nurses: the impact of a game-based triage educational app. Int Emerg Nurs 2024 Feb;72:101398. [doi: 10.1016/j.ienj.2023.101398] [Medline: 38198949]
21. Zhu S, Li Z, Sun Y, et al. A serious game for enhancing rescue reasoning skills in tactical combat casualty care: development and deployment study. JMIR Form Res 2024 Aug 12;8(1):e50817. [doi: 10.2196/50817] [Medline: 39133911]
22. Kowalewski KF, Hendrie JD, Schmidt MW, et al. Validation of the mobile serious game application Touch Surgery for cognitive training and assessment of laparoscopic cholecystectomy. Surg Endosc 2017 Oct;31(10):4058-4066. [doi: 10.1007/s00464-017-5452-x] [Medline: 28281111]
23. Sharifzadeh N, Tabesh H, Kharrazi H, et al. Play and learn for surgeons: a serious game to educate medical residents in uterine artery ligation surgery. Games Health J 2021 Aug;10(4):220-227. [doi: 10.1089/g4h.2020.0220] [Medline: 34264757]

XSL•FO
RenderX

24. IJgosse W, van Goor H, Rosman C, Luursema JM. Construct validity of a serious game for laparoscopic skills training: validation study. JMIR Serious Games 2020 May 7;8(2):e17222. [doi: 10.2196/17222] [Medline: 32379051]
25. Min A, Min H, Kim S. Effectiveness of serious games in nurse education: a systematic review. Nurse Educ Today 2022 Jan;108:105178. [doi: 10.1016/j.nedt.2021.105178] [Medline: 34717098]
26. Demircan B, Kıyak Y, Kaya H. The effectiveness of serious games in nursing education: a meta-analysis of randomized controlled studies. Nurse Educ Today 2024 Nov;142:106330. [doi: 10.1016/j.nedt.2024.106330] [Medline: 39128400]
27. Nabecker S, Huwendiek S, Theiler L, Huber M, Petrowski K, Greif R. The effective group size for teaching cardiopulmonary resuscitation skills – a randomized controlled simulation trial. Resuscitation 2021 Aug;165:77-82. [doi: 10.1016/j.resuscitation.2021.05.034] [Medline: 34107336]
28. Ghazali DA, Ragot S, Breque C, et al. Randomized controlled trial of multidisciplinary team stress and performance in immersive simulation for management of infant in shock: study protocol. Scand J Trauma Resusc Emerg Med 2016 Mar 25;24(1):36. [doi: 10.1186/s13049-016-0229-0] [Medline: 27012938]
29. Pourmohammadi S, Maleki A. Stress detection using ECG and EMG signals: a comprehensive study. Comput Methods Programs Biomed 2020 Sep;193:105482. [doi: 10.1016/j.cmpb.2020.105482] [Medline: 32408236]
30. Chang TP, Beshay Y, Hollinger T, Sherman JM. Comparisons of stress physiology of providers in real-life resuscitations and virtual reality-simulated resuscitations. Simul Healthc 2019 Apr;14(2):104-112. [doi: 10.1097/SIH.0000000000000356] [Medline: 30730471]
31. Tramèr L, Becker C, Hochstrasser S, Marsch S, Hunziker S. Association of electrocardiogram alterations of rescuers and performance during a simulated cardiac arrest: a prospective simulation study. PLOS One 2018;13(6):e0198661. [doi: 10.1371/journal.pone.0198661] [Medline: 29902264]
32. Dias RD, Scalabrini Neto A. Acute stress in residents during emergency care: a study of personal and situational factors. Stress 2017 May;20(3):241-248. [doi: 10.1080/10253890.2017.1325866] [Medline: 28462597]
33. Hosseini S, Gottumukkala R, Katragadda S, et al. A multimodal sensor dataset for continuous stress detection of nurses in a hospital. Sci Data 2022 Jun 1;9(1):255. [doi: 10.1038/s41597-022-01361-y] [Medline: 35650267]
34. Crosswell AD, Lockwood KG. Best practices for stress measurement: how to measure psychological stress in health research. Health Psychol Open 2020;7(2). [doi: 10.1177/2055102920933072] [Medline: 32704379]
35. Campanella S, Altaleb A, Belli A, Pierleoni P, Palma L. A method for stress detection using Empatica E4 bracelet and machine-learning techniques. Sensors (Basel) 2023 Jan;23(7):3565. [doi: 10.3390/s23073565]
36. Gjoreski M, Gjoreski H, Luštrek M, Gams M. Continuous stress detection using a wrist device: in laboratory and real life. In: UbiComp '16: The 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Association for Computing Machinery; 2016:1185-1193. [doi: 10.1145/2968219.2968306]
37. Ollander S, Godin C, Campagne A, Charbonnier S. A comparison of wearable and stationary sensors for stress detection. In: 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC): IEEE; 2016:004362-004366. [doi: 10.1109/SMC.2016.7844917]
38. Chandra V, Priyarup A, Sethia D. Comparative study of physiological signals from Empatica E4 wristband for stress classification. In: Singh M, Tyagi V, Gupta PK, Flusser J, Ören T, Sonawane VR, editors. Advances in Computing and Data Sciences: Springer; 2021:218-229. [doi: 10.1007/978-3-030-88244-0_21]
39. Menghini L, Gianfranchi E, Cellini N, Patron E, Tagliabue M, Sarlo M. Stressing the accuracy: wrist-worn wearable sensor validation over different conditions. Psychophysiology 2019 Nov;56(11):e13441. [doi: 10.1111/psyp.13441] [Medline: 31332802]
40. Umair M, Chalabianloo N, Sas C, Ersoy C. HRV and stress: a mixed-methods approach for comparison of wearable heart rate sensors for biofeedback. IEEE Access 2021;9:14005-14024. [doi: 10.1109/ACCESS.2021.3052131]
41. Noble M. Adult Coloring Celestial Mandalas Coloring Book: Dover Publications; 2016. URL: https://www.goodreads.com/book/show/27778040-adult-coloring-celestial-mandalas-coloring-book [accessed 2025-08-08]
42. Fong JSY, Hui ANN, Ho KM, Chan AKM, Lee A. Brief mindful coloring for stress reduction in nurses working in a Hong Kong hospital during COVID-19 pandemic: a randomized controlled trial. Medicine (Baltimore) 2022 Oct 28;101(43):e31253. [doi: 10.1097/MD.0000000000031253] [Medline: 36316873]
43. Fijačko N, Masterson Creber R, Metličar Š, et al. Effects of a serious smartphone game on nursing students' theoretical knowledge and practical skills in adult basic life support: randomized wait list-controlled trial. JMIR Serious Games 2024 Apr 5;12:e56037. [doi: 10.2196/56037] [Medline: 38578690]
44. Neset A, Birkenes TS, Furunes T, et al. A randomized trial on elderly laypersons' CPR performance in a realistic cardiac arrest simulation. Acta Anaesthesiol Scand 2012 Jan;56(1):124-131. [doi: 10.1111/j.1399-6576.2011.02566.x] [Medline: 22092097]
45. Fijačko N, Greif R, Metličar Š, et al. A five-finger mnemonic for teaching schoolchildren the theoretical components of adult basic life support: a modified reactive Delphi-guided development and memorability pilot test with schoolchildren. BMC Med Educ 2024 Dec 30;24(1):1554. [doi: 10.1186/s12909-024-06519-3] [Medline: 39736674]
46. Olasveengen TM, Semeraro F, Ristagno G, et al. European Resuscitation Council Guidelines 2021: basic life support. Resuscitation 2021 Apr;161:98-114. [doi: 10.1016/j.resuscitation.2021.02.009] [Medline: 33773835]

XSL•FO
RenderX

47. Fijačko N, Metličar Š, Vinojčić D, Greif R, Masterson Creber R. Learn how to save the life of a victim of out-of-hospital cardiac arrest by playing a serious smartphone game. Resuscitation 2023 Jul;188:109859. [doi: 10.1016/j.resuscitation.2023.109859] [Medline: 37270089]
48. de Looff P, Duursma R, Noordzij M, et al. Wearables: an R package with accompanying Shiny application for signal analysis of a wearable device targeted at clinicians and researchers. Front Behav Neurosci 2022;16:856544. [doi: 10.3389/fnbeh.2022.856544] [Medline: 35813597]
49. Villanueva RAM, Chen ZJ. ggplot2: Elegant Graphics for Data Analysis (2nd ed.) [Book review]. Measurement 2019 Jul 3;17(3):160-167. [doi: 10.1080/15366367.2019.1565254]
50. Stiglic G, Kocbek P, Fijacko N, Zitnik M, Verbert K, Cilar L. Interpretability of machine learning‐based prediction models in healthcare. WIREs Data Min Knowl 2020 Sep;10(5):e1379. [doi: 10.1002/widm.1379]
51. Haouij NE, Poggi JM, Sevestre-Ghalila S, Ghozi R, Jaïdane M. AffectiveROAD system and database to assess driver's attention. In: SAC '18: Proceedings of the 33rd Annual ACM Symposium on Applied Computing: Association for Computing Machinery; 2018:800-803. [doi: 10.1145/3167132.3167395]
52. Wang HC, Lin YJ, Shih FY, et al. The role of serial oxidative stress levels in acute traumatic brain injury and as predictors of outcome. World Neurosurg 2016 Mar;87:463-470. [doi: 10.1016/j.wneu.2015.10.010] [Medline: 26481337]
53. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126. [doi: 10.2196/jmir.1923] [Medline: 22209829]
54. E4 data - BVP expected signal. Empatica Inc. URL: https://web.archive.org/web/20240414084722/https://support.empatica.com/hc/en-us/articles/360029719792-E4-data-BVP-expected-signal [accessed 2025-8-22]
55. van Sas J. The effect of team leader stress on teams practicing cardiopulmonary resuscitation in a simulation room: an exploratory study into the effect of team leader stress on team leader behaviour, closed-loop communication, and team performance of a simulated medical emergency team [Master's thesis]. : University of Twente; 2017 URL: https://web.archive.org/web/20240622085449/https://essay.utwente.nl/73881/ [accessed 2025-08-08]
56. Harvey A, Bandiera G, Nathens AB, LeBlanc VR. Impact of stress on resident performance in simulated trauma scenarios. J Trauma Acute Care Surg 2012 Feb;72(2):497-503. [doi: 10.1097/ta.0b013e31821f84be] [Medline: 22439221]
57. Ahmadi N, Sasangohar F, Nisar T, et al. Quantifying occupational stress in intensive care unit nurses: an applied naturalistic study of correlations among stress, heart rate, electrodermal activity, and skin temperature. Hum Factors 2022 Feb;64(1):159-172. [doi: 10.1177/00187208211040889] [Medline: 34478340]
58. Zhang Q, Sasangohar F, Saravanan P, et al. Exploring cognitive stress-related tasks of intensive care unit nurses – a comparison study. Proc Hum Factors Ergon Soc Annu Meet 2023 Sep;67(1):2205-2209. [doi: 10.1177/21695067231192711]
59. Romine WL, Schroeder NL, Graft J, et al. Using machine learning to train a wearable device for measuring students' cognitive load during problem-solving activities based on electrodermal activity, body temperature, and heart rate: development of a cognitive load tracker for both personal and classroom use. Sensors (Basel) 2020 Aug 27;20(17):4833. [doi: 10.3390/s20174833] [Medline: 32867055]
60. Lee CM, Jin SP, Doh EJ, Lee DH, Chung JH. Regional variation of human skin surface temperature. Ann Dermatol 2019 Jun;31(3):349-352. [doi: 10.5021/ad.2019.31.3.349] [Medline: 33911607]
61. Koželj A, Šikić Pogačar M, Fijan S, Strauss M, Poštuvan V, Strnad M. Exploring the feelings of nurses during resuscitation—a cross-sectional study. Healthcare (Basel) 2022 Jan;10(1):5. [doi: 10.3390/healthcare10010005]
62. Koželj A, Strauss M, Strnad M. Experiencing stress among nurses during resuscitation procedure and legal obligations for healthcare workers in Slovenia. Med Law Soc 2020 Oct;13(2):237-262. [doi: 10.18690/mls.13.2.237-262.2020]
63. Rudland JR, Golding C, Wilkinson TJ. The stress paradox: how stress can be good for learning. Med Educ 2020 Jan;54(1):40-45. [doi: 10.1111/medu.13830]
64. Dichev C, Dicheva D. Gamifying education: what is known, what is believed and what remains uncertain: a critical review. Int J Educ Technol High Educ 2017 Dec;14:9. [doi: 10.1186/s41239-017-0042-5]
65. Fijačko N, Masterson Creber R, Gosak L, et al. Evaluating quality, usability, evidence-based content, and gamification features in mobile learning apps designed to teach children basic life support: systematic search in app stores and content analysis. JMIR Mhealth Uhealth 2021 Jul 20;9(7):e25437. [doi: 10.2196/25437] [Medline: 34283034]
66. Binotti M, Cavallin F, Ingrassia PL, et al. Heart rate assessment using NeoTapAdvancedSupport: a simulation study. Arch Dis Child Fetal Neonatal Ed 2019 Jul;104(4):F440-F442. [doi: 10.1136/archdischild-2018-315408] [Medline: 30262463]
67. Marsh-Armstrong BP, Seng E, Ting-Wei F, Saka S, Greenberg M. Effectiveness of Rescue Me CPR! smartphone app providing real-time guidance to untrained bystanders performing CPR. Heliyon 2023 Oct;9(10):e20908. [doi: 10.1016/j.heliyon.2023.e20908] [Medline: 37867873]
68. Chen TS, Hsieh PL, Tung CC, Wu CH, Cheng YC. Evaluation of registered nurses' interprofessional emergency care competence through the gamification of cardiopulmonary resuscitation training: a cross-sectional study. BMC Med Educ 2023 May 22;23(1):359. [doi: 10.1186/s12909-023-04332-y] [Medline: 37217994]
69. Gugura R, Fischer P, Tan ău M, Tefas C. Just five more minutes, mom: why video games could make you a better endoscopist. Surg Endosc 2023 Sep;37(9):6901-6907. [doi: 10.1007/s00464-023-10167-x] [Medline: 37316677]

XSL•FO
RenderX

70. Alcázar Artero PM, Pardo Rios M, Greif R, et al. Efficiency of virtual reality for cardiopulmonary resuscitation training of adult laypersons: a systematic review. Medicine (Baltimore) 2023 Jan 27;102(4):e32736. [doi: 10.1097/MD.0000000000032736] [Medline: 36705392]
71. Groombridge CJ, Kim Y, Maini A, Smit DV, Fitzgerald MC. Stress and decision-making in resuscitation: a systematic review. Resuscitation 2019 Nov;144:115-122. [doi: 10.1016/j.resuscitation.2019.09.023] [Medline: 31562904]
72. Tanaka M, Ishii S, Matsuoka A, et al. Perspectives of Japanese elders and their healthcare providers on use of wearable technology to monitor their health at home: a qualitative exploration. Int J Nurs Stud 2024 Apr;152:104691. [doi: 10.1016/j.ijnurstu.2024.104691] [Medline: 38262231]
73. Missel AL, Donnelly JP, Tsutsui J, et al. Effectiveness of lay bystander hands-only cardiopulmonary resuscitation on a mattress versus the floor: a randomized cross-over trial. Ann Emerg Med 2023 Jun;81(6):691-698. [doi: 10.1016/j.annemergmed.2023.01.012] [Medline: 36841661]
74. Gasteiger N, van der Veer SN, Wilson P, Dowding D. Virtual reality and augmented reality smartphone applications for upskilling care home workers in hand hygiene: a realist multi-site feasibility, usability, acceptability, and efficacy study. J Am Med Inform Assoc 2023 Dec 22;31(1):45-60. [doi: 10.1093/jamia/ocad200]
75. Affanni A, Aminosharieh Najafi T, Guerci S. Development of an EEG headband for stress measurement on driving simulators. Sensors (Basel) 2022 Jan;22(5):1785. [doi: 10.3390/s22051785]
76. Li X, Zhu W, Sui X, Zhang A, Chi L, Lv L. Assessing workplace stress among nurses using heart rate variability analysis with wearable ECG device – a pilot study. Front Public Health 2022;9. [doi: 10.3389/fpubh.2021.810577]
77. Giannakakis G, Koujan MR, Roussos A, Marias K. Automatic stress analysis from facial videos based on deep facial action units recognition. Pattern Anal Applic 2022 Aug;25(3):521-535. [doi: 10.1007/s10044-021-01012-9]

Abbreviations

BLS: basic life support
BT: body temperature
BVP: blood volume pulse
CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
CPR: cardiopulmonary resuscitation
EDA: electrodermal activity
HR: heart rate
IG: intervention group
OHCA: out-of-hospital cardiac arrest
WL-CG: waitlist control group

Edited by A Coristine; submitted 18.10.24; peer-reviewed by A Gazis, D Kitapçıoğlu; revised version received 31.05.25; accepted 01.06.25; published 29.08.25.

Please cite as:
Fijačko N, Abella BS, Metličar Š, Kopitar L, Greif R, Štiglic G, Skok P, Strnad M
Exploring the Impact of Home-Based Serious Smartphone Resuscitation Gaming on Stress Among Nursing Students Practicing Simulated Adult Basic Life Support: Randomized Waitlist Controlled Trial
JMIR Serious Games 2025;13:e67623
URL: https://games.jmir.org/2025/1/e67623
doi:10.2196/67623

© Nino Fijačko, Benjamin S Abella, Špela Metličar, Leon Kopitar, Robert Greif, Gregor Stiglic, Pavel Skok, Matej Strnad. Originally published in JMIR Serious Games (https://games.jmir.org), 29.8.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Virtual Reality for Workplace Violence Training of Health Care Workers: Pilot Mixed Methods Usability Study

Daniel Isaiah Jackson[1], BSc; Thipkanok Wongphothiphan[2], PhD; John Luna[2], MFA; Tensing Maa[3,4], MD; Mary A Fristad[5,6], PhD; Yungui Huang[2], MBA, PhD; Brittany Schaffner[5,6], MEd; Jennifer Reese[5,6], PsyD; Jason Wheeler[2], AS; Brandon Abbott[2], BA; Emre Sezgin[1,4], PhD

[1]Center for Biobehavioral Health, Abigail Wexner Research Institute, Nationwide Children's Hospital, 431 S 18th Street, Columbus, OH, United States

[2]IT Research and Innovation, Abigail Wexner Research Institute, Nationwide Children's Hospital, Columbus, OH, United States

[3]Center for Clinical Excellence, Nationwide Children's Hospital, Columbus, OH, United States

[4]College of Medicine, The Ohio State University, Columbus, OH, United States

[5]Division of Child and Family Psychiatry, Nationwide Children's Hospital, Columbus, OH, United States

[6]Department of Psychiatry and Behavioral Health, The Ohio State University, Columbus, OH, United States

Corresponding Author:
Emre Sezgin, PhD
Center for Biobehavioral Health, Abigail Wexner Research Institute, Nationwide Children's Hospital, 431 S 18th Street, Columbus, OH, United States

Abstract

Background: Workplace violence (WPV) is a growing concern in health care, adversely impacting frontline providers, patients, and visitors. Traditional training programs have demonstrated limited long-term effectiveness in equipping health care professionals with de-escalation and crisis management skills. Virtual reality (VR) may offer an opportunity to create an innovative, immersive, and engaging platform for WPV training that could address the limitation of conventional methods.

Objective: The aim of this pilot usability study was to assess the user experience of a prototype VR training course designed to prepare frontline health care staff for WPV scenarios. We evaluated the VR system's practicality, engagement, and perceived value and identified areas for improvement.

Methods: A cross-sectional, mixed methods study was conducted with 13 frontline health care providers. Four pilot-training modules were developed and deployed in a VR environment on stand-alone headsets to address a variety of topics around WPV: Situational Awareness, Self-Awareness and Self-Regulation, Team Dynamics, and Evasive Maneuvers. Participants engaged with each module while providing qualitative feedback during the training. Qualitative feedback was analyzed using a rapid qualitative analysis technique. After completing the pilot training courses, participants completed surveys on usability (System Usability Scale) and user experience (mini Player Experience Inventory) and shared first impressions via Reaction Cards.

Results: Participants found the pilot VR training to be engaging (mini Player Experience Inventory; mean 5.23, SD 1.34), with 89% of Reaction Card responses reflecting positive impressions such as "valuable," "creative," and "accessible." However, the overall System Usability Scale score (mean 63.30, SD 9.53) indicated room for improvement in usability. Although participants identified the VR system as manageable and intuitive, first-time users experienced challenges navigating the virtual environment. We identified four themes from qualitative feedback: (1) Perceived Value, (2) Technical and Navigational Barriers, (3) User Preferences, and (4) Vision. Participants described the VR training modules as refreshing due to the immersion in complex environments and noted areas for improvements in the tone and emotional expressiveness of nonplayer characters.

Conclusions: Despite reported limitations, VR training has the potential to be a useful WPV training tool. It offers an immersive, hands-on, and safe environment for health care professionals to practice but may present challenges in engaging learners with the training objectives initially. While overall engagement and value in the training were high, refining dialogue realism and technical usability will support wider adoption. Future iterations of the pilot material may benefit from exploring role-specific content, multiplayer functionality, and integration of artificial intelligence–driven interactions to enhance responsiveness. Further research should compare VR WPV trainings with traditional trainings to evaluate differences in short- and long-term training effectiveness.

(JMIR Serious Games 2025;13:e70817) doi:10.2196/70817

KEYWORDS

virtual reality; usability; interprofessional; frontline health care staff; workplace violence; mixed methods

XSL•FO
RenderX

Introduction

Overview

Workplace violence (WPV) has increasingly been recognized as a significant public health issue, especially in health care settings [1]. Nonfatal injuries among health care workers have surged in recent years [2,3]. A common form of WPV includes client-initiated violence—aggression from patients or clients toward employees [4,5]. A systematic review of studies published in 2013‐2023 shows that the prevalence of client-initiated violence against health care workers varies from 7.5% to 74%, with verbal abuse as the most prevalent form, making up 31.1%-89.7% [6].

To address the rise in violence, the American Hospital Association in 2021 compiled resources and evidence-based practices aimed at identifying and mitigating risk factors for WPV [7]. This guidance builds on decades of research by organizations such as the Occupational Safety and Health Administration [8], Centers of Disease Control and Prevention [9], and The Joint Commission [5]. Unfortunately, a recent literature review of violence education highlighted unclear and mixed results regarding their effectiveness [10]. This outcome may be due to the fact that most programs rely on either web-based courses or in-person training. While Geoffrion et al [10] highlighted the immediate benefits of such training for both patients and providers, there is little evidence to suggest that conventional textbook, in-person, or web-based training yields long-term reductions in injury rates.

One of the inherent challenges in providing education related to managing WPV is the need to provide an established, standardized set of skills that then needs to be applied to dynamic situations based on judgment. Furthermore, these skills are typically taught in environments (eg, classrooms) that lack the real-life stress experienced in actual situations (eg, inpatient units). In other words, education is often lacking in realistic portrayals of events with which learners will be faced. Identifying ways to bridge this gap between education and application is critical in advancing our effectiveness in equipping the workforce to deal with WPV.

Virtual Reality in WPV Education

Virtual reality (VR) offers unique strengths as an adaptable platform for enhancing violence prevention training [11]. VR could enable health care staff to practice real-time communication strategies, using existing guidance and resources in dynamic simulated settings [12]. Its major advantage lies in the real-time, fully immersive experience it offers, accessible through mobile headsets and controllers [13,14]. With VR's tracking capabilities for body posture, gaze direction, and verbal interactions, it could create a personalized environment for skill building, accommodating the staff member's preferred time and location—whether at home or in a clinical environment [15]. Systematic reviews have demonstrated VR's ability to enhance both social and safety skills within health care contexts [16,17]. Recent findings also highlight VR's efficacy for improving practical skills and long-term retention among health care students, further supporting its use as an educational tool [18]. By allowing controlled, tailored exposure to various scenarios, VR presents potential as a flexible and engaging alternative to traditional training, positioning health care staff to respond effectively to diverse workplace challenges.

Objective

This study aims to assess the usability of a WPV VR training course with frontline health care workers, including physicians, nurses, and allied health care professionals. We investigated a prototype VR training course designed to augment the existing WPV training curriculum and capacity for health care staff.

Methods

We used a cross-sectional, mixed methods study design to assess current usability and gather input on future enhancements to our early-stage VR training platform for health care staff.

VR Training Content

Overview

The VR content and modules developed for this study were informed by existing literature [19-26] and insights gathered from a preceding qualitative study that explored staff expectations, personal experiences, challenges, and opportunities in WPV training programs [27]. Our multicomponent VR WPV course consists of 4 modules: Situational Awareness, Self-awareness and Self-regulation, Team Dynamics, and Evasive Maneuvers [28]. The training content targets significant challenges in de-escalation, where health care staff are responsible for managing physical, emotional, and communication issues. To ensure clinical relevance, these modules were designed to foster adaptive problem-solving skills, which often are overlooked in other WPV intervention programs [29]. Anticipating the interprofessional backgrounds of our participant health care workers, these prototype VR training modules were tailored to address challenges encountered in pediatric and adult behavioral health settings. The learning objectives for VR training modules included recognizing environmental triggers, regulating emotions under stress, and encouraging teamwork behaviors during crisis.

All of the modules, except Evasive Maneuvers, were structured as a narrative drawn from events occurring over the course of 1 day to simulate real-world scenarios commonly encountered in the workplace. To minimize distractions and help learners focus on the training content rather than adjusting to new VR environments, the scenarios were set in a single hospital location with recurring characters across the modules. The Evasive Maneuvers module, on the other hand, was independent of the overarching narrative and designed as an interactive physical training experience.

Situational Awareness

Situational awareness is a key component of crisis prevention and management. This module focuses on educating participants about individual and environmental risk factors that could increase the likelihood of escalating behaviors in patients. In the narrative, the learner is immersed in a hospital waiting room, observing a scenario where Riley (hypothetical patient) arrives and disrupts the check-in process of another patient by yelling at the receptionist. Learners are tasked with identifying potential

XSL•FO
RenderX

risk factors in the environment and behavioral cues from characters that may signal an impending crisis or escalation. The system incorporates gaze tracking to monitor where the learner is looking, and they can indicate awareness of various contributing environmental factors by pressing the trigger button on their controller (Figure 1). A debrief session on which risk factors participants noticed or missed was provided.

Figure 1. The learner observing a hospital waiting room with potentially identifiable risk factors.

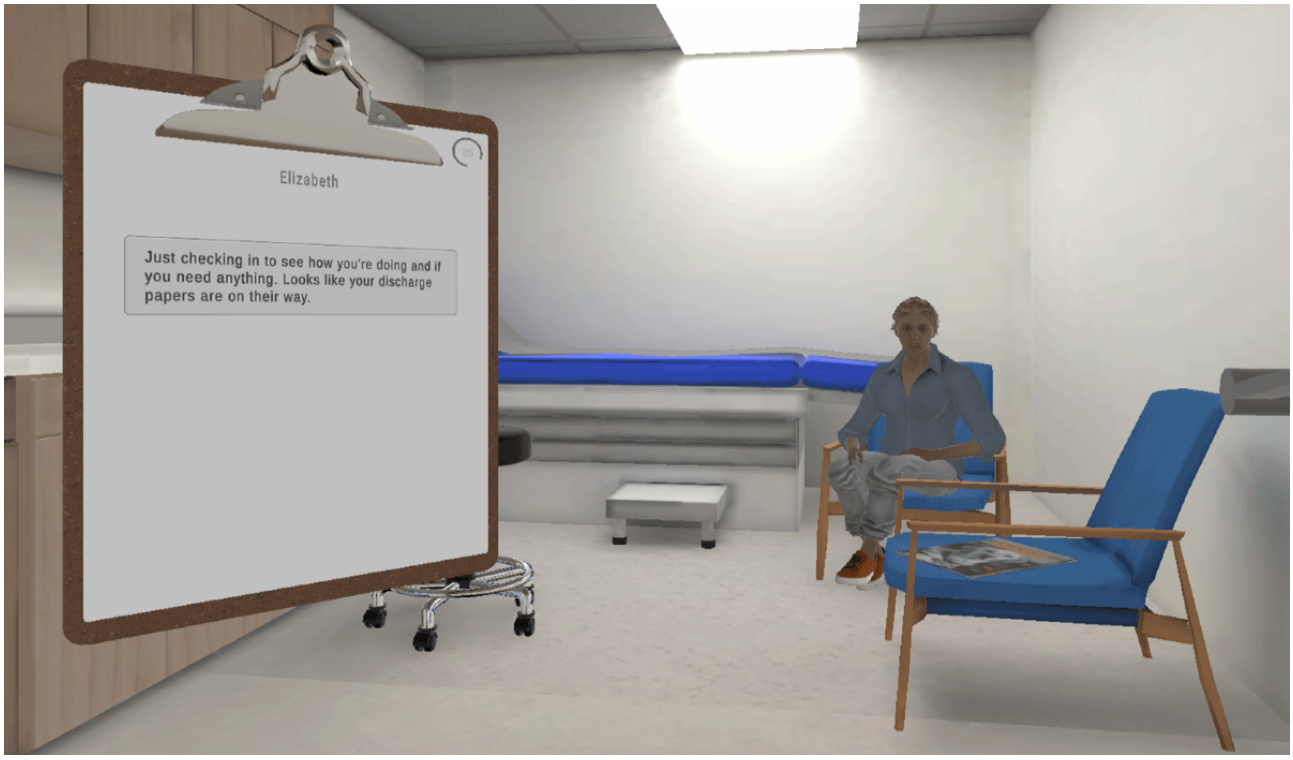


Self-Awareness and Self-Regulation

This module is designed to help learners identify and reflect on how their own emotions and stress levels influence decision-making during crisis situations. In the narrative, the learner assumes the role of a staff member conducting clinical rounds to check on patients, including the hypothetical patient, named Riley, who is exhibiting agitated behavior such as psychomotor agitation and verbal aggression. The learner navigates multiple interaction paths, each leading to different outcomes. Throughout the scenario, learners are presented with various verbal and behavioral response options, including the option to call Code Violet—an emergency code calling for assistance in a violent or combative situation. While calling a Code Violet is not the primary objective of this module, the exercise provides an opportunity to practice determining when it is appropriate to seek assistance (Figure 2).

XSL•FO
RenderX

Figure 2. A learner's virtual reality perspective conducting clinical rounds.



Team Dynamics

The Team Dynamics module focuses on the importance of effective teamwork during crisis situations. The scenario takes place during the restraint of Riley, who displayed aggressive behavior. To enhance perspective-taking, the learner can choose to either observe the scenario as a bystander or take on the role of the patient during the restraint. If learners choose the role of the restrained patient, they have the option to either sit in a chair or lie down on the ground for a fully immersive experience. Based on their chosen perspective, learners are exposed to 2 contrasting versions of team dynamics. Version 1 demonstrates mistakes in communication and noncohesive teamwork during a crisis, highlighting the negative impacts on patient experience and outcomes. Version 2 showcases best practices in team dynamics, including clear role assignment, effective communication, and mutual team support, illustrating the positive effects of a well-coordinated team. In both versions, learners gain insights into how team dynamics influence the patient experience and the effectiveness of crisis management (Figure 3).

XSL•FO
RenderX

Figure 3. A learner's virtual reality perspective as a bystander of a simulated patient being restrained.

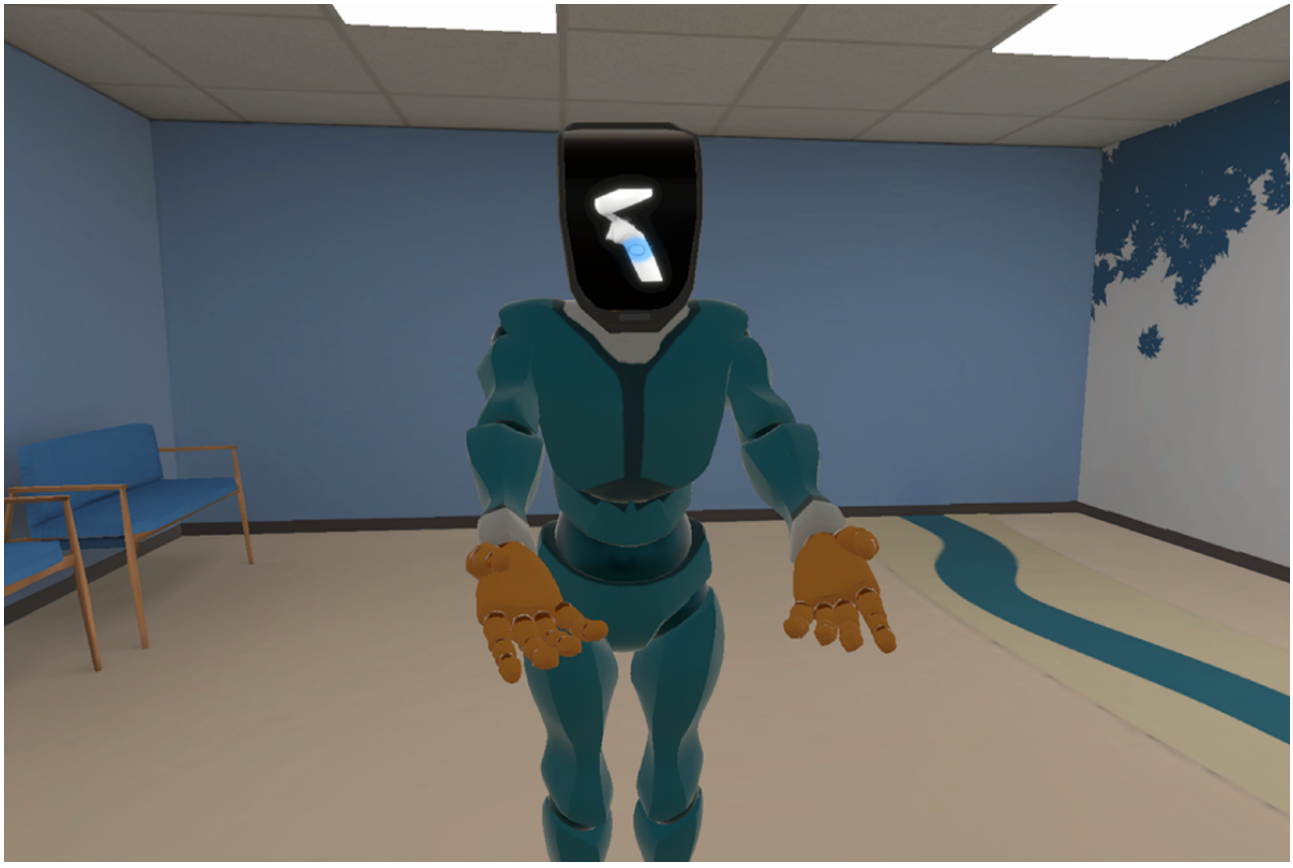


Evasive Maneuvers

This stand-alone refresher module focuses on teaching evasive maneuvers—protective poses and techniques used to safely disengage from holds, avoid or escape from harm, or reduce risk of injury by minimizing the impact of physical aggression. The system uses the built-in motion tracking of the VR headset and controllers to approximate the learner's body position and hand movements. At the start of the training, learners watch a refresher video demonstrating the techniques with real people. Following the video, a robot character guides them through physical practice of these techniques, while the system tracks their movements. For instance, if a technique requires the learner to form a fist, the system uses buttons on the side of the controller to detect a successful "fist" pose (Figure 4). This module's narrative content is disconnected from the overarching story of the other modules to clearly communicate exact poses and hand gestures to newer VR users.

XSL•FO
RenderX

Figure 4. A robot guide teaching the learner about defensive techniques.



VR Headset

The PICO 3 headset, a VR device by Pico Interactive [30], was chosen for VR training due to its stand-alone design, processing and eye-tracking capabilities, and enterprise-level software development support. Each headset included 2 motion-tracked controllers, each with a joystick, trigger, face buttons, and a menu button, enabling an interactive training experience. The VR content was developed using Unity 2021.3.28f1 [31-37], using all available interactive components, including head-based tracking to monitor user attention.

Target Participants

Target participants were health care workers recruited between March and June 2024 from hospitals in Central Ohio through email and phone calls using convenience sampling. Eligible participants were aged 18 years or older with health care expertise or frontline experience in interprofessional roles. Additionally, we aimed to recruit a diverse set of perspectives to represent multiple roles within health care (>2 staff with different roles), which guided recruitment efforts until reaching thematic saturation.

Data Collection

Each study session lasted approximately 1 hour and included three activities: (1) Introduction to VR, (2) Testing Session, and (3) Semistructured Discussion. All participants provided written consent and received compensation for their participation.

Introduction to VR

Participants were given 10 minutes to practice using VR headsets and controllers until they felt comfortable navigating a 3D environment. Study staff provided guidance on menu navigation, attention-sensitive actions, and safety protocols, including trip hazard checks, proper fitting and calibration, headset hygiene, and guardian system setup. Any technical issues were addressed during this session.

Testing Session

The testing session consisted of 4 training modules. During testing, participants were encouraged to think aloud, ask questions, and provide feedback, helping the study team gather valuable insights into their experiences and thought processes. First, participants completed Self-Awareness and Self-Regulation and Team Dynamics modules. In those modules, participants completed the tasks with step-by-step instructions provided through in-modules tutorials, supported by technical assistance from the study team while users engaged with the modules lasting 20 minutes. Modules were selected based on difficulty and structured frameworks to standardize learning objectives and then observe user behaviors [38].

After that, participants were checked for their comfort. Then, they interacted with the Situational Awareness and Evasive Maneuvers modules. These 7- to 9-minute modules also contained tutorial elements, allowing participants to explore freely at their own pace and rely on their existing knowledge and instincts to develop relevant skills. The testing session took approximately 30 minutes total.

XSL•FO
RenderX

Semistructured Discussion

Participants provided feedback on their experience with VR technology, discussing perceived facilitators and barriers related to VR design, functionality, educational value, and the potential advantages or limitations this technology could bring to WPV training. Those with prior conflict management training in VR and WPV trainings compared their previous impressions with the current VR experience, evaluating factors such as realism, immersive feelings, and discussing other topics they would be interested in learning through VR. All responses were audio-recorded and transcribed for qualitative analysis.

Measurements

After the semistructured discussion, participants completed 3 evaluation measures: mini Player Experience Inventory (miniPXI), System Usability Scale (SUS), and Reaction Cards.

Mini Player Experience Inventory

MiniPXI is a validated 11-item measure (7-point scale) derived from the popular Player Experience Inventory (Table S1 in Multimedia Appendix 1) [39]. The Likert scale ranges from "Strongly Disagree" (1) to "Strongly Agree" (7). It is designed to evaluate the user experience of a product, focusing on the emotional and experiential aspects of interacting with it. The scale is widely applied across multiple digital health interventions, including mobile apps and VR training apps [40]. It assesses aspects of user experience, such as engagement, functionality, esthetics, and information quality, helping to understand how well a VR training app meets the needs and expectations of its users.

System Usability Scale

The SUS is a widely used tool to assess the usability of a system (Table S2 in Multimedia Appendix 1) [41]. It consists of 10 items rated on a 5-point Likert scale from "Strongly Disagree" to "Strongly Agree." The SUS provides a measure of user satisfaction and helps identify areas for improvement [11]. It is designed and validated as a benchmark for usability with extensive use in both academic and industry-focused research. SUS scores are calculated out of 100, where the target usability score is 68 or greater.

Reaction Cards

Reaction Cards is a method initially developed by Microsoft in 2002 for user experience research to gather quantitative feedback on users' emotional reactions to a system or product (Table S3 in Multimedia Appendix 1) [42]. Users select words or phrases from a predefined set (118 words) that best describe their feelings about the product. For the purpose of this study, the target benchmark was set at 50% or greater positive word selection in the sample.

Data Analysis

We used transcriptions of participant audio recordings, analyzed through Rapid Qualitative Analysis [43]. This approach is streamlined to minimize analysis time while maximizing thematic saturation ("Interview Guide" in Multimedia Appendix 1). To ensure the confirmability of study analysis, coding of participant comments was triangulated with the corresponding quantitative results. A structured template was created to consistently extract main discussion points about the VR experience into representative quotes ("Structured Template for Rapid Qualitative Analysis" in Multimedia Appendix 1). To assess the saturation of themes, we continued conducting focus groups until no new codes emerged from the transcripts and we met the recommended range of focus groups for data quality based on other empirical studies in qualitative research [44]. Cohen κ was used to calculate interrater reliability between coders [45].

Quantitative data from the SUS and the Player Experience Inventory (miniPXI) survey were analyzed with descriptive statistics (means and SDs) calculated with Microsoft Excel. Additionally, frequencies from reaction card data were visualized with a word cloud, created using the wordcloud2 package in *R* version 4.3.0 (2023-04-21 Universal C Runtime) through RStudio (2023.12.1 Ocean Storm Build 402 by Posit Software) [46-48].

Ethical Considerations

The institutional review board (FWA000002860) at Nationwide Children's Hospital reviewed and approved the protocol (STUDY00004035) for this investigation. The procedures followed were in accordance with ethical standards of the responsible committee on human participants' research and federal guidelines 45 CFR 46 from the United States Department of Health and Human Services. All participants signed an informed consent form that explained the objective of the study, location of study sites, study activities, risks of participation (eg, physical discomfort and motion sickness), direct benefits of participation (eg, improved skill development and postintervention knowledge handling WPV), compensation, efforts for study data privacy, contact information of the research team, and other important information. Participants were compensated with a US $50 gift card for their time and commitment. To minimize risks to participants, a guardian system was implemented to automatically detect the user's position relative to set boundaries to avoid physical collisions. If the participant approaches or exceeds the boundaries, the VR training application is minimized and pass-through cameras allow the participant to see the real world to return to the play space. Research staff members also actively monitored participants for physical risks both inside and outside of the play space.All participant data were securely stored on password-protected institutional servers, and accessed only by authorized study personnel to ensure privacy and confidentiality.

Results

Demographic Information

All participants (N=13) were frontline health care workers, comprising nurses (n=6), physicians (n=4), social workers (n=2), and a security officer. To maintain participant anonymity, we did not collect additional demographic information of our participants. Five of them had previous exposure to VR, and 3 owned a VR headset at home.

XSL•FO
RenderX

Descriptive Statistics

The VR training modules received above-average ratings on the miniPXI (Table 1), with a mean score of 5.23 (SD 1.34). The mean scores of functional and psychological constructs are 4.77 (SD 1.41) and 5.62 (SD 1.14), respectively. The VR training was perceived to provide limited feedback throughout the modules (miniPXI-Progress Feedback: mean 4.08, SD 1.27), yet the training motivated participants to explore the simulations (miniPXI-Curiosity: mean 6.15, SD 0.66). This score indicates strong user engagement and suggests that participants found the VR training to operate as expected while meeting practical needs, without causing frustration.

Table . Descriptives of the mini Player Experience Inventory.

Constructs		Mean (SD)	Median (IQR)
Functional constructs			
	Audiovisual Appeal	5.46 (1.08)	6.00 (5.00-6.00)
	Challenge	5.46 (1.08)	6.00 (6.00-6.00)
	Ease of Control	4.54 (1.39)	4.00 (4.00-6.00)
	Clarity of Goals	4.31 (1.54)	4.00 (3.00-6.00)
	Progress Feedback	4.08 (1.27)	4.00 (3.00-5.00)
Psychosocial constructs			
	Autonomy	5.08 (1.14)	6.00 (4.00-6.00)
	Curiosity	6.15 (0.66)	6.00 (6.00-7.00)
	Immersion	6.00 (1.18)	6.00 (5.00-7.00)
	Mastery	5.00 (1.47)	5.00 (5.00-6.00)
	Meaning	5.58 (0.86)	6.00 (5.00-6.00)
	Enjoyment	5.92 (0.73)	6.00 (6.00-6.00)

In terms of usability, the SUS score (Table 2) for the VR modules was slightly below the industry benchmark [41,49,50], with a mean of 63.3/68 (SD 9.53), suggesting room for improvements to better align with user expectations for usability of features and content. Overall, participants found the VR training manageable (mean 2.92, SD 0.46) and learned to operate VR quickly (mean 2.85, SD 0.51) but may need technical support to initiate VR training (mean 1.92, SD 0.70). Participants (n=3) who encountered navigational issues during the VR experience scored the overall usability of the training lower (mean 53.1, SD 2.07) than other users. Leading usability concerns from this subgroup identified the VR intervention as overly complex (mean 1.67, SD 0.47) and involved a high learning curve (mean 1.67, SD 0.47).

Table . Descriptives of the System Usability Scale.

Item definitions	Mean (SD)	Median (IQR)
Frequency Use (1)	2.54 (0.61)	3.00 (2.00-3.00)
Complexity (2)	2.46 (0.81)	3.00 (2.00-3.00)
Ease of Use (3)	2.62 (0.60)	3.00 (2.00-3.00)
Support Needs (4)	1.92 (0.70)	2.00 (1.00-2.00)
Functionality Integration (5)	2.31 (0.58)	2.00 (2.00-3.00)
Inconsistency (6)	2.54 (0.72)	3.00 (2.00-3.00)
Confidence (7)	2.85 (0.51)	3.00 (3.00-3.00)
Cumbersomeness (8)	2.92 (0.46)	3.00 (3.00-3.00)
Learnability (9)	2.54 (0.61)	2.00 (2.00-3.00)
Learning Curve (10)	2.62 (0.97)	3.00 (2.00-3.00)

Participant reactions to the VR content were highly positive, as indicated by the Reaction Card responses. Out of 140 total selected cards by participants, 89% (124) reflected favorable impressions of the VR experience. The 6 most frequently chosen reaction words were “valuable,” “engaging,” “creative,” “accessible,” “useful,” and “fun” (“Participant’s Reaction Cards” in Multimedia Appendix 1). A small number of negative reactions—such as “disconnected” and “distracting”—highlighted the areas for potential enhancement to increase user satisfaction and immersion, which aligns with usability score feedback. Overall, the poststudy survey results suggest that while participants found the VR modules highly

XSL•FO
RenderX

engaging and beneficial for certain aspects of training, improvements in usability and immersive quality could enhance the overall experience further.

Qualitative Analysis

Overview

The qualitative analysis revealed four themes: (1) Perceived Value, (2) Technical and Navigational Barriers, (3) User Preferences, and (4) Vision. These themes expanded our understanding of quantitative findings. Data saturation was reached after 5 focus groups. Interrater reliability between authors (DIJ and ES) indicated substantial agreement (κ=815; P<.001).

Perceived Value

The Perceived Value theme described how the immersive environment influenced participants' engagement with the training content and experience of the training. Participants frequently compared the value of VR with traditional training settings, noting the unique advantages of the VR modality. Overall, feedback was positive, with many praising the VR training modules for providing a fresh and impactful perspective on WPV training.

> *It was more helpful to do that [training] in a VR environment than clicking through modules on a computer.* [Participant 2]
>
> *We can watch as many [procedural videos on] codes as we want to, but this [VR training module] can mimic [real world events].* [Participant 5]
>
> *I think it was super helpful to hear and to get the patient perspective of what conversations are going on around you. The patient is definitely paying attention too, but we might not be.* [Participant 8]

Participants frequently compared the VR training tasks with their previous experiences and routine job responsibilities in health care, noting a limitation in the available response options during training. This feedback informed refinements to dialogues and VR environments, ensuring closer alignment with participant expectations and real-life decision-making processes. Moreover, interprofessional aspects of the training were well received, especially by participants less familiar with the roles of other health care staff, highlighting the potential of VR to enhance interprofessional understanding in training programs.

> *I think VR training could be really useful interdisciplinary, like security [personnel included].* [Participant 1]
>
> *The training was pretty realistic [compared to] other [procedural color code trainings] that I've been a bystander of in clinic.* [Participant 2]
>
> *I feel like the waiting room was pretty real feeling aside from the security guard not intervening.* [Participant 7]
>
> *We as physicians, we never really end up doing [self-defense] so we don't know how to do a lot of these basics. We're never really taught that part.* [Participant 11]
>
> *This training is for the nurses and technicians–for sure, definitely! How many times do we walk back into the room and we're like, "Why is that technician speaking like that to the patient this way?"* [Participant 12]

Technical and Navigational Barriers

Participants identified several barriers related to accessibility and the social design features of the VR experience in the training module. Navigating challenges was common, especially for first-time users, with some finding the teleportation mechanic difficult to use. While many described the experience as refreshing and highly engaging, some personal preferences, such as sound effects, did not align with the designed VR environment.

Attention-based mechanisms, which required users to focus on specific objects, characters, or dialogue to progress through the narrative, were a point of friction. Participants, primed to view the tool as a pilot, suggested incorporating more environmental cues to guide navigation and instruction.

> *[The patient] didn't give us any feedback. He wasn't vocal for us to gauge where he was, like is the patient calm?* [Participant 2]
>
> *I could hear that [non-player characters] were saying stuff, but I couldn't really understand it.* [Participant 6]
>
> *I wasn't sure if I was doing [situational awareness training] incorrectly or just kind of clicking around.* [Participant 11]
>
> *I had a lot of trouble jumping to the teleport. I couldn't get it to focus well.* [Participant 12]

A notable portion of participants, primarily new users, encountered challenges adapting to the VR controllers. Participants offered their own solutions to these navigational difficulties.

> *I had a lot of difficulty getting into the situations, working the controllers for [training tasks].* [Participant 5]
>
> *It would be kinda cool if there was a see-through map of where your fingers are.* [Participant 9]
>
> *It would be nice if you were to hit the wrong button and you get a text popup that asks [to confirm] because at one point, I thumbed the wrong button.* [Participant 10]

User Preferences

The User Preferences theme highlighted interactive properties that participants valued in VR, emphasizing the importance of convincing simulations to enhance the realism of training. Feedback focused primarily on the language and emotional intensity of VR nonplayer characters (NPCs, ie, virtual characters controlled by the system rather than the user). Opinions were mixed regarding the tone and emotional responses portrayed by the NPCs; while some participants felt that the NPC behaviors, particularly those of virtual patients, aligned well with their professional experiences, others noted

XSL•FO
RenderX

that emotions and behavioral symptoms were underrepresented, especially within specific health care domains.

> *I want to hear some "oomph" behind [explicit language]. It was not being used as a power word. It was being used as like a generic [filler word].* [Participant 3]

> *If you're going to use it in a psychiatric setting, putting in some psychotic or manic flavoring [to show] not just [irritated but] they're also disorganized mentally. That would probably mirror more of what we deal with.* [Participant 5]

> *The staff sounded more human to me than the patients. Make the patients more mad. The staff being despondent kinda works.* [Participant 7]

> *It felt very realistic where somebody comes barging in before you have time to decide.* [Participant 8]

Participants highlighted discrepancies between the VR training scenarios and the realities of their occupational responsibilities. They noted that some environmental factors and response options lacked the nuance required to accurately reflect real-life situations.

> *Could use more realistic complaints. We're rarely giving out patients coffee, [instead] there's a lot of complaints of they want their phone, why do they have to change said gown, and why do they have to be in this room with no curtain privacy.* [Participant 8]

> *When the escalated patient came into the room, I'm not just gonna sit there waiting for my clipboard to say something. I need to do something. I think I might have said out loud, "Can I close the door?" [because] he went into a patient room with a baby! The lag was not reflective of real life.* [Participant 10]

Vision

The Vision theme captured participants' expectations for how VR could be used in future training topics. Suggestions included addressing additional types of WPV (eg, criminal intent and personal relationship) and incorporating training for other challenging conversations. The Team Dynamics scenario received particularly strong approval, with participants noting that it fostered a sense of unity and purpose similar to in-person training. Many recommended adding a multiplayer feature to future modules, enabling team-oriented exercises and collaborative debrief sessions similar to traditional training programs.

> *Maybe at the end, there is a scenario that all players can join in and have a virtual patient that they're holding. It would be helpful and interactive.* [Participant 2]

> *These headsets are so isolating. A multiplayer game would be more reflective experience.* [Participant 3]

> *I think its very suitable for staff that don't have really any experience de-escalating [other participants nod].* [Participant 8]

> *It would be cool to link the headsets and have us practice as a team instead of having to pick responses separately.* [Participant 13]

Participants highlighted new educational opportunities for VR training, particularly its potential for simulating low-frequency but high-impact situations. They emphasized the value of providing health care workers with opportunities to practice challenging conversations, which are rarely encountered in routine training. This feedback underscores the potential for VR to evolve further by incorporating collaborative features that enhance both realism and educational impact.

> *It would be nice to have an active shooter drill [in an emergency room]. It's hard to do that in the real setting because there's people there, and it's always busy.* [Participant 5]

> *[VR] would be useful right after self-defense training. Let's reinforce training by throwing some situations at you that would be new and novel to let you practice your actual reactions.* [Participant 6]

> *You could also use it to teach people interviewing skills. For example, how to ask people about their psychotic symptoms, suicidal ideation, and violent history. [To teach] how to segway into difficult topics.* [Participant 9]

> *Practicing [lists many procedural color codes] would be nice as patient scenarios.* [Participant 13]

Discussion

Principal Findings

This study evaluated the usability of a prototype VR training program aimed at augmenting existing WPV mitigation training in health care settings. To ensure broad applicability and relevance of the current training modules, health care staff from diverse occupational roles were recruited to represent those who may encounter WPV.

The mixed methods results provide a comprehensive understanding of areas for improvement in the current VR training modules. Quantitative findings reveal high levels of engagement, with participants rating the VR modules positively for their immersive and motivational qualities, highlighting VR's potential to adapt to real-world situations. Despite usability scores (mean 63.3, SD 9.53) indicating room for improvement, these findings aligned with qualitative feedback that identified challenges related to navigation, realism, and mismatches between training scenarios and real-life occupational responsibilities. Negative Reaction Cards underscored a subgroup of users that saw the VR training as "disconnected" or "distracting," revealing accessibility concerns for first-time or struggling users. These findings were aligned with lower usability scores hinting at higher complexity, which was acknowledged in the thematic analysis (teleporting mechanics and controller difficulties). Nonetheless, participants found these VR training modules easy to use, which allowed them to focus on training objectives. Reaction Card responses further demonstrated the creative and engaging nature of the VR modules, with 89% of selected terms reflecting positive

XSL•FO
RenderX

impressions such as “valuable,” “engaging,” “creative,” and “accessible.”

These insights are particularly valuable for guiding modification and additions to WPV VR training, particularly in refining the environment, enhancing NPC interactions, and tailoring the content to better align with real-world scenarios. Nonetheless, qualitative data provided comparisons to traditional training formats, emphasizing the unique value of VR in creating an engaging, accessible, and innovative training experience. Together, these findings offer valuable guidance for the iterative development of the WPV VR training environment, ensuring that it engages learners in WPV prevention curricula.

Advancing Role-Specific Training in VR

In alignment with VR health care education interventions [51], our prototype VR training underscores the importance of refining features to improve usability. Maltby et al [52] deployed a similar modular VR training focused on telehealth health care staff. The user feedback emphasized a need for role-based divisions in the training curricula. To address this gap, evolving our interprofessional content to replay the same modules from different roles may improve the usability of VR training. Additionally, reusability of the scenario content offered a method to lower the scaling cost while maintaining a professional adaptation of traditional video-based training. Usability components found in other studies underscore the potential of VR in health care education, where immersive experiences can build transferable skills. Although the literature on role-specific training in VR is limited, we anticipate that tailoring VR objectives to specific health care roles may address this feedback and enhance engagement further.

Enhancing Feedback and Interaction Mechanics in VR Training

The use of tracking to record and adapt learner interactions received mixed feedback, particularly regarding attention-based mechanics. Our findings highlight the importance of heuristic feedback, where the learner can intrinsically gauge the success of a task through VR interactive elements. Some participants, especially first-time users, also encountered challenges using VR controllers. These observations align with the findings by Khundam et al [53], who noted similar needs for feedback mechanisms during endotracheal intubation training with medical students. Their study highlighted the value of incorporating verbal or visual cues to affirm throughout the simulation. To overcome problems with VR controls, they used hand tracking to replace physical controllers. Similarly, integrating hand-tracking capabilities into our VR modules could improve usability and make the training more accessible for novice users.

Natural Dialogue in VR

Balancing the authenticity of staff and patient dialogue in VR scenarios proved challenging. Some participants encountered choices or behaviors that did not align with their own decision-making processes, which led to negative perceptions of the VR training. This limitation may stem from the current branching design, which offers limited response pathways.

Conversely, other participants found certain dialogue options closely aligned with previous training or personal experience, indicating the potential of well-designed dialogue to reinforce learning. To address different user expectations, future iterations of the VR training may offer unique pathways tailored to specific health care roles, allowing trainees to review protocols and policies from various perspectives. Additionally, incorporating artificial intelligence, particularly large language models, could allow users to respond freely while enabling NPCs to deliver more personalized and adaptive interactions, fostering a fluid and realistic interaction experience for each user. Observed barriers and potential solutions are summarized in Table 3.

Table . Observed obstacles and barriers to virtual reality with recommendations.

Obstacle and barrier to VR[a]	Recommendation
High learning curve for new users	Enable controller-free navigation using hand tracking.
Difficulty with teleportation mechanics	Leverage a guardian system to replace teleportation to navigate physically throughout the play space.
Lack of feedback on actions (sounds, environmental cues, “disconnected”)	Add haptic, auditory, or visual effects based on successful or unsuccessful interactions.
Confusing UI[b] (menu options, “distracting,” attention-based)	Increase space between menu options and relabel for better clarity.
Difficulty with VR controllers or user misinputs	Highlight finger overlays within the headset over simulated controllers.
NPC[c] expressions	Expand the emotional range of characters with voice acting and co-designing scripts with clinicians.
Limited response choices	Detect open-ended responses using the participant’s voice to simulate reactive, live NPC responses.
Mismatched care setting	Allow users to select their health care environment based on their workplace experience.

[a]VR: virtual reality.

[b]UI: user interface.

[c]NPC: nonplayer character.

XSL•FO
RenderX

Future Directions

Regarding the realism of VR experience, in future iterations, technical improvements might be required to align the virtual environment with the real world. VR with a wider binocular field of view (200° - 220°) may enhance immersion, creating a more complete visual experience [54], further expanding with cinematic techniques (360° videos) [55] toward meeting expectations of immersion, as the content itself is reused in favor of presenting new perspectives for the learner. However, the cost of high-end VR headsets with advanced graphic capabilities presents a financial barrier for widespread health care training adoption. Given this constraint, enhancing the VR experience may require focusing on immersive elements that foster a stronger sense of presence rather than relying solely on visual fidelity. Improving the depth of immersion, especially through increased interactivity and realistic environmental cues, could yield greater improvements in user comprehension and training usability [56,57]. This observation aligned with user feedback as we discussed differences between simulated environments and VR training. Furthermore, sound plays an important role in the perception of "presence" in immersive VR experiences [58]. The successful implementation of audio can create the "illusion of nonmediation" [59] where the marked barriers between the user and the medium of communication are diminished. In line with feedback from our participants regarding challenges with the audio volume, VR training simulations may benefit from participants wearing headphones or earbuds during the training. This would potentially address future concerns with audio volume and create a more immersive sense of presence.

Limitations

The usability of VR training may vary depending on the specific needs and training environments of different health care institutions. While the sample size (N=13) was too small for statistical analysis, it exceeded the recommended range (5 - 10 participants) for pilot usability assessments. Regardless, quantitative measures were reported descriptively and were not powered for inferential testing. These findings may not be broadly generalizable to all health care settings or populations due to this limited sample size and potential variations in training needs across institutions. Our work reported usability of a VR simulation, without a control group, that scored below the benchmark (mean 63.3, SD 9.53) due to navigational mechanics. We did not investigate eye-tracking data as a direct measure of trainee behavior due to technical limitations. Additionally, participant feedback may have been influenced by the novelty effect, as many were first-time users, or by navigation challenges related to controller use, potentially skewing responses positively or negatively. Future studies should account for this potential bias by incorporating participants with prior exposure to VR to capture a broader range of perspectives.

Another limitation is the long-term retention of knowledge and skills acquired through VR training. While this study focused on immediate usability and participant feedback, the sustainability of the training's impact remains unknown. A follow-up study is underway to evaluate the longitudinal effects of VR training on learning outcomes, including knowledge retention, behavioral changes, and real-world application in managing WPV scenarios. We will use a precomparison and postcomparison of skill retention with a control group to understand the effectiveness of VR for workplace safety.

Finally, the study did not examine potential barriers to adoption, such as costs, technological infrastructure, or institutional readiness, which could influence the feasibility and scalability of VR training in diverse health care settings. Addressing these factors in future research will help ensure that VR training can be effectively integrated into routine health care education and violence prevention programs.

Further research is needed to compare the effectiveness of VR training against traditional training methods longitudinally. Investigating the long-term benefits of immersive training, especially within the scope of WPV, would provide insights into enhanced employee preparedness and safety. Such studies could potentially identify the specific elements of scenarios that most effectively reduce incidents of violence, improve response protocols, and bolster overall staff confidence in handling aggressive behaviors in high-risk environments. In addition, more research is needed to identify and verify the usability of teaching circumstances of niche cases with WPV into training curricula as well as development of learning assessment techniques for VR WPV training. Anticipated improvements to the VR training program may include the introduction of multitrainee scenarios, modules targeting communication skills, and other interactive features that enhance training realism and relevance.

Conclusions

VR shows mixed reception as a training tool for health care staff managing WPV. The immersive quality of VR allowed participants to engage in realistic, hands-on scenarios within a safe, simulated environment but added complexity to engaging with the training material. The participants' feedback was predominantly positive, although suggestions for improvement highlight impactful refinement opportunities in user interface and character design to enhance usability. Participants' experiences further identified with the immersive nature of the training, as well as technical and navigational challenges. Taken together, the study findings demonstrate that VR can be a usable and engaging modality for delivering educational WPV training content if necessary enhancements are made appropriately.

Acknowledgments

This study was funded by the Ohio Bureau of Workers' Compensation Workforce Safety Innovation Center.

XSL•FO
RenderX

Data Availability

The data underlying this paper cannot be shared publicly due to the privacy of individuals who participated in the study. However, they may be available from the corresponding author on reasonable request.

Conflicts of Interest

MAF receives royalties from American Psychiatric Publishing and Guilford Press and travel support/an editorial stipend from the Society of Clinical Child and Adolescent Psychology. ES is an editorial board member of JMIR. No other authors have competing interests to declare.

Multimedia Appendix 1
Study instruments and materials.
[DOC File, 143 KB - games_v13i1e70817_app1.doc]

References

1. National Institute for Occupational Safety and Health. Workplace violence prevention strategies and research needs: report from the conference partnering in workplace violence prevention: translating research to practice. : Centers for Disease Control and Prevention; 2004 Nov 17 11418 URL: https://stacks.cdc.gov/view/cdc/11418/cdc_11418_DS1.pdf [accessed 2025-08-28]
2. Workplace violence SBREFA. Occupational Safety and Health Administration. URL: https://www.osha.gov/workplace-violence/sbrefa [accessed 2024-09-19]
3. Fact sheet: health care workplace violence and intimidation, and the need for a federal legislative response. American Hospital Association. URL: https://www.aha.org/fact-sheets/2022-06-07-fact-sheet-workplace-violence-and-intimidation-and-need-federal-legislative [accessed 2024-09-19]
4. Merchant JA, Lundell JA. Workplace violence: a report to the nation. : University of Iowa Injury Prevention Research Center; 2001 URL: https://iprc.public-health.uiowa.edu/wp-content/uploads/2015/09/workplace-violence-report-1.pdf [accessed 2025-08-28]
5. Arnetz JE. The Joint Commission's new and revised workplace violence prevention standards for hospitals: a major step forward toward improved quality and safety. Jt Comm J Qual Patient Saf 2022 Apr;48(4):241-245. [doi: 10.1016/j.jcjq.2022.02.001] [Medline: 35193809]
6. Yusoff HM, Ahmad H, Ismail H, et al. Contemporary evidence of workplace violence against the primary healthcare workforce worldwide: a systematic review. Hum Resour Health 2023 Oct 13;21(1):82. [doi: 10.1186/s12960-023-00868-8] [Medline: 37833727]
7. Kingston MB, Apolinar A, Ardabell T, et al. Building a safe workplace and community. American Hospital Association. 2022. URL: https://www.aha.org/system/files/media/file/2023/04/Building-a-Safe-Workplace-and-Community-Mitigating-the-Risk-of-Violence.pdf [accessed 2025-08-28]
8. Healthcare—workplace violence. Occupational Safety and Health Administration. URL: https://www.osha.gov/healthcare/workplace-violence [accessed 2024-09-19]
9. Prioritizing our healthcare workers: the importance of addressing the intersection of workplace violence and mental health and wellbeing. Centers for Disease Control and Prevention. 2024. URL: https://blogs.cdc.gov/niosh-science-blog/2024/05/29/hcw_violence_mh/ [accessed 2024-09-19]
10. Geoffrion S, Hills DJ, Ross HM, et al. Education and training for preventing and minimizing workplace aggression directed toward healthcare workers. Cochrane Database Syst Rev 2020 Sep 8;9(9):CD011860. [doi: 10.1002/14651858.CD011860.pub2] [Medline: 32898304]
11. Moore N, Ahmadpour N, Brown M, Poronnik P, Davids J. Designing virtual reality-based conversational agents to train clinicians in verbal de-escalation skills: exploratory usability study. JMIR Serious Games 2022 Jul 6;10(3):e38669. [doi: 10.2196/38669] [Medline: 35793129]
12. Liaw SY, Ooi SW, Rusli KDB, Lau TC, Tam WWS, Chua WL. Nurse-physician communication team training in virtual reality versus live simulations: randomized controlled trial on team communication and teamwork attitudes. J Med Internet Res 2020 Apr 8;22(4):e17279. [doi: 10.2196/17279] [Medline: 32267235]
13. Kyaw BM, Saxena N, Posadzki P, et al. Virtual reality for health professions education: systematic review and meta-analysis by the digital health education collaboration. J Med Internet Res 2019 Jan 22;21(1):e12959. [doi: 10.2196/12959] [Medline: 30668519]
14. Clay CJ, Hochmuth JM, Wirth O. Virtual reality training to reduce workplace violence in healthcare. Issues Ment Health Nurs 2025 Jan;46(1):2-11. [doi: 10.1080/01612840.2024.2429712] [Medline: 39761223]
15. Antoniou PE, Arfaras G, Pandria N, et al. Biosensor real-time affective analytics in virtual and mixed reality medical education serious games: cohort study. JMIR Serious Games 2020 Sep 2;8(3):e17823. [doi: 10.2196/17823] [Medline: 32876575]

XSL•FO
RenderX

16. Chang AA, Kazemi E, Esmaeili V, Davies MS. The effectiveness of virtual reality training: a systematic review. J Organ Behav Manage 2024 Jul 2;44(3):214-232. [doi: 10.1080/01608061.2023.2240767]
17. Abich J IV, Parker J, Murphy JS, Eudy M. A review of the evidence for training effectiveness with virtual reality technology. Virtual Real 2021 Dec;25(4):919-933. [doi: 10.1007/s10055-020-00498-8]
18. Liu K, Zhang W, Li W, Wang T, Zheng Y. Effectiveness of virtual reality in nursing education: a systematic review and meta-analysis. BMC Med Educ 2023 Sep 28;23(1):710. [doi: 10.1186/s12909-023-04662-x] [Medline: 37770884]
19. Accinni T, Papadogiannis G, Orso L. De-escalation techniques in various settings empathy, normalization and de-escalation. In: Biondi M, Pasquini M, Tarsitani L, editors. Empathy, Normalization and de-Escalation: Management of the Agitated Patient in Emergency and Critical Situations: Cham: Springer International Publishing; 2021:65-91. [doi: 10.1007/978-3-030-65106-0_5]
20. D. Croce N, Mantovani C. Using de-escalation techniques to prevent violent behavior in pediatric psychiatric emergencies: it is possible. Pediatr Dimensions 2017;2(1). [doi: 10.15761/PD.1000138]
21. Ponnuthurai S, Moghraby OS, Reynolds S, Marikar D, Roland D. Initial assessment and management of agitation and aggression in the paediatric setting. Paediatr Child Health (Oxford) 2019 Dec;29(12):543-546. [doi: 10.1016/j.paed.2019.09.005]
22. Holloman GH Jr, Zeller SL. Overview of Project BETA: best practices in evaluation and treatment of agitation. West J Emerg Med 2012 Feb;13(1):1-2. [doi: 10.5811/westjem.2011.9.6865] [Medline: 22461914]
23. Richmond JS, Berlin JS, Fishkind AB, et al. Verbal DE-escalation of the agitated patient: consensus statement of the American Association for Emergency Psychiatry Project BETA DE-escalation Workgroup. West J Emerg Med 2012 Feb;13(1):17-25. [doi: 10.5811/westjem.2011.9.6864] [Medline: 22461917]
24. Roppolo LP, Morris DW, Khan F, et al. Improving the management of acutely agitated patients in the emergency department through implementation of Project BETA (Best Practices in the Evaluation and Treatment of Agitation). J Am Coll Emerg Physicians Open 2020 Oct;1(5):898-907. [doi: 10.1002/emp2.12138] [Medline: 33145538]
25. Luck L, Jackson D, Usher K. STAMP: components of observable behaviour that indicate potential for patient violence in emergency departments. J Adv Nurs 2007 Jul;59(1):11-19. [doi: 10.1111/j.1365-2648.2007.04308.x] [Medline: 17543010]
26. Simpson SA, Pidgeon M, Nordstrom K. Using the behavioral activity rating scale as a vital sign in the psychiatric emergency service. Colo J Psychiatry Psychol 2017 Aug;2(2):1-66. [doi: 10.25677/10968/4333]
27. Jackson DI, Maa T, Luna J, Huang Y, Fristad MA, Sezgin E. Challenges and opportunities in healthcare workplace violence training: a qualitative study of staff experiences. J Workplace Behav Health 2025:1-14. [doi: 10.1080/15555240.2025.2474055]
28. Somani R, Muntaner C, Hillan E, Velonis AJ, Smith P. A systematic review: effectiveness of interventions to de-escalate workplace violence against nurses in healthcare settings. Saf Health Work 2021 Sep;12(3):289-295. [doi: 10.1016/j.shaw.2021.04.004] [Medline: 34527388]
29. AbuAlRub RF, Al Khawaldeh AT. Workplace physical violence among hospital nurses and physicians in underserved areas in Jordan. J Clin Nurs 2014 Jul;23(13-14):1937-1947. [doi: 10.1111/jocn.12473] [Medline: 24354354]
30. Virtual real. PICO Immersive Pte ltd.: PICO Global; 2023. URL: https://www.picoxr.com/global [accessed 2024-12-19]
31. Unity. Unity Technologies. 2023. URL: https://unity.com/releases/editor/whats-new/2021.3.28 [accessed 2025-01-21]
32. Character creator. Reallusion Inc. 2023. URL: https://www.reallusion.com/character-creator/download.html [accessed 2025-01-21]
33. Index of /release/. Blender Foundation. 2023. URL: https://download.blender.org/release/ [accessed 2025-01-21]
34. Photoshop. Adobe. 2023. URL: https://www.adobe.com/products/photoshop.html [accessed 2025-01-21]
35. Substance 3D painter. Adobe. 2023. URL: https://www.adobe.com/products/substance3d/apps/painter.html [accessed 2025-01-21]
36. Premiere Pro. Adobe. 2023. URL: https://www.adobe.com/products/premiere.html [accessed 2025-01-21]
37. Visual Studio 2019. Microsoft. 2023. URL: https://visualstudio.microsoft.com/vs/older-downloads/ [accessed 2025-01-21]
38. Kolb DA. Experiential Learning: Experience as the Source of Learning and Development: Financial Times Prentice Hall; 1983.
39. Haider A, Harteveld C, Johnson D, et al. miniPXI: development and validation of an eleven-item measure of the Player Experience Inventory. Proc ACM Hum-Comput Interact 2022 Oct 25;6(CHI PLAY):1-26. [doi: 10.1145/3549507]
40. Camacho E, Cohen A, Torous J. Assessment of mental health services available through smartphone apps. JAMA Netw Open 2022 Dec 1;5(12):e2248784. [doi: 10.1001/jamanetworkopen.2022.48784] [Medline: 36576737]
41. Brooke J. SUS: a “quick and dirty” usability scale. In: Usability Evaluation in Industry, 1st edition: CRC Press; 1996:207-212. [doi: 10.1201/9781498710411-35]
42. Barnum C. Product reaction words. UXFirm. 2022. URL: https://www.uxfirm.com/microsofts-product-reaction-cards-unlock-user-satisfaction-part-1 [accessed 2024-09-06]
43. Taylor B, Henshall C, Kenyon S, Litchfield I, Greenfield S. Can rapid approaches to qualitative analysis deliver timely, valid findings to clinical leaders? A mixed methods study comparing rapid and thematic analysis. BMJ Open 2018 Oct 8;8(10):e019993. [doi: 10.1136/bmjopen-2017-019993] [Medline: 30297341]

XSL•FO
RenderX

44. Hennink M, Kaiser BN. Sample sizes for saturation in qualitative research: a systematic review of empirical tests. Soc Sci Med 2022 Jan;292(114523):114523. [doi: 10.1016/j.socscimed.2021.114523] [Medline: 34785096]
45. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas 1960 Apr;20(1):37-46. [doi: 10.1177/001316446002000104]
46. RStudio releases. Posit Software. 2023. URL: https://dailies.rstudio.com/release/ [accessed 2024-10-29]
47. Index of /src/base/R-4. R Foundation. 2023. URL: https://cran.r-project.org/src/base/R-4/ [accessed 2024-10-29]
48. Lang C. Wordcloud2 introduction. 2018. URL: https://cran.r-project.org/web/packages/wordcloud2/vignettes/wordcloud.html [accessed 2024-11-25]
49. Sauro J. Measuring usability with the system usability scale (SUS). Measuring U. 2011. URL: https://measuringu.com/sus/ [accessed 2025-11-25]
50. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: adding an adjective rating scale. J Usability Stud 2009;4(3):114-123 [FREE Full text]
51. Birrenbach T, Wespi R, Hautz WE, et al. Development and usability testing of a fully immersive VR simulation for REBOA training. Int J Emerg Med 2023 Oct 6;16(1):67. [doi: 10.1186/s12245-023-00545-6] [Medline: 37803269]
52. Maltby S, Garcia-Esperon C, Jackson K, et al. TACTICS VR stroke telehealth virtual reality training for health care professionals involved in stroke management at telestroke spoke hospitals: module design and implementation study. JMIR Serious Games 2023 Dec 7;11:e43416. [doi: 10.2196/43416] [Medline: 38060297]
53. Khundam C, Sukkriang N, Noël F. No difference in learning outcomes and usability between using controllers and hand tracking during a virtual reality endotracheal intubation training for medical students in Thailand. J Educ Eval Health Prof 2021;18(22):22. [doi: 10.3352/jeehp.2021.18.22] [Medline: 34403577]
54. Kamińska D, Zwoliński G, Laska-Leśniewicz A. Usability testing of virtual reality applications—the pilot study. Sensors (Basel) 2022 Feb 10;22(4):1342. [doi: 10.3390/s22041342] [Medline: 35214246]
55. Beverly EA, Love C, Love M, Williams E, Bowditch J. Using virtual reality to improve health care providers' cultural self-efficacy and diabetes attitudes: pilot questionnaire study. JMIR Diabetes 2021 Jan 27;6(1):e23708. [doi: 10.2196/23708] [Medline: 33502335]
56. Grassini S, Laumann K, Rasmussen Skogstad M. The use of virtual reality alone does not promote training performance (but sense of presence does). Front Psychol 2020;11(1743):1743. [doi: 10.3389/fpsyg.2020.01743] [Medline: 32765384]
57. Logishetty K, Rudran B, Cobb JP. Virtual reality training improves trainee performance in total hip arthroplasty: a randomized controlled trial. Bone Joint J 2019 Dec;101-B(12):1585-1592. [doi: 10.1302/0301-620X.101B12.BJJ-2019-0643.R1]
58. Dubois E, Gray P, Nigay L, editors. The Engineering of Mixed Reality Systems: Springer London; 2010. [doi: 10.1007/978-1-84882-733-2]
59. Lombard M, Ditton T. At the heart of it all: the concept of presence. J Comput Mediat Commun 2006;3(2):0-0. [doi: 10.1111/j.1083-6101.1997.tb00072.x]

Abbreviations

MiniPXI: mini Player Experience Inventory
NPC: nonplayer character
SUS: System Usability Scale
VR: virtual reality
WPV: workplace violence

Edited by A Coristine; submitted 06.02.25; peer-reviewed by CJ Clay, Y Xiao, Z Aghaei; revised version received 25.06.25; accepted 30.07.25; published 15.09.25.

Please cite as:
Jackson DI, Wongphothiphan T, Luna J, Maa T, Fristad MA, Huang Y, Schaffner B, Reese J, Wheeler J, Abbott B, Sezgin E
Virtual Reality for Workplace Violence Training of Health Care Workers: Pilot Mixed Methods Usability Study
JMIR Serious Games 2025;13:e70817
URL: https://games.jmir.org/2025/1/e70817
doi:10.2196/70817

© Daniel Isaiah Jackson, Thipkanok Wongphothiphan, John Luna, Tensing Maa, Mary A Fristad, Yungui Huang, Brittany Schaffner, Jennifer Reese, Jason Wheeler, Brandon Abbott, Emre Sezgin. Originally published in JMIR Serious Games (https://games.jmir.org), 15.9.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic

XSL•FO
RenderX

information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Effectiveness of Augmented Reality in the Teaching of Health University Students: Quasi-Experimental Study

Rocío Martín-Valero[1], MSc, PhD; Alejandro Vega-Morales Sr[1], MSc; Francisco Javier Martín-Vega[2], MSc, PhD; Manuel Rodriguez-Huguet[2], MSc, PhD; Maria Carmen Rodríguez-Martínez[1,3], MSc, PhD; Maria Jesus Vinolo-Gil[2], MSc, PhD

[1]Department of Physiotherapy, Faculty of Health Science, University of Málaga, Arquitecto Francisco Peñalosa 3, Ampliación de Campus de Teatinos, CTS-1071 Research Group, Málaga, Spain

[2]Department of Nursing and Physiotherapy, University of Cádiz, Cádiz, Spain

[3]Biomedical Research Institute of Malaga-Nanomedicine Platform (IBIMA-BIONAND Platform), Málaga, Spain

Corresponding Author:
Rocío Martín-Valero, MSc, PhD
Department of Physiotherapy, Faculty of Health Science, University of Málaga, Arquitecto Francisco Peñalosa 3, Ampliación de Campus de Teatinos, CTS-1071 Research Group, Málaga, Spain

Abstract

Background: The exponential growth of new technologies has resulted in the need for updating the field of education. From the educational point of view, there are some studies that have promoted the implementation of new technologies. These facts have raised the need to implement augmented reality in the university environment, especially among students of health sciences. The use of augmented reality can mean a new approach to teaching by teachers and better learning by students.

Objective: We aimed to analyze the degree of usability of two augmented reality applications and to compare the academic performance between the control group and the experimental group at the Universities of Cádiz and Málaga. The students at the University of Málaga used the Zapworks augmented reality software, while those at the University of Cádiz used the Aumentaty augmented reality software for their respective experimental groups. The secondary objective was to measure the relationships between all the studied variables.

Methods: This was a quasi-experimental design with a posttest as the only evaluation measure. We followed the SPIRIT (Standard Protocol Items: Recommendations for Interventional Trials) statement and the ethical and legal aspects of the Principles of the Declaration of Helsinki. An intervention was carried out using two augmented reality applications on the subject of General Procedures in Physiotherapy II at the Universities of Málaga and Cádiz.

Results: A total of 199 participants took part in the study. Demographic variables, ratings, and usability were assessed, followed by statistical analysis, with the results and their interpretation being described in the study. Significant differences (P<.001) were found in the ratings at both the universities. In addition, significant differences (P<.001) were found between the experimental group and the control group. Regarding the degree of usability in the univariate analysis, no significant differences were found (P=.049). A multiple regression analysis of the rating and usability was performed. The rating showed significant differences, with a beta of 1.4 (P<.001), and usability was also significant (P=.03) in favor of the Aumentaty group.

Conclusions: Significant differences were observed in those who used augmented reality compared to the control group, with higher values observed in the University of Cádiz. There are no correlations between the variables of usability and qualifications.

Trial Registration: ClinicalTrials.gov NCT05798468; https://clinicaltrials.gov/study/NCT05798468

(JMIR Serious Games 2025;13:e54312) doi:10.2196/54312

KEYWORDS

augmented reality; qualifications; usability; university; teaching; education; implementation; academic performance; quasi-experimental design; control group; applications; experimental group

Introduction

Currently, there is exponential development of new technologies in all areas (education, work, hospitals, etc). Therefore, given this obvious technological development, it is necessary to cover the needs that arise in the areas where they are being implemented, especially in the field of education [1]. New technologies such as computers, telecommunications, and all such devices and materials that enable advances in computer-mediated communication have not only exponentially broadened and improved our ability of handling, processing, and communicating information and knowledge on their own,

XSL•FO
RenderX

but have also assimilated all the other resources that are now integrated into them [2].

Education must not remain on the side lines considering the evolution of new technologies, as they are a part of the basis of today's society. Moreover, education is characterized by the use of the media employed in social communication, which are currently the new technologies [3]. Therefore, there is a need to investigate the use of new technologies in the field of education in order to analyze their use and effectiveness and therefore be able to implement new tools in the university education system.

Augmented reality (AR), understood as the combination of digital and physical information in real time through different technological devices, is becoming one of the emerging technologies with the closest penetration in university studies [4]. AR combines real and virtual objects in a real environment, aligns them together, and operates interactively in real time. It encompasses various senses beyond sight and is not limited to specific display technologies [5]. Additionally, AR includes the concept of overlaying virtual objects to remove or diminish real-world elements, known as mediated or diminished reality [6]. AR is presented as an attractive technology that promises to offer the tools needed to create engaging and motivating content [7-9]. Moreover, AR devices show potential in alleviating adverse health effects like blurred vision, disorientation, and cybersickness typically associated with prolonged utilization of virtual reality [10]. This allows different generic and transversal competences within the university environment to become effective. Furthermore, the use of this type of technology leads to a significant increase in motivation levels [11-14].

In a previous meta-review, the positive effects mentioned were enhanced comprehension of content, acquisition of spatial and language skills, improved long-term memory retention, and increased collaboration and motivation, while the negative effects included encompass narrowed attention, challenges in usability, ineffective integration in the classroom setting, and individual learner variations [15]. The importance of acquiring new abilities, working in enriched and motivating environments and, especially, assessing the possible usefulness of this tool in the performance of university students makes us reflect on the relevance of doing research on this aspect [11]. The significance has been explored in the field of health care, where some studies [16,17] have investigated the use of AR glasses for simulating venous catheter procedures in medical students. Furthermore, Logishetty et al [18] recognized AR as a valuable tool for other surgical procedures related to arthroplasty, as well as for novice ophthalmology [19]. Additionally, Barteit et al [20] presented the notion that incorporating virtual or AR teaching methods in medical education had advantageous outcomes, including heightened enthusiasm and enjoyment for learners.

For physiotherapy students, acquiring foundational knowledge and specific skills in subjects like body structure, biological functions, movement analysis, and motion characteristics through traditional teaching methods may be insufficient [21]. Moreover, it is crucial for them to acquire a diverse range of specialized competencies prior to commencing their professional journey in health care environments [22].

Mastering these skills demands the integration of intricate elements, hypothesis generation, analytical reasoning, and more. Traditional instructional approaches such as lectures, case study demonstrations, tutorials, and practical classes may present challenges in effectively acquiring these proficiencies [23]. However, research has demonstrated that incorporating audio-visual materials can enhance physiotherapy students' enthusiasm for learning while also improving their grasp of theoretical concepts and practical aptitude [24]. Therefore, the use of AR technologies in physiotherapy education holds the potential for addressing challenges and providing innovative ways to enhance learning experiences for students [25-27].

In order to effectively integrate AR into education, it is crucial to design the teaching and learning process to fit the context. Teachers play a vital role in maximizing its benefits by taking on roles as planners, designers, and facilitators of the learning process. Additionally, when designing AR objects, the modules, content, and activities should be concise, interactive, and practical, allowing for student participation and action. The methodology should prioritize interactivity and practicality [28]. Regarding the types of AR according to the classification by Wojciechowski et al [29], it can be categorized into marker-based AR, markerless AR, and location-based AR. Marker-based AR is the most commonly used type (59.3%) in educational settings due to its superior tracking capabilities provided by markers. Marker-less AR, although less prevalent (12.5%), has potential with technologies like Microsoft Kinect for object tracking [30].

Moreover, it is worth noting that there are different software programs that can be used for AR. For the execution of this quasi-experimental design, two AR applications, common in teaching, were used. These applications are Zapworks and Aumentaty. Both types of AR utilize markers. There are other AR studies that address the importance of AR in the university environment within the field of health sciences; in the majority of them, the most important variable was the understanding of knowledge [7,31-33] together with motivation [7,33], qualifications [31,33], and attention and training [7,33]. Other variables considered were the autonomous work provided to students [7,33] and the assessment and use of new technologies [32]. One of the studies used Aumentaty as an AR tool in the same way as it was used in our design [31].

Zapworks is a software that allows the generation and delivery of digital content, which has made the use of AR more accessible. In contrast, Aumentaty is a platform focused on AR projects that allow the possibility of creating this type of technology in a simple way [34,35].

The aim was to compare the degree of usability of the two AR applications mentioned above. Moreover, the purpose was to compare the academic performance of the control group and the experimental group of the Universities of Cádiz and Málaga as a result of the use of the AR applications. Therefore, the null hypothesis will be the presence of significant differences between the control group and the experimental group in the

XSL•FO
RenderX

variables of rating and usability, and the alternative hypothesis would be the absence of significant differences.

Methods

Study Design and Participants

This research was conducted through a quasi-experimental design with a posttest as the only evaluation measure. The design followed the SPIRIT (Standard Protocol Items: Recommendations for Interventional Trials) declaration where the protocols to be followed were established. The initial sample consisted of 203 students of health sciences who were enrolled in the Physiotherapy degree course from the Universities of Málaga and Cádiz. However, 4 participants over the age of 30 years were excluded from the analysis to homogenize the study. The inclusion criteria were students who were attending the Physiotherapy degree course and were enrolled in the subject of General Procedures in Physiotherapy II. The results of the study were measured in the 2021-2022 academic year.

Ethical Considerations

The legal ethical aspects followed the principles of the World Medial Association and the Declaration of Helsinki for ensuring the principles and well-being of the study participants. Before the execution of the research intervention, participants signed an informed consent form in which they were provided with detailed information about the research. Their confidentiality and data protection were guaranteed at all times. Ethical approval for this educational research was obtained from the Andalusian Human Research Ethics Committee (approval number 001/2022). In addition, this study was conducted according to the ethical guidelines for educational research [36]. Participants in this research did not receive compensation.

Study Procedures

The study was carried out in the subject of General Procedures in Physiotherapy II at the Universities of Málaga and Cádiz. Each teacher in charge of the subject explained to the students the possibility of participation in the study and, therefore, the use of the AR applications, with which the general and transversal competences of the subject would be worked on. The topics of the AR program were about electrotherapy, phototherapy, ultrasounds, shock waves, and tecartherapy.

The students were informed and filled in the informed consent form before creating the control group and the experimental group, which would use the AR applications. A control and an experimental group were formed for both universities.

All students belonging to the experimental and control groups participated in the same lecture on electrotherapy, which was part of the second year of the Physiotherapy degree, both at the University of Cádiz and the University of Málaga. The experimental groups used AR applications to complement learning, integrating interactive 3D models that allowed the visualization of anatomical structures and simulations of electrode placement to apply electrotherapy procedures. At the University of Cádiz, the Aumentaty platform was used, while at the University of Málaga, Zapworks was used, although the content was the same in both cases, with the markers being adjusted to each technology. In addition, the applications provided step-by-step instructions for the correct configuration and application of electrotherapy.

In the experimental group, students received a lesson that included the use of AR applications. This complemented the subject content by integrating interactive elements such as links, videos, and 3D models. Students interacted with bookmarks that allowed them to access additional information, which enriched their learning experience.

Both applications used markers (physical or digital) to trigger AR experiences. They are compatible with mobile devices (iOS and Android), allowing users to interact with the content on their own phones. Both allow the visualization of 3D models and additional multimedia content (audio and text). They are used to enrich educational experiences. Nevertheless, there are some differences between Zapworks and Aumentaty; Zapworks is more versatile and advanced, with a visual interface that allows the creation of more customized and complex experiences, while Aumentaty is simpler and oriented to users with less technical experience. Zapworks relies on internet connection to access experiences through a browser or app, while Aumentaty can work offline. Aumentaty is more focused on the educational and cultural environments, with a more restricted use in terms of interactive possibilities, while Zapworks offers more flexibility to create different types of interactive experiences.

The control group received the same content via a traditional lesson without AR. The control group participated in a traditional lecture that consisted of a PowerPoint presentation covering the same content that was presented to the experimental group. This approach allowed for a direct comparison of the effectiveness of traditional teaching with AR-enriched teaching.

The data gathering procedure was carried out in the same way at both universities in order not to influence the results obtained. The System Usability Scale (SUS) [37] was used for this purpose. It is a methodological tool similar to the Likert Scale that measures the usability of an object, device, and application. The scale consists of 10 questions that can be scored from 1 to 5, where 1 means “Strongly disagree” and 5 means “Strongly agree” (Multimedia Appendix 1). First, the student was introduced to the two AR applications and was then asked a series of questions, which were the ones translated from the original test on Usability.gov [37]. The objective was to involve the students in the application of AR as well as to see the degree of agreement or disagreement with it in order to use it correctly and thus guarantee their agreement with the software and their correct learning.

A descriptive analysis of the control and the experimental group was conducted, considering the variables of age and sex. In addition, an individual analysis was conducted for the Rating and Usability variables in each of the Universities. In this analysis, the results of each population were obtained, and subsequently, analysis and interpretation of both variables were carried out together in order to contrast the previously mentioned hypotheses.

XSL•FO
RenderX

Variables Studied

Demographic Variables

Participants were interviewed to obtain information about age and sex. These variables were obtained and analyzed separately according to whether the students belonged to the Universities of Málaga or Cádiz.

Rating Variable

The students' academic performance was assessed by means of multiple-choice examinations at both the study sites. The rating variable considered the academic performance of the students participating in the study. These examinations were designed in a uniform manner to ensure comparability between the two programs, assessing the same key competencies and knowledge. The questions for the examinations were developed in conjunction with the teachers at each site, ensuring that the items reflected the established learning objectives. The grading of the examinations was performed in a standardized manner, using the same scoring criteria at both sites. The information was first analyzed and interpreted separately for the population of the University of Málaga and the University of Cádiz. This allowed for a direct comparative analysis of academic performance between the experimental and control groups.

Usability Variable

The Usability variable considered the use of AR applications in the populations of the Universities of Málaga and Cádiz. To evaluate the usability of AR applications, the SUS, a standard scale widely used in technology studies, was used. The SUS consists of 10 items that measure the perceived ease of use through a 5-point Likert scale, ranging from "Strongly Disagree" to "Strongly Agree" (Multimedia Appendix 1). The final score is obtained by summing the item responses, adjusting the values, and converting the result to a scale from 0 to 100.

In terms of its psychometric properties, the SUS has proven to be a reliable measure, with a Cronbach α coefficient typically exceeding 0.90, reflecting high internal consistency [38]. Its validity has been confirmed in a variety of studies and contexts, including educational settings, making it a suitable tool for the evaluation of the AR applications used in this study.

As with the rating variable, the analysis and interpretation were conducted individually for each population and then jointly using the Student *t* test and the Wilcoxon test. It should be noted that for the University of Cádiz, as the survey was anonymous, the date of birth and sex were used to assign the surveys to the ratings.

Usability and Rating Variables

Aiming to conduct an interpretation and analysis of the correlation between the Usability and Rating variables, the Spearman rank correlation coefficient was determined to establish the existence or absence of a correlation between variables.

Statistical Analysis

The main aim was to compare the degree of usability of the two AR applications used during teaching. Analysis was performed using the R software (version 4.1.2; R Foundation for Statistical Computing). In addition, we also aimed to compare the academic performance of the groups using AR versus the control group and between the two experimental groups, which used two different AR applications.

It was accepted that the control group and the experimental group were composed of different people based on the sex and age variables determined by means of the contingency tables. As the dropout rate was 0% in the usability survey, owing to the same dates of birth, it was not possible to assign a value in the surveys and it was given as a lost value. Therefore, a total sample size of 199 participants was used for the design. First, the hypothesis (H0: $\mu1=\mu2$) of the existence of significant differences between the control group and the experimental group in academic performance was accepted as a general idea, given the significant differences in the variables of Rating and Usability due to the use of AR applications. Otherwise, the hypothesis (H1: $\mu1\neq\mu2$) that there are no significant differences in academic performance due to the use of AR was assumed.

To evaluate the main hypothesis, the probability μ2 was set to the distribution μ1. In this case, the mean μ2 was within the 95% confidence limits, noting that the mean μ3 and μ4 were not within the limits. This indicated that there were significant differences at 95%.

For the analysis of the data obtained from the Rating and Usability variables, the data were checked for incompleteness, normality, and homoscedasticity. In the event that there was no normality or homoscedasticity, the pertinent analyses would be carried out for subsequent interpretation in order to be able to contrast the main hypothesis put forward.

An ANOVA with the Welch correction was performed for the variables age and rating. Two-by-two comparisons were performed using the Holm significance correction because it is the most robust with respect to homoscedasticity. In addition, comparisons of the ratings obtained by the four groups were performed using the Kruskal-Wallis test. For the statistical analysis of the Usability variable, the rating obtained and the usability survey (SUS) were used. A parametric T-Welch test analysis was performed for the Usability variable. Finally, a multiple regression analysis was conducted on the scores obtained by the participants, including sex and age as variables.

Results

Participant Characteristics

The initial sample consisted of 203 participants. However, 4 participants over the age of 30 years were excluded from the analysis to homogenize the study. The total sample analyzed was 199 participants. Out of the total number of participants, 103 belonged to the experimental group and therefore had access to the AR applications. A total of 49 students from the University of Cádiz used the Aumentaty AR artifact (ie, the Aumentaty group), and 54 students from the University of Málaga tested the Zapworks AR artifact (ie, the Zapworks group). A total of 96 participants belonged to the control group, of which 37 were from the University of Cádiz and 59 were from the University of Málaga.

XSL•FO
RenderX

Regarding Demographic Variables

Regarding the population sample from the University of Cádiz, the mean (SD) age was 21.5 (1.2) years, with no significant difference between the control group (mean 21.4, SD 1.0) and the experimental group (mean 21.6, SD 1.3).

In the sample from the University of Málaga, there was a significant difference in the population, with a total mean (SD) age of 20.5 (1.2) years. The control group had a mean (SD) age of 20.9 (0.9) years, while the experimental group had a mean (SD) age of 20.2 (1.3) years. The demographic data are presented in Table 1.

Table . Demographic data (N=199).

Variable	Zapworks control group[a] (n=59)	Aumentaty control group[b] (n=37)	Zapworks group[a] (n=54)	Aumentaty group[b] (n=49)	*P* value[c]
Sex, n (%)					.40
Male	27 (46)	15 (41)	24 (44)	28 (57)	
Female	32 (54)	22 (59)	30 (56)	21 (43)	
Age (years)					<.001
Mean (SD)	20.9 (0.9)	21.4 (1.0)	20.2 (1.3)	21.6 (1.4)	
Median (IQR)	20.8 (20.3-21.2)	21.1 (20.8-22.0)	19.8 (19.5-20.5)	20.9 (20.6-22.4)	
Min, Max	19.2, 24.6	20.3, 25.7	19.0, 25.7	20.3, 26.3	

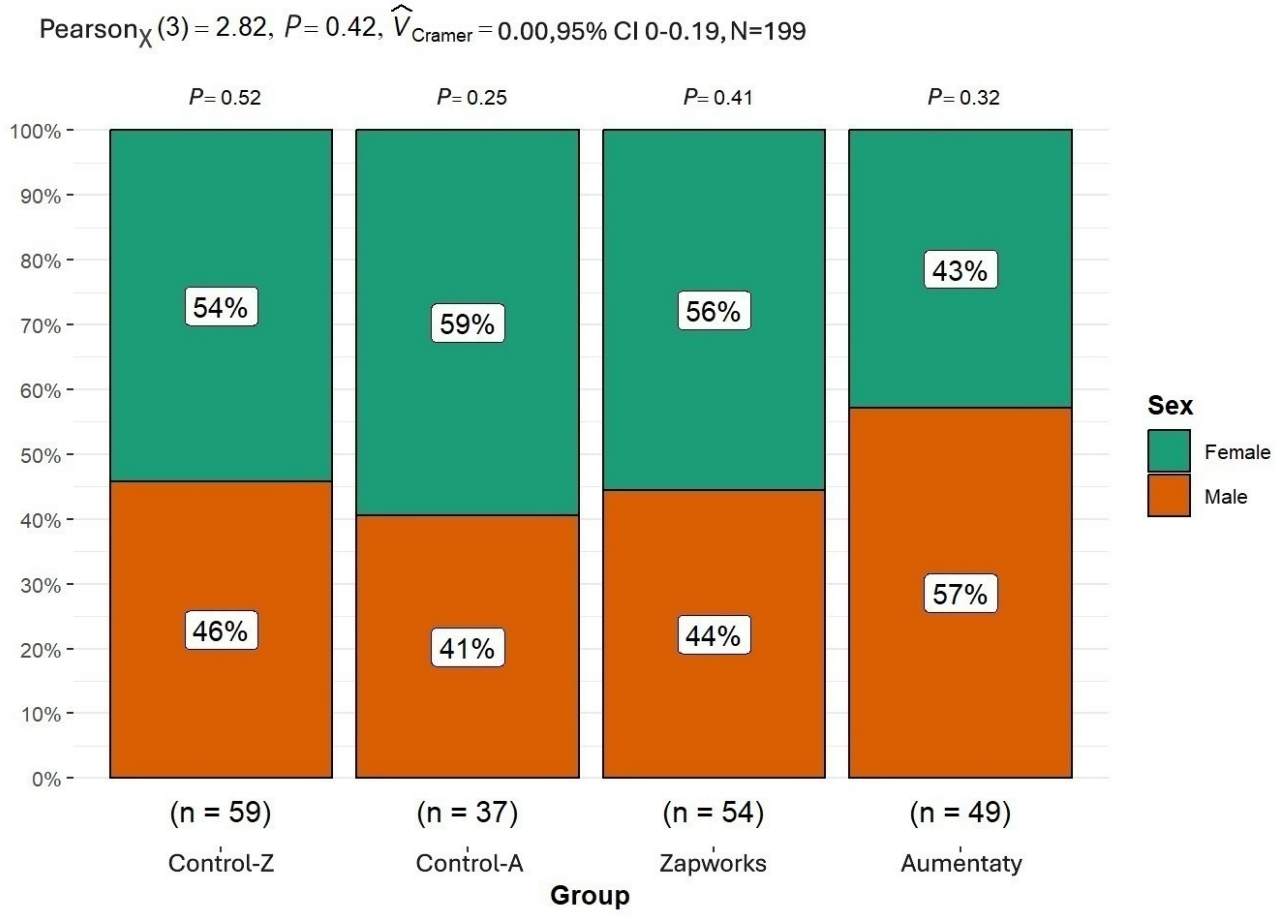
[a]Zapworks: The students of the experimental group at the University of Málaga used Zapworks augmented reality software.

[b]Aumentaty: The students of the experimental group at the University of Cádiz used the Aumentaty augmented reality software.

[c]Pearson $\chi 2$ test or Kruskal-Wallis rank sum test.

Figure 1 shows the distribution of the sexes according to the group, and there were no differences in the distribution of sex among the 4 groups. There were slightly more men in the Aumentaty group, but the Cramer V indicated that the effect size was very small.

Figure 1. Sex difference in the distribution of participants across each group. Control-Z: Zapworks control group; Control-A: Aumentaty control group.



XSL•FO
RenderX

Regarding the Rating Variables

The rating variables showed a significant difference individually in each of the samples in both Universities (P<.001), as presented in Table 2. In addition, Figure 2 shows the comparisons of the rating variables obtained by the 4 groups using the Kruskal-Wallis test. As shown in Figure 2, the ages of the control groups were similar, while the ages of both the control groups differed from those of the intervention groups. Additionally, the ages of the intervention groups differed from each other.

Table . Data for the rating variables.

Variable	N	Zapworks control group (n=59), median (IQR)	Aumentaty control group (n=37), median (IQR)	Zapworks group (n=54), median (IQR)	Aumentaty group (n=49), median (IQR)	*P* value[a]
Age (years)	199	20.86 (20.37-21.17)	21.06 (20.81-22.00)	19.83 (19.46-20.47)	20.99 (20.65-22.45)	<.001
Rating	199	7.40 (6.90-8.10)	7.40 (6.80-8.00)	8.40 (8.10-8.70)	9.00 (8.80-9.00)	<.001
Usability	93	N/A[b]	N/A	61.0 (59.0-64.0)	63.0 (60.0-64.0)	.049

[a]Kruskal-Wallis rank sum test.

[b]N/A: not applicable.

Figure 2. Comparing the data for the ratings variables. Control-Z: Zapworks control group; Control-A: Aumentaty control group.

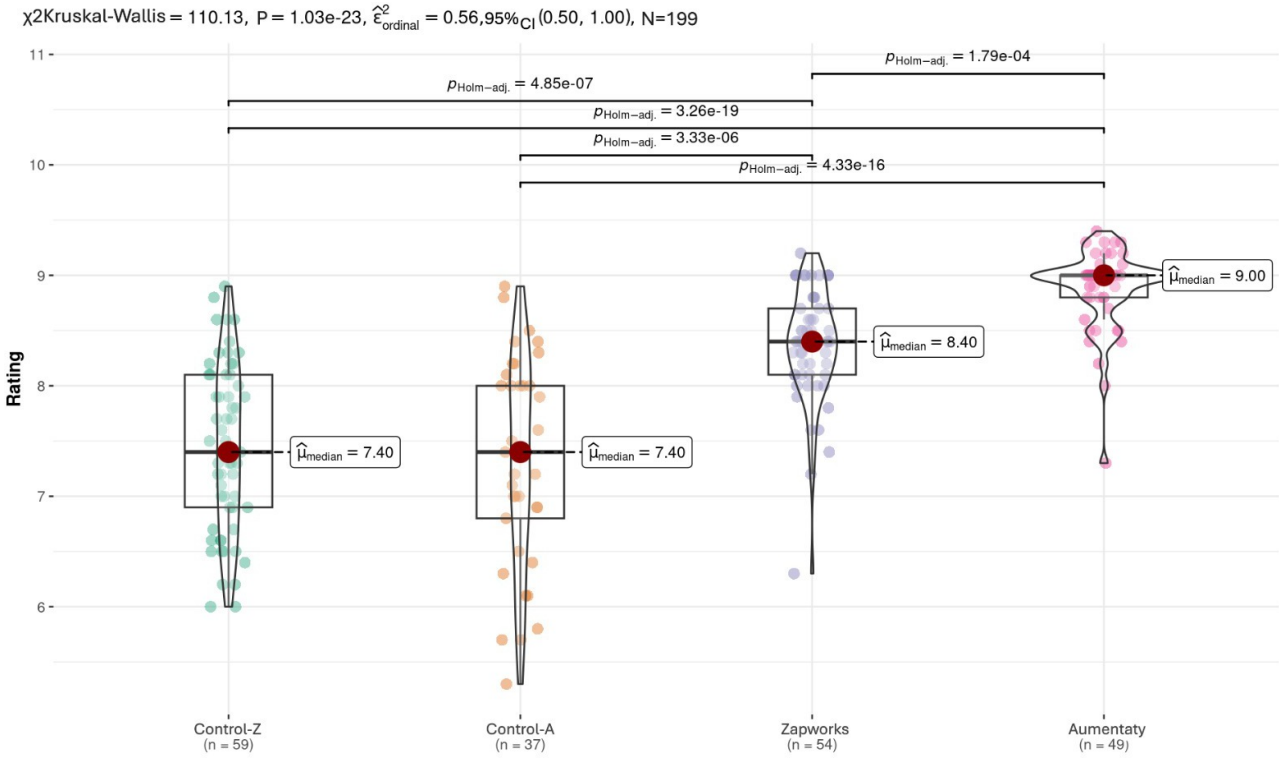
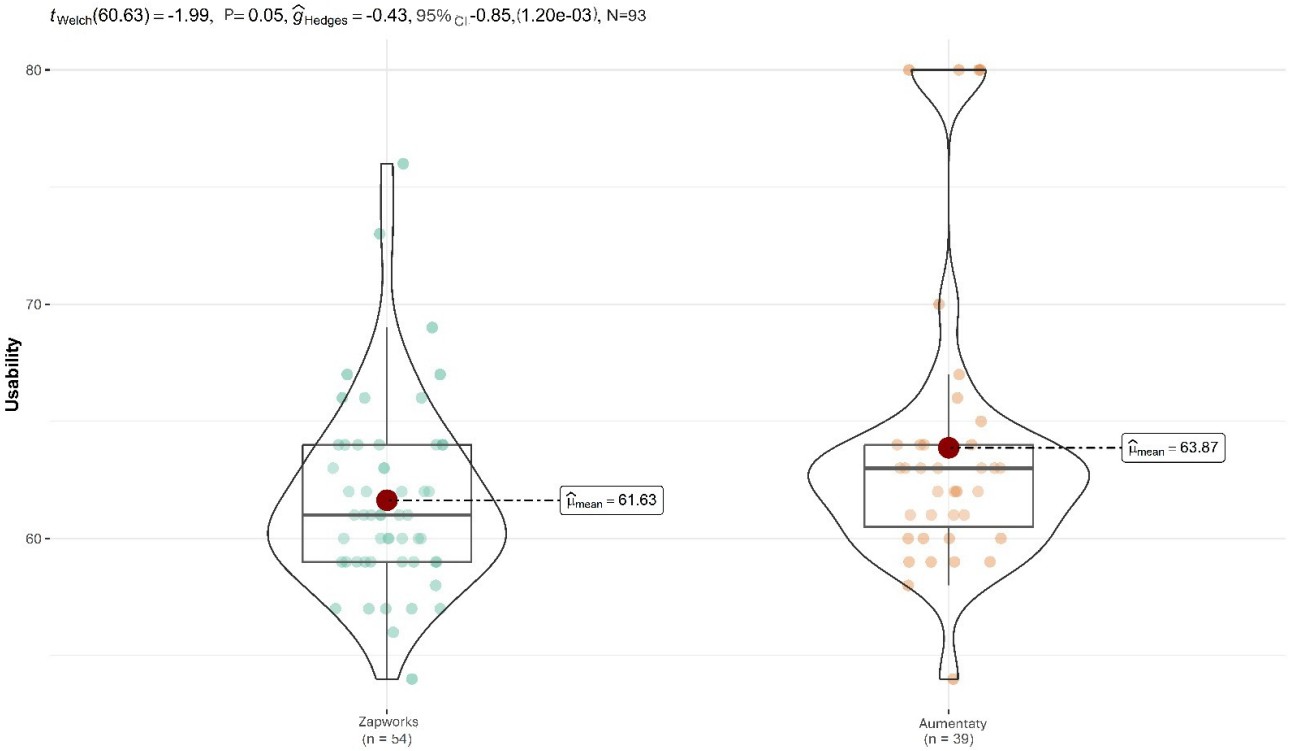


Figure 3 shows the data on the difference between the Usability scores between groups. A moderate difference was observed between the University of Cádiz (mean 63.8, SD 6.15) and the University of Málaga (mean 61.6, SD 3.99), with a Hedges' g of 0.43. However, this difference was marginally non-significant (P=.05)

XSL•FO
RenderX

Figure 3. Univariate analysis of the usability variable.



Finally, a multiple regression analysis was conducted on the scores obtained by the participants, including sex and age as variables. The results from both parametric and nonparametric analyses indicated that neither sex nor age had a significant effect on the participants' scores. However, significant differences were found between students from the University of Málaga and the University of Cádiz (*P*<.001), with a beta coefficient of 1.4 favoring the Aumentaty group. Additionally, significant differences were observed between the control-A groups and the Zapworks group (*P*<.001), with a beta coefficient of 0.92 for the experimental group from Málaga, as presented in Table 3.

Table . Multiple regression analysis for the rating variables.

Characteristic	N	β	SE	95% CI	*P* value
Group					
Zapworks control	59	__[a]	—	—	—
Aumentaty control	37	−0.14	0.142	−0.42 to 0.15	.30
Zapworks	54	.92[b]	0.129	0.66 to 1.2	<.001
Aumentaty	49	1.4[b]	0.133	1.2 to 1.7	<.001
Sex					
Male	94	—	—	—	—
Female	105	.07	0.096	−0.12 to 0.26	.40
Age	199	.00	0.041	−0.08 to 0.08	>.99

[a]Not available.

[b]*P*<.001.

In Table 4, a multiple regression analysis was performed for the Usability variable, including the variable of sex. In this case, the differences between groups, which were on the borderline of significance, became more clearly significant. The Usability variable showed a significant difference (*P*=.03) in favor of the Aumentaty group as shown in Table 4.

XSL•FO
RenderX

Table . Multiple regression analysis for the Usability variable.

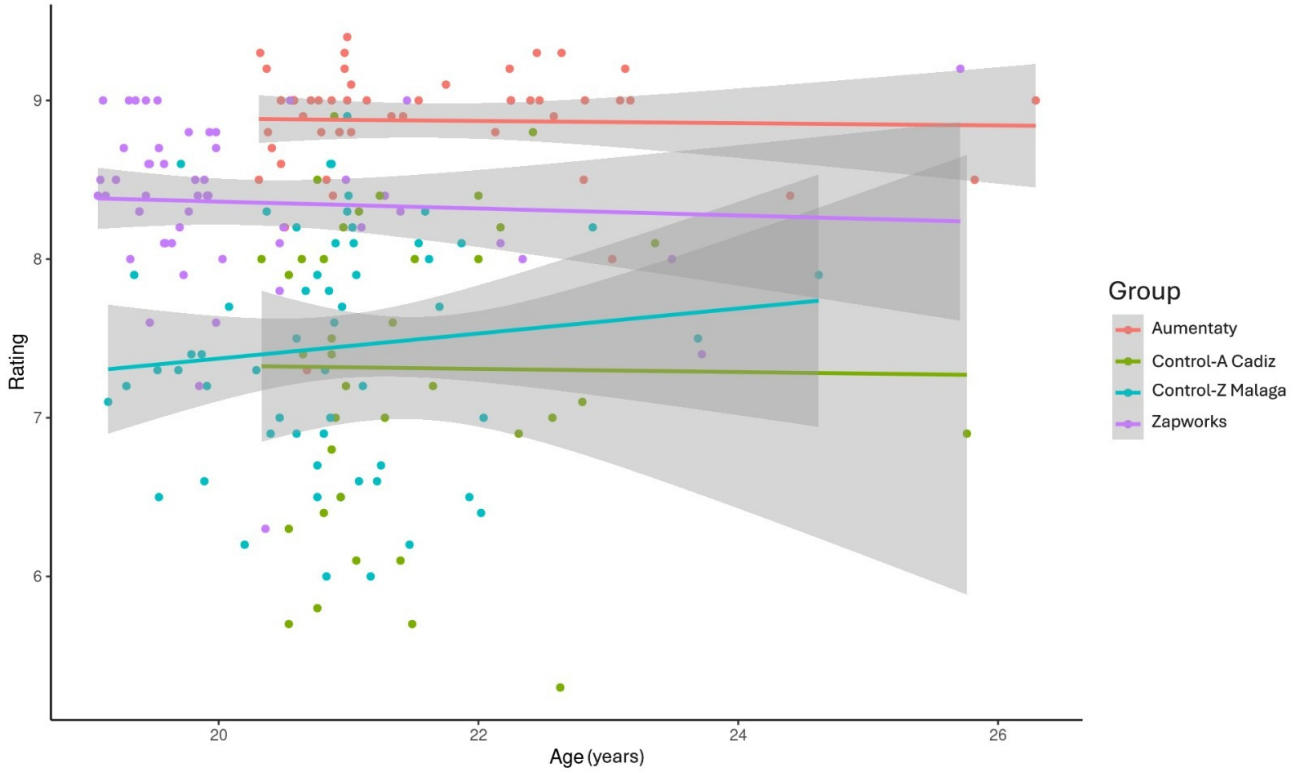
Characteristic	N	Beta	SE	95% CI	*P* value[a]
Group					
Zapworks	54	—[b]	—	—	—
Aumentaty	39	2.7	1.19	0.3 to 5.0	.02
Sex					
Male	46	—	—	—	—
Female	47	.66	1.06	−1.4 to 2.8	.50
Age	93	−.26	0.42	−1.1 to 0.6	.50

[a]*P*<.05.

[b]Not applicable.

Figure 4 shows how age affects the scores; in terms of scores, age differences between individuals did not matter. A similar performance in scores was observed among participants from the control groups of both the universities. However, it should be noted that including students over the age of 26 years in the analysis might have influenced the scores obtained, depending on the AR technique used.

Figure 4. Relationship between age, group, and rating. Control-Z: Zapworks control group; Control-A: Aumentaty control group.



Discussion

Study Findings and Comparison With Previous Studies

The aim of this quasi-experimental design with the posttest was only to assess the use of two AR applications in two university populations and their influence on academic performance. It also sought to compare both populations in terms of the applications used and the academic performance.

The results of this study confirm the theoretical hypothesis that AR can enhance the learning experience compared to traditional methods. AR allows for greater interactivity with the content, which, as indicated by previous studies [39], can foster greater retention of information and increased student motivation. In contrast, the control group, who used traditional PowerPoint presentations, showed a more passive approach to learning, which supports previous research pointing to the limitations of traditional methodologies in capturing students' continued interest [40].

From a theoretical point of view, comparison with the control group is fundamental to understand how AR changes the dynamics of learning. By offering a more immersive and participatory experience, AR can be aligned with constructivist theories of learning [41], which emphasize the active role of the learner in the construction of knowledge. This study provides evidence that AR not only enriches visual and contextual

XSL•FO
RenderX

content, but can also alter the role of the learner within the learning process, making it more dynamic and proactive. However, it is important to note that the effectiveness of AR may depend on factors such as lesson design and students' prior technological experience, which raises areas for future research.

When specifically examining the studies that utilized AR, positive results were found. First, the projection condition in physiotherapy education showed superior performance across various cognitive load dimensions compared to traditional teaching and tablet conditions; additionally, the use of AR and the projection of anatomical images resulted in a significant increase in the questionnaire score for the experimental group [42].

Second, significant differences were noted in students' self-confidence regarding their learning abilities [39]. In another study, Kandasamy et al [43] implemented an AR application for teaching spine anatomy, which resulted in significant differences in students' perceived comprehension and level of engagement [43].

Other studies have also yielded positive results in physiotherapy students; the researchers utilized a questionnaire based on the DeLone and McLeanís model [44] to evaluate the effectiveness of the educational experience. The main findings revealed that participants had a positive perception of the Augmented Studio experience, with high ratings in satisfaction, learner engagement, and system effectiveness [27].

In another study conducted by Pérez et al [28], the majority of students (89.8%) considered AR to be highly relevant in their educational processes. Regarding its usefulness and ease of use, 51.3% of respondents highlighted initial difficulties with familiarizing themselves with the AR app design and digital object creation. However, 82.1% viewed the incorporation of AR as highly positive for learning, finding it interesting (83.4%) and engaging, fostering active participation and reflection. Despite potential challenges, 79.5% of students recommended the use of AR and its apps as valuable resources for learning, promoting creativity and innovation [28]. Regarding the limitations of AR, a study conducted by Bacca et al [30] highlighted the significant challenge of maintaining superimposed information, particularly in marker-based AR applications; this limitation, reported at a rate of 9.3%, can result in student frustration and necessitates the need for improvements.

In a previous study about AR in anatomy learning, it was shown that there were significant differences (P<.0001) in the grades of students who belonged to the group that had used AR, showing an improvement in knowledge about anatomy [7]. This truly relevant fact is one of the results obtained in our design, where it was observed that the grades obtained in each sample (Universities of Málaga and Cádiz) were also significant (P<.001). These results showed a significant improvement in grades when AR is used in university teaching.

In a study similar to ours [31], which used Aumentaty with students from the University of Cádiz in the Physiotherapy degree, satisfactory results were observed in the improvement of knowledge comprehension, which translated into an improvement in the grades. The same happened in this study, when an interpretation of the Rating variable was conducted jointly for both universities through univariate and multiple regression analyses [31]. In the previous study, on the first univariate analysis for rating, the significant differences (P<.001) with a beta of 1.4 were observed in the sample from the University of Cádiz [31]. Furthermore, when comparing the control group with the Zapworks group between the two universities, the significant differences (P<.001) with a beta of .92 was observed, similar to the study where the significant results were in the group that used Aumentaty [31].

In another study that evaluated the use of AR in learning, a significant improvement in the visualization of the content was observed (P<.0001) in the group that had used AR. In addition, 3D compression, which is fundamental for the acquisition of knowledge in anatomy, was significantly greater in the experimental group with higher percentages compared to the control group [33]. Therefore, the significant differences were found in the group using AR and the results found were similar to the results (P<.001) observed in this study where an improvement in ratings was observed.

In a previous study where the usability was studied, the use of AR, one of the variables related to our design, was addressed [32]. The sample was taken from the participants who were obtaining a Degree in Nursing at the University of Valladolid, where 85% of the students who were familiar with AR believed that it offers many educational possibilities and, therefore, 86.67% thought that it could be implemented in the Faculty of Nursing [32]. Using these data, 93.33% responded that it would enhance their learning, which is why it is important to analyze and interpret the usability of this tool in Health Sciences students [32]. However, in our design, no significant differences were obtained (P=.11) in the degree of usability, which is why it was not possible to obtain conclusive facts about this variable.

Limitations and Strengths

This design has a number of strengths. During the development of the quasi-experimental design, the sample size was adequate and there were similarities in the subject matter at both universities. The main limitation was that the control group received conventional teaching and therefore cannot be considered a pure control group.

Conclusion

Significant differences are observed between the group using AR applications compared to the control group in both universities and in the two univariate contrasts. In the multiple regression analysis, which explains the greatest variability regarding the main effects, a greater effect is detected in the use of AR at the University of Cádiz compared to the University of Málaga. Finally, the results of the multiple regression analysis indicates that there are no effects of sex and age on the scores obtained by the participants in the study.

XSL•FO
RenderX

Acknowledgments

The quasi-experimental design was funded by a grant from the University of Málaga's own research initiation grant scheme. This research was partially funded by the “Plan Propio” program of the University of Cádiz. The authors would like to thank all participants for their contribution to the design presented. The authors are grateful for Javier Baron’s statistical support. We would also like to thank Marina Heredia for her supervision of the design in English.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questions from Usability.gov
[DOCX File, 18 KB - games_v13i1e54312_app1.docx]

References

1. Correa JM, de Pablos J. Nuevas tecnologías e innovación educativa [Article in Spanish]. Revista de Psicodidáctica 2009;14:133-145 [FREE Full text]
2. Lorido MP. Nuevas tecnologías y educación [Article in Spanish]. Cad Psicopedag 2005;5(9) [FREE Full text]
3. Sánchez AM. El uso de las nuevas tecnologías en el profesorado universitario [Article in Spanish]. Revista de Medios y Educación 2007 Jul 30:61-72 [FREE Full text]
4. Almenara JC, Osuna JB, Obrador M. Realidad aumentada aplicada a la enseñanza de la medicina [Article in Spanish]. Educación Médica 2017 Jul;18(3):203-208. [doi: 10.1016/j.edumed.2016.06.015]
5. Azuma R, Baillot Y, Behringer R, Feiner S, Julier S, MacIntyre B. Recent advances in augmented reality. IEEE Comput Grap Appl 2001;21(6):34-47. [doi: 10.1109/38.963459]
6. Van Krevelen DWF, Poelman R. A survey of augmented reality technologies, applications and limitations. IJVR 2010;9(2):1-20. [doi: 10.20870/IJVR.2010.9.2.2767]
7. Lorido MP. Nuevas tecnologías y educación [Article in Spanish]. Cadernos de Psicopedagogia 2005;5(9) [FREE Full text]
8. Liu TY, Tan TH, Chu YL. Outdoor natural science learning with an RFID-supported immersive ubiquitous learning environment. Educ Technol Soc 2009;12:161-175.
9. Gonzalez-Vidal IM, López BC, Otero LC. La Revista Iberoamericana de La Educación Digital | Método de enseñanza | Educación a distancia [Article in Spanish]. Scribd. URL: https://es.scribd.com/document/600651372/142 [accessed 2023-06-10]
10. Moro C, Štromberga Z, Raikos A, Stirling A. The effectiveness of virtual and augmented reality in health sciences and medical anatomy. Anat Sci Educ 2017 Nov;10(6):549-559. [doi: 10.1002/ase.1696] [Medline: 28419750]
11. Cabero-Almenara J, Vázquez-Cano E, Villota-Oyarvide WR, López-Meneses E. La innovación en el aula universitaria a través de la realidad aumentada. Análisis desde la perspectiva del estudiantado español y latinoamericano [Article in Spanish]. Rev Electr Educare 2021 Jun 2;25(3):1-17. [doi: 10.15359/ree.25-3.1]
12. Cordova DI, Lepper MR. Intrinsic motivation and the process of learning: Beneficial effects of contextualization, personalization, and choice. J Educ Psychol 1996;88(4):715-730. [doi: 10.1037//0022-0663.88.4.715]
13. Kaufmann H, Dünser A. Summary of usability evaluations of an educational augmented reality application. In: Shumaker R, editor. Virtual Reality: Springer; 2007:660-669. [doi: 10.1007/978-3-540-73335-5_71]
14. Juan CM, Toffetti G, Abad F, Cano J. Tangible cubes used as the user interface in an augmented reality game for edutainment. Presented at: 10th IEEE International Conference on Advanced Learning Technologies; Jul 5-7, 2010; Sousse, Tunisia.
15. Radu I. Augmented reality in education: a meta-review and cross-media analysis. Pers Ubiquit Comput 2014 Aug;18(6):1533-1543. [doi: 10.1007/s00779-013-0747-y]
16. Huang CY, Thomas JB, Alismail A, et al. The use of augmented reality glasses in central line simulation: “see one, simulate many, do one competently, and teach everyone”. Adv Med Educ Pract 2018;9:357-363. [doi: 10.2147/AMEP.S160704] [Medline: 29785148]
17. Rochlen LR, Levine R, Tait AR. First-person point-of-view-augmented reality for central line insertion training: a usability and feasibility study. Simul Healthc 2017 Feb;12(1):57-62. [doi: 10.1097/SIH.0000000000000185] [Medline: 27930431]
18. Logishetty K, Western L, Morgan R, Iranpour F, Cobb JP, Auvinet E. Can an augmented reality headset improve accuracy of acetabular cup orientation in simulated THA? A randomized trial. Clin Orthop Relat Res 2019 May;477(5):1190-1199. [doi: 10.1097/CORR.0000000000000542] [Medline: 30507832]
19. Rai AS, Rai AS, Mavrikakis E, Lam WC. Teaching binocular indirect ophthalmoscopy to novice residents using an augmented reality simulator. Can J Ophthalmol 2017 Oct;52(5):430-434. [doi: 10.1016/j.jcjo.2017.02.015] [Medline: 28985799]

XSL•FO
RenderX

20. Barteit S, Lanfermann L, Bärnighausen T, Neuhann F, Beiersmann C. Augmented, mixed, and virtual reality-based head-mounted devices for medical education: systematic review. JMIR Serious Games 2021 Jul 8;9(3):e29080. [doi: 10.2196/29080] [Medline: 34255668]
21. Kurul R, Ögün MN, Neriman Narin A, Avci Ş, Yazgan B. An alternative method for anatomy training: immersive virtual reality. Anat Sci Educ 2020 Sep;13(5):648-656. [doi: 10.1002/ase.1959] [Medline: 32163659]
22. Hurst KM. Using video podcasting to enhance the learning of clinical skills: A qualitative study of physiotherapy students' experiences. Nurse Educ Today 2016 Oct;45:206-211. [doi: 10.1016/j.nedt.2016.08.011] [Medline: 27552715]
23. Hoang T, Reinoso M, Joukhadar Z, Vetere F, Kelly D. Augmented studio: projection mapping on moving body for physiotherapy education. Presented at: CHI '17: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems; May 6-11, 2017; Denver, Colorado. [doi: 10.1145/3025453.3025860]
24. Lucena-Anton D, Fernandez-Lopez JC, Pacheco-Serrano AI, Garcia-Munoz C, Moral-Munoz JA. Virtual and augmented reality versus traditional methods for teaching physiotherapy: a systematic review. Eur J Investig Health Psychol Educ 2022 Dec 2;12(12):1780-1792. [doi: 10.3390/ejihpe12120125] [Medline: 36547026]
25. Blum T, Kleeberger V, Bichlmeier C, Navab N. Mirracle: an augmented reality magic mirror system for anatomy education. Presented at: 2012 IEEE Virtual Reality Workshops (VRW); Mar 4-8, 2012; Costa Mesa, CA. [doi: 10.1109/VR.2012.6180909]
26. Johnson AS, Sun Y. Spatial augmented reality on person: exploring the most personal medium. In: Shumaker R, editor. Virtual Augmented and Mixed Reality Designing and Developing Augmented and Virtual Environments: Springer; 2013:169-174. [doi: 10.1007/978-3-642-39405-8_20]
27. Kelly D, Hoang TN, Reinoso M, Joukhadar Z, Clements T, Vetere F. Augmented reality learning environment for physiotherapy education. Phys Ther Rev 2018 Jan 2;23(1):21-28. [doi: 10.1080/10833196.2018.1447256]
28. Pérez SM, Robles BF, Osuna JB. La realidad aumentada como recurso para la formación en la educación superior [Article in Spanish]. Campus Virtuales 2021 Jan 31;10(1):9-19.
29. Wojciechowski R, Cellary W. Evaluation of learners' attitude toward learning in ARIES augmented reality environments. Comput Educ 2013 Oct;68:570-585. [doi: 10.1016/j.compedu.2013.02.014]
30. Bacca J, Baldiris S, Fabregat R, Graf S. Mobile augmented reality in vocational education and training. Procedia Comput Sci 2015;75:49-58. [doi: 10.1016/j.procs.2015.12.203]
31. Viñolo-Gil MJ, Vega FJM, Rodríguez-Huguet M. Uso de la realidad aumentada para mejorar el aprendizaje en estudiantes del grado de fisioterapia [Article in Spanish]. Innovación y Mejora Docente 2020 [FREE Full text]
32. García Herrero SB. Tics en el proceso de enseñanza-aprendizaje en enfermería: posibilidades de la realidad aumentada [Article in Spanish]. Universidad de Valladolid. 2018. URL: https://uvadoc.uva.es/handle/10324/30508 [accessed 2025-03-06]
33. Torregrosa JF. Desarrollo, estudio y evaluación de contenidos didácticos mediante realidad aumentada en la formación de graduados de podología universidad católica de valencia san vicente Mártir [Article in Spanish]. Fundación Dialnet. 2014. URL: https://dialnet.unirioja.es/servlet/tesis?codigo=118075 [accessed 2025-03-06]
34. Reeves LE, Bolton E, Bulpitt M, et al. Use of augmented reality (AR) to aid bioscience education and enrich student experience. Research in Learning Technology 2021 Jan 15;29. [doi: 10.25304/rlt.v29.2572]
35. Blázquez Sevilla A. Manual Aumentaty: Herramienta de realidad aumentada [Article in Spanish]. Archivo Digital UPM. 2018. URL: https://oa.upm.es/51588/ [accessed 2025-03-06]
36. Otero LR, Casado-Muñoz R. Guía Ética Para La Investigación Educativa [Book in Spanish]: Asociación Británica de Investigación Educativa [BERA]; 2019. URL: https://investigacion.ubu.es/documentos/623326365fa07802f2a55012 [accessed 2025-03-17]
37. Sevilla-Gonzalez MDR, Moreno Loaeza L, Lazaro-Carrera LS, et al. Spanish version of the System Usability Scale for the assessment of electronic tools: development and validation. JMIR Hum Factors 2020 Dec 16;7(4):e21161. [doi: 10.2196/21161] [Medline: 33325828]
38. Lewis JR, Sauro J. The factor structure of the System Usability Scale. In: Kurosu M, editor. Human Centered Design: Springer; 2009, Vol. 5619:94-103 URL: http://link.springer.com/10.1007/978-3-642-02806-9_12 [accessed 2025-03-17] [doi: 10.1007/978-3-642-02806-9_12]
39. Billinghurst M, Duenser A. Augmented reality in the classroom. Computer (Long Beach Calif) 2012;45(7):56-63. [doi: 10.1109/MC.2012.111]
40. Keengwe J, Georgina D, Wachira P. Faculty training strategies to enhance pedagogy-technology integration. Int J Inf Commun Technol Educ 2010 Jul;6(3):1-10 [FREE Full text] [doi: 10.4018/jicte.2010070101]
41. Blikstein P. Digital fabrication and 'making' in education. In: Walter-Herrmann J, Büching C, editors. FabLab: transcript Verlag; 2014:203-222. [doi: 10.1515/transcript.9783839423820.203]
42. Ferdous HS, Hoang T, Joukhadar Z, Reinoso MN, Vetere F, Kelly D, et al. "What's happening at that hip?": evaluating an on-body projection based augmented reality system for physiotherapy classroom. Presented at: CHI '19: CHI Conference on Human Factors in Computing System; May 4-9, 2019; Glasgow, Scotland, UK p. 1-12. [doi: 10.1145/3290605.3300464]
43. Kandasamy G, Bettany-Saltikov J, Cordry J, McSherry R. Use of vision-based augmented reality to improve student learning of the spine and spinal deformities. An exploratory study. S Afr J Physiother 2021;77(2):1579. [doi: 10.4102/sajp.v77i2.1579] [Medline: 34859163]

XSL•FO
RenderX

44. Delone WH, McLean ER. The DeLone and McLean model of information systems success: a ten-year update. Journal of Management Information Systems 2003 Apr;19(4):9-30. [doi: 10.1080/07421222.2003.11045748]

Abbreviations

AR: augmented reality
SPIRIT: Standard Protocol Items: Recommendations for Interventional Trials
SUS: System Usability Scale

Edited by A McDaniel; submitted 05.11.23; peer-reviewed by G Myreteg, J Buchner, MJ Ibarra-Cabrera, R Baeza; revised version received 16.10.24; accepted 18.10.24; published 27.03.25.

Please cite as:
Martín-Valero R, Vega-Morales Sr A, Martín-Vega FJ, Rodriguez-Huguet M, Rodríguez-Martínez MC, Vinolo-Gil MJ
Effectiveness of Augmented Reality in the Teaching of Health University Students: Quasi-Experimental Study
JMIR Serious Games 2025;13:e54312
URL: https://games.jmir.org/2025/1/e54312
doi:10.2196/54312

© Rocío Martín-Valero, Alejandro Vega-Morales Sr, Francisco Javier Martín-Vega, Manuel Rodriguez-Huguet, Maria Carmen Rodríguez-Martínez, Maria Jesus Vinolo-Gil. Originally published in JMIR Serious Games (https://games.jmir.org), 27.3.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

A Mixed Reality–Based Telesupervised Ultrasound Education Platform on 5G Network Compared to Direct Supervision: Prospective Randomized Pilot Trial

Minha Kim[1,2*], MD; Meong Hi Son[1,3,4*], MD; Suhyeon Moon[5], MS; Won Chul Cha[1,3,4], MD, PhD; Ik Joon Jo[1], MD, PhD; Hee Yoon[1*], MD, PhD

[1]Department of Emergency Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea

[2]Department of Medical Sciences, Graduate School of Kangwon National University, Chuncheon, Republic of Korea

[3]Department of Digital Health, Samsung Advanced Institute for Health Science & Technology, Sungkyunkwan University, Seoul, Republic of Korea

[4]Data Innovation Center, Samsung Medical Center, Seoul, Republic of Korea

[5]Division of Biostatistics, Department of Academic Research, Kangbuk Samsung Hospital, Seoul, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Hee Yoon, MD, PhD
Department of Emergency Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea

Related Article:

This is a corrected version. See correction statement: https://games.jmir.org/2025/1/e77586

Abstract

Background: Ultrasound education is transitioning from in-person training to remote methods using mixed reality (MR) and 5G networks. Previous studies are mainly experimental, lacking randomized controlled trials in direct training scenarios.

Objective: This study aimed to compare an MR-based telesupervised ultrasound education platform on private 5G networks with traditional in-person training for novice doctors.

Methods: Conducted at a tertiary academic hospital from November to December 2023, the prospective unblinded randomized controlled pilot study assigned doctors without prior abdominal ultrasound education experience to either the telesupervision group (TG; n=20) or direct supervision group (DG; n=20). Participants received a 15-minute video lecture, conducted ultrasound on a phantom, and had 18 images scored by 2 blinded experts. Additionally, the TG received 5 minutes of training on the basic operation of a head-mounted display. Communication between doctors in the TG and supervisors was facilitated through a head-mounted display, whereas those in the DG interacted directly with supervisors. Primary outcomes were image quality scores, while secondary outcomes included procedure time, number of supervisor interventions, user experience using National Aeronautics and Space Administration-Task Load Index (NASA-TLX), System Usability Scale (SUS), and self-confidence through pre- and postsurveys.

Results: Image quality scores and procedure times showed no significant differences between the groups (TG: 66.8 [SD 10.3] vs DG: 66.8 [SD 10.4], *P*=.84; TG: 23.8 [SD 8.0] min vs DG: 24.0 [SD 8.1] min, *P*=.95, respectively). However, the TG engaged in more educational interventions (TG: 4.0 [SD 2.5] vs DG: 0.8 [SD 1.1], *P*<.001), reflecting a more interactive training environment. TG participants reported lower NASA-TLX scores for mental demand (43.8 [SD 24.8] vs 60.6 [SD 22.4], *P*=.03), effort (43.1 [SD 22.9] vs 67.9 [SD 17], *P*<.001), and frustration (26.9 [SD 20.3] vs 45.2 [SD 27.8], *P*=.02), indicating a reduced cognitive load compared to the DG. The mean SUS score was also higher in the TG (66.6 [SD 9.1] vs 60.2 [SD 10.4], *P*=.046), suggesting better usability. Both groups showed significant improvements in confidence, with the TG showing notably greater improvement in abdominal ultrasound proficiency (pre-education TG: 1.6 [SD 0.9] vs DG: 1.7 [SD 0.9], *P*=.73; post-education TG: 3.8 [SD 0.9] vs DG: 2.8 [SD 1.0], *P*=.006).

Conclusions: Although no significant differences in image quality scores were observed between groups, considerable differences in positive educational interactions, workload, and usability were evident. These findings emphasize the platform's potential to enhance the ultrasound training experience, suggesting more interactive and efficient learning.

Trial Registration: ClinicalTrials.gov NCT06171828; https://clinicaltrials.gov/study/NCT06171828

XSL•FO
RenderX

(JMIR Serious Games 2025;13:e63448) doi:10.2196/63448

KEYWORDS

ultrasonography; telemedicine; medical education; distance learning; fifth-generation network; mixed reality; ultrasound education; hospital; randomized pilot trial; pilot study; doctor; telesupervision; head-mounted display; primary outcomes; user experience; self-confidence; image quality; educational intervention; training experience; South Korea; telehealth

Introduction

Ultrasounds, a low-cost, noninvasive, real-time imaging modality, have faced a paradigm shift in training methods due to the COVID-19 pandemic [1-6]. Ultrasonography is skill-intensive and operator-dependent, typically requiring hands-on practice with supervision [1]. However, technological advancements, the mismatch between global medical technology supply and demand, and social distancing measures due to infectious diseases have accelerated the transition of ultrasound education from traditional in-person training to digital training [7]. Some studies have shown no significant differences between ultrasound novices learning basic hands-on skills and basic ocular ultrasound skills through online self-learning compared to hands-on learning [8,9]. However, the real-time imaging characteristics of ultrasound necessitate real-time supervision during training [10]. Current tele-education methods, such as mobile phone apps, videoconferencing devices, and telemedicine software, have been used to connect learners with supervisors, enabling them to share and discuss ultrasound images [2,10]. Nevertheless, these methods present several issues. They transmit the ultrasound image but fail to show how the learner is holding the probe, thus hindering the supervisor's ability to provide effective feedback on probe handling. Moreover, because the learner needs one hand to operate the probe, using the other to hold a communication device prevents effective manipulation of any additional materials or instruments.

To address these challenges, mixed reality (MR) systems have recently gained prominence in medical education [11-13]. MR involves merging physical objects with virtual information to facilitate interaction between the two [14]. Previous studies have explored the application of MR in contactless ultrasound education, demonstrating its potential to enhance the realism of educational experiences [15-17]. MR technology applies a virtual overlay to ultrasound images, allowing learners to study anatomy and pathology in a more immersive and interactive manner [18]. It also allows for immediate feedback and correction within a simulated clinical setting without requiring the supervisor to be physically present [19]. A key element of incorporating MR into ultrasound education is minimizing the delay in ultrasound image transmission between learners and supervisors. This can be achieved by using fifth-generation (5G) networks to ensure optimal transmission speeds [20]. 5G wireless networks reliably transmit data at speeds of up to 100 times faster than previous generations (10 gigabits per second), reducing latency to 1 - 2 ms and multiplying simultaneous device connectivity by 100 [21]. The use of a 5G network is critical for minimizing latency during the transmission of ultrasound images, thereby enabling effective remote supervision.

Previous studies using MR systems and 5G networks have been mainly experimental, and there is a lack of randomized controlled trials evaluating the educational efficacy of these technologies in direct training scenarios [22]. Using MR on a 5G network, we aimed to assess the educational effectiveness of remote telesupervision with head-mounted display (HMD) versus traditional direct supervision, analyze the feasibility of the tele-education platform by surveying the workload and usability of novice sonographers, and address the prevalent educational challenges of ultrasound, which are marked by a scarcity of faculty, inadequate equipment, and limited training time. This study can be used as a reference to improve ultrasound training that emphasizes the tactile and manual skills essential to the discipline.

Methods

Study Design

This prospective randomized controlled pilot study was conducted at a tertiary academic hospital in South Korea from November 2023 to December 2023. The study aimed to compare an MR-based telesupervised ultrasound education platform using private 5G networks with traditional in-person training for novice doctors.

Participants

We recruited medical doctors from the Samsung Medical Center interested in ultrasound education with supervision through online and offline bulletin boards. Participants had no prior experience with abdominal ultrasonography. To oversee the ultrasound examinations performed by each participant, the study engaged 2 faculty members from the emergency medicine (EM) department to provide supervision.

Study Protocol

Overall, 40 participants were randomly assigned to the direct supervision group (DG) and telesupervision group (TG) using the lottery method. First, all participants completed a pre-survey and watched a 15-minute video lecture created using PowerPoint (Microsoft) for abdominal ultrasound education. The researcher provided an additional 5-minute training for the TG on basic HMD operations, including essential finger movements for HMD use, finger taps for navigating pages, and screen movement and resizing, allowing participants to adapt to the HMD while performing an ultrasound scan. Following the video lecture and HMD guidance, each participant was tasked with performing an abdominal ultrasound on the phantom and saving the 18 scanned images required. The DG were then provided with a printed handout, and the TG were provided with a PDF version of the handout presented as a hologram. During ultrasound scanning, the participants could ask questions and communicate with the faculty according to the assigned group's

XSL•FO
RenderX

supervision method. If the participants in either group needed assistance, they could seek help from their supervisors at any time. Next, the entire ultrasound procedure performed by the participants was recorded using a 360° camera. The time frame between the moment participants held the probe and the moment they returned the probe to the machine or pressed the "end exam" button to save images was measured. Finally, all participants completed a post-survey regarding their workload and usability.

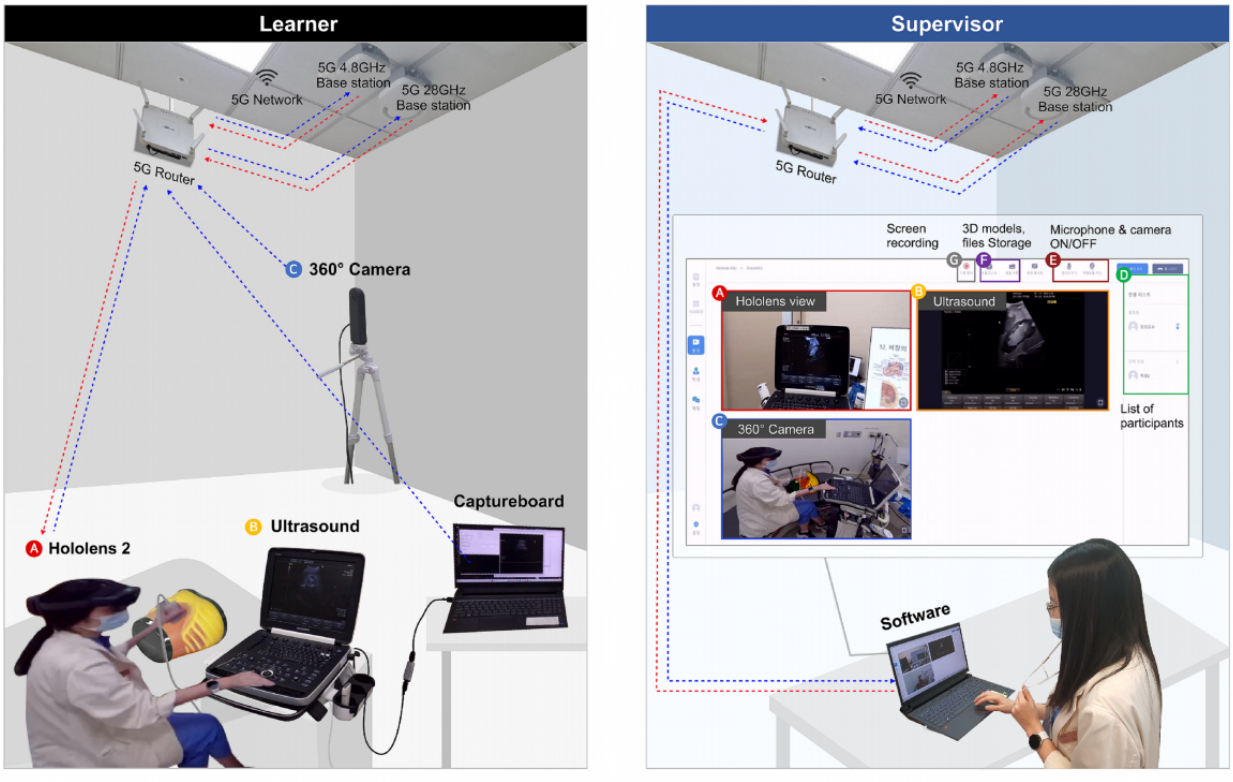
MR Telesupervised Ultrasound Education Platform

The MR telesupervised ultrasound education platform consists of 3 components: the network, the supervisor (a laptop for supervision), and the learner (an ultrasound device, a laptop for capturing ultrasound images, an HMD, and a 360° camera) (Figure 1). The hospital installed 4.7 and 28 GHz 5G base stations, along with a private 5G core [23]. The 28 GHz band offers low latency and high speeds but has a limited range, making it suitable for hotspot use. Therefore, 4.7 GHz base stations were installed in multiple designated practice areas, including some operating rooms, wards, and emergency rooms, to provide wide-area coverage, while 28 GHz stations were used as hotspots. Download and upload throughput measured by iperf were 1453 and 147 Mbps, respectively.

The software used for telesupervision mainly consists of remote assistance services (facilitating MR communication between HMD users and supervisors), media services (media processing and streaming 360° camera, HMD, and ultrasound images), authentication (account authentication), image conversion, and 3-dimensional model services. Ultrasound images were converted to 3840×2160 NV12 at 30 frames per second via the capture board, converted to 640×480 on the laptop, and forwarded to the service platform via real-time streaming protocol. The HMD device used was a HoloLens 2 (Microsoft), which captures images at a resolution of 1088×612 and delivers them to the service platform as a web real-time communication media stream. The 360° camera used was a QooCam 8K (Kandao Technology Company Ltd), which captured a video at 4K (3840×2160) resolution at 30 frames per second and sent it to the service platform via the real-time streaming protocol.

Using this platform, the supervisor could simultaneously observe clear ultrasound images, the learner's perspective via the HMD, and the learner's ultrasound scan posture using a 360° camera. The supervisor manipulated the 360° camera image to observe the learner's probe manipulation and offered assistance using visual and auditory cues through the HMD display. In the study hospital, communication latency from the HMD, ultrasound device, and 360° camera to the supervisor's laptop was measured in 10 iterations, with mean values of 196 (SD 5) ms, 49 (SD 36) ms, and 797 (SD 26) ms, respectively.

Figure 1. System architecture of MR telesupervised ultrasound education platform. The learner's part consists of (A) an HMD, (B) an ultrasound device, and a laptop to capture ultrasound images through the capture board, and (C) a 360° camera. The software used by the supervisor displays the video streamed from (A), (B), and (C). Additionally, there is a (D) list of participants, (E) communication with the learner via microphone and camera that can be turned on and off by the supervisor, (F) the ability to send 3D objects, files, and images from the storage, and (G) a recording function. HMD: head-mounted display; MR: mixed reality.



XSL•FO
RenderX

Abdominal Ultrasonography

The study materials were developed based on the abdominal ultrasound certification examination of the Korean Association of Clinical Ultrasound [24,25]. For novices, general ultrasonography information and knobology were introduced, and 18 abdominal ultrasound images were taught with the corresponding anatomical images: liver (7 images), biliary system (2), pancreas (2), spleen (1), kidney (4), and aorta (2) (Figure S1 in Multimedia Appendix 1). The ultrasound was performed using a Samsung Medison HM70A machine with a 1 - 7 MHz curved array probe. Ultrasound settings were set by the researcher. Participants used an abdominal ultrasound phantom (ABDFAN, Kyoto Kagaku Co.). After the test, the 2 EM faculty members evaluated the images in a blinded manner. They assigned scores ranging from 1 to 5 to each image based on predefined evaluation criteria, and the mean score was used [26]. Missing images received a score of zero (Table S1 in Multimedia Appendix 2).

User Experience and Confidence

To evaluate the feasibility of the MR-based telesupervised ultrasound education platform and its practical use, subjective user experience must be assessed. The National Aeronautics and Space Administration Task Load Index (NASA-TLX) is the most widely accepted and validated multidimensional tool that evaluates the overall workload across 6 dimensions: 3 dimensions measure the demands imposed on the subjects (mental demand, physical demand, and temporal demand), and 3 dimensions focus on how the subject deals with the task at hand (self-rated performance, effort, and frustration level). We used the raw NASA-TLX score, which is the arithmetic mean of all subscales, and did not weigh them [27,28].

The System Usability Scale (SUS) is a simple and widely adopted usability test tool based on questionnaires that can be used to evaluate software, mobile devices, and medical systems. The final scores were interpreted based on a well-established reference standard [29,30]. The SUS contains 10 questions based on a 5-point Likert scale. The odd questions were positive, and the even questions were negative. A higher SUS value indicates better product usability. Additionally, a pre- and post-training confidence survey was conducted using a 5-point Likert scale to assess the participants' confidence in each organ and overall abdominal ultrasound before and after training [31,32].

Evaluation of Outcomes

The primary outcome was the overall scan score of the 18 ultrasound images, with a total score of 90. Secondary outcomes included assessing the user experience through the NASA-TLX and SUS [29,33]. Additionally, we measured the time taken for the ultrasound procedure, the number of interventions made by supervisors, and self-confidence in performing abdominal ultrasound pre- and post-surveys.

Statistical Analysis

All continuous variables were described as mean (SD), while categorical variables were described as numbers and percentages. For continuous variables, the analysis employed either the 2-sample *t* test or Wilcoxon rank-sum test, and for categorical variables, the χ^2 test or Fisher exact test was used. Statistical significance was set at $P<.05$. All statistical analyses were conducted using R software (version 4.3.1; R Foundation for Statistical Computing).

Ethical Considerations

The experimental protocol of the study involving human participants was carefully designed and implemented in strict compliance with the ethical standards of the Declaration of Helsinki. The study protocol was approved by the Institutional Review Board of the Samsung Medical Center (IRB file number 2023-10-015) and registered with ClinicalTrials.gov (NCT06171828). Each participant provided written informed consent before registration and received a small monetary compensation for their time and involvement in the study. All study data were deidentified to ensure privacy and confidentiality, in accordance with institutional guidelines. No identifiable images or personal details of individual participants are included in this manuscript.

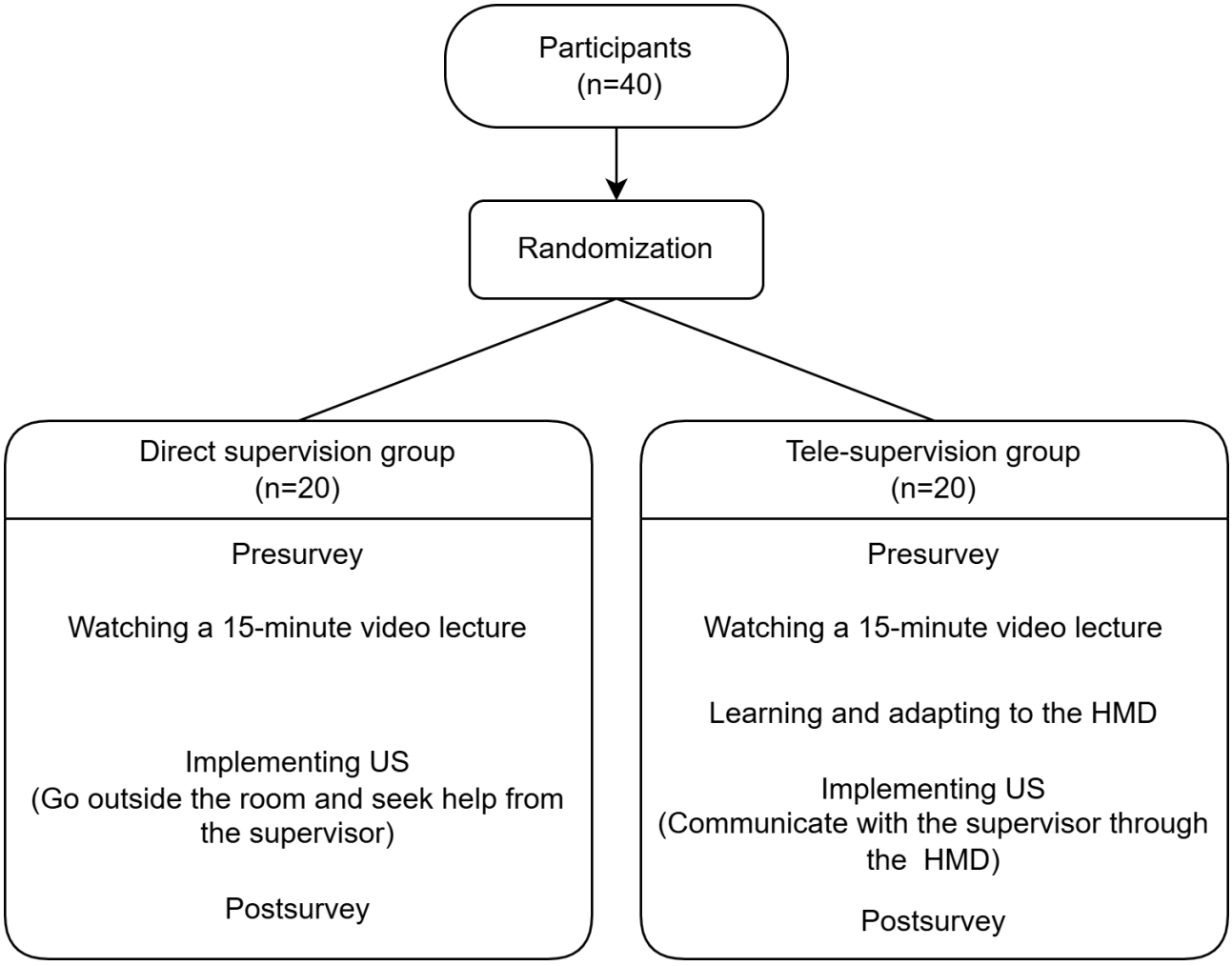
Results

Demographics

From November to December 2023, 40 doctors who had not received prior training in abdominal ultrasonography participated in the study and completed the entire process, including subsequent surveys (Figure 2). The participants had a mean age of 28.3 years, and 25% (n=10) were men. The participants were comprised of interns (n=14, 35%), family medicine physicians (n=10, 25%), and pediatricians (n=6, 15%). No significant differences were observed in age, sex, physician grade, specialty, frequency of ultrasound (excluding abdominal ultrasound) use, or prior use of an HMD between the groups (Table 1).

XSL•FO
RenderX

Figure 2. Flowchart of the study. HMD: head-mounted display; US: ultrasound.



XSL•FO
RenderX

Table . Demographic and basic characteristics of study participants.

Participant characteristics	Direct supervision group (n=20)	Telesupervision group (n=20)	*P* value
Age (years), mean (SD)	29.9 (3.9)	28.3 (3.5)	.19
Gender, n (%)			>.99
Men	5 (25)	5 (25)	
Women	15 (75)	15 (75)	
Physician-grade, n (%)			.57
Intern	6 (30)	8 (40)	
First-year resident	4 (20)	6 (30)	
Second-year resident	2 (10)	3 (15)	
Third-year resident	5 (25)	1 (5)	
Fourth-year resident	1 (5)	1 (5)	
Attending physician	2 (10)	1 (5)	
Specialty, n (%)			.27
Dermatology	1 (5)	0 (0%)	
Family medicine	7 (35)	3 (15)	
General surgery	1 (5)	1 (5)	
Internal medicine	0 (0)	2 (10)	
Intern	6 (30)	8 (40)	
Orthopedic surgery	0 (0)	2 (10)	
Pathology	0 (0)	1 (5)	
Pediatrics	3 (15)	3 (15)	
Radiology	2 (10)	0 (0)	
Frequency of ultrasound use, n (%)			.06
None	14 (70)	6 (30)	
1 - 2 times per year	1 (5)	5 (25)	
1 - 2 times per month	5 (25)	6 (30)	
1 - 2 times per week	0 (0)	2 (10)	
More than 3 times per week	0 (0)	1 (5)	
Prior use of HMD[a], n (%)	1 (5)	2 (10)	>.99

[a]HMD: head-mounted display.

Overall Performance Between Groups

There were no significant differences in scores between the 2 groups (Table 2). The total scores for DG and TG were 67.5 (SD 10.4) and 66.8 (SD 10.4), respectively. The intraclass correlation coefficient between the scores given by the 2 EM faculty members was 0.9. The duration of the ultrasound between the 2 groups was not significantly different, with a mean time of 24.0 (SD 8.1) and 23.8 (SD 8.0) minutes for DG and TG, respectively. However, the number of interventions was significantly higher in the TG, with 4.0 (SD 2.5) interventions, compared to 0.8 (SD 1.1) in the DG.

XSL•FO
RenderX

Table . Analysis of abdominal ultrasound score between 2 groups[a].

	Direct supervision group (n=20), mean (SD)	Telesupervision group, (n=20), mean (SD)	*P* value
Score			
Total (out of 90)	67.5 (10.4)	66.8 (10.4)	.84
Liver (out of 35)	24.7 (4.6)	24.6 (3.7)	.94
Gallbladder (out of 10)	7.5 (1.6)	8.2 (1.5)	.19
Pancreas (out of 10)	9.0 (1.0)	7.9 (2.2)	.06
Spleen (out of 5)	3.1 (1.5)	3.1 (1.5)	.96
Kidney (out of 20)	15.4 (2.8)	15.5 (2.7)	.86
Procedure time (min)	24.0 (8.1)	23.8 (8.0)	.95
Intervention	0.8 (1.1)	4.0 (2.5)	<.001

[a]The intraclass correlation coefficient between the scores given by 2 emergency medicine faculty members was 0.9.

User Experience, Workload, and Confidence

All participants completed the NASA-TLX, SUS questions, and pre- and post-confidence surveys. The TG reported significantly lower scores in NASA-TLX for mental demand (mean 43.8 [SD 24.8] vs mean 60.6 [SD 22.4]), effort (mean 43.1 [SD 22.9] vs mean 67.9 [SD 17]), and frustration level (mean 26.9 [SD 20.3] vs mean 45.2 [SD 27.8]) compared to the DG, indicating reduced workload (Table 3).

Table . Analysis of National Aeronautics and Space Administration-Task Load Index (NASA-TLX) score between 2 groups.

Measure	Direct supervision group (n=20), mean (SD)	Telesupervision group (n=20), mean (SD)	*P* value
Mental demand	60.6 (22.4)	43.8 (24.8)	.03
Physical demand	40.1 (24.5)	33.1 (26.9)	.40
Temporal demand	47.6 (22.0)	36.9 (26.0)	.17
Self-rated performance	51.4 (22.5)	45.4 (26.3)	.44
Effort	67.9 (17)	43.1 (22.9)	<.001
Frustration level	45.2 (27.8)	26.9 (20.3)	.02

The mean SUS score of TG was significantly higher than that of DG (mean 66.6 [SD 9.1] vs mean 60.2 [SD 10.4]). Among all statements, “frequent use” (mean 4.3 [SD 0.6] vs mean 3.9 [SD 0.6]), “well-integrated system” (mean 4.1 [SD 0.6] vs mean 3.3 [SD 0.9]), and “learn quickly” (mean 4.2 [SD 0.7] vs mean 3.5 [SD 0.6]) were significantly higher in TG than in DG (Table 4).

XSL•FO
RenderX

Table . Analysis of standardized System Usability Scale (SUS) score between the 2 groups and each SUS statement[a].

SUS questions (5-point Likert scale)	Direct supervision group (n=20), mean (SD)	Telesupervision group (n=20), mean (SD)	*P* value
I think that I would like to use this system frequently.	3.9 (0.6)	4.3 (0.6)	.02
I found the system unnecessarily complex.	2.1 (0.6)	2.0 (0.7)	.60
I thought the system was easy to use.	3.5 (0.6)	3.8 (0.6)	.13
I think that I would need the support of a technical person to be able to use this system.	3.5 (1.1)	4.1 (0.8)	.06
I found that the various functions in this system were well integrated.	3.3 (0.9)	4.1 (0.6)	.001
I thought there was too much inconsistency in this system.	1.9 (0.7)	1.7 (0.5)	.30
I would imagine that most people would learn to use this system very quickly.	3.6 (0.6)	4.2 (0.7)	.005
I found the system cumbersome to use.	2.1 (0.7)	2.3 (0.7)	.27
I felt very confident using the system.	3.3 (0.8)	3.7 (0.6)	.08
I needed to learn a lot of things before I could get going with this system.	3.8 (1.1)	3.3 (1.0)	.14
Total SUS score	60.2 (10.4)	66.6 (9.1)	.046

[a]Each statement has a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree).

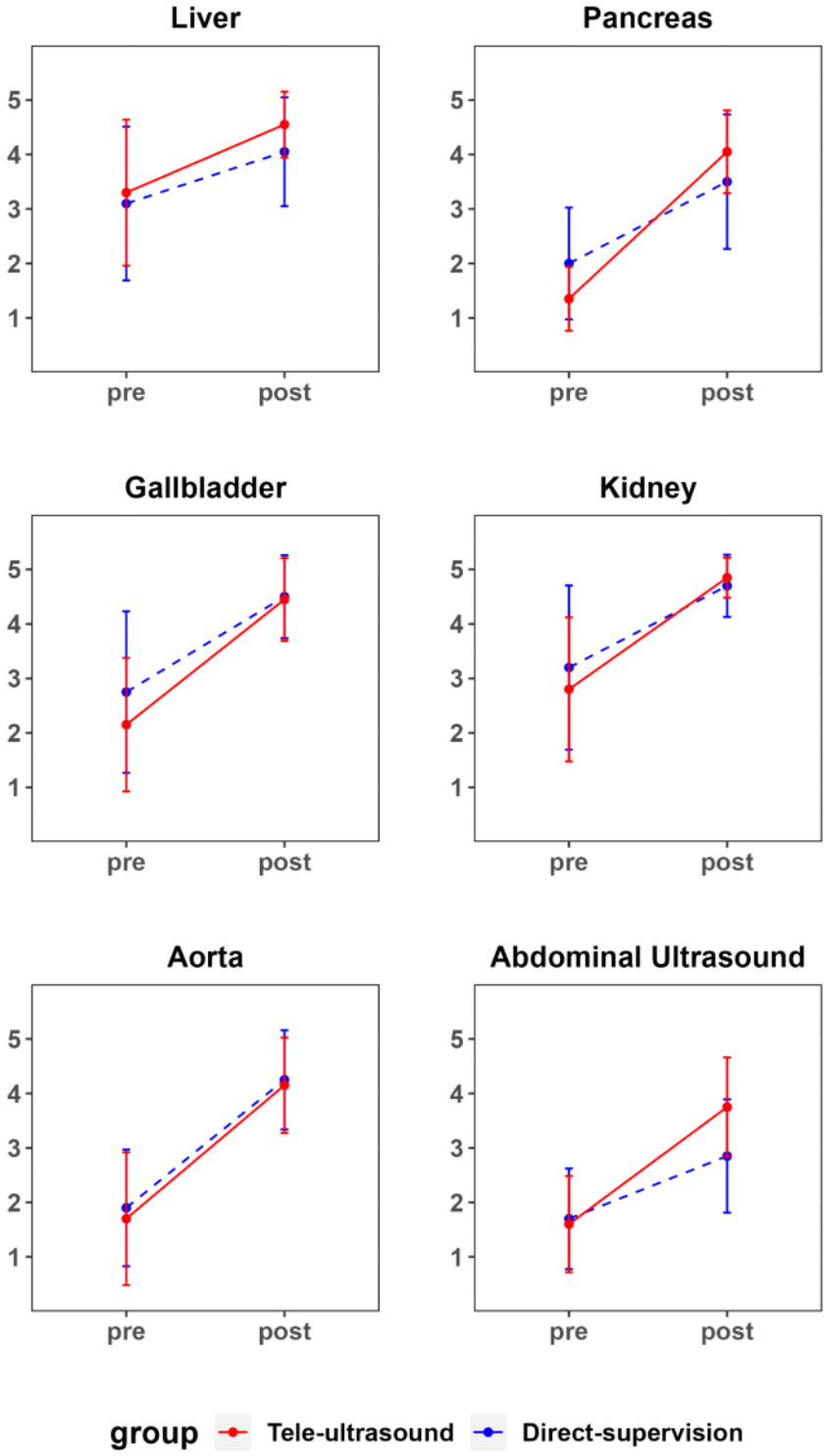
The change in confidence for both groups showed significant increases in organ-specific measures from pre- to post-intervention (Table 5). Before the study, both groups generally had confidence levels below 3 points, with no significant differences except for the pancreas. After the study, most scores increased to the 4-point range, with the TG showing a significantly greater improvement, especially in abdominal ultrasound, referring to overall skill confidence in abdominal ultrasound (Figure 3).

Table . Confidence of each group before and after the study.

Confidence (5-likert scale)	Pre-education			Posteducation		
	Direct supervision group (n=20), mean (SD)	Telesupervision group (n=20), mean (SD)	*P* value	Direct supervision group (n=20), mean (SD)	Telesupervision group (n=20), mean (SD)	*P* value
Liver	3.1 (1.4)	3.3 (1.3)	.65	4.0 (1.0)	4.5 (0.6)	.07
Pancreas	2.0 (1.0)	1.4 (0.6)	.02	3.5 (1.2)	4.0 (0.8)	.10
Gallbladder	2.8 (1.5)	2.1 (1.2)	.17	4.5 (0.8)	4.4 (0.8)	.84
Kidney	3.2 (1.5)	2.8 (1.3)	.38	4.7 (0.6)	4.8 (0.4)	.33
Aorta	1.9 (1.1)	1.7 (1.2)	.59	4.2 (0.9)	4.2 (0.9)	.73
Abdominal ultrasound	1.7 (0.9)	1.6 (0.9)	.73	2.8 (1.0)	3.8 (0.9)	.006

XSL•FO
RenderX

Figure 3. Pre-post assessment confidence levels by organ for each group. The confidence levels are displayed as mean values with standard deviations on a 5-point Likert scale.



Discussion

Principal Findings

This study compared an MR-based telesupervised ultrasound education platform on private 5G networks with traditional in-person training for novice doctors. Conducted through a randomized controlled trial with a relatively small sample size, the results indicated that the quality of abdominal ultrasound scans was consistent between the TG and the DG, indicating the platform's ability to maintain educational integrity remotely. Additionally, telesupervision was associated with decreased mental demand, effort, and frustration, and showed positive impacts on usability and confidence levels. This study provides an important reference for advancing ultrasound training that focuses on the tactile and manual skills necessary for the field, even when using remote learning platforms.

XSL•FO
RenderX

Our findings affirm the feasibility, equivalence, and noninferiority of telehealth in ultrasound education, with outcomes comparable to in-person training [8,34-36]. Similar to other studies using HMD and MR guidance, telesupervision led to reduced cognitive loads [33]. Participants noted enhanced skill acquisition, easier command following, and greater confidence compared to traditional methods [37]. Challenges within the TG included the difficulty of managing dual tasks—performing ultrasound procedures while adapting to the HMD use. These issues could be distracting and potentially induce cybersickness, adversely affecting the learning experience [38]. As learners grow more familiar with HMD technology, these distractions are expected to decrease, improving the learning environment by minimizing psychological and physical barriers.

To address common technical issues like internet interruptions [39], delays, image freezes, and audio or video transmission problems [37], our study leveraged advanced 5G technology. This significantly mitigated latency issues and enhanced the potential for future applications of augmented reality in telehealth [40]. Unlike studies that reported complications with smartphone use during examinations [41], our implementation of an HMD freed the operator's hands, enabling more realistic and immersive educational experiences.

In the study, several technical issues arose, including 2 instances of program crashes due to duplicate logins, 3 app crashes (with reconnection times within 5 s), and 1 microphone error, all of which required reconnection and extended the examination time for some in the TG group. Additionally, 2 delays in ultrasound image transmission and 1 hologram screen displacement caused by the HMD's hand recognition required time adjustments but did not interrupt the session. Continuous updates to the system have been implemented to address these challenges and enhance the program's stability, thus reducing the likelihood of similar disruptions in future studies.

Our system integrates an audiovisual cue system for real-time mentoring, addressing previous limitations related to the lack of physical guidance and verbal explanation difficulties [42]. The supervisor could use visual holographic cues such as arrows, lines, and images projected to the user's view through an HMD, along with audio cues for precise probe manipulation. These cues were found to be the most effective means of conveying information to the operator [22]. However, it was challenging for supervisors to swiftly send holograms while engaging in conversation. Proficiency takes time for supervisors, and appropriate 3D objects need to be created and prepared.

To further enhance remote ultrasound education methods, our study emphasizes the importance of speed, stability, and security. In ultrasound training, the promptness of feedback is critical, making low latency essential. Maintaining latency under the 700 ms threshold is recommended for effective supervision [43-45]. Also, our platform includes 3 screens (HMD, ultrasound, and 360° camera), offering redundancy to maintain effective supervision even if one display encounters delays or freezes. This feature is essential for telemedicine, providing comprehensive observations from multiple angles. Furthermore, concerns about personal information leakage due to cyberattacks, particularly those targeting network and communication connections, are addressed by our system's use of a 5G private network accessible only to registered devices [46,47]. This enhances security and protects sensitive data.

The developed platform's versatility extends to various medical practices, including endoscopy and laparoscopy, using HDM interface for image transmission. With 5G base stations installed in hospital areas like operating rooms and emergency departments, it supports diverse skill training across medical fields. Its established infrastructure and innovative design facilitate a dynamic role swap between educators and learners, particularly enhancing surgical education by overcoming traditional constraints such as limited visual fields during surgery. Equipped with HMDs, surgeons can perform and broadcast procedures simultaneously, expanding learners' perspectives and fostering interactive, immersive training experiences. This adaptability and functionality underscore the platform's potential to revolutionize medical education, positioning it as a crucial tool for metaverse education platforms.

Limitations

First, our study's generalizability is limited by its single-institution setting within a private network, as technological and infrastructural conditions may differ across institutions. Additionally, this study did not include noninferiority testing due to the lack of a prior sample size calculation and statistical planning, which limits the robustness and generalizability of our findings. Future studies should address this aspect with appropriate sample sizes and statistical designs. Second, several technical challenges arose during the study, including program crashes, app errors, and ultrasound image transmission delays, which required additional time to resolve. While these issues were promptly addressed and did not affect the overall results or session flow, they highlight the need for further improvements in system stability to ensure consistent operation in future studies. Third, the lack of an assessment of the supervisor's user experience creates a gap in the comprehensive evaluation of a platform's overall effectiveness. Fourth, while a 5-minute HMD training session was provided, it may have been brief for participants with no prior HMD experience, potentially affecting initial ease of use. Although only a few HMD-related interventions were required, a longer training session could improve user familiarity and comfort in future studies (Table S2 in Multimedia Appendix 3). Fifth, the environmental conditions in the emergency department treatment room, with its inherent noise and commotion, might have affected the participants' concentration and performance, adding an external variable to the study's context. Finally, the use of a phantom model for simulation, although practical, may not fully capture the complexities of real patient interactions, potentially limiting the applicability of the training in actual clinical practice. Therefore, future studies should broaden participant bases, ensure technological robustness, evaluate all user experiences, and incorporate real patient scenarios to enhance the applicability of the findings and the platform's utility in clinical education.

XSL•FO
RenderX

Conclusions

This study demonstrated the potential of a novel MR-based ultrasound education platform using a 5G private network. Although no significant differences in scores were observed between groups, considerable differences in positive educational interactions, workload, and usability were evident. These findings emphasize the platform's potential to significantly enhance the ultrasound training experience, suggesting a pivotal shift toward more interactive and efficient learning.

Acknowledgments

ABDFAN (Kyoto Kagaku Co.), an ultrasound phantom, and HoloLens 2 (Microsoft) were provided by the Education and Human Resources Development Department of the Samsung Medical Center. Kyungmi Min from the Research Institute for Future Medicine, Samsung Medical Center, Seoul, Korea, helped with participant enrollment and study conduction. This research was supported by a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute, funded by the Ministry of Health and Welfare, Republic of Korea (grant HI23C0460), and by the National Research Foundation of Korea, funded by the Ministry of Science and ICT (grant NRF-2022R1C1C1011864).

Data Availability

The datasets generated during and analyzed during this study are not publicly available due to privacy restrictions. However, they are available from the corresponding author on reasonable request.

Authors' Contributions

HY and MHS conceptualized the study. MK led the investigation, curated the data, conducted formal analyses, created visualizations, and wrote the initial draft. HY engaged in the investigation, formal analysis, funding acquisition, and contributed to the initial draft. SM was involved in formal analysis and visualization. WCC secured funding for the project. MHS, WCC, and IJJ participated in the review and editing of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Standard images for abdominal ultrasound.
[DOCX File, 1349 KB - games_v13i1e63448_app1.docx]

Multimedia Appendix 2
Evaluation criteria of ultrasonography.
[DOCX File, 17 KB - games_v13i1e63448_app2.docx]

Multimedia Appendix 3
Total number of interventions during experiment.
[DOCX File, 14 KB - games_v13i1e63448_app3.docx]

Checklist 1
Consolidated Standards of Reporting Trails (CONSORT) e-health checklist.
[PDF File, 1038 KB - games_v13i1e63448_app4.pdf]

References

1. Whitson MR, Mayo PH. Ultrasonography in the emergency department. Crit Care 2016 Aug 15;20(1):227. [doi: 10.1186/s13054-016-1399-x] [Medline: 27523885]
2. Uschnig C, Recker F, Blaivas M, Dong Y, Dietrich CF. Tele-ultrasound in the era of COVID-19: a practical guide. Ultrasound Med Biol 2022 Jun;48(6):965-974. [doi: 10.1016/j.ultrasmedbio.2022.01.001] [Medline: 35317949]
3. Rose S. Medical student education in the time of COVID-19. J Am Med Assoc 2020 Jun 2;323(21):2131-2132. [doi: 10.1001/jama.2020.5227] [Medline: 32232420]
4. Trafton J, Martins S, Michel M, et al. Evaluation of the acceptability and usability of a decision support system to encourage safe and effective use of opioid therapy for chronic, noncancer pain by primary care providers. Pain Med 2010 Apr;11(4):575-585. [doi: 10.1111/j.1526-4637.2010.00818.x] [Medline: 20202142]

XSL•FO
RenderX

5. Eke OF, Henwood PC, Wanjiku GW, Fasina A, Kharasch SJ, Shokoohi H. Global point-of-care ultrasound education and training in the age of COVID-19. Int J Emerg Med 2021 Feb 18;14(1):12. [doi: 10.1186/s12245-021-00338-9] [Medline: 33602112]
6. Choi W, Cho YS, Ha YR, et al. Role of point-of-care ultrasound in critical care and emergency medicine: update and future perspective. Clin Exp Emerg Med 2023 Dec;10(4):363-381. [doi: 10.15441/ceem.23.101] [Medline: 38225778]
7. Samannodi M, Bulkhi A, Alwafi H, et al. Impact of COVID-19 pandemic on medical education: a cross-sectional study in the western region of Saudi Arabia. Adv Med Educ Pract 2022;13:741-754. [doi: 10.2147/AMEP.S369213] [Medline: 35903321]
8. Kang SY, Yoo J, Park S, et al. Online learning versus hands-on learning of basic ocular ultrasound skills: a randomized controlled non-inferiority trial. Med Bogota Colomb 2022 Jul 20;58(7):960. [doi: 10.3390/medicina58070960]
9. Situ-LaCasse E, Acuña J, Huynh D, et al. Can ultrasound novices develop image acquisition skills after reviewing online ultrasound modules? BMC Med Educ 2021 Mar 20;21(1):175. [doi: 10.1186/s12909-021-02612-z] [Medline: 33743680]
10. Bui C, Parange N, Bezak E, Bidner A. The role of telehealth in ultrasound training for remote learners: a systematic review. Telemed J E Health 2024 Apr;30(4):963-975. [doi: 10.1089/tmj.2023.0396] [Medline: 38052050]
11. Barteit S, Lanfermann L, Bärnighausen T, Neuhann F, Beiersmann C. Augmented, mixed, and virtual reality-based head-mounted devices for medical education: systematic review. JMIR Serious Games 2021 Jul 8;9(3):e29080. [doi: 10.2196/29080] [Medline: 34255668]
12. Munzer BW, Khan MM, Shipman B, Mahajan P. Augmented reality in emergency medicine: a scoping review. J Med Internet Res 2019 Apr 17;21(4):e12368. [doi: 10.2196/12368] [Medline: 30994463]
13. Chen L, Day TW, Tang W, John NW. Recent developments and future challenges in medical mixed reality. Presented at: 2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR); Oct 9-13, 2017; Nantes, France. [doi: 10.1109/ISMAR.2017.29]
14. Milgram P, Colquhoun H. A taxonomy of real and virtual world display integration. Mix Real: Merg Real Virt Worlds 1999;1:1-26 [FREE Full text]
15. Hu HZ, Feng XB, Shao ZW, et al. Application and prospect of mixed reality technology in medical field. Curr Med Sci 2019 Feb;39(1):1-6. [doi: 10.1007/s11596-019-1992-8] [Medline: 30868484]
16. Nguyen T, Plishker W, Matisoff A, Sharma K, Shekhar R. HoloUS: augmented reality visualization of live ultrasound images using HoloLens for ultrasound-guided procedures. Int J Comput Assist Radiol Surg 2022 Feb;17(2):385-391. [doi: 10.1007/s11548-021-02526-7] [Medline: 34817764]
17. Nylund ME, Jain S, Tegnander E, et al. Mixed reality training application to perform obstetric pulsed-wave Doppler ultrasound. Educ Inf Technol 2024 Apr;29(6):7519-7551. [doi: 10.1007/s10639-023-12069-w]
18. Stojanovska M, Tingle G, Tan L, et al. Mixed reality anatomy using Microsoft HoloLens and cadaveric dissection: a comparative effectiveness study. Med Sci Educ 2020 Mar;30(1):173-178. [doi: 10.1007/s40670-019-00834-x] [Medline: 34457656]
19. Stark PW, Borger van der Burg BLS, van Waes OJF, et al. Telemedicine-guided two-incision lower leg fasciotomy performed by combat medics during tactical combat casualty care: a feasibility study. Mil Med 2024 Feb 27;189(3-4):e645-e651. [doi: 10.1093/milmed/usad364]
20. Jell A, Vogel T, Ostler D, et al. 5th-Generation mobile communication: data highway for Surgery 4.0. Surg Technol Int 2019 Nov 10;35:36-42. [Medline: 31694061]
21. Lacy AM, Bravo R, Otero-Piñeiro AM, et al. 5G-assisted telementored surgery. Br J Surg 2019 Nov;106(12):1576-1579. [doi: 10.1002/bjs.11364] [Medline: 31483054]
22. Shabir D, Anjum A, Hamza H, et al. Development and evaluation of a mixed-reality tele-ultrasound system. Ultrasound Med Biol 2023 Aug;49(8):1867-1874. [doi: 10.1016/j.ultrasmedbio.2023.04.017] [Medline: 37263893]
23. Lee SY. Private 5G (e-Um 5G) trend and promotion policy. J Korean Inst Electromagn Eng Sci 2022 Jul;33(7):531-540. [doi: 10.5515/KJKIEES.2022.33.7.531]
24. Kim HC, Kim PN, Lee JS, Choi JI, Lee MW. Test for the accreditation of ultrasound practices using abdominal phantom: preliminary study intended for radiology residents. J Korean Soc Ultrasound Med 2013;32(3):171-181 [FREE Full text]
25. Park SE, Lee SH, Lee DJ, Kim KM, Park SB, Kim BT, et al. Abdominal ultrasonography: procedure and training. Korean J Fam Prac 2014;4(1):28-37 [FREE Full text]
26. Liu R, Blaivas M, Moore C, et al. Emergency Ultrasound Standard Reporting Guidelines: American College of Emergency Physicians; 2018.
27. Said S, Gozdzik M, Roche TR, et al. Validation of the Raw National Aeronautics and Space Administration Task Load Index (NASA-TLX) Questionnaire to assess perceived workload in patient monitoring tasks: pooled analysis study using mixed models. J Med Internet Res 2020 Sep 7;22(9):e19472. [doi: 10.2196/19472] [Medline: 32780712]
28. Dias RD, Ngo-Howard MC, Boskovski MT, Zenati MA, Yule SJ. Systematic review of measurement tools to assess surgeons' intraoperative cognitive workload. Br J Surg 2018 Apr;105(5):491-501. [doi: 10.1002/bjs.10795] [Medline: 29465749]
29. Helin K, Kuula T, Vizzi C, Karjalainen J, Vovk A. User experience of augmented reality system for astronaut's manual work support. Front Robot AI 2018;5(106):106. [doi: 10.3389/frobt.2018.00106] [Medline: 33500985]

XSL•FO
RenderX

30. Liang J, Xian D, Liu X, et al. Usability study of mainstream wearable fitness devices: feature analysis and system usability scale evaluation. JMIR Mhealth Uhealth 2018 Nov 8;6(11):e11066. [doi: 10.2196/11066] [Medline: 30409767]
31. Crouch AK, Dawson M, Long D, Allred D, Madsen T. Perceived confidence in the FAST exam before and after an educational intervention in a developing country. Int J Emerg Med 2010 Feb 27;3(1):49-52. [doi: 10.1007/s12245-009-0144-5] [Medline: 20414382]
32. Morrison RG, Halpern SA, Brace EJ, et al. Open-source ultrasound trainer for healthcare professionals: a pilot randomized control trial. Simul Healthc 2022;19(1):113-121. [doi: 10.1097/SIH.0000000000000697]
33. Birlo M, Edwards PJE, Yoo S, et al. CAL-Tutor: a HoloLens 2 application for training in obstetric sonography and user motion data recording. J Imaging 2022 Dec 29;9(1):6. [doi: 10.3390/jimaging9010006] [Medline: 36662104]
34. Drake AE, Hy J, MacDougall GA, et al. Innovations with tele-ultrasound in education sonography: the use of tele-ultrasound to train novice scanners. Ultrasound J 2021 Feb 14;13(1):6. [doi: 10.1186/s13089-021-00210-0] [Medline: 33586112]
35. Brisson AM, Steinmetz P, Oleskevich S, Lewis J, Reid A. A comparison of telemedicine teaching to in-person teaching for the acquisition of an ultrasound skill - a pilot project. J Telemed Telecare 2015 Jun;21(4):235-239. [doi: 10.1177/1357633X15575446] [Medline: 25766853]
36. Lim H, Kim MJ, Park JM, et al. Use of smart glasses for ultrasound-guided peripheral venous access: a randomized controlled pilot study. Clin Exp Emerg Med 2019 Dec;6(4):356-361. [doi: 10.15441/ceem.19.029] [Medline: 31910508]
37. Dyer D, Cusden J, Turner C, et al. The clinical and technical evaluation of a remote telementored telesonography system during the acute resuscitation and transfer of the injured patient. J Trauma 2008 Dec;65(6):1209-1216. [doi: 10.1097/TA.0b013e3181878052] [Medline: 19077603]
38. Moro C, Phelps C, Redmond P, Stromberga Z. HoloLens and mobile augmented reality in medical and health science education: a randomised controlled trial. Br J Educ Tech 2021 Mar;52(2):680-694. [doi: 10.1111/bjet.13049]
39. Burckett-St Laurent DA, Cunningham MS, Abbas S, Chan VW, Okrainec A, Niazi AU. Teaching ultrasound-guided regional anesthesia remotely: a feasibility study. Acta Anaesthesiol Scand 2016 Aug;60(7):995-1002. [doi: 10.1111/aas.12695] [Medline: 26860837]
40. Douglas TM, Levine AR, Olivieri PP, et al. Brief training increases nurses' comfort using tele-ultrasound: a feasibility study. Intensive Crit Care Nurs 2019 Apr;51:45-49. [doi: 10.1016/j.iccn.2018.11.004] [Medline: 30514602]
41. Smith A, Addison R, Rogers P, et al. Remote mentoring of point-of-care ultrasound skills to inexperienced operators using multiple telemedicine platforms: is a cell phone good enough? J Ultrasound Med 2018 Nov;37(11):2517-2525. [doi: 10.1002/jum.14609] [Medline: 29577381]
42. Soni NJ, Boyd JS, Mints G, et al. Comparison of in-person versus tele-ultrasound point-of-care ultrasound training during the COVID-19 pandemic. Ultrasound J 2021 Sep 6;13(1):39. [doi: 10.1186/s13089-021-00242-6] [Medline: 34487262]
43. Kim T, Zimmerman PM, Wade MJ, Weiss CA. The effect of delayed visual feedback on telerobotic surgery. Surg Endosc 2005 May;19(5):683-686. [doi: 10.1007/s00464-004-8926-6]
44. Korte C, Nair SS, Nistor V, Low TP, Doarn CR, Schaffner G. Determining the threshold of time-delay for teleoperation accuracy and efficiency in relation to telesurgery. Telemed J E Health 2014 Dec;20(12):1078-1086. [doi: 10.1089/tmj.2013.0367] [Medline: 25290465]
45. Xu S, Perez M, Yang K, Perrenot C, Felblinger J, Hubert J. Determination of the latency effects on surgical performance and the acceptable latency levels in telesurgery using the dV-Trainer(®) simulator. Surg Endosc 2014 Sep;28(9):2569-2576. [doi: 10.1007/s00464-014-3504-z] [Medline: 24671353]
46. Jin ML, Brown MM, Patwa D, Nirmalan A, Edwards PA. Telemedicine, telementoring, and telesurgery for surgical practices. Curr Probl Surg 2021 Dec;58(12):100986. [doi: 10.1016/j.cpsurg.2021.100986] [Medline: 34895561]
47. Bonaci T, Herron J, Yusuf T, Yan J, Kohno T, Chizeck HJ. To make a robot secure: an experimental analysis of cyber security threats against teleoperated surgical robots. arXiv. Preprint posted online on Apr 16, 2015. [doi: 10.48550/arXiv.1504.04339]

Abbreviations

DG: direct supervision group
EM: emergency medicine
HMD: head-mounted display
MR: mixed reality
NASA-TLX: National Aeronautics and Space Administration Task Load Index
SUS: System Usability Scale
TG: telesupervision group

XSL•FO
RenderX

Edited by A Coristine; submitted 25.06.24; peer-reviewed by A Bayro, CI Sartorão Filho; revised version received 16.11.24; accepted 20.11.24; published 12.06.25.

Please cite as:
Kim M, Son MH, Moon S, Cha WC, Jo IJ, Yoon H
A Mixed Reality–Based Telesupervised Ultrasound Education Platform on 5G Network Compared to Direct Supervision: Prospective Randomized Pilot Trial
JMIR Serious Games 2025;13:e63448
URL: https://games.jmir.org/2025/1/e63448
doi:10.2196/63448

© Minha Kim, Meong Hi Son, Suhyeon Moon, Won Chul Cha, Ik Joon Jo, Hee Yoon. Originally published in JMIR Serious Games (https://games.jmir.org), 12.6.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Evaluating the Efficacy of a Serious Game to Deliver Health Education About Invasive Meningococcal Disease: Clustered Randomized Controlled Equivalence Trial

Lauren Bloomfield[1*], BHlthSc, MPH, DrPH; Julie Boston[2*], BSc, Grad Dip, PhD; Martin Masek[2*], BEng, PhD; Lesley Andrew[2*], BSc, MSc, PhD; Donna Barwood[2], BEd, MEd, PhD; Amanda Devine[2], BAppSc, BEd, PhD

[1]School of Medicine, The University of Notre Dame Australia, Fremantle, Australia

[2]Edith Cowan University, Joondalup, Australia

[*]these authors contributed equally

Corresponding Author:
Julie Boston, BSc, Grad Dip, PhD
Edith Cowan University
270 Joondalup Drive
Joondalup, 6027
Australia
Phone: 61 0863045702
Email: julie.boston@ecu.edu.au

Abstract

Background: Invasive meningococcal disease (IMD) is a serious, vaccine-preventable infectious disease that can be life-threatening. Teaching adolescents about the early detection and prevention of IMD can be challenging in a school environment, with educators reporting they lack confidence or expertise to cover this in the classroom environment. Professional guest educators are an alternative to cover specialist topics such as IMD; however, time and resourcing constraints can mean that these educators are not always available. Serious games may be an alternative to face-to-face education, where complex health information may be delivered via self-directed gameplay.

Objective: This study aims to develop a serious game that can replace a face-to-face educator in a classroom setting to educate adolescents aged 12 years to 15 years. This study evaluates the efficacy of the Meningococcal Immunisation Awareness, Prevention and Protection app (MIApp), a serious game designed to replicate the information provided in a 30-minute face-to-face presentation provided by a trained educator.

Methods: This clustered, randomized controlled equivalence trial involved students (Years 7-10) from 6 secondary schools across metropolitan Western Australia who completed pre- and postintervention questionnaires with a follow-up at 3 months postintervention to measure the primary outcome of IMD knowledge acquisition following this self-guided intervention. The findings were compared with changes in an active control (comparison) group who received an in-class educational presentation about IMD transmission and protection. A questionnaire was developed to assess 9 key areas of knowledge. Median scores for knowledge pre- and postintervention were collected from a self-administered assessment of this questionnaire and, at 3 months postintervention, were compared between groups. A knowledge score of +/–2/16 was determined a priori to meet the criteria for equivalence. Participants who used MIApp were also asked a series of questions to assess the enjoyment of and engagement with the game.

Results: Of the 788 participating students, the median postintervention correct score in both the MIApp and control cohorts was 14/16 (87.5% correct responses), compared with the median pre-intervention correct score of 6/16 (37.5% correct responses), representing a significant (P<.001) increase in IMD knowledge in both groups. Improvements were retained in both groups 3 months after the initial intervention (median correct score: 11/16 in the intervention group; 12/16 in the control group; P=.86), demonstrating the efficacy of MIApp to deliver health education about IMD transmission and protection, although response rates in the follow-up cohort were low (255/788, 32.4%).

Conclusions: MIApp met the predetermined threshold for equivalence, demonstrating similar improvements in knowledge posttrial and at the 3-month follow-up. Participating adolescents considered the MIApp game more enjoyable than a presentation, with equivalent improvements in knowledge. Serious games could represent a constructive tool to help teachers impart specialized health education.

XSL•FO
RenderX

(JMIR Serious Games 2025;13:e60755) doi:10.2196/60755

KEYWORDS

serious games; meningococcal disease; immersive digital applications; health promotion; gaming; meningitis; infection; bacteria; contagious; infectious; immersive; education; mHealth; mobile health; applications; youth; adolescents

Introduction

An estimated 94% of Australian teenagers are smartphone users, with research showing 82% of teenagers frequently use these devices for digital gaming [1]. The emergence of digital technologies has led to a generation with new ways of thinking, learning, and interacting with each other [2]. Game-based learning occurs through "an environment where game content and gameplay enhance knowledge and skill acquisition, and where game activities involve problem solving spaces and challenges that provide players/learners with a sense of achievement" [3]. In the school environment, digital games are increasingly being considered a constructive learning tool to captivate and intrinsically motivate adolescents' engagement [4-6]. The concept of "serious games" generally refers to those digital games in which the intention is to provide meaningful learning while simulating real-world scenarios that are typically aligned with educational standards [7-9].

Research suggests serious games can promote active learning, build knowledge, engage and motivate students, support higher-level thinking, and increase learner independence [10]. Studies have also demonstrated the attractiveness of digital games as a learning environment to support social, emotional, and cognitive outcomes in young people [11-13], with evidence illustrating their influence on the long-term retention of knowledge [4,11]. Serious games appear to be well suited to specific mandated curriculum learning areas like science, technology, engineering, and mathematics [14-16] and offer additional learning opportunities to engage adolescents in new emerging fields of study such as health promotion [17,18].

Given their growing appeal, serious games are increasingly being explored as a novel alternative to the traditional classroom delivery of health promotion messages to adolescents [2,5,19,20]. Serious games for health promotion can stimulate real-life situations, providing adolescents with explorative learning environments [21], and can be adapted to the developmental stage, educational level, and personal interest, allowing the player to self-manage their educational goals [22].

Moreover, serious games can offer support for the delivery of health education around issues that may not be easy for teachers to address due to time, stigma, and their own lack of knowledge [22,23]. In addition, a lack of resources for expert presenters to deliver face-to-face content across large geographical areas means that regional and remote areas are less likely to receive such interventions, furthering the inequities between metropolitan and rural populations. A large body of research exists to illustrate how serious games can be designed specifically for topics such as nutrition [14,24,25], sexual health [26,27], substance use [28], and mental health [29], among others. This ranges from understanding strategies that can be used to explain complex concepts to pedagogical approaches that capture the social and cultural nuances of communities or age-related groups. However, despite 2 meta-analyses suggesting serious games represent promising learning tools [5,6], limited and inconsistent evidence of serious game efficacy in delivering health education content remains an issue [27,30].

For digital games to become a key tool for the support of traditional modalities of teaching and learning, there is a need to evaluate the efficacy of these games. The challenge for teachers is to ensure the serious game is curriculum-aligned, fit-for-purpose, and effective as a resource to meet the needs of a range of student ages and settings and that it can extend or complement learning. Developing serious games for health education, however, is time-consuming and challenging. It requires multidisciplinary team members to address stated objectives, encompassing elements of design, content, format, and pedagogy to the psychology of behavioral change [31]. Adapting complex health information for suitability with younger audiences within an educational environment can pose several further challenges for game developers. These challenges include maintaining player engagement and ensuring the game effectively reflects national curriculum standards and learning goals and outcomes.

The Meningococcal Immunisation Awareness, Prevention and Protection app (MIApp) is a serious game developed in Western Australia (WA) to deliver education via a digital platform about the transmission of meningococci and protection targeted at adolescents aged 12 years to 15 years. Invasive meningococcal disease (IMD), caused by *Neisseria meningitidis* bacteria, is a serious, vaccine-preventable infectious disease that is rare but life-threatening [32]. IMD can result in ongoing disability ranging from major neurological complications, including hearing loss, cognitive deficit, and seizure, to severe non-neurological effects, such as the amputation of a limb and death [33]. Adolescents are at an increased risk of meningococcal carriage, a precursor for the development of invasive disease [34]. Multilevel public health measures to reduce the risk of IMD include vaccination; minimization of behaviors known to increase transmission, such as intimate kissing and openly sneezing and coughing; and education for the early identification and antibiotic administration for suspected cases [35]. Anecdotal experience of educators suggests teachers in WA are hesitant to introduce the subject of IMD to students without adequate support from a reliable information source. Current educational strategies for IMD are limited by the availability of (often volunteer) staff to provide face-to-face presentations to school-aged children and by the coordination of organizing these, especially in rural and remote areas.

The primary aim of the study was to evaluate the efficacy of MIApp, a serious game that delivers health education for adolescents about IMD, compared with that of a face-to-face presentation.

XSL•FO
RenderX

The associated objectives were as follows: (1) evaluate the efficacy of the serious game MIApp on knowledge about meningococcal carriage, disease signs and symptoms, and vaccination among adolescents; (2) assess the acceptability of, usability of, and overall student experience with MIApp. The overall purpose of the investigation was to determine whether the delivery via an iPad of a fun, engaging, age-appropriate, self-navigating MIApp could be used as a valuable pedagogical tool to help educators better engage with young people around health issues they may find challenging to address. The secondary aim was to assess the retention of knowledge 3 months postintervention.

Methods

Design

The development of the educational MIApp was part of a broader initiative funded by Lotterywest, the WA Department of Health, the Amanda Young Foundation (AYF), and Edith Cowan University (ECU) between 2019 and 2022. The project aimed to raise awareness and promote IMD prevention among adolescents, a high-risk group. MIApp, a serious game, integrates thoughtfully chosen gamification mechanics and educational components. The app was developed using an iterative co-design approach, with collaboration among the funding partners and a steering committee of experts in IMD, public health, serious game development, and learning design, as well as input from high school students and end users. The app was developed with the education team of the AYF and was designed to reflect the information in the current face-to-face presentation. Health Promotion Officers developed these resources and designed them to deliver key health messages at an age-appropriate level.

The resulting narrative emerged as an overworld that brings to life the reality of meningococcal infection through interactive gameplay where the players are tasked with solving the source of IMD infection. Through MIApp, the students are introduced to the content including communicable disease, IMD, the concept of “healthy carriers,” signs and symptoms of infection, and preventive health care strategies (“ways” to be safe and remain safe). An extended MIApp user guide outlining all the game's features is available from the AYF website [36] [Multimedia Appendix 6]. Briefly, players collect items while solving the source of the meningococcal infection to increase their work experience points [Multimedia Appendices 2-4]. They interact with the dog character Buddy, who in real life was the pet of Amanda Young. Amanda tragically passed away at 18 years old from IMD, and her legacy is honored through the AYF. Pilot testing took place across 2 sites, involving students from years 7 through 10, with 68 pre- and postsurvey dyads included in the pilot study analysis. Insights from the pilot study guided adjustments to the gameplay and survey tools, leading to the formal evaluation of MIApp. These data were not included in the trial analysis.

Following the launch of MIApp in mid-2021, formal evaluation began in 2021 and was completed in 2022. We compared 2 education delivery models in the study: the intervention group, who used MIApp [Multimedia Appendices 2-4], and the active control (comparison) group, who received an in-class educational presentation [Multimedia Appendix 5]. This in-class presentation covered the same key messages and concepts on which MIApp was based. However, the students in the comparison group listened to a lecture from a qualified AYF presenter while viewing a PowerPoint presentation, followed by a classroom question and answer session. Participants played MIApp independently, without further instruction from the researchers.

Both models of delivery took approximately 50 minutes. Pre- and postintervention questionnaires were conducted to examine student responses to 9 key knowledge areas of IMD (see Textbox 1). The MIApp research design was a cluster randomized controlled trial, following data analysis from the pilot. Students were recruited across 4 year groups from a range of secondary schools. Recruitment and initial pre-post testing occurred between April 2021 and March 2021, with follow-up surveys sent 3 months after the school visit. The number of participating schools recruited was adjusted depending on the number of enrolled students per school and, consequently, the consent rate at each site, with recruitment ceasing when the prespecified sample size was reached.

Textbox 1. Key 9 knowledge areas that were assessed before and after the intervention.

Key knowledge areas to be assessed

1. Meningococcal infection is life-threatening but rare
2. Early signs and symptoms, progression of illness
3. Common symptoms of invasive meningococcal disease to look out for
4. What to do if you are experiencing symptoms
5. The importance of vaccination as a prevention tool
6. Which strains are covered by vaccination
7. The significance of “healthy carriers” in disease transmission
8. How the bacteria spread and behaviors that can result in transmission
9. The importance of avoiding behaviors that could spread the bacteria

XSL•FO
RenderX

Participants

Adolescents, being at increased risk for meningococcal carriage, were the target demographic for this study. Participants were high school students in year levels 7, 8, 9, and 10; aged 12 years to 15 years; and in the metropolitan area of Perth, WA. Schools were purposely selected using an online directory of the School Index of Community Socio-Educational Advantage (ICSEA) [37]. The ICSEA is a scale of socioeducational advantage computed for each school, allowing comparisons between schools based on the level of educational advantage or disadvantage that students bring to their academic studies. School facilities, resources, and staffing do not influence ICSEA. Sampling and recruitment aimed to acquire an equal proportion of male and female students between years 7 and 10 and across the ICSEA school values in the intervention and control groups.

The sample size calculation was based on an expected cluster size of 100 participants, with an intracluster correlation coefficient of 0.02 and an anticipated effect size of 0.4 (Cohen *d*). The design effect was calculated as 1 + (expected average cluster size – 1) × intracluster correlation coefficient. This resulted in a design effect of 2.98. After adjusting for clustering, an adequate sample size of 201 participants was required. To account for an expected 20% attrition, we aimed for an initial total sample size of 750 participants, leading to an expected final sample size of 600.

The effect size of 0.4 was selected as a moderate expected difference between the intervention and control arms, given the nature of the intervention and previous literature on similar studies. The design effect calculation and the adjustment for clustering follow standard methods for cluster randomized trials as described by Donner and Klar [38]. Cohen *d* was used during the design phase to guide the sample size estimation, though nonparametric methods were used in the analysis phase due to the skewness of the data.

A total of 6 secondary schools across metropolitan Perth, WA, gave consent and participated in the research study. Parent and student information sheets informed participants that the trial would compare MIApp with a face-to-face presentation. The consent process was undertaken by the schools, with the research team only engaging with students who had returned consent. All students who returned consent were eligible to participate in the study. Those who did not return consent participated in a different teacher-led activity chosen at the school's discretion. It was assumed that all students had a base level of computer literacy suitable for participation in the trial.

Due to mandatory lockdowns and COVID-19 rules, the study team could not travel to regional areas within the time limit of the study, and consequently, only metropolitan schools were included. Recruited schools were asked to provide an aggregate number of classes per year group (years 7, 8, 9, and 10) and students per class who were eligible to participate. Despite differences in student ICSEA scores between the MIApp and control groups, with the MIApp group having more significant variation in socioeducational backgrounds, the median was the same for each group.

Randomization Procedure

This was a clustered randomized controlled equivalence trial. Randomization was performed at a class level within each school. Following school enrollment, each site provided a class list to the research team indicating the number of classes available for participation at each level. Paper consent was collected by the school, who determined the day on which students were eligible to participate based on the provision of consent. Classes within each school were given a unique identifier and stratified by year level. These were then stratified by class level and randomly allocated to receive the intervention or control in sequential order by research team members with no affiliation to the school or knowledge of the class composition. Due to the nature of the intervention, blinding or allocation concealment was not applied. The arm of participation was recorded in the data as a numerical code (1, 2) that was not revealed to the analyst during the statistical analysis before preparing the data for reporting.

Instrumentation

All participating students were required to complete pre- and postintervention questionnaires designed to measure the primary outcome: change in knowledge of IMD. To assess the 9 key knowledge areas, 16 questions were developed (Multimedia Appendix 1). These included questions on the symptoms of IMD, where it lives in and on the body, how to prevent transmission, and what to do if you suspect you or someone you know may have IMD (Textbox 1). Participants developed a unique identifier based on elements of their name and date of birth. The participants in the intervention group were also asked to rank on a 5-point scale a series of questions about the study's secondary outcomes: the participants' perception of usability and engagement with the game and, more generally, about the delivery of educational messages using serious games as a platform [39]. Data on knowledge acquisition, timing, and behavior were also obtained to gain further insight into the overall usability, learnability, and experience of the MIApp game.

The validity and reliability of the research instrument were established in several ways during the project. First, the instrument was closely examined by the research team and the steering committee to ensure alignment with formative research, curriculum text, clarity, and ease of language for the age group and the functioning and facilities within the instrument, such as response mechanisms. The instrument was then pilot tested for the construct validity of the topics included. Minor revisions were made to discordant wording. Test-retest reliability was not performed for the questionnaire items.

Procedures

Following agreement to participate in the MIApp study and written informed consent provided by the school principal, schools were asked to provide an aggregate number of classes per year group (years 7, 8, 9, and 10) and students per class eligible to participate. A parent or carer and student (participant) information sheet and consent form were disseminated by the schools before the provision of consent and content delivery. This information sheet outlined the rationale for the study, the

XSL•FO
RenderX

risks and benefits to participants, and the proposed activity undertaken as part of the research.

The educational content for the intervention and control groups was based on the AYF guidelines to support safer, healthier, and more physically active living by young Australians by creating meaningful, valuable, and strengths-based learning focused on IMD awareness and prevention. They reflect WA's Healthy and Physical Education (HPE) learning outcomes via the WA P-10 Syllabus for HPE and the Australian Curriculum for HPE [40]. The intervention, MIApp, is a new software application developed by ECU. MIApp is a self-directed, serious game designed to be completed in 30 minutes. It targets students in years 7 through 10 to raise awareness and bring to life the realities of meningococcal infection. At the Detect and Protect Agency, players investigate a case to solve the source of a mystery infection. Players engage with the learning content and key messages delivered in the AYF presentation through a combination of games, quizzes, simulations, and interactive activities. The control group received a face-to-face educational presentation delivered by a presenter from the AYF. This is the primary mode of education developed by the AYF and includes a secondary school PowerPoint presentation adapted to fit the allocated research time of 30 minutes. The presentation delivers 9 key messages about IMD including prevention and identification of symptoms and preventative health care strategies (see Textbox 1). Embedded in the presentation is a short interactive quiz and a video highlighting real-life stories of young people who have caught and survived the disease.

Classes within each participating school were randomized to the intervention or control arm stratified by year group, with one-half of the classes at each school receiving either the app or the face-to-face presentation. At the discretion of individual schools, classrooms or locations were set up for the research trials. A requirement was that rooms be separate, so no interference was experienced with either mode of delivery. All questionnaires were completed by the participants using the Qualtrics survey platform [41] on an iPad during the initial onsite visit and via an emailed link for the 3-month follow-up.

Tablets were distributed to all the participants, followed by a briefing on the scope of the research and instructions for the session. Each session (control and intervention) took 50 minutes comprised of 5 minutes to 10 minutes for the pre-intervention questionnaire, 30 minutes for either the intervention (MIApp) or control (presentation), and a further 5 minutes to 10 minutes for the postintervention questionnaire. There was limited interaction between the researchers and participants in order to mirror the use in a classroom environment. Participants were first asked to create a unique identifier that could be used to link the questionnaire data collected before and after their session and again at the 3-month follow-up. The identifier used the first 2 letters of their first name, the day of their date of birth, and the last 2 letters of their last name. MIApp cohort participants also used this unique identifier to link survey responses to in-app data gathered by MIApp. Participants were not required to remember the username they created for the follow-up survey, and no identifying information about the student was collected by the app.

Data Analysis

Statistical analysis was conducted using R version 4.0.3 [42]. “Differences in correct” scores of the 16 pre and postintervention questions were used to compare the MIApp and control cohorts (Multimedia Appendix 1). Comparative subanalysis of the differences was conducted by age and sex to assess differences in median score improvement between groups. The Kruskal-Wallis test was used for continuous variables [43]. Differences in the proportions of the baseline median included a Pearson chi-square test for categorical variables. Statistically significant differences were considered at $P<.05$. Statistical equivalence testing was also performed using the method by Rusticus and Lovato [44]. First, 95% CIs were calculated using Games-Howell post hoc tests (which account for unequal group sizes and violations in homogeneity of variance). The equivalence interval was defined as ±2 points in the pre-post score difference, which we considered a priori to represent a similar level of improvement. Data were analyzed using intention-to-treat analysis, whereby all participants were analyzed according to the intervention they received on the day of the trial.

Ethics Approval

Ethics approval for the study was obtained from the Human Research Ethics Committee at ECU (reference: 2019-00736-BLOOMFIELD). The parents of all participants were provided with participant information and a consent form outlining the rationale for the trial and the activities of the intervention and control groups.

Trial Registration

As this was not clinical research, it was not registered on the Australian New Zealand Clinical Trials Registry. A protocol was not published prior to commencement.

Results

Approximately 1500 students were invited to participate in the trial by the 6 schools invited to participate. A total of 788 participants from the 6 trial schools consented and completed the pre- and postintervention questionnaires and were eligible for inclusion in the primary analysis: 442 in the intervention arm and 346 in the control arm (Figure 1). Although different demographic patterns were observed across the 6 participating schools, approximately one-half of the students were in year 7 (397/788, 50.4%), were female (397/788, 50.4%), and self-reported not knowing anything about meningococcal disease at the start of the trial (395/788, 50.1%) (Table 1). The demographic and self-reported pre-knowledge characteristics of the study population by group are shown in Table 2. There were no statistically significant differences between the groups, which suggests a good balance of these properties following group randomization.

XSL•FO
RenderX

Figure 1. CONSORT Flow Diagram.

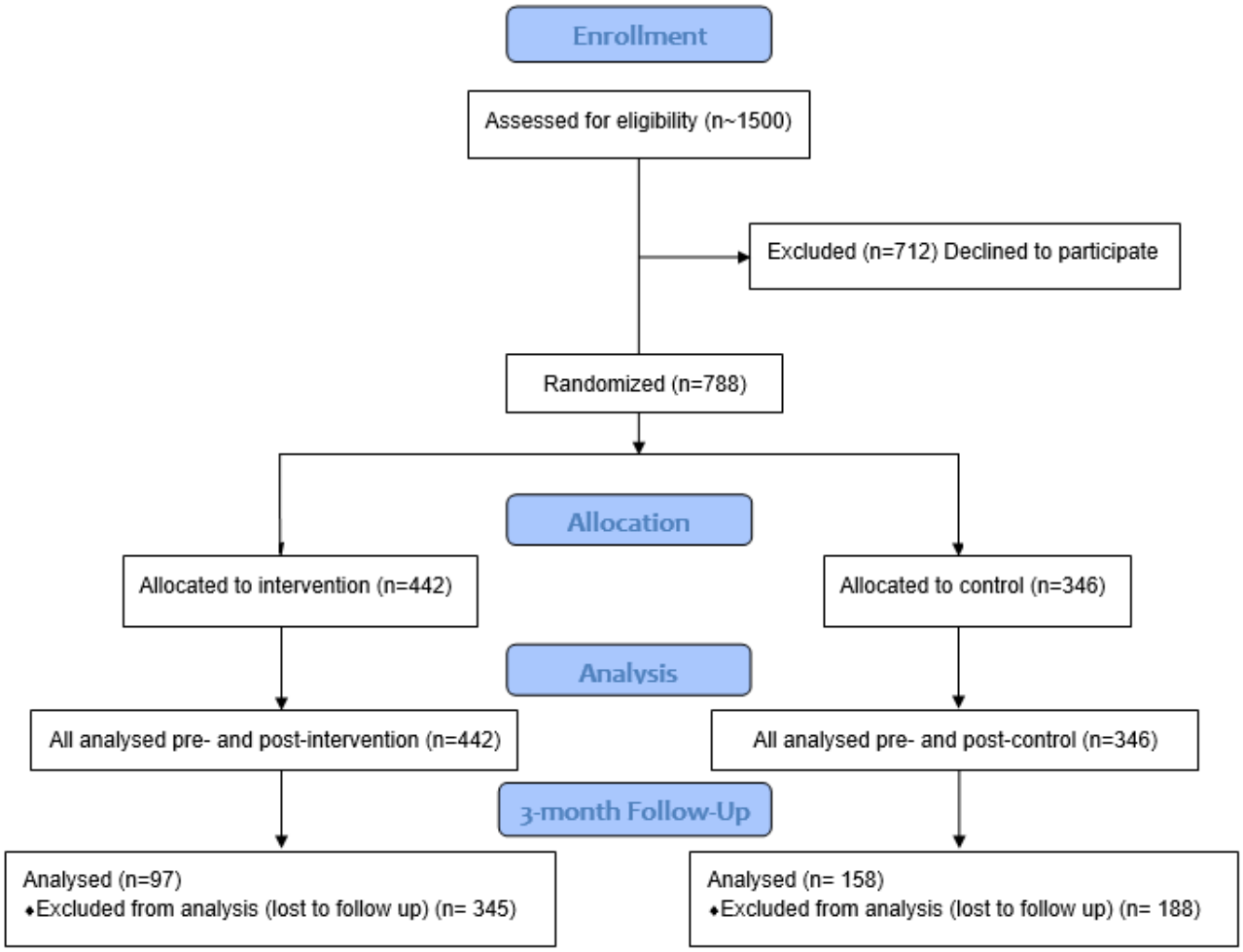


Table 1. Demographics of the participants.

Demographics	School 1 (n=173), n (%)	School 2 (n=56), n (%)	School 3 (n=138), n (%)	School 4 (n=198), n (%)	School 5 (n=214), n (%)	School 6 (n=9), n (%)	Total sample (N=788), n (%)
ICSEA[a]	1100	1013	1032	1110	1120	890	—[b]
Year							
7	111 (64.2)	26 (46.4)	0 (0)	46 (23.2)	214 (100)	0 (0)	397 (50.4)
8	62 (35.8)	5 (8.9)	0 (0)	31 (15.7)	0 (0)	6 (66.7)	104 (13.2)
9	0 (0)	25 (44.6)	138 (100)	24 (12.1)	0 (0)	3 (33.3)	190 (24.1)
10	0 (0)	0 (0)	0 (0)	97 (49)	0 (0)	0 (0)	97 (12.3)
Gender							
Male	67 (38.7)	31 (55.4)	76 (55.1)	93 (47)	91 (42.5)	3 (33.3)	361 (45.8)
Female	102 (59)	21 (37.5)	56 (40.6)	97 (49)	116 (54.2)	5 (55.6)	397 (50.4)
Other	4 (2.3)	4 (7.1)	6 (4.3)	8 (4)	7 (3.3)	1 (11.1)	30 (3.8)

[a]ICSEA: Index of Community Socio-Educational Advantage.

[b]Not applicable.

XSL•FO
RenderX

Table 2. Demographic characteristics and self-reported preknowledge of meningococcal disease in the Meningococcal Immunisation Awareness, Prevention and Protection app (MIApp) group versus the control presentation group.

Demographic characteristics and preknowledge	MIApp (n=442)	Control (n=346)	Total sample (n=788)	*P* value[a]
Year, n (%)				.19
7	215 (48.6)	182 (52.6)	397 (50.4)	
8	60 (13.6)	44 (12.7)	104 (13.2)	
9	118 (26.7)	72 (20.8)	190 (24.1)	
10	49 (11.1)	48 (13.9)	97 (12.3)	
Gender, n (%)				.94
Male	200 (45.2)	161 (46.5)	361 (45.8)	
Female	225 (50.9)	172 (49.7)	397 (50.4)	
Other	17 (3.8)	13 (3.8)	30 (3.8)	
Baseline knowledge[b], n (%)				.62
Nothing	218 (49.3)	177 (51.2)	395 (50.1)	
Some	217 (49.1)	166 (48)	383 (48.6)	
A lot	7 (1.6)	3 (0.9)	10 (1.3)	
ICSEA[c], median (IQR)	1100 (1032-1120)	1100 (1100-1120)	—[d]	.02

[a]Pearson chi-square test for categorical variables, and the Kruskal-Wallis test for continuous variables.

[b]Students were asked to nominate which of these 3 statements was the most accurate: I don't know anything about meningococcal disease, I know a little about meningococcal disease, and I know a lot about meningococcal disease.

[c]ICSEA: Index of Community Socio-Educational Advantage.

[d]Not applicable.

Because they answered “I’m not sure/I don’t know” to all the questions on the survey, 33 participants in the control group (33/346, 9.5%) and 18 participants in the MIApp group (18/442, 4.1%) had a score of 0 for the pre-intervention survey. The median pre-intervention correct score in both the MIApp and control cohorts was 6/16 (MIApp: IQR 4-8; control: IQR 3-9), and the median postintervention correct score in both the MIApp and control cohorts was 14/16 (MIApp: IQR 12-15; control: IQR 13-15). The median improvement in score for the total cohort was 7/16 (IQR 5-10). The median improvements in score were 7/16 (IQR 5-9) for the MIApp arm and 8/16 (IQR 5-11) for the control arm (P<.001). Statistical equivalence in mean improvement between arms was demonstrated, with the confidence intervals for the mean difference falling within 2 points (MIApp vs control: mean difference=–1.036, lower limit=–1.569, upper limit=–0.503). Score improvement differed by year group, with students in years 9 and 10 having slightly lower median improvements in score (7 and 6, respectively) than those in years 7 and 8, who both had median improvements of 8 (P<.001). No statistically significant differences in median score improvement were noted by gender (P=.20; Table 3).

XSL•FO
RenderX

Table 3. Median number of correct scores (from a total score of 16) pre- and postintervention and changes in the median score by year and gender.

Characteristic	MIApp[a] (n=442)			F2F[b] (n=346)			Total sample (N=788)			*P* value[c]
	Pre-intervention, median (IQR)	Postintervention, median (IQR)	Change in the score, median (IQR)	Pre-intervention, median (IQR)	Postintervention, median (IQR)	Change in the score, median (IQR)	Pre-intervention, median (IQR)	Postintervention, median (IQR)	Change in the score, median (IQR)	
Year										<.001
7	6 (3.5-8)	14 (12-15)	7 (5-9.5)	6 (3-8)	14 (13-15)	8 (6-11)	6 (3-8)	14 (13-15)	8 (5-10)	
8	6 (3.75-8)	14 (13-15)	7 (4.75-10)	6 (2-8)	15 (14-15)	9 (6-12)	6 (3-8)	14 (13-15)	8 (5-10.25)	
9	7 (4-8)	14 (13-15)	7 (5-10)	6 (4-9)	14 (13-15)	8 (5-11)	6 (4-8)	14 (13-15)	7 (5-10)	
10	7 (6-9)	14 (12-15)	5 (4-8)	7 (5.75-9.25)	15 (14-15)	7 (4-9)	7 (6 9)	14 (13-15)	6 (4-8)	
Gender										.20
Male	7 (4-9)	14 (12-15)	7 (4-9)	6 (3-8)	14 (13-15)	8 (5-11)	6 (3-8)	14 (13-15)	7 (5-10)	
Female	4 (4-8)	14 (13-15)	7 (5-10)	6 (3-9)	14 (13-15)	8 (5-11)	6 (4-9)	14 (13-15)	8 (5-10)	
Other	6 (5-8)	13 (8-14)	5 (2-8)	6 (2-7)	14 (13-15)	8 (6-11)	6 (4-7.75)	13 (11-14)	6.5 (2.5-9)	

[a]MIApp: Meningococcal Immunisation Awareness, Prevention and Protection app.

[b]F2F: face to face.

[c]Kruskal-Wallis test.

Table 4 shows the percentage of MIApp participants responding to questions about attitudes toward and engagement with the game; because these questions were not compulsory, not all students responded. Most participants responded positively: “I thought MIApp was easy to use” (348/426, 81.7%); “It was clear when I was playing what I was supposed to do next” (342/427, 80.1%); “MIApp was more enjoyable than a presentation” (326/426, 76.5%); “I felt very confident using MIApp” (315/423, 74.5%); and “I would like to use educational games like this more often” (313/424, 73.8%). Few found it difficult to use or needed help: “I needed to learn a lot more before I could understand MIApp” (58/426, 13.6%); “Finding things in MIApp was too hard” (45/428, 10.5%); “I found using MIApp a bit overwhelming” (37/424, 8.7%); and “I needed help from my teacher to be able to use MIApp” (27/427, 6.3%).

XSL•FO
RenderX

Table 4. Student attitudes about the Meningococcal Immunisation Awareness, Prevention and Protection app (MIApp) by year.

Statement	Agreement with the statement, n (%)				
	Year 7 (n=215)	Year 8 (n=60)	Year 9 (n=118)	Year 10 (n=49)	Total sample (n=442)
I would like to use educational games like this more often	162 (79.4)[a]	45 (76.3)[b]	73 (64)[c]	33 (70.2)[d]	313 (73.8)[e]
Finding things in MIApp was too hard	23 (11.1)[f]	3 (5.1)[b]	10 (8.8)[c]	9 (19.1)[d]	45 (10.5)[g]
It was clear when I was playing what I was supposed to do next	166 (80.2)[h]	48 (81.4)[b]	91 (79.8)[c]	37 (78.7)[d]	342 (80.1)[i]
I needed help from my teacher to be able to use MIApp	15 (7.2)[h]	2 (3.4)[b]	5 (4.4)[c]	5 (10.6)[d]	27 (6.3)[i]
MIApp was more enjoyable than a presentation	159 (76.8)[h]	46 (78)[b]	86 (75.4)[c]	35 (76.1)[j]	326 (76.5)[k]
I thought the features of on the screen on MIApp were confusing	27 (13.1)[l]	3 (5.1)[b]	7 (6.2)[m]	12 (25.5)[d]	49 (11.5)[n]
I thought MIApp was easy to use	168 (81.6)[l]	48 (81.4)[b]	92 (80.7)[c]	40 (85.1)[d]	348 (81.7)[k]
I found using MIApp a bit overwhelming	17 (8.3)[o]	3 (5.1)[b]	12 (10.6)[m]	5 (10.6)[d]	37 (8.7)[e]
I felt very confident using MIApp	151 (73.7)[o]	44 (74.6)[b]	85 (75.9)[p]	35 (74.5)[d]	315 (74.5)[q]
I needed to learn a lot more before I could understand MIApp	29 (14.1)[l]	9 (15.3)[b]	14 (12.3)[c]	6 (12.8)[d]	58 (13.6)[k]

[a]n=204.
[b]n=59.
[c]n=114.
[d]n=47.
[e]n=424.
[f]n=208.
[g]n=428.
[h]n=207.
[i]n=427.
[j]n=46.
[k]n=426.
[l]n=206.
[m]n=113.
[n]n=425.
[o]n=205.
[p]n=212.
[q]n=423.

A total of 255 participants from both groups completed nonduplicated (by unique identifier) responses for the 3-month follow-up survey, with a lower response rate for the MIApp arm (MIApp: 97/442, 22%; control: 158/346, 45.7%; *P*<.001). Of these, only 89 responses could be linked to the primary questionnaire results using the unique study identifier (MIApp: 50/442, 11.3%; control: 39/346, 11.3%; *P*=.99). The inability to match responses to the original survey for the majority of respondents precludes an analysis at follow-up by demographic factors.

The 3-month follow-up survey asked participants to answer the same set of questions related to the 9 key knowledge areas and indicate if they participated in the MIApp or control arm on the day of the trial. The median total correct scores at the 3-month follow-up were 11/16 (IQR 9-13) and 12/16 (IQR 9-13) for the MIApp and control arms, respectively (*P*=.86). A sensitivity analysis only including the 89 participants who could be matched to the trial surveys produced similar results (correct scores of 12/16 and 13/16 for the intervention and control arms, respectively).

Although there was a slight decline in knowledge at the 3-month mark, this again demonstrates that there was similar knowledge retention in both groups 3 months after the initial lesson. Overall, these results suggest very similar improvements in the pre-post intervention scores by intervention group, year, and gender. This suggests that students either receiving an educational session via an AYF presenter or playing the MIApp can expect to improve their knowledge of IMD in the 9 key areas outlined in the previous sections.

XSL•FO
RenderX

Discussion

Principal Findings

This study demonstrated the efficacy of a serious game, MIApp, in improving knowledge about IMD in high school–aged children at equivalent levels to that of a face-to-face presentation delivered by a content expert. Serious games are attracting significant attention for promoting health education among children and adolescents [4-6,45], yet there is still a need to understand their effectiveness [22,27,30]. Our literature review suggested that, although serious games may offer a viable alternative to traditional pedagogical methods in terms of knowledge improvement, behavioral and attitudinal change, and engagement, the varied nature of serious games used in health promotion and the lack of standardized tools to measure outcomes mean that the results are heterogeneous, often with small sample sizes and high attrition [18].

This study compared the efficacy of the serious game MIApp with a traditional classroom health education model on knowledge about IMD in a sample of secondary school students. The results demonstrated that this serious game (1) significantly improved knowledge about the disease among adolescents, (2) had an outcome statistically equivalent to that of the current face-to-face educational approach, and (3) was easy to use and more enjoyable. Importantly, the study showed that gains in knowledge from playing the serious game MIApp were sustained 3 months later at equivalent levels to those in the control arm.

In line with recent research, this study supports the notion that age-appropriate, self-navigating serious games may be considered a positive learning tool for delivering school-based health education [25,26,46,47]. One fundamental feature of serious games is that they are “fun” and pedagogically sound [7-9]. As illustrated by the MIApp study, adolescents find digital game–based learning more enjoyable than a traditional classroom presentation, characteristics that are acknowledged as essential for effective engagement [48]. By providing a captivating learning environment, serious games can intrinsically and extrinsically motivate adolescents to promote active learning to understand health issues better [17,30,49]. Research also suggests that, through the provision of interactive engagement and immediate feedback, serious games can maintain students’ involvement for a longer period [50-52], supporting higher-level thinking and increasing learner independence [10].

With increasing numbers of children and adolescents in Australia using digital games [1], together with the immediate transition to digital learning during the COVID-19 pandemic [53], serious games offer a unique educational platform to improve the reach of health education [5,28,54,55]. As a self-guided resource not limited to staff facilitating an education session, serious games can support teachers to provide health education around topics they have limited knowledge about or may find difficult to discuss [7]. Moreover, this study revealed that students using MIApp also retained their IMD knowledge over 3 months, as did those who attended a face-to-face educational presentation. Although research has suggested serious games might be more beneficial for long-term retention of health information due to their enjoyable format [4], to date, most studies have focused primarily on short-term pre- and posttests [25].

Although this study was conducted with schools in primarily medium-to-high ICSEA areas of Perth, WA, as shown in Table 2, its relevance needs to be assessed in rural settings where cases of IMD are higher; the preliminary findings of this study suggest MIApp could represent an emerging resource in the space of meningococcal education for adolescents and young adults in Australia. Key features underlying MIApp are known to promote motivational learning, improve knowledge, and promote engagement, which can increase educational achievement [56,57], and are likely responsible for the positive findings observed in our study [46,52,58]. MIApp is freely downloadable from the App store [59] on personal devices or is available on the AYF website [60] as a desktop version, accessible to all, including those regional and rural communities that previously would not have had the resources or access to receive the face-to-face guest presentations. Findings of this study can help teachers to feel more assured that students are gaining the same learning opportunity through an interactive and engaging app as they would receive from a qualified presenter.

As shown in previous research [26,31,47,56], engaging key stakeholders in the co-design, development, testing, and refining of the serious game MIApp was essential; adopting a co-design approach with all stakeholders is useful in promoting health communication interventions and enhancing the acceptability, usability, and utility of the serious game [26]. For MIApp, collaborators included an interdisciplinary team of researchers, experts, and community partners with varied expertise in education, health and computer science, and game development. This collaboration facilitated important development and design improvements of the game interface, mechanism, health content, and pedagogy for students and ensured that the social and cultural interests of the targeted adolescent group were included.

For greater influence, serious games for health education and promotion should be designed in a way that enables them to be embedded in national curriculum, as this gives teachers the opportunity to combine gameplay with the mandated curriculum to address learning outcomes [15,16,18]. MIApp was aligned specifically with the teaching and learning HPE curriculum as outlined by the Australian curriculum and, more specifically, the lower secondary year-level syllabi for the HPE Learning Area in WA [40]. Additionally, a free downloadable teaching and learning curriculum companion of activities was developed for teachers to assist them to cognitively build and support student understanding of concepts and health-enhancing skills and dispositions related to meningococcal awareness and prevention and is available from the AYF website. These learning activities have also been aligned to meet the teaching and learning requirements of the HPE curriculum.

The overall educational package draws on 5 interrelated ideas underpinning the pedagogy in the delivery of HPE in Australia. The broader dissemination of this tool by teachers as a public health intervention could support the digital implementation of national efforts to reduce IMD among adolescents. The demonstrated efficacy of MIApp means that educators,

XSL•FO
RenderX

particularly in rural and remote areas without previous access to specialist education sessions about IMD, can deploy this game in the classroom to provide their students with self-guided education supported by teaching resources.

Limitations

There were several limitations to this research. First, a major limitation to this study was the low response rate and match rate for the 3-month follow-up survey. There was a relatively low rate of complete responses (255/788, 32.4%) for the 3-month follow-up; the response rate was significantly lower for the MIApp group (97/442, 21.9%) than the control group (158/346, 45.7%). Attrition for this follow-up survey was expected to be low, and our literature review highlighted significant issues with attrition experienced by researchers running similar interventions [18]. Although attrition was built into the sample size calculation, the 3-month follow-up response rate was still much lower than anticipated. Future projects may consider methods to improve response rates, including incentives for completing follow-up surveys.

Further, only 35% of completed surveys could be linked to the identifiers for the pre-postintervention arm, due to students entering an identifier that did not match the one created on the day of trial. The method for creating a unique identifier involved using parts of key demographic information (name and date of birth) that were considered robust and unlikely to change over the study period. Despite this assumption, 65% of identifiers and school codes did not match those created on the day of the trial. A more robust identifier may be considered for future trials where follow-up data need to be linked to data collected on the day of the trial.

Finally, a significantly lower proportion of responses in the 3-month follow-up was received from those in the intervention arm than those in the control arm. The implications of this difference are that, potentially, respondents from the intervention arm were systematically different from those who did not respond. If respondents reflect those who were more likely to retain information, this may have biased the results. The overall low response rate in both groups may suggest this bias would be present in both; however, residual bias from the difference in response rates between groups may still be present. A sensitivity analysis using only surveys that were able to be linked to the primary trial data, which also allows independent confirmation that the correct trial arm was selected on the follow-up survey, yielded a response rate that was *not* significantly different between groups. In addition, the data from these groups confirmed the 1-point difference in correct responses between the intervention and control arms reported in the larger follow-up cohort, which was within the predetermined equivalence threshold. This increases our confidence that the results reported in the larger follow-up sample were reflective of the larger cohort.

Despite these limitations, the completed surveys suggested that many key messages delivered via either MIApp or in a face-to-face presentation are retained 3 months after a single 30-minute session. Future studies assessing the longer-term efficacy of serious games for health promotion may consider outcome measures of behavior change that do not rely on participant response rate (eg, vaccination and infection rates).

The overall findings of this study suggest that serious games, such as MIApp, represent an effective strategic avenue for the use of digital technology to engage adolescents in learning about key health issues. The research found both delivery methods to be similar in improving knowledge and attitudes toward meningococcal carriage and disease signs, symptoms, and prevention. Coupled with greater reported enjoyment, if well-designed, serious games could be used in national health education. The strengths of this study included the engagement of a range of key stakeholders in its planning and design to improve outcomes and ensure specificity of knowledge relating to the curriculum and possibilities for content inclusion in the teaching and learning of HPE in Australia. Although further investigations are needed to explore any minor changes required for the delivery in the diverse socioeconomic, cultural, and geographic settings of Australia, this study has illustrated the efficacy of the MIApp game. Moreover, data generated in this study can support evidence for the development of future serious games for adolescent health education worldwide. Today, in our digital landscape, greater understanding of the game learning interests of adolescent students is paramount to actively promote their active engagement in health education.

Acknowledgments

Funding for this study was provided by a Lotterywest Grant (2019–00736-BLOOMFIELD), with additional funding from the Western Australia Department of Health, the Amanda Young Foundation (AYF), and Edith Cowan University (ECU).

Data Availability

The data sets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Authors' Contributions

LB, AD, LA, DB, and MM conceptualized the study. LB and AD curated the data, acquired funding, and designed the methodology. LB and JB performed the formal analysis. JB, AD, and DB performed the investigation. AD, JB, and DB provided project administration. AD provided the resources. MM provided the software. AD, JB, and MM supervised the study. LB created the

XSL•FO
RenderX

visualizations. LB, JB, and DB wrote the original draft of the manuscript. LB, AD, LA, JB, DB, and MM reviewed and edited the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
MiApp Post Study Questionnaire.
[DOCX File , 32 KB - games_v13i1e60755_app1.docx]

Multimedia Appendix 2
Screenshots from MiApp game.
[PNG File , 545 KB - games_v13i1e60755_app2.png]

Multimedia Appendix 3
Screenshots #2 from MiApp game.
[PNG File , 545 KB - games_v13i1e60755_app3.png]

Multimedia Appendix 4
Screenshot #1 from MiApp game.
[PNG File , 459 KB - games_v13i1e60755_app4.png]

Multimedia Appendix 5
Amanda Young Foundation classroom presentation.
[PDF File (Adobe PDF File), 843 KB - games_v13i1e60755_app5.pdf]

Multimedia Appendix 6
MIApp Extended User Guide.
[PDF File (Adobe PDF File), 11572 KB - games_v13i1e60755_app6.pdf]

Multimedia Appendix 7
CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) form for the Meningococcal Immunisation Awareness, Prevention and Protection app (MIApp) study.
[PDF File (Adobe PDF File), 894 KB - games_v13i1e60755_app7.pdf]

References

1. 9 in 10 Aussie teens now have a mobile (and most are already on to their second or subsequent handset). Roy Morgan. 2016 Aug 22. URL: http://www.roymorgan.com/findings/6929-australian-teenagers-and-their-mobile-phones-june-2016-201608220922 [accessed 2024-12-07]
2. Emmerich K, Masuch M, Schmidt R. Digital Entertaining Health Applications: Researching the Fundamentals of a Cause-Effect Relationship. 2014 Nov 11 Presented at: Workshops on Advances in Computer Entertainment Conference; November 11-14, 2014; Funchal, Portugal. [doi: 10.1145/2693787.2693800]
3. Qian M, Clark K. Game-based learning and 21st century skills: A review of recent research. Computers in Human Behavior 2016 Oct;63:50-58. [doi: 10.1016/j.chb.2016.05.023]
4. Blakely G, Skirton H, Cooper S, Allum P, Nelmes P. Educational gaming in the health sciences: systematic review. J Adv Nurs 2009 Feb;65(2):259-269. [doi: 10.1111/j.1365-2648.2008.04843.x] [Medline: 19032512]
5. DeSmet A, Van Ryckeghem D, Compernolle S, Baranowski T, Thompson D, Crombez G, et al. A meta-analysis of serious digital games for healthy lifestyle promotion. Prev Med 2014 Dec;69:95-107 [FREE Full text] [doi: 10.1016/j.ypmed.2014.08.026] [Medline: 25172024]
6. Wouters P, van Nimwegen C, van Oostendorp H, van der Spek E. A meta-analysis of the cognitive and motivational effects of serious games. Journal of Educational Psychology 2013 May;105(2):249-265. [doi: 10.1037/a0031311]
7. Ronimus M, Eklund K, Pesu L, Lyytinen H. Supporting struggling readers with digital game-based learning. Education Tech Research Dev 2019 Feb 20;67(3):639-663. [doi: 10.1007/s11423-019-09658-3]
8. Zainuddin Z. Students' learning performance and perceived motivation in gamified flipped-class instruction. Computers & Education 2018 Nov;126:75-88. [doi: 10.1016/j.compedu.2018.07.003]
9. Zyda M. From visual simulation to virtual reality to games. Computer 2005 Sep;38(9):25-32. [doi: 10.1109/mc.2005.297]

XSL•FO
RenderX

10. Newhouse C. STEM the boredom: engage students in the Australian curriculum using ICT with problem-based learning and assessment. J Sci Educ Technol 2016 Sep 14;26(1):44-57. [doi: 10.1007/s10956-016-9650-4]
11. Annetta LA, Minogue J, Holmes SY, Cheng M. Investigating the impact of video games on high school students' engagement and learning about genetics. Computers & Education 2009 Aug;53(1):74-85. [doi: 10.1016/j.compedu.2008.12.020]
12. Plass JL, Homer BD, Kinzer CK. Foundations of game-based learning. Educational Psychologist 2016 Feb 06;50(4):258-283. [doi: 10.1080/00461520.2015.1122533]
13. Young MF, Slota S, Cutter AB, Jalette G, Mullin G, Lai B, et al. Our princess is in another castle. Review of Educational Research 2012 Mar 01;82(1):61-89. [doi: 10.3102/0034654312436980]
14. Barwood D, Smith S, Miller M, Boston J, Masek M, Devine A. Transformational game trial in nutrition education. Australian Journal of Teacher Education 2020 Apr;45(4):18-29. [doi: 10.14221/ajte.2020v45n4.2]
15. Li M, Tsai C. Game-based learning in science education: a review of relevant research. J Sci Educ Technol 2013 Feb 6;22(6):877-898. [doi: 10.1007/s10956-013-9436-x]
16. Masek M, Boston J, Lam C, Corcoran S. Improving mastery of fractions by blending video games into the Math classroom. Computer Assisted Learning 2017 Apr 27;33(5):486-499. [doi: 10.1111/jcal.12194]
17. Papastergiou M. Exploring the potential of computer and video games for health and physical education: A literature review. Computers & Education 2009 Nov;53(3):603-622. [doi: 10.1016/j.compedu.2009.04.001]
18. Andrew L, Barwood D, Boston J, Masek M, Bloomfield L, Devine A. Serious games for health promotion in adolescents - a systematic scoping review. Educ Inf Technol (Dordr) 2023;28(5):5519-5550 [FREE Full text] [doi: 10.1007/s10639-022-11414-9] [Medline: 36373044]
19. Barber NC, Stark LA. Online resources for understanding outbreaks and infectious diseases. CBE Life Sci Educ 2015 Mar 02;14(1):fe1 [FREE Full text] [doi: 10.1187/cbe.14-12-0221] [Medline: 25678620]
20. Fontes P, Boaventura P, Azevedo J. Mobile games to improve healthier lifestyle amongst youth: "are we there yet" or should we move forward? 2014 Nov 11 Presented at: Workshops on Advances in Computer Entertainment Conference; November 11-14, 2014; Funchal, Portugal. [doi: 10.1145/2693787.2693799]
21. Wattanasoontorn V, Boada I, García R, Sbert M. Serious games for health. Entertainment Computing 2013 Dec;4(4):231-247. [doi: 10.1016/j.entcom.2013.09.002]
22. Espinosa-Curiel IE, Pozas-Bogarin EE, Hernández-Arvizu M, Navarro-Jiménez ME, Delgado-Pérez EE, Martínez-Miranda J, et al. HelperFriend, a serious game for promoting healthy lifestyle behaviors in children: design and pilot study. JMIR Serious Games 2022 May 06;10(2):e33412 [FREE Full text] [doi: 10.2196/33412] [Medline: 35522474]
23. Ricciardi F, De Paolis LT. A comprehensive review of serious games in health professions. International Journal of Computer Games Technology 2014;2014:1-11. [doi: 10.1155/2014/787968]
24. Baranowski T, Ryan C, Hoyos-Cespedes A, Lu AS. Nutrition education and dietary behavior change games: a scoping review. Games Health J 2019 Jun;8(3):153-176 [FREE Full text] [doi: 10.1089/g4h.2018.0070] [Medline: 30339086]
25. Holzmann SL, Schäfer H, Groh G, Plecher DA, Klinker G, Schauberger G, et al. Short-term effects of the serious game "Fit, Food, Fun" on nutritional knowledge: a pilot study among children and adolescents. Nutrients 2019 Aug 30;11(9):A. [doi: 10.3390/nu11092031] [Medline: 31480257]
26. Chu SKW, Kwan ACM, Reynolds R, Mellecker RR, Tam F, Lee G, et al. Promoting sex education among teenagers through an interactive game: reasons for success and implications. Games Health J 2015 Jun;4(3):168-174. [doi: 10.1089/g4h.2014.0059] [Medline: 26182060]
27. DeSmet A, Shegog R, Van Ryckeghem D, Crombez G, De Bourdeaudhuij I. A systematic review and meta-analysis of interventions for sexual health promotion involving serious digital games. Games Health J 2015 Apr;4(2):78-90. [doi: 10.1089/g4h.2014.0110] [Medline: 26181801]
28. Rodriguez DM, Teesson M, Newton NC. A systematic review of computerised serious educational games about alcohol and other drugs for adolescents. Drug Alcohol Rev 2014 Mar;33(2):129-135. [doi: 10.1111/dar.12102] [Medline: 24329810]
29. David OA, Costescu C, Cardos R, Mogoaşe C. How effective are serious games for promoting mental health and health behavioral change in children and adolescents? A systematic review and meta-analysis. Child Youth Care Forum 2020 Jul 31;49(6):817-838. [doi: 10.1007/s10566-020-09566-1]
30. Kato PM. Evaluating efficacy and validating games for health. Games Health J 2012 Feb;1(1):74-76. [doi: 10.1089/g4h.2012.1017] [Medline: 26196436]
31. Sharifzadeh N, Kharrazi H, Nazari E, Tabesh H, Edalati Khodabandeh M, Heidari S, et al. Health education serious games targeting health care providers, patients, and public health users: scoping review. JMIR Serious Games 2020 Mar 05;8(1):e13459 [FREE Full text] [doi: 10.2196/13459] [Medline: 32134391]
32. Meningococcal Infection Notifications 2018. 32 WA Department of Health. 2018. URL: http://ww2.health.wa.gov.au/Articles/N_R/Notifiable-infectious-disease [accessed 2020-09-01]
33. Sadarangani M, Scheifele DW, Halperin SA, Vaudry W, Le Saux N, Tsang R, investigators of the Canadian Immunization Monitoring Program, ACTive (IMPACT). Outcomes of invasive meningococcal disease in adults and children in Canada between 2002 and 2011: a prospective cohort study. Clin Infect Dis 2015 Apr 15;60(8):e27-e35. [doi: 10.1093/cid/civ028] [Medline: 25605282]

XSL•FO
RenderX

34. Burman C, Serra L, Nuttens C, Presa J, Balmer P, York L. Meningococcal disease in adolescents and young adults: a review of the rationale for prevention through vaccination. Hum Vaccin Immunother 2019;15(2):459-469 [FREE Full text] [doi: 10.1080/21645515.2018.1528831] [Medline: 30273506]
35. Invasive Meningococcal Disease: CDNA National Guidelines for Public Health Units. Communicable Diseases Network Australia. 2024 Jul 01. URL: https://www.health.gov.au/sites/default/files/2024-07/invasive-meningococcal-disease-cdna-national-guidelines-for-public-health-units.pdf [accessed 2024-12-07]
36. MIApp Years 7 to 10. The Amanda Young Foundation. URL: https://amandayoungfoundation.org.au/miapp-years-7-to-10/ [accessed 2024-12-07]
37. Find a School. Australian Curriculum, Assessment and Reporting Authority. URL: https://www.myschool.edu.au/ [accessed 2024-12-07]
38. Klar N, Donner A. Design and Analysis of Cluster Randomization Trials in Health Research. London, England: Wiley; 2000.
39. Brooke J. SUS - A quick and dirty usability scale. In: Jordan PW, Thomas B, McClelland IL, Weerdmeester B, editors. Usability evaluation in industry (Vol. 189). New York, NY: Routledge; 1996:4-7.
40. Health and Physical Education. School Curriculum and Standards Authority. URL: https://k10outline.scsa.wa.edu.au/home/teaching/curriculum-browser/health-and-physical-education [accessed 2024-12-07]
41. Qualtrics. URL: https://www.qualtrics.com/ [accessed 2024-12-07]
42. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria; 2021. URL: https://www.R-project.org/ [accessed 2024-12-07]
43. Kruskal WH, Wallis WA. Use of ranks in one-criterion variance analysis. Journal of the American Statistical Association 1952 Dec;47(260):583-621. [doi: 10.1080/01621459.1952.10483441]
44. Rusticus SA, Lovato CY. Applying tests of equivalence for multiple group comparisons: demonstration of the confidence interval approach. Practical Assessment, Research, and Evaluation 2011;16(1):7. [doi: 10.7275/d5wf-5p77]
45. Charlier N, Zupancic N, Fieuws S, Denhaerynck K, Zaman B, Moons P. Serious games for improving knowledge and self-management in young people with chronic conditions: a systematic review and meta-analysis. J Am Med Inform Assoc 2016 Jan;23(1):230-239 [FREE Full text] [doi: 10.1093/jamia/ocv100] [Medline: 26186934]
46. Froome HM, Townson C, Rhodes S, Franco-Arellano B, LeSage A, Savaglio R, et al. The effectiveness of the Foodbot Factory mobile serious game on increasing nutrition knowledge in children. Nutrients 2020 Nov 06;12(11):1. [doi: 10.3390/nu12113413] [Medline: 33172094]
47. Haruna H, Hu X, Chu SKW, Mellecker RR. Initial validation of the MAKE framework: a comprehensive instrument for evaluating the efficacy of game-based learning and gamification in adolescent sexual health literacy. Ann Glob Health 2019 Feb 28;85(1):A [FREE Full text] [doi: 10.5334/aogh.1110] [Medline: 30873788]
48. Starks K. Cognitive behavioral game design: a unified model for designing serious games. Front Psychol 2014;5:28 [FREE Full text] [doi: 10.3389/fpsyg.2014.00028] [Medline: 24550858]
49. Sailer M, Hense JU, Mayr S, Mandl H. How gamification motivates: An experimental study of the effects of specific game design elements on psychological need satisfaction. Computers in Human Behavior 2017 Apr;69:371-380. [doi: 10.1016/j.chb.2016.12.033]
50. Gee JP. Pleasure, learning, video games, and life: the projective stance. E-Learning and Digital Media 2005 Sep 01;2(3):211-223. [doi: 10.2304/elea.2005.2.3.2]
51. Prensky M. Digital game-based learning. Comput. Entertain 2003 Oct;1(1):21-21. [doi: 10.1145/950566.950596]
52. Shute VJ, Ventura M, Bauer M, Zapata-Rivera D. Melding the power of serious games and embedded assessment to monitor and foster learning: Flow and grow. In: Ritterfeld U, Cody M, Vorderer P, editors. Serious games: Mechanisms and effects. New York, NY: Routledge; 2009:317-343.
53. Budd J, Miller BS, Manning EM, Lampos V, Zhuang M, Edelstein M, et al. Digital technologies in the public-health response to COVID-19. Nat Med 2020 Aug;26(8):1183-1192. [doi: 10.1038/s41591-020-1011-4] [Medline: 32770165]
54. Ravyse WS, Seugnet Blignaut A, Leendertz V, Woolner A. Success factors for serious games to enhance learning: a systematic review. Virtual Reality 2016 Sep 20;21(1):31-58. [doi: 10.1007/s10055-016-0298-4]
55. Chow CY, Riantiningtyas RR, Kanstrup MB, Papavasileiou M, Liem GD, Olsen A. Can games change children's eating behaviour? A review of gamification and serious games. Food Quality and Preference 2020 Mar;80:103823. [doi: 10.1016/j.foodqual.2019.103823]
56. Arnab S, Lim T, Carvalho M, Bellotti F, de Freitas S, Louchart S, et al. Mapping learning and game mechanics for serious games analysis. Brit J Educational Tech 2014 Jan 05;46(2):391-411. [doi: 10.1111/bjet.12113]
57. Cugelman B. Gamification: what it is and why it matters to digital health behavior change developers. JMIR Serious Games 2013 Dec 12;1(1):e3 [FREE Full text] [doi: 10.2196/games.3139] [Medline: 25658754]
58. DeSmet A, Thompson D, Baranowski T, Palmeira A, Verloigne M, De Bourdeaudhuij I. Is Participatory Design Associated with the Effectiveness of Serious Digital Games for Healthy Lifestyle Promotion? A Meta-Analysis. J Med Internet Res 2016 Apr 29;18(4):e94 [FREE Full text] [doi: 10.2196/jmir.4444] [Medline: 27129447]
59. MIApp. App Store. URL: https://apps.apple.com/us/app/miapp/id1623282445 [accessed 2024-12-07]
60. The Amanda Young Foundation. URL: https://www.amandayoungfoundation.org.au/ [accessed 2024-12-07]

XSL•FO
RenderX

Abbreviations

AYF: Amanda Young Foundation
ECU: Edith Cowan University
HPE: Healthy and Physical Education
ICSEA: Index of Community Socio-Educational Advantage
IMD: invasive meningococcal disease
MIApp: Meningococcal Immunisation Awareness, Prevention and Protection app
WA: Western Australia

Edited by A Coristine; submitted 20.05.24; peer-reviewed by K Erenli, M McMillan; comments to author 29.07.24; revised version received 20.09.24; accepted 19.11.24; published 11.02.25.

Please cite as:
Bloomfield L, Boston J, Masek M, Andrew L, Barwood D, Devine A
Evaluating the Efficacy of a Serious Game to Deliver Health Education About Invasive Meningococcal Disease: Clustered Randomized Controlled Equivalence Trial
JMIR Serious Games 2025;13:e60755
URL: https://games.jmir.org/2025/1/e60755
doi:10.2196/60755
PMID:39932769

©Lauren Bloomfield, Julie Boston, Martin Masek, Lesley Andrew, Donna Barwood, Amanda Devine. Originally published in JMIR Serious Games (https://games.jmir.org), 11.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Exploring Nursing Students' Experiences of Empathy and User Experiences in an Immersive Virtual Reality Simulation Game: Cross-Sectional Study

Jaana-Maija Koivisto[1,2*], PhD; Sanna Kämäräinen[2*], MNSc; Katri Mattsson[3], MNSc; Satu Jumisko-Pyykkö[4], PhD; Riikka Ikonen[2], PhD; Elina Haavisto[2,5], PhD

[1]Department of Public Health, Faculty of Medicine, University of Helsinki, Helsinki, Finland

[2]Department of Health Sciences/Nursing Science, Faculty of Social Sciences, Tampere University, Tampere, Finland

[3]Department of Health and Well-Being, Turku University of Applied Sciences, Turku, Finland

[4]Department of HAMK Tech, Häme University of Applied Sciences, Hämeenlinna, Finland

[5]Tampere University Hospital, Tampere, Finland

[*]these authors contributed equally

Corresponding Author:
Jaana-Maija Koivisto, PhD
Department of Public Health
Faculty of Medicine
University of Helsinki
PO Box 20
Helsinki, 00014
Finland
Phone: 358 503207111
Email: jaana-maija.koivisto@helsinki.fi

Abstract

Background: Empathy is associated with better clinical outcomes and patient-care experiences, and it has been demonstrated that training can improve nursing students' empathy. The use of virtual reality (VR) as an experiential learning strategy may increase the empathetic behavior of caregivers. Although much research exists on the use of VR in education, there is still little research on learning empathy in nursing education through immersive VR games that include a head-mounted display and hand controllers. In addition, it is important to study both learning and user experiences in nursing education that utilizes VR technology.

Objective: This study aims to explore nursing students' experiences of empathy and user experiences in an immersive VR simulation game.

Methods: A cross-sectional design was used. A total of 52 graduating nursing students from 3 universities of applied sciences in Finland participated in the study. The immersive VR simulation game employed in the study was played with a head-mounted display and hand controllers. The instruments used were the Basic Empathy Scale in Adults (BES-A) before the VR simulation gaming session and the Comprehensive State Empathy Scale (CSES) and AttrakDiff 2.0 Scale after the session.

Results: The students' overall level of empathy experienced in the immersive VR simulation game was favorable (CSES; mean 2.9, SD 0.57). Participants who had a higher level of empathy (BES-A) before playing the immersive VR simulation game also experienced slightly more feelings of empathy after playing (CSES). However, the association between the measures was not statistically significant (r=0.187, *P*=.18). The overall empathy (CSES) experienced in the immersive VR simulation game was positively correlated with its subscales. The use of the VR simulation provided a positive user experience in all 4 factors of the AttrakDiff 2.0 Scale. Overall User Experience and Emotion Sharing correlated negatively (r=−0.248, *P*=.042), as did Attractiveness and Emotion Sharing (r=−0.327, *P*=.018). Hedonic Quality Stimulation correlated negatively with Cognitive Empathy (r=–0.279, *P*=.045).

Conclusions: The results of this study indicate that the use of an immersive VR simulation game in nursing education as a means of increasing empathy seems promising and justified. The immersive VR simulation game offered positive user experiences, which further supported the idea of implementing it in education. However, more research is needed on what kinds of VR

XSL•FO
RenderX

environments are the most effective in promoting empathy among nursing students. Furthermore, when using VR technology in learning, one should consider that the VR setting must not be too technical but rather simple, straightforward, and predictable.

(JMIR Serious Games 2025;13:e62688) doi:10.2196/62688

KEYWORDS

education; nursing; learning; empathy; virtual reality; simulation; user experience; cross sectional

Introduction

Empathy in Nursing

Empathy is a key factor in quality of care and patient-centered care [1]. Greater empathy is associated with better clinical outcomes and patient-care experiences [2]. Empathy is an innate characteristic of the individual, and it is dynamic, constantly evolving, and built under the influence of various personal and environmental factors [3]. Empathy can be considered to include 3 dimensions: affective, cognitive, and behavioral [4,5]. The affective dimension consists of caring and a sincere and unconditional acceptance of each health care client, whereas the cognitive dimension is related to interpersonal sensitivity, intuition, and the ability to understand the point of view of another person [4,5]. The behavioral dimension is associated with altruism and therapeutic relationships, which develop empathy in practice; it can be understood as an intention to respond compassionately to the needs and concerns of another person [4,5].

Health care professionals with a high capacity for empathy work more effectively in favor of promoting the patient's condition [1,4,6]. The empathetic attitude of health care professionals toward health care users strengthens collaboration between them and increases patient satisfaction and commitment to treatment [1,4]. Furthermore, empathy enables appropriate support for patients, increases patient satisfaction, improves patient outcomes, reduces errors, and leads to better overall care [1,4]. An empathetic professional can understand the needs of health care users because it is easier for patients to bring their problems and thoughts to them [6].

Health care workers have felt that work experience and a strong professional identity strengthen empathy, even during busy and stressful times [3,4,7]. Empathy is also linked to job satisfaction among health care professionals and a reduced risk of burnout, but also to compassion fatigue and depression [4,5]. It has been observed that with a decrease in anxiety and other negative emotions, empathy toward the patient increases [7]. In addition, positive interaction in the work community and with patients and relatives increases empathy momentarily [7].

Characteristics of empathy with high or medium stability, such as a strong capacity for empathy developed in childhood, a work environment that supports empathy, and a strong professional identity, have been found to have protective effects against factors that lower empathy [3]. Factors that negatively affect empathy are increased technology use [8]; feelings of anxiety, fear, or uncertainty; a large number of patients; lack of time; negative interaction; and a lack of training in empathy [4,5,7]. While the importance of empathy for nursing is undeniable, many health care professionals face challenges in adopting an empathic communication model for their actions [5].

A variety of personal factors (such as inherent traits, physiological and mental conditions, and professional identity) and external factors (including the work environment, life experiences, and situational stressors) influence the development of empathy [3]. The value and importance of empathy should be emphasized more in health care education [3,6], and empathy should be developed during nursing education [9]. In the health care sector, a significant decrease in empathy has been observed, especially in education [10]. The decline in empathy has been attributed to many factors, such as changed curricular requirements and tightened time constraints that have led to the prioritization of technical and clinical knowledge over humanistic values such as empathy [11].

Several factors affect nursing students' empathy, including engaging, efficiency, unpredictable, and burdensome elements. These factors are a combination of personal, patient, and environmental influences [12]. Strategies that may foster empathy in nursing students are enhancing self-esteem, boosting self-efficacy, and improving interpersonal relationships [13]. Research has shown that, without specific training, personal traits and factors related to social and family environments significantly influence the development of empathy in nursing students [14].

Training improves individuals' empathy [2], and empathy training seems to have an impact on the empathy skills of health care students [4,9,15]. Conversely, a review revealed that educational programs did not increase health care students' empathic concerns [16]. However, nursing education seems to lack systematic training in empathy, although its importance has been identified [17].

Virtual Reality in Empathy Education

Experiential learning has shown good results in increasing empathy [17] because it focuses on learning through lived or shared experience, and it has been utilized using simulations, role-playing games, and virtual reality (VR) in teaching [5,10,17,18]. A recent systematic review revealed that the application of different modalities of simulation promotes empathy in nursing education [19], while another systematic review demonstrated that for improving empathy, the most effective interventions were immersive and experiential simulations providing opportunities for reflection [20]. Simulation experiences can reduce the decline in experiencing empathy during training because it gives the student a sense of control and helps identify professional behavior in patient relationships [21]. In VR simulations, students have been more in the role of a patient [5,10,22,23] than a nurse [18]. Research has shown that the most effective simulation exercises for

XSL•FO
RenderX

learning empathy are those in which the learner takes on the role of a patient [21]. However, Levett-Jones et al [5] found that campus point-of-view simulations, where students experience the world "through the eyes" of a patient with hemiparesis, positively impacted nursing students' empathy toward persons with disabilities. Furthermore, they found a greater increase in the empathy levels of participants who played the role of a caregiver in a VR simulation compared with participants in the role of a patient.

The use of VR as an experiential learning strategy may increase the empathetic behavior of caregivers [1]. VR pertains to a computer-generated 3D environment that replicates the facets of the physical world [24]. Within health care education, prevalent VR technologies encompass computer-based simulations, haptic simulators, and head-mounted display (HMD) systems, with HMD being the least frequently used [25]. In HMD systems, the sensations of immersion, presence, and interaction [24] are the most pronounced. The level of immersion is determined by the extent to which the VR system supports the user's perception and use of their body in VR. High-level immersion commonly refers to a 3D experience with a virtual interface, such as HMD, that provides the user with a wide field of view, high-resolution image, sound, and motion detection [1,26]. The concept of presence in VR refers to the user's experience of a sense of being in a virtual environment rather than in their real physical environment [26]. The concept of interactivity refers to the interaction between the user and the VR environment [25].

The integration of immersive technologies into education offers novel avenues for nursing education, with their adoption having gained momentum during and after the COVID-19 pandemic. Virtual simulations are useful because they enable the combination of theoretical clinical knowledge and practice through realistic patient situations in online learning [27,28]. Students' experiences of using VR in learning have been mostly positive, and the teaching method has been considered motivating [10,29]. Havola et al [30] found that students spending more time in a VR simulation with an HMD and hand controllers achieved better learning outcomes in patient scenarios than students spending less time in VR. VR has mainly been perceived as an easy-to-use and functional teaching tool [22]. However, nursing students have experienced some technical difficulties in VR simulations, and therefore prior technical practice is required before entering a VR simulation session [18,31].

In VR simulations, factors affecting empathy include a sense of presence and the illusion of being in the body of another person [15]. A qualitative study describing undergraduate nursing students' empathy in an immersive VR simulation game found that nursing students experienced empathic concern toward a virtual patient, and they recognized the virtual patient's emotions and responded to those [18]. A systematic review by Bas-Sarmiento et al [17] found variations in empathy levels regarding the different characteristics of health care students, such as gender, age, or cultural background. They stated that younger participants' empathy levels were higher compared with older participants, and women's empathy levels were higher than men's [17].

In a good user experience (UX), a VR simulation game supports learning while keeping the technology in the background. UX is defined as "a consequence of the user's internal state […], the characteristics of a designed system […], and the context […] within which interaction occurs" [32]. In effective interactions with technology, users feel that they achieve their goals and that their needs regarding technology in a certain situation are met [33]. UX has 4 main qualities: perceived pragmatic quality (PQ; eg, simple, practical, clearly structured), hedonic quality identification (HQI; eg, stylish, connective, presentable), hedonic quality stimulation (HQS; eg, innovative, inventive, creative), and attractiveness (ATT; eg, pleasant, good, motivating) [33,34]. UX determines the overall judgment of a product, the choices made, and user behavior [35]. In designing a good UX, participatory and human-centered design methods are used, and the evaluation of experience is an essential part of this [36,37]. Recently, Law and Heintz [38] have underlined the need to study both learning and UX in education that utilizes technology.

Although there has been much research on the use of VR in education, there is still little research on learning empathy in nursing education through immersive VR games that include HMDs and hand controllers. The objective of this study was to explore nursing students' experiences of empathy and UX in a VR simulation. The overall aim was to contribute to the increased discussion related to the educational use of VR by creating new information about learning empathy in nursing education by interacting with virtual patients in a VR simulation. This study is part of a larger research project that aims to develop methods that utilize immersive technology in nursing education and to investigate their impact on students' competencies. The following research questions were addressed:

- What is the extent to which nursing students experience feelings of empathy in a patient scenario in an immersive VR simulation game?
- What factors are associated with the experience of empathy during an immersive VR simulation game?
- How is nursing students' UX in an immersive VR simulation game?
- How is UX associated with experiencing empathy during an immersive VR simulation game?

Methods

Research Design and Participant Recruitment

A cross-sectional design was used. Purposive sampling [39] was used to recruit undergraduate nursing students in their final academic year from 3 universities of applied sciences (UASs) on the southern, western, and southwestern coast of Finland. Each UAS had a contact teacher who was informed about the study and who helped the research group with organizing the data collection. The researchers did not know the students beforehand and were not involved in their teaching or evaluation. The inclusion criteria were as follows: (1) graduating nursing students, (2) participation in a VR simulation, and (3) voluntary participation in the study. Exclusion criteria were medical conditions (such as migraine) that prevented participation in the VR simulation.

XSL•FO
RenderX

Immersive Virtual Reality Simulation Game

The immersive VR simulation game used in this study utilized an HMD and VR software featuring audio-visual enhancements, encompassing graphics, animations, and haptic feedback. In addition, game elements, such as points and feedback systems, were used to promote learning in a nursing context [40,41]. Developed using the Unity development platform (Unity Technologies), the game was developed for compatibility with Oculus Quest devices. This configuration enables players to engage with the virtual environment using hand controllers, maneuvering within a simulated hospital setting. The features of the VR simulation game are described in Textbox 1.

Textbox 1. Features of the virtual reality simulation game.

1. The patient scenario
   - A 59-year-old man with no previous illnesses is transferred from the emergency department to the department of inner medicine.
   - The patient has a deteriorating condition involving pneumonia.
2. Learning goals
   - Assess the patient's clinical state using the Airway, Breathing, Circulation, Disability, Exposure (ABCDE) approach.
   - Recognize patients care needs.
   - Implement nursing interventions.
3. The technology
   - 3D patient and hospital environments developed with the Unity game engine.
   - Single-player game.
   - Oculus Quest head-mounted display and hand controllers.
4. Interaction
   - The user views with a head-mounted display 360-degree 3D environment.
   - The user moves around the virtual hospital room by natural walking or teleporting. The user interacts with the patient by choosing options with hand controllers from the options menu.
   - The patient responds with multisensory feedback (audio, visual, and physical).

Presented from a first-person view, participants assume the role of a nurse tasked with attending to a virtual patient with pneumonia. Equipped with an Oculus Quest HMD and hand controllers, users conduct clinical assessments and alleviate the patient's symptoms (Figure 1). Navigation within the virtual hospital room is facilitated through a 360-degree viewing capability, permitting users to either walk naturally or utilize teleportation, a technique employing handheld controllers for movement. The immersive VR simulation game emerged from a collaborative effort across multiple professions within a Finnish UAS. This VR simulation game has been utilized in various studies, including those by Havola et al [30], Mattsson et al [18], and Mäkinen et al [42], demonstrating its efficacy as a learning tool.

Figure 1. Screenshot of the immersive virtual reality simulation game where a student is placing a pulse oximeter on the virtual patient.



XSL•FO
RenderX

Data Collection

Overview

Data were collected by 2 researchers (SK and KM) between May 2021 and February 2022. The individual VR simulation sessions were facilitated by 2 researchers (SK and KM). Students received brief information about the immersive VR simulation game storyline, how to wear the HMD, and how to move and grab things in the virtual world using hand controllers. The researcher verbally instructed students on technical issues if needed when they were immersed in the virtual world. The sessions lasted about 30-45 minutes. The researchers ensured that the gameplay always remained safe by monitoring the situation on-site. They were prepared to end the gameplay if they observed any student experiencing significant dizziness or nausea. However, participants did not report, nor did the researchers observe, any physical adverse effects during the study.

Nursing students filled out an electronic questionnaire before (PRE-Q) and after (POST-Q) the VR simulation session. The PRE-Q included demographic questions (age, gender, educational background, work experience in social and health services, previous gaming activity, and previous experience with VR) and the Basic Empathy Scale in Adults (BES-A) [43]. The POST-Q included the Comprehensive State Empathy Scale (CSES) [5] and the AttrakDiff 2.0 Scale [34]. The instruments used in the study were internationally validated, and permission was obtained for their use in the study. To ensure cross-cultural validation of the instruments, the validation process followed the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) guidelines [44].

Basic Empathy Scale in Adults

The BES-A consisted of 20 items [43] assessing nursing students' empathy with a 5-point Likert Scale from 1 (completely disagree) to 5 (completely agree). Higher scores indicated higher levels of empathy.

Comprehensive State Empathy Scale

The CSES [5] measured nursing students' experience of empathy in the immersive VR simulation game. Students were asked to rate the extent to which they experienced the emotions listed in the CSES in their VR patient scenario. The CSES consists of 6 subscales (Empathic Concern, Distress, Shared Affect, Empathic Imagination, Helping Motivation, and Cognitive Empathy) and 30 items. Each CSES item was scored using a 5-point Likert Scale from 1 (completely disagree) to 5 (completely agree), with higher scores reflecting higher empathy levels.

AttrakDiff 2.0 Scale

The AttrakDiff 2.0 Scale measured the UX of the immersive VR simulation game with 4 subscales, including PQ, HQI, HQS, and ATT, and with 28 items evaluated on a 7-point semantic differential scale [34]. AttrakDiff 2.0 is a widely used and validated method for studying UX with interactive products [45-47].

Statistical Analysis

Data analyses were performed using SPSS Statistics 27.0 (IBM Corp.). In evaluating empathy, descriptive statistics and frequency tables were used to characterize variables. Subscales were formed based on a previously determined instrument structure [5,43]. No severe violations against the prerequisites of parametric tests were observed in skewness and kurtosis evaluations; therefore, parametric tests were chosen: an independent-sample *t* test and 1-way analysis of variance (or Welch test) were used to compare the mean scores of scales and subscales. The correlations between scales and subscales were examined using the Pearson correlation coefficient. The reliability of scales and subscales was evaluated with Cronbach α coefficients.

The AttrakDiff 2.0 questionnaire was analyzed with descriptive statistics. Four factors summarize all their items: (1) PQ (eg, simple/complicated, practical/impractical), (2) HQI (eg, stylish/tacky, isolating/connective), (3) HQS (eg, inventive/conventional, repelling/appealing), and (4) ATT (eg, ugly/attractive, repulsive/inviting) [34]. The internal consistency was acceptable for all factors: PQ (Cronbach α=0.76), HQI (Cronbach α=0.63), HQS (Cronbach α=0.74), and ATT (Cronbach α=0.80). The correlation between UX and CSES evaluations was analyzed using the Pearson correlation coefficient.

Ethical Considerations

At all stages of the study, the responsible conduct of research guidelines by the Finnish National Board on Research Integrity [48] were followed. Ethical approval was received from the Human Sciences Ethics Committee of Helsinki Region UAS (24.3.2021). In addition, permissions were obtained from the higher-education organizations from which the data were collected. The study participants were informed about the study both orally and in writing. They were told that although playing the game was integrated into their studies, participation in the study was voluntary and that participation did not affect the participants' grades or academic evaluation. Participants were also informed about the mild physical symptoms such as dizziness, eye fatigue, and nausea that wearing VR headsets may cause. After receiving this information, all participants gave their written consent to participate in the study.

Results

User Statistics

A total of 52 graduating nursing students participated in the study (Table 1). Most of the students were aged 21-25 years (27/52, 51.9%). The participants mainly had 1-5 years (22/52, 42.3%) or less than a year (22/52, 42.3%) of work experience in the social and health sector. Most of the participants (43/52, 82.7%) had no previous VR experience in the past year, and no one had used an Oculus Quest headset and hand controllers previously.

XSL•FO
RenderX

Table 1. Background variables (n=52).

Characteristics	Values, n (%)
Gender	
Female	42 (81)
Male	9 (17)
I do not want to tell	1 (2)
Age	
21-25	27 (52)
26-30	15 (29)
31-50	10 (19)
Educational background	
High-school degree	27 (52)
Practical nursing	11 (21)
Vocational education (other than practical nursing)	3 (6)
Bachelor's degree (other than nursing)	11 (21)
Work experience in social and health services	
Not at all to 1 year	22 (42)
1-5 years	22 (42)
Over 5 years	8 (15)
Previous virtual reality experience in the past year	
Less than once a month	9 (17)
Not at all	43 (83)
Experience with the Oculus Quest headset and hand controllers	
Not at all	52 (100)

Nursing Students' Experiences of Empathy in the Immersive Virtual Reality Simulation Game

The students' overall level of empathy experienced in the immersive VR simulation game was favorable (CSES; mean 2.9, SD 0.57). The highest level of empathy experienced by nursing students measured by CSES after the VR simulation session was in the subscale Helping Motivation (mean 4.0, SD 0.80), while the lowest were in the subscales Distress (mean 2.1, SD 0.65) and Shared Affect (mean 2.4, SD 0.75; Table 2). At the item level, the highest score was in the item "I found myself thinking about what could be done to help the patient" (mean 4.4, SD 0.85), indicating a high level of empathy. The lowest level of empathy was experienced in the item Upset (mean 1.4, SD 0.64), which, in turn, indicated that the students experienced this emotion in the immersive VR simulation game to a lesser extent.

XSL•FO
RenderX

Table 2. Nursing students' experience of empathy in a virtual reality simulation game (n=52).

Comprehensive State Empathy Scale	Mean (SD)[a]	Cronbach α
Subscale 1: Empathic Concern	3.2 (0.74)	0.86
Compassionate	3.5 (1.16)	
Moved	2.1 (0.88)	
Soft-hearted	3.4 (0.93)	
Sympathetic	3.5 (0.85)	
Tender	3.6 (0.91)	
Warm	3.3 (1.07)	
Subscale 2: Distress	2.1 (0.65)	0.75
Distressed	2.5 (1.24)	
Disturbed	1.8 (0.88)	
Grieved	1.9 (1.03)	
Troubled	3.4 (1.05)	
Upset	1.4 (0.64)	
Afraid	1.7 (0.85)	
Subscale 3: Shared Affect	2.4 (0.75)	0.68
I found that the scenario affected my mood	3.1 (1.17)	
I was very affected by the emotions in this story	1.9 (0.93)	
I actually felt the patient's distress	2.6 (1.13)	
I experienced the patient's feelings as if they were my own	1.9 (0.98)	
Subscale 4: Empathic Imagination	2.6 (1.06)	0.87
I found myself imagining how I would feel in the patient's situation	2.5 (1.31)	
I found myself imagining myself in the patient's shoes	2.2 (1.21)	
I found myself trying to imagine how things looked to the patient	2.8 (1.28)	
I found myself trying to imagine what the patient was experiencing	3.0 (1.24)	
Subscale 5: Helping Motivation	4.0 (0.80)	0.80
I would really focus on the patient's emotions if I were caring for him	3.5 (1.29)	
I experienced a strong urge to help the patient	3.9 (1.01)	
I would get really involved in trying to help the patient	4.2 (0.82)	
I found myself thinking about what could be done to help the patient	4.4 (0.85)	
Subscale 6: Cognitive Empathy	2.9 (0.81)	0.86
I feel confident that I could accurately describe the patient's experience from his point of view	3.1 (1.01)	
I found it easy to understand the patient's reactions	3.4 (0.97)	
I found it easy to see how the situation looked from the patient's point of view	3.0 (1.09)	
Even though the patient's life experiences are different from mine, I can really see things from their perspective	3.0 (1.07)	
I am sure that I know how the patient was feeling	2.5 (1.11)	
I feel confident that I could accurately describe how the patient felt	2.7 (1.08)	
Overall Comprehensive State Empathy Scale score	2.9 (0.57)	0.92

[a]A 5-point Likert Scale with responses ranging from 1 (completely disagree) to 5 (completely agree), with higher scores reflecting higher empathy levels.

XSL•FO
RenderX

Factors Associated With the Experience of Empathy During the Immersive VR Simulation Game

Participants who had a higher level of empathy (BES-A) before playing the immersive VR simulation game also experienced slightly more feelings of empathy after playing (CSES). However, the association between the measures was not statistically significant (r=0.187, P=.18). The overall empathy experienced in the immersive VR simulation game was positively correlated with its subscales. The strongest positive correlation in the subscales of empathy was between Empathic Concern and Empathic Imagination (r=0.666, P<.001; Table 3).

Table 3. The correlations between CSES[a] subscales and BES-A[b].

Variables	Empathic Concern		Distress		Shared Affect		Empathic Imagination		Helping Motivation		Cognitive Empathy		BES-A	
	r	*P* value	*r*	*P* value	*r*	*P* value	*R*	*P* value	*r*	*P* value	*r*	*P* value	*r*	*P* value
CSES	0.807	<.001	0.466	<.001	0.747	<.001	0.834	<.001	0.623	<.001	0.778	<.001	0.187	.18
Empathic Concern	—[c]	—	0.164	.24	0.479	<.001	0.666	<.001	0.517	<.001	0.552	<.001	0.270	.05
Distress	—	—	—	—	0.490	<.001	0.227	.11	0.049	.73	0.145	.31	0.126	.37
Shared Affect	—	—	—	—	—	—	0.571	<.001	0.367	.007	0.418	.002	0.146	.30
Empathic Imagination	—	—	—	—	—	—	—	—	0.395	.004	0.631	<.001	0.090	.52
Helping Motivation	—	—	—	—	—	—	—	—	—	—	0.434	.001	0.240	.09
Cognitive Empathy	—	—	—	—	—	—	—	—	—	—	—	—	−0.020	.89

[a]CSES: Comprehensive State Empathy Scale.

[b]BES-A: Basic Empathy Scale in Adults.

[c]Not applicable.

The relationship between the background variables and overall empathy experienced in the immersive VR simulation game is presented in Table 4. Background variables did not have a statistically significant association with the overall empathy experienced in the VR game (Table 4). Examined by subscales, age was associated with cognitive empathy ($F_{2,49}$=3.926, P=.03) and distress ($F_{2,49}$=2.833, P=.07) experienced in the game as follows: The experience of cognitive empathy was greater in the 31-50-year age group (mean 3.6, SD 0.78) than in the 21-25- (mean 2.8, SD 0.79) and 26-30-year (mean 2.9, SD 0.73) age groups. Those aged 21-25 years experienced more distress (mean 2.3, SD 0.60) than respondents aged 26-30 (mean 1.9, SD 0.68) or 31-50 (mean 1.9, SD 0.60) years.

When examining the subscales of empathy, the educational background had a weak association with distress experienced in the immersive VR simulation game ($F_{3,48}$=2.593, P=.06). Students with higher-degree education experienced more feelings of distress (mean 2.3, SD 0.59) than other participants. Work experience seemed to be related to shared affect during the game ($F_{2,25}$=3.307, P=.05). Those who had worked for more than 5 years experienced less (mean 2.0, SD 0.43) shared affect in the game than other participants (Table 5).

XSL•FO
RenderX

Table 4. The relationship between the background variables and the experience of empathy in the immersive virtual reality simulation game.

Variables	n	Mean (SD)	*t* test/*F* test[a,b] (*df*)	*P* value
Gender			0.137 (49)[c]	.89[c]
Male	9	2.9 (0.61)		
Female	42	2.9 (0.57)		
Age (years)				.59[a]
21-25	27	2.8 (0.62)		
26-30	15	2.8 (0.51)		
31-50	10	3.0 (0.53)		
Educational background			0.441 (3, 48)	.73[a]
High-school degree	27	2.8 (0.61)		
Practical nursing	11	2.9 (0.40)		
Vocational education (other than practical nursing)	3	2.5 (0.35)		
Bachelor's degree (other than practical nursing)	11	2.9 (0.67)		
Work experience in social and health services			1.274 (2, 18)	.30[b]
Not at all to 1 year	22	2.7 (0.72)		
1-5 years	22	3.0 (0.35)		
Over 5 years	8	2.8 (0.56)		

[a]One-way analysis of variance.

[b]Welch *t* test.

[c]*t* test.

XSL•FO
RenderX

Table 5. The relationship between background variables and empathy subscales in a virtual reality simulation game (n=52).

Variables	Empathic Concern			Distress			Shared Affect			Empathic Imagination			Helping Motivation			Cognitive Empathy		
	Mean (SD)	*P* value	*F*/*t* test (*df*)	Mean (SD)	*P* value	*F*/*t* test (*df*)	Mean (SD)	*P* value	*F*/*t* test (*df*)	Mean (SD)	*P* value	*F*/*t* test (*df*)	Mean (SD)	*P* value	*F*/*t* test (*df*)	Mean (SD)	*P* value	*F*/*t* test (*df*)
Gender		.91[a]	0.110 (49)		.71[a]	0.373 (49)		.84[a]	0.207 (9)		.86[a]	0.176 (49)		.93[a]	0.093 (49)		.78[a]	0.278 (49)
Male	3.2 (0.55)			2.0 (0.75)			2.4 (1.14)			2.7 (1.25)			4.0 (0.72)			3.0 (1.00)		
Female	3.2 (0.79)			2.1 (0.61)			2.4 (0.67)			2.6 (1.04)			4.0 (0.83)			2.9 (0.78)		
Age (years)		.55[b]	0.612 (2, 49)		.07[b]	2.833 (2, 49)		.98[b]	0.017 (2, 49)		.41[b]	0.896 (2, 49)		.71[b]	0.345 (2, 49)		.03[b]	3.926 (2, 49)
21-25	3.1 (0.77)			2.3 (0.60)			2.4 (0.72)			2.5 (1.04)			4.0 (0.83)			2.8 (0.79)		
26-30	3.4 (0.74)			1.9 (0.68)			2.4 (0.65)			2.5 (1.00)			4.0 (0.63)			2.9 (0.73)		
31-50	3.3 (0.69)			1.9 (0.60)			2.4 (1.04)			3.0 (1.23)			4.2 (0.97)			3.6 (0.78)		
Educational background		.64[b]	0.567 (3, 48)		.06[b]	2.593 (3, 48)		.69[b]	0.492 (3, 48)		.36[b]	1.090 (3,48)		.77[b]	0.379 (3, 48)		.40[b]	1.014 (3, 48)
High-school degree	3.1 (0.87)			2.3 (0.59)			2.5 (0.65)			2.6 (1.01)			3.9 (0.84)			2.8 (0.83)		
Practical nursing	3.3 (0.49)			1.8 (0.68)			2.4 (0.85)			2.8 (0.92)			4.2 (0.89)			3.3 (0.52)		
Vocational education (other than practical nursing)	3.2 (0.50)			1.7 (0.33)			1.9 (0.63)			1.7 (0.95)			3.8 (0.14)			2.9 (0.38)		
Bachelor's degree other than practical nursing)	3.4 (0.68)			2.0 (0.67)			2.3 (0.96)			2.8 (1.30)			4.1 (0.73)			3.0 (1.04)		
Work experience in social and health services		.23[c]	1.557 (2, 27)		.13[b]	2.240 (2, 19)		.05[b]	3.307 (2, 25)		.31[b]	1.258 (2, 20)		.14[b]	2.235 (2, 17)		.43[b]	0.886 (2, 18)
0-1 year	3.0 (0.95)			2.2 (0.52)			2.4 (0.93)			2.4 (1.18)			3.8 (0.85)			2.8 (0.99)		
1-5 years	3.4 (0.56)			2.2 (0.70)			2.5 (0.62)			2.9 (0.93)			4.3 (0.58)			2.9 (0.56)		
Over 5 years	3.4 (0.34)			1.6 (0.66)			2.0 (0.43)			2.4 (1.05)			4.0 (1.06)			3.3 (0.85)		

[a]*t* test.

[b]One-way analysis of variance.

XSL•FO
RenderX

[c]Welch.

User Experience in the Immersive VR Simulation Game

The use of the VR simulation provided a positive UX in all 4 factors (Figures 2 and 3). The simulation was experienced as stimulating (mean 1.6, SD 0.72), and this was highlighted in items such as being novel, captivating, or innovative (mean 2.0-2.2, SD 0.87-1.11). However, the VR simulation was experienced as neither challenging nor undemanding (mean 0.1, SD 1.4). The simulation was experienced as being attractive (mean 1.6, SD 0.8) and, in more detail, as being good, motivating, and inviting (mean 1.9-2.1, SD 1.1-1.2). Regarding hedonic quality, the participants were able to positively identify themselves with a situation of simulation (mean 1.1, SD 0.6). This was experienced positively as being presentable, integrating, and professional (mean 1.3-1.9, SD 1.1-1.2). The VR simulation was also mildly positive in its PQ (mean 0.5, SD 0.86). Although the VR simulation was experienced as practical and manageable (mean 1.3-1.9 SD 1.2-1.3), it was seen as slightly technical (mean 0.4, SD 1.6) and neutral in the scales of Complicated-Simple and Cumbersome-Straightforward (mean 0.1, SD 1.4).

Figure 2. Mean user experience of the virtual reality simulation. The error bars show 95% CI of the mean.

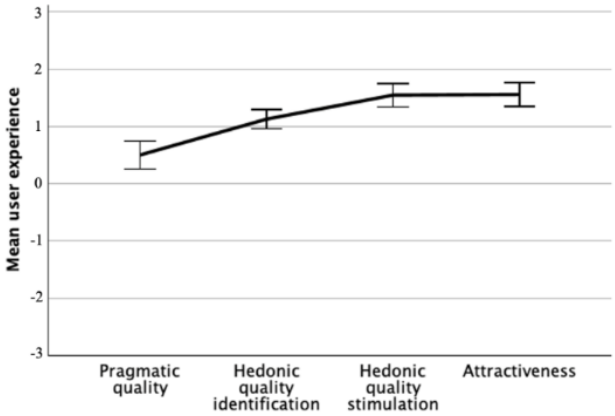
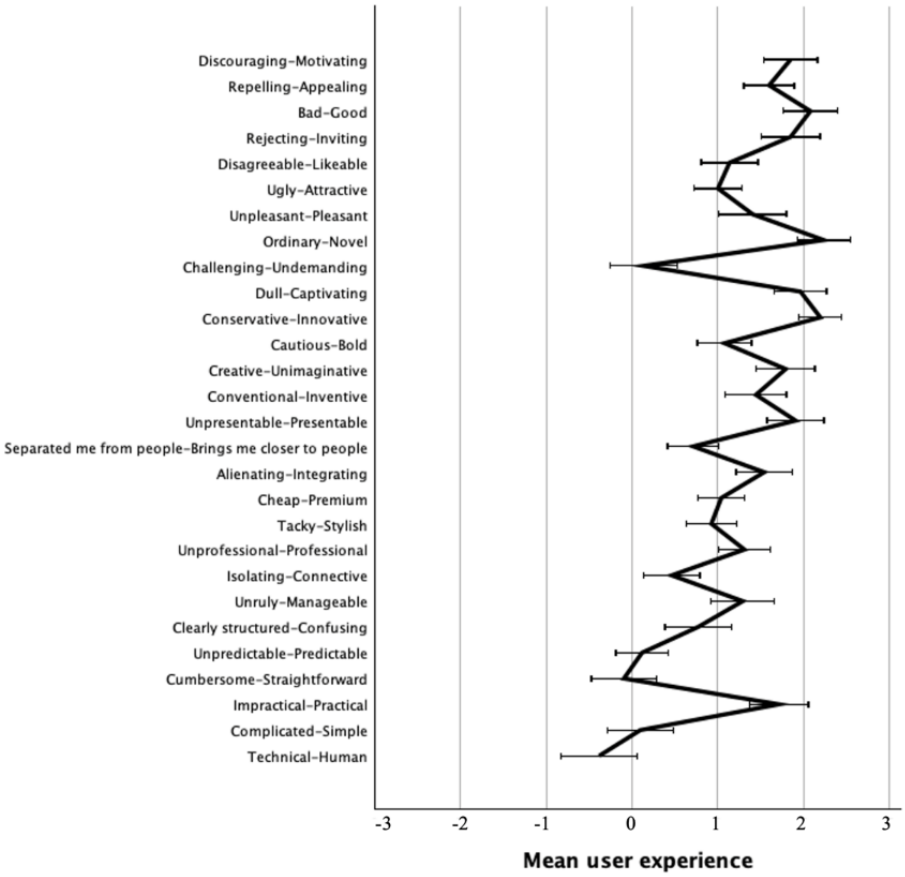


Figure 3. Mean user experience per item of the virtual reality simulation. The error bars show 95% CI of the mean.



XSL•FO
RenderX

Correlation Between Overall UX and CSES

The correlation between the overall UX and CSES ratings was not statistically significant (r=−0.166, P=.24, Table 6). In the subscale analysis, overall UX and Shared Affect correlated negatively (r=−0.248, P=.042), as did ATT and Shared Affect (r=−0.327, P=.018). HQS correlated negatively with Cognitive Empathy (r=–0.279, P=.045). All other correlations between the subscales were nonsignificant (Table 6).

Table 6. The correlations between CSES[a] and its subscales and AttrakDiff 2.0 and its subscales[b].

Variables	UX (AttrakDiff 2.0)		PQ		HQI		HQS		ATT	
	r	*P* value	*r*	*P* value	*r*	*P* value	*R*	*P* value	*r*	*P* value
CSES	–0.166	0.240	–0.090	0.524	–0.087	–0.206	0.834	0.143	–0.179	0.205
Empathic Concern	0.039	0.782	0.157	0.267	0.167	0.267	–0.135	0.342	0.011	0.936
Distress	–0.021	0.881	–0.161	0.254	0.098	0.492	0.106	0.454	–0.068	0.631
Shared Affect	–0.284	<0.05	–0.222	0.114	–0.195	0.166	–0.211	0.133	–0.327	<0.05
Empathic Imagination	–0.149	0.293	–0.057	0.689	–0.062	0.661	–0.208	0.139	0.178	0.207
Helping Motivation	–0.105	0.457	0.034	0.813	–0.189	0.181	–0.133	0.349	–0.107	0.448
Cognitive Empathy	–0.166	0.240	–0.166	0.240	–0.159	0.260	–0.279	<0.05	–0.151	–0.179

[a]CSES: Comprehensive State Empathy Scale.

[b]UX measured with AttrakDiff 2.0 User Experience Scale.

Discussion

Principal Findings

The purpose of this study was to investigate nursing students' experiences of empathy by playing an immersive VR simulation game. The overall aim was to contribute to increased discussion related to the educational use of VR by creating new information about learning empathy in nursing education by interacting with virtual patients in a game. The findings presented herein corroborate prior research, indicating a pronounced inclination among nursing students toward the utilization of VR simulation as an educational modality [18].

The main results of this study were that nursing students experienced some feelings of empathy when playing the immersive VR simulation game, which is in line with previous evidence [18,49]. Nursing students experienced the simulation as helping especially motivation and empathic concern. These results are consistent with those of Mattsson et al [18], who showed that nursing students experienced compassion and feelings of concern during a VR simulation game. In addition, the results further support the idea that acting in the role of a nurse might arouse willingness to help the patient and feelings of empathetic concern [5,18].

In this study, playing the immersive VR simulation game clearly evoked more positive than negative emotions in nursing students. Positive emotions experienced by the participants included tenderness, compassion, and sympathy. Negative emotions, such as being upset, afraid, or disturbed, were experienced clearly less, confirming previous study results [18]. This result is positive with regard to learning because feelings of anxiety, insecurity, and fear negatively affect empathy [3,4], whereas a strong professional identity, a sense of control at work, and putting oneself in another person's shoes increase empathy [3-5,17]. Feeling less distress could be explained by the suitability of the immersive VR simulation game for the stage of the participants' studies and their current competence level.

The participants felt a strong motivation to help during the immersive VR simulation game. It manifested itself in thinking about what could be done to help the patient, trying to get really involved in helping the patient, experiencing a strong urge to help the patient, and focusing on the patients' emotions when they were caring for them. These results seem to be consistent with previous research which has indicated that VR simulation increases empathy [1,10,15,29] and represents a usable method for teaching and learning empathy [18]. These results further support the idea that simulation training increases empathy during nursing education and helps identify professional behavior in patient relationships [21].

In our study, nursing students played the role of a nurse in the VR simulation, which differs from many other studies where students have played more the role of a patient [10,22,23]. Beforehand, there has been evidence that the most effective simulation exercises for learning empathy are those in which the learner takes on the role of a patient [21], but the results of this study support the results by Levett-Jones et al [5], who demonstrated an increase in empathy among participants who played the role of a caregiver.

Our results revealed that the participants felt that they understood the patient's perspective and emotions to some extent, and emotional sharing was fairly limited between the virtual patient and the participants. The cognitive and affective dimensions of empathy can be improved by interventions in which the student represents a specific role, such as the role of a nurse [17], and thus this immersive VR simulation game may not have been optimally successful in promoting the cognitive and affective dimensions of empathy.

Background variables had no statistically significant association with overall empathy (CSES). However, cognitive empathy was greater in the older age group than in the younger age group.

XSL•FO
RenderX

In addition, younger participants experienced more distress than older ones. These differences can be explained in part by the life experiences of older participants [3]. Work experience seemed to influence shared affect, such as how the scenario impacted students' mood and their ability to empathize with the patient's feelings. Those who had worked for more than 5 years experienced less shared affect than other participants.

The results of the UX assessment showed that the immersive VR simulation game offered a positive UX. Overall, this indicated a good game quality, enabling users to achieve set goals with it, and supporting stimulation, identification, and being attractive. The results also showed that the UX within a simulation can correlate with some aspects of empathy, such as affect sharing and cognitive empathy. The experienced PQ of the VR simulation, and participants' limited prior experience with VR technology, might have had a small influence on the results. In previous studies conducted with participants with little or no prior experience with VR technology, technology has been reported to capture the participants' attention to some extent [18,31].

Limitations

One of the limitations of our study was the frequency of participation in VR simulations. We acknowledge that more frequent participation in VR simulations could potentially yield better and more lasting results. The data set was relatively small for a quantitative study, and therefore, these results need to be interpreted with caution. Usually, cross-sectional studies are quite quick and inexpensive to conduct [50], but our data collection with the immersive VR simulation game during 2021-2022 required the researchers to make considerable practical arrangements due to restrictions set forth by the COVID-19 pandemic and the constant changes it caused in teaching arrangements. This brought about a lot of challenges for data collection. Therefore, the data can be considered reasonable, especially considering that the participants came from 3 different UASs and 3 different cities. However, the relatively small number of participants impacts the reliability of our statistical tests and, consequently, the robustness of our analyses. A larger sample size would provide more statistical power and potentially more reliable and generalizable results. However, as participation was voluntary and the VR simulation sessions were not part of normal teaching, students with an interest in VR were more likely to participate than those not interested in such technology. This could cause sampling bias [50], which could undermine the reliability and generalizability of the study. However, the strength of this study is that multiple outcomes were studied with previously validated instruments, improving the validity of our results. Regarding CSES, there are no previously published research findings on the use of the Finnish version of the instrument. Therefore, this study pilots its application in terms of international equivalence and validity. The values ranged from 0.68 to 0.92, indicating good internal consistency of the instrument.

Conclusions

The results of this study on the use of an immersive VR simulation game in nursing education as a means of increasing empathy seem promising, as the students mainly experienced positive emotions strengthening their own clinical competence in the VR simulation environment. Based on the results, the use of an immersive VR simulation game to practice empathy skills seems justified in nursing education. However, more research is needed on what kinds of VR environments are the most effective in promoting empathy among nursing students, and on whether playing a VR simulation game repeatedly improves the game's positive effects on learning. To improve the game in the future, more attention needs to be paid to PQ, such as making the game less technical and more simple, straightforward, and predictable.

Conflicts of Interest

None declared.

References

1. Brydon M, Kimber J, Sponagle M, MacLaine J, Avery J, Pyke L, et al. Virtual reality as a tool for eliciting empathetic behaviour in carers: an integrative review. Journal of Medical Imaging and Radiation Sciences 2021 Sep;52(3):466-477. [doi: 10.1016/j.jmir.2021.04.005] [Medline: 34023219]
2. Nembhard IM, David G, Ezzeddine I, Betts D, Radin J. A systematic review of research on empathy in health care. Health Serv Res 2023 Apr 15;58(2):250-263 [FREE Full text] [doi: 10.1111/1475-6773.14016] [Medline: 35765156]
3. Yu CC, Tan L, Tang B, Liaw SY, Tierney T, Ho YY, et al. The development of empathy in the healthcare setting: a qualitative approach. BMC Med Educ 2022 Apr 04;22(1):245-245 [FREE Full text] [doi: 10.1186/s12909-022-03312-y] [Medline: 35379249]
4. Moudatsou M, Stavropoulou A, Philalithis A, Koukouli S. The role of empathy in health and social care professionals. Healthcare (Basel, Switzerland) 2020;8(1):26-26. [doi: 10.3390/healthcare8010026] [Medline: 32019104]
5. Levett-Jones T, Lapkin S, Govind N, Pich J, Hoffman K, Jeong SY, et al. Measuring the impact of a 'point of view' disability simulation on nursing students' empathy using the Comprehensive State Empathy Scale. Nurse Education Today 2017 Dec;59:75-81. [doi: 10.1016/j.nedt.2017.09.007]
6. Patterson J. Empathy: A Concept Analysis. International Journal for Human Caring 2018;22(4):217-223. [doi: 10.20467/1091-5710.22.4.217]
7. Webster D. Promoting empathy through a creative reflective teaching strategy: a mixed-method study. J Nurs Educ 2010 Mar;49(2):87-94. [doi: 10.3928/01484834-20090918-09] [Medline: 19810667]

XSL•FO
RenderX

8. Konrath SH, O'Brien EH, Hsing C. Changes in dispositional empathy in American college students over time: a meta-analysis. Pers Soc Psychol Rev 2011 May 05;15(2):180-198. [doi: 10.1177/1088868310377395] [Medline: 20688954]
9. Gholamzadeh S, Khastavaneh M, Khademian Z, Ghadakpour S. The effects of empathy skills training on nursing students' empathy and attitudes toward elderly people. BMC Med Educ 2018 Aug 15;18(1):198 [FREE Full text] [doi: 10.1186/s12909-018-1297-9] [Medline: 30111312]
10. Ward A, Mandrusiak A, Levett-Jones T. Cultural empathy in physiotherapy students: a pre-test post-test study utilising virtual simulation. Physiotherapy 2018 Dec;104(4):453-461. [doi: 10.1016/j.physio.2018.07.011] [Medline: 30318129]
11. Neumann M, Edelhäuser F, Tauschel D, Fischer MR, Wirtz M, Woopen C, et al. Empathy decline and its reasons: a systematic review of studies with medical students and residents. Acad Med 2011 Aug;86(8):996-1009. [doi: 10.1097/ACM.0b013e318221e615] [Medline: 21670661]
12. Carol B Dulay M, A Domingo J, Domingo K, Domondon H, Dumangon L, D Duran R, et al. An exploratory study of factors influencing student nurses' empathy. J Health Edu Res Dev 2018;06(02). [doi: 10.4172/2380-5439.1000259]
13. Kim J. Factors influencing nursing students' empathy. Korean J Med Educ 2018 Sep;30(3):229-236. [doi: 10.3946/kjme.2018.97] [Medline: 30180510]
14. Berduzco-Torres N, Medina P, San-Martín M, Delgado Bolton RC, Vivanco L. Non-academic factors influencing the development of empathy in undergraduate nursing students: a cross-sectional study. BMC Nurs 2021 Dec 08;20(1):245 [FREE Full text] [doi: 10.1186/s12912-021-00773-2] [Medline: 34876108]
15. Ventura S, Badenes-Ribera L, Herrero R, Cebolla A, Galiana L, Baños R. Virtual reality as a medium to elicit empathy: a meta-analysis. Cyberpsychology, Behavior, and Social Networking 2020 Oct 01;23(10):667-676. [doi: 10.1089/cyber.2019.0681]
16. Everson N, Levett-Jones T, Pitt V. The impact of educational interventions on the empathic concern of health professional students: a literature review. Nurse Educ Pract 2018 Jul;31:104-111. [doi: 10.1016/j.nepr.2018.05.015] [Medline: 29852474]
17. Bas-Sarmiento P, Fernández-Gutiérrez M, Baena-Baños M, Correro-Bermejo A, Soler-Martins PS, de la Torre-Moyano S. Empathy training in health sciences: A systematic review. Nurse Education in Practice 2020 Mar;44:102739. [doi: 10.1016/j.nepr.2020.102739]
18. Mattsson K, Haavisto E, Jumisko-Pyykkö S, Koivisto JM. Nursing students experiences of empathy in a virtual reality simulation game: a descriptive qualitative study. Comput Inform Nurs CIN 2024 Apr;42(7):537-545. [doi: 10.1097/cin.0000000000001132]
19. Juan S, O'Connell KA. A systematic review of the application of simulation to promote empathy in nursing education. Sim Healthcare 2024 Jan 25;19(6):379-387. [doi: 10.1097/sih.0000000000000775]
20. Levett-Jones T, Cant R, Lapkin S. A systematic review of the effectiveness of empathy education for undergraduate nursing students. Nurse Education Today 2019 Apr;75:80-94. [doi: 10.1016/j.nedt.2019.01.006]
21. Bearman M, Palermo C, Allen LM, Williams B. Learning empathy through simulation: a systematic literature review. Simulation in Healthcare: journal of the Society for Simulation in Healthcare 2015;10(5):308-319. [doi: 10.1097/sih.0000000000000113]
22. Brammer SV, Regan SL, Collins CM, Gillespie GL. Developing innovative virtual reality simulations to increase health care providers' understanding of social determinants of health. J Contin Educ Health Prof 2021 Nov 17;42(1):60-65. [doi: 10.1097/ceh.0000000000000400]
23. Dyer E, Swartzlander BJ, Gugliucci MR. Using virtual reality in medical education to teach empathy. Journal of the Medical Library Association: JMLA 2018 Oct 04;106(4):498-500. [doi: 10.5195/jmla.2018.518]
24. Mandal S. Brief introduction of virtual reality and its challenges. International Journal of Scientific & Engineering Research 2013;4(4):303-309 [FREE Full text]
25. Mäkinen H, Haavisto E, Havola S, Koivisto J. User experiences of virtual reality technologies for healthcare in learning: an integrative review. Behaviour & Information Technology 2020 Jul 01;41(1):1-17. [doi: 10.1080/0144929x.2020.1788162]
26. Slater M. Immersion and the illusion of presence in virtual reality. Br J Psychol 2018 Aug 21;109(3):431-433. [doi: 10.1111/bjop.12305] [Medline: 29781508]
27. Koivisto J, Buure T, Engblom J, Rosqvist K, Haavisto E. The effectiveness of simulation game on nursing students' surgical nursing knowledge—a quasi-experimental study. Teaching and Learning in Nursing 2024 Jan;19(1):e22-e29. [doi: 10.1016/j.teln.2023.08.006]
28. Shorey S, Pereira TL, Teo WZ, Ang E, Lau TC, Samarasekera DD. Navigating nursing curriculum change during COVID-19 pandemic: a systematic review and meta-synthesis. Nurse Educ Pract 2022 Nov;65:103483 [FREE Full text] [doi: 10.1016/j.nepr.2022.103483] [Medline: 36327596]
29. Benedict N, Schonder K, McGee J. Promotion of self-directed learning using virtual patient cases. Am J Pharm Educ 2013 Sep 12;77(7):151 [FREE Full text] [doi: 10.5688/ajpe777151] [Medline: 24052654]
30. Havola S, Haavisto E, Mäkinen H, Engblom J, Koivisto JM. The Effects of Computer-Based Simulation GameVirtual Reality Simulation in Nursing Students Self-evaluated Clinical Reasoning Skills. Comput. Inform. Nurs. CIN 2021 May;39(11):725-735. [doi: 10.1097/cin.0000000000000748]

XSL•FO
RenderX

31. Saab MM, Landers M, Murphy D, O'Mahony B, Cooke E, O'Driscoll M, et al. Nursing students' views of using virtual reality in healthcare: a qualitative study. J Clin Nurs 2022 May 22;31(9-10):1228-1242. [doi: 10.1111/jocn.15978] [Medline: 34296483]
32. Hassenzahl M, Tractinsky N. User experience - a research agenda. Behaviour & Information Technology 2006 Mar;25(2):91-97. [doi: 10.1080/01449290500330331]
33. Hassenzahl M, Burmester M, Koller F. User experience is all there is: twenty years of designing positive experiences and meaningful technology. i-com 2021 Dec;20(3):197-213. [doi: 10.1515/icom-2021-0034]
34. Hassenzahl M. The interplay of beauty, goodness, and usability in interactive products. Human-Comp Interaction 2004 Dec 1;19(4):319-349. [doi: 10.1207/s15327051hci1904_2]
35. Mahlke S. User Experience of Interaction with Technical Systems. Computer Science, Engineering. 2008 May 26. URL: https://www.semanticscholar.org/paper/User-Experience-of-Interaction-with-Technical-Mahlke/79896db42ecc9decaab9c550039cec3f844db956
36. ISO 9241-210 (2019) Ergonomics of human-system interaction – Part 210: Human-centred design for interactive systems. International Organization for Standardization (ISO). URL: https://www.iso.org/standard/77520.html [accessed 2024-05-05]
37. Sanders EB, Stappers PJ. Co-creation and the new landscapes of design. CoDesign 2008 Mar;4(1):5-18. [doi: 10.1080/15710880701875068]
38. Law EL, Heintz M. Augmented reality applications for K-12 education: a systematic review from the usability and user experience perspective. International Journal of Child-Computer Interaction 2021 Dec;30:100321. [doi: 10.1016/j.ijcci.2021.100321]
39. Grove S. In: Grey J, Grove S, editors. Burn’s & Grove’s The practice of Nursing Research. Appraisal, Synthesis, and Generation of Evidence, 9th Edition. St. Louis, MO: Elsevier; 2021:310-449.
40. Koivisto J, Haavisto E, Niemi H, Katajisto J, Multisilta J. Elements explaining learning clinical reasoning using simulation games. IJSG 2016 Dec 21;3(4):29-43. [doi: 10.17083/ijsg.v3i4.136]
41. Koivisto J, Multisilta J, Niemi H, Katajisto J, Eriksson E. Learning by playing: A cross-sectional descriptive study of nursing students' experiences of learning clinical reasoning. Nurse Educ Today 2016 Oct;45:22-28. [doi: 10.1016/j.nedt.2016.06.009] [Medline: 27429399]
42. Mäkinen H, Haavisto E, Havola S, Koivisto J. Graduating nursing students' user experiences of the immersive virtual reality simulation in learning - a qualitative descriptive study. Nurs Open 2023 May 04;10(5):3210-3219 [FREE Full text] [doi: 10.1002/nop2.1571] [Medline: 36598872]
43. Carré A, Stefaniak N, D'Ambrosio F, Bensalah L, Besche-Richard C. The Basic Empathy Scale in adults (BES-A): factor structure of a revised form. Psychol Assess 2013 Sep;25(3):679-691. [doi: 10.1037/a0032297] [Medline: 23815121]
44. Wild D, Grove A, Martin M, Eremenco S, McElroy S, Verjee-Lorenz A, ISPOR Task Force for TranslationCultural Adaptation. Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: report of the ISPOR Task Force for Translation and Cultural Adaptation. Value Health 2005;8(2):94-104 [FREE Full text] [doi: 10.1111/j.1524-4733.2005.04054.x] [Medline: 15804318]
45. Hassenzahl M, Platz A, Burmester M, Lehner K. Hedonic and ergonomic quality aspects determine a software’s appeal. 2000 Apr Presented at: Proceedings of the SIGCHI conference on Human Factors in Computing Systems, Association for Computing Machinery; 2000; New York, NY, USA p. 201-208. [doi: 10.1145/332040.332432]
46. Sassatelli L, Winckler M, Fisichella T, Dezarnaud A, Lemaire J, Aparicio-Pardo R, et al. New interactive strategies for virtual reality streaming in degraded context of use. Computers & Graphics 2020 Feb;86:27-41. [doi: 10.1016/j.cag.2019.10.005]
47. Takahashi L, Nebe K. Observed differences between lab and online tests using the AttrakDiff semantic differential scale. Journal of Usability Studies 2019;14(2):65-75 [FREE Full text]
48. The ethical principles of research with human participants and ethical review in the human sciences in Finland. Finnish National Board on Research Integrity TENK guidelines 2019. URL: https://tenk.fi/sites/default/files/2021-01/Ethical_review_in_human_sciences_2020.pdf [accessed 2025-01-28]
49. Hannans JA, Nevins CM, Jordan K. See it, hear it, feel it: embodying a patient experience through immersive virtual reality. ILS 2021 Jun 21;122(7/8):565-583. [doi: 10.1108/ils-10-2020-0233]
50. Wang X, Cheng Z. Cross-sectional studies: strengths, weaknesses, and recommendations. Chest 2020 Jul;158(1S):S65-S71. [doi: 10.1016/j.chest.2020.03.012] [Medline: 32658654]

Abbreviations

ATT: attractiveness
BES-A: Basic Empathy Scale in Adults
CSES: Comprehensive State Empathy Scale
HMD: head-mounted display
HQS: hedonic quality stimulation
ISPOR: International Society for Pharmacoeconomics and Outcomes Research

XSL•FO
RenderX

PQ: pragmatic quality
UAS: university of applied sciences
UX: user experience
VR: virtual reality

Edited by N Ahmadpour; submitted 31.05.24; peer-reviewed by R Kaul, H Lee, Q Zhou; comments to author 17.10.24; revised version received 06.11.24; accepted 16.12.24; published 13.02.25.

Please cite as:
Koivisto JM, Kämäräinen S, Mattsson K, Jumisko-Pyykkö S, Ikonen R, Haavisto E
Exploring Nursing Students' Experiences of Empathy and User Experiences in an Immersive Virtual Reality Simulation Game: Cross-Sectional Study
JMIR Serious Games 2025;13:e62688
URL: https://games.jmir.org/2025/1/e62688
doi:10.2196/62688
PMID:

©Jaana-Maija Koivisto, Sanna Kämäräinen, Katri Mattsson, Satu Jumisko-Pyykkö, Riikka Ikonen, Elina Haavisto. Originally published in JMIR Serious Games (https://games.jmir.org), 13.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Immersive Virtual Reality and AI (Generative Pretrained Transformer) to Enhance Student Preparedness for Objective Structured Clinical Examinations: Mixed Methods Study

Shaniff Esmail[1*], PhD; Brendan Concannon[1*], MSc

Department of Occupational Therapy, University of Alberta, Edmonton, AB, Canada

[*]all authors contributed equally

Corresponding Author:
Shaniff Esmail, PhD
Department of Occupational Therapy
University of Alberta
8205 - 114 St 2-64 Corbett Hall
Edmonton, AB, T6G 2G4
Canada
Phone: 1 780 492 2499
Email: shaniff.esmail@ualberta.ca

Abstract

Background: Immersive virtual reality (VR) and artificial intelligence have been used to determine whether a simulated clinical exam setting can reduce anxiety in first-year occupational therapy students preparing for objective structured clinical examinations (OSCEs). Test anxiety is common among postsecondary students, leading to negative outcomes such as increased dropout risk, lower grades, and limited employment opportunities. Students unfamiliar with specific testing environments are particularly prone to anxiety. VR simulations of OSCEs may allow students to become familiar with the exam setting and reduce anxiety.

Objective: This study aimed to assess the efficacy of a VR simulation depicting clinical settings to reduce student anxiety about a clinical exam while gathering perspectives on their first-year coursework experiences to better understand their learning environment.

Methods: An experimental, nonrandomized controlled trial compared state anxiety, trait test anxiety, and OSCE grades in 2 groups of first-year occupational therapy students analyzed using independent *t* tests (2-tailed). Group 1 (NoVR) was not exposed to the VR simulation and acted as a control group for group 2 (YesVR), who were exposed to the VR simulation. The VR used artificial intelligence in the form of a generative pretrained transformer to generate responses from virtual patients as students interacted with them in natural language. Self-reported psychometric scales measured anxiety levels 3 days before the OSCE. YesVR students completed perceived preparation surveys at 2 time points—3 weeks and 3 days before the OSCE—analyzed using dependent *t* tests. Semistructured interviews and focus groups were conducted within 1 week after the OSCE. Student perspectives on their classes and VR experiences were summarized using interpretative thematic analysis.

Results: In total, 60 students—32 (53%) in the NoVR group and 28 (47%) in the YesVR group—participated in the study, and the YesVR group showed a significant reduction in state anxiety (t_{58}=3.96; P<.001; Cohen d=1.02). The mean difference was 11.96 units (95% CI 5.92-18.01). Trait test anxiety and OSCE scores remained static between groups. There was an increase in all perceived preparedness variables in the YesVR group. In total, 42% (25/60) of the participants took part in interviews and focus groups, providing major themes regarding factors that affect OSCE performance, including student experience and background, feedback and support, fear of unknown, self-consciousness, and knowledge of the exam environment.

Conclusions: Intolerance of uncertainty may lead students to interpret ambiguous exam situations as overly precarious. Findings suggest that VR simulation was associated with reduced state anxiety, although results from this small, nonrandomized sample should be interpreted cautiously. Qualitative data indicated that VR helped students gain familiarity with clinical exam settings, potentially decreasing uncertainty-based anxiety. Future research with larger or randomized samples is needed to confirm these findings and explore advanced VR tools offering feedback to enhance learning.

(JMIR Serious Games 2025;13:e69428) doi:10.2196/69428

XSL•FO
RenderX

KEYWORDS

virtual reality; head-mounted display; immersive technology; artificial intelligence; generative pretrained transformer; occupational therapy; objective structured clinical examination; simulation; psychology; anxiety

Introduction

Background

This investigation was a mixed methods replication study expanding on a previous study by Concannon et al [1]. Their original work examined an immersive virtual reality (VR) simulation for occupational therapy (OT) students preparing for a clinical exam, observing a notable reduction in state anxiety 1 week before the objective structured clinical examination (OSCE). However, the aforementioned study did not collect qualitative feedback on VR's benefits or examine how first-year coursework factors might influence performance, leaving open questions about which aspects of VR students found most useful and whether the results would generalize to new cohorts. Therefore, this investigation updated the technology using enhanced artificial intelligence (AI)–driven avatar responses and incorporated qualitative methods to better understand the student experience and confirm the consistency of the previous findings.

The purpose of this study was to replicate the results of the study by Concannon et al [1], expanding on the concept of student academic self-efficacy by measuring individual components of perceived preparedness, including student levels of confidence, awareness, knowledge, understanding, and skill level. This investigation used an improved VR simulation featuring the use of AI, which allowed the virtual patient to provide dynamic responses to student questions. This investigation also included qualitative components—semistructured interviews and focus groups were used to obtain student perspectives regarding their experiences with their first-year coursework and the VR simulation. These qualitative additions may offer context and student recommendations to better understand how immersive VR benefits first-year OT coursework.

Campus Anxiety: Still a Prevalent Problem

Anxiety is an innate and typical response, preparing the body for an impending situation perceived to be dangerous [2]. It is a theoretical construct capable of manifesting in either general or specific situations, with *predisposition* (ie, trait anxiety) representing the frequency or intensity of the response and *transitory* (ie, state anxiety) representing an acute response to a specific situation and time [3]. Trait anxiety is associated with individual personality traits, impacting the level of state anxiety response within a specific context [3]. Test anxiety is also differentiated between *state test anxiety* [4], showing transitory anxiety affects provoked by a specific evaluative situation, and *trait test anxiety* [5], considered to be a general condition with varying ranges of intensity and contexts. Test anxiety has been defined by cognitive components, including negative thinking patterns (eg, *worry*) that may manifest negative thought outcomes and reduced attention, including the shifting of focus to irrelevant cues [6,7]. Worry thought processes include *catastrophizing*, where individuals have intrusive thoughts of failure or doubt in their capability [7].

There are also affective components, comprising *emotionality* aspects with physiological symptoms, including arousal and uneasiness (eg, stomach sensations, muscle tension, increased heart rate, headache, and sweating) [6,7]. These negative cognitive and emotional responses are strongly associated with reduced student performance on intellectual-based tasks [5]. There is a positive correlation between distracting contextual factors (ie, lack of dedicated study space and lack of access to reliable technology) and test anxiety [8]. Students who are unfamiliar with the test-taking environment for a particular exam are more likely to have symptoms of test anxiety as well [9]. It would be expected for students with previous experience in clinical practical exams, such as an OSCE, to show reduced anxiety due to improved familiarity. However, OSCEs have been identified as one of the most stressful exam types, with additional student exposure not necessarily lessening their anxiety about them in general [10,11].

The long-term effects of anxiety in postsecondary education include increased risk of dropout, reduced grades, reduced chance of employment, and the loss of billions in government dollars each year [12]. Test anxiety is associated with student burnout [13]. Campus anxiety is associated with depression and may be linked to increased student suicide rates [14-16]. Between 2008 and 2016, emergency department visits from students of US universities showed that anxiety complaints had increased by 260% [17]. In addition, test anxiety may predict the increase of student cheating behavior on exams [18]. One psychological model explaining why test anxiety occurs includes the intolerance of uncertainty model [19]. The intolerance of uncertainty model theorizes that people may inherently have an intolerance of uncertainty, resulting in ambiguous situations being perceived as threatening, leading to associated worry and anxiety [19,20]. While test anxiety is a normal phenomenon, it becomes problematic when students are unable to stop worrying or when the anxiety disrupts their practice [20].

VR Versus Anxiety

A common method used to condition individuals against anxiety-inducing situations is in vivo exposure, in which an individual is exposed to situations or stimuli in real-world environments until distress is decreased [21]. VR extends this approach by reproducing real-world sights and sounds digitally, creating what is known as VR exposure therapy [21]. It is also known as *in virtuo* exposure [22]. VR may be preferable to traditional cognitive behavioral therapy as a means of exposure, functioning as an effective, affordable, and practical alternative to treatment avoidance [22].

A 3-step conceptual description of VR includes elements of user portrayal, immersion, and level of fidelity [23]. VR includes a broad range of computer-based programs, often used to render immersive and detailed 3D environments, allowing users to navigate within digital representations of the real world [24].

XSL•FO
RenderX

The goal of immersive VR is to replace the sensory flow of information from the real world with those created by a computer, promoting the illusion that the virtual world is real. The study and review by Concannon et al [1,25] provide more information about the immersive, interactive, imaginative, and fidelity abilities of immersive VR.

For cognitive adaptation (eg, improvement of memory, information processing, problem-solving, and logical sequencing), VR training has shown improvements in the brain's frontal lobe and its cognitive characteristics, such as prompt recall of prospective memory tasks and precise performance of event- and time-based objectives [26]. VR learning may improve procedural memory because of changes in neural plasticity, improving working memory [27]. VR may improve a user's cognitive and selective attention [28]. Enhancing problem-solving abilities and selective memory processes may improve social interactive elements such as those used in interview situations [29].

VR in Health Science

A meta-analysis by Kyaw et al [30] found that health science students (ie, physicians, nurses, and students pursuing medical degrees) showed improvements in knowledge and skill outcomes when VR interventions were used (ie, 3D anatomical models, virtual patients, and surgical simulations) in comparison to traditional educational programs. Within OT education, VR simulations have gradually shifted toward more realistic, ecologically valid environments, especially for client intervention scenarios [31,32]. There is a need for mixed methods research that captures the nuanced perspectives and motivations of OT students, informing how VR can be most effectively integrated into their coursework.

Aim of This Investigation

This study aimed to replicate the effectiveness of immersive VR in lowering state anxiety in OT students preparing for an OSCE. To uncover qualitative factors influencing student performance in coursework and VR simulations, interviews and focus groups were conducted. As an extension of academic self-efficacy, the students' perceived preparedness factors for their OSCEs were tested to determine which had improved the most (ie, confidence, awareness, knowledge, understanding, and skill).

The objectives of this study were 2-fold: first, to quantitatively evaluate whether an AI-enhanced VR simulation was associated with reduced OSCE-related anxiety and, second, to gather student perceptions regarding VR's usefulness, limitations, and potential enhancements. By integrating these perspectives, we aimed to refine the simulation's design and determine how VR might be embedded into the broader OT curriculum to address both practical and emotional aspects of exam preparation.

Research Questions

This investigation was designed to answer the following research questions: (1) does immersive VR simulation of a clinical practical exam (ie, OSCE) show a reduction in state anxiety in OT students when compared to a control group? (2) To better understand and improve learning systems, what are the student perceptions regarding VR simulation and OT coursework experiences? (3) Which variables of student perceived preparedness for their OSCE improved the most?

Methods

Experimental Design

This investigation was a mixed, cross-sectional, nonrandomized controlled trial involving 2 groups of participants, each comprising first-year OT students from the same class. This investigation attempted to replicate the difference in mean state anxiety between 2 groups of participants: students not exposed to a VR simulation (NoVR group), who acted as a control group for students who were exposed to VR (YesVR group). The University of Alberta Department of Occupational Therapy program had eased remote learning practice, which was established between the years of 2020 and 2022 in response to COVID-19 restrictions. First-year OT coursework remained consistent in terms of faculty infrastructure, course content, curricular sequencing, and instructors since the study by Concannon et al [1].

The mixed methods design was required for the two objectives: (1) to quantify any changes in state anxiety or exam performance following VR exposure and (2) to explore student perceptions of VR's usefulness, limitations, and potential enhancements. This approach ensured that any observed quantitative shifts in anxiety could be contextualized with the qualitative insights essential for refining the simulation and shaping future curricular strategies.

This investigation assumed that student study time and strategies would not be significantly different between the 2 groups, omitting the need for students to provide written logs of their study strategies and time spent in preparation for the OSCE. This decision was implemented to minimize participant burden, forgoing the need for students to provide detailed study logs in addition to their course workload. The study by Concannon et al [1] used such logs and found that student time spent on study strategies for the OSCE was similar between groups, yet participants remarked that it was difficult to maintain the written logs in addition to regular coursework.

Recruitment

At the start of the term, a speaker who was neutral to the investigation's outcome made in-class announcements providing details of the investigation. The speaker was not an instructor from the faculty; this was to minimize the compulsory pressure on students to participate. The announcement was made during a lecture to an OT class of 125 first-year students. All OT students were invited and eligible to participate. Students were informed of the class email announcements and invitations to web-based surveys, which would be made available for them to accept and participate. The speaker concluded the announcement with a brief demonstration of the VR hardware that was available, including basics on how to use it and the risks typically associated with VR use.

XSL•FO
RenderX

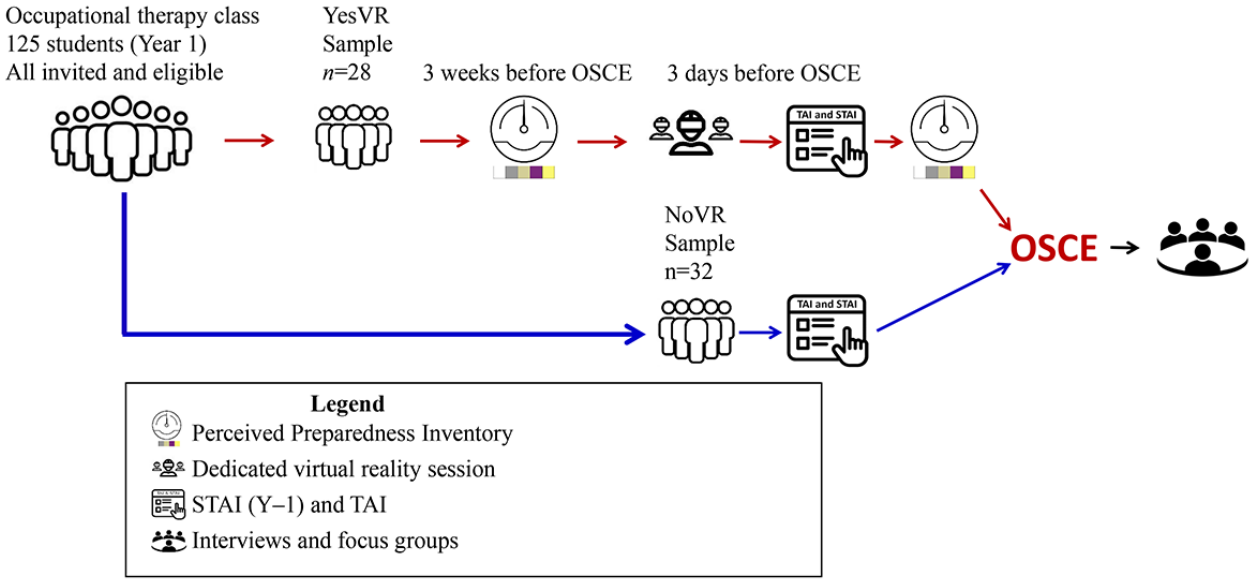
Ethical Considerations

This study was reviewed and approved by the University of Alberta's Research Ethics Office (Pro00107738) in accordance with the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans. All participants received an information letter outlining the study purpose, procedures, and potential risks, and they provided informed consent via signed or electronic forms. To protect privacy, data were deidentified at collection and stored in encrypted files accessible only to designated research assistants; any identifying details were removed from transcripts or replaced with aliases to ensure anonymity in all reports and publications. No compensation was offered, and use of the VR simulation was made available regardless of participation. This manuscript does not include images that could identify individuals. Any similarity to actual persons is purely coincidental. Although demographic data on sex are often collected in studies of this nature, the reviewing ethics board restricted these measures to protect individuals' anonymity given the program's enrollment composition (approximately 90% female, 8% male, and 1% identifying otherwise). As a result, sex-based data were not gathered.

Experimental Process

In this investigation, while the VR simulation was available as a self-serve tool, it was important to have a consistent measure taken at the peak anxiety time point. This investigation used class email announcements to offer a consolidated VR session appointment. Those who accepted the offer were designated as YesVR participants and administered a web-based Perceived Preparedness Inventory survey 3 weeks before the OSCE. The appointed YesVR students were given their scheduled VR sessions 3 days before the OSCE, followed by a Perceived Preparedness Inventory survey again. There were separate rooms, each equipped with VR headsets, allowing multiple participants to privately complete the simulation simultaneously. If one headset had a technical difficulty, participants could be assigned to another room to keep the process expeditious. After their VR appointments, a class email announcement was sent out to invite all students to complete web-based State-Trait Anxiety Inventory (STAI) and Test Anxiety Inventory (TAI) surveys measuring student state anxiety and trait test anxiety levels, respectively. These surveys asked students whether they had completed the VR simulation. Students who reported not having used the VR simulation yet chose to complete the STAI were designated as the NoVR group. The timing of the web-based STAI, TAI, and Perceived Preparedness Inventory surveys was set to take place after the VR sessions and class hours so that students could complete them outside the presence of instructors or researchers. Focus groups and interviews were scheduled within a 1-week time frame after the students completed their OSCEs via class email announcements and web-based sign-up sheets. Figure 1 provides a summary of this investigation's experimental process.

Figure 1. Experimental process of this investigation. NoVR: participants not exposed to the virtual reality simulation; OSCE: objective structured clinical examination; STAI: State-Trait Anxiety Inventory; TAI: Test Anxiety Inventory; YesVR: participants exposed to the virtual reality simulation.



Primary Outcome Measures

STAI Forms

The STAI is divided into 2 forms: the Y-1 (S-Anxiety) scale, which measures a participant's level of anxiety at a specific moment in time, and the Y-2 (T-Anxiety) scale, which measures a participant's predisposition and long-standing level of anxiety. Each form comprises 20 items, with final scores ranging from 20 to 80 and higher scores representing greater levels of anxiety [3]. The STAI S-Anxiety scale was developed to measure college students' levels of state anxiety, showing a reliability stability of $r<0.62$. Reliability coefficients for this instrument are expected to be low to moderate due to its capability to show differences in participant anxiety levels between each testing-retesting situation [3]. The STAI S-Anxiety scale normative Cronbach α coefficients for college students have been reported to be 0.91 and 0.93 for male and female individuals, respectively [3]. For validity, the STAI S-Anxiety

XSL•FO
RenderX

scale was compared to the Institute of Personality and Ability Testing anxiety scale and Taylor Manifest Anxiety Scale, showing validity correlation coefficients of r=0.75 and r=0.80, respectively [3]. Overall, the STAI shows consistency in measuring defined components of anxiety, including apprehension, tension, nervousness, and worry [3]. For this investigation, the STAI S-Anxiety scale was administered as a web-based survey to both the YesVR and NoVR participants at the same time point, 3 days before the OSCE.

TAI Forms

The TAI is a subjective psychometric scale measuring individual differences in anxiety about testing situations as a personality trait. The instrument consists of 20 items, with final scores ranging from 20 to 80 and higher scores being associated with stronger symptoms of test anxiety [33]. The TAI scale has shown a reliability stability of r=0.80 for periods varying between 2 weeks and 6 months [33]. The TAI showed normative Cronbach α coefficients for college students of 0.92 for both male and female individuals [33]. For validity, the TAI was compared to the Test Anxiety Scale and Worry-Emotionality Questionnaire, showing validity coefficients of r=0.83 and r=0.85, respectively [33]. For this investigation, the TAI was administered as a web-based survey to both the YesVR and NoVR participants at the same time point, 3 days before the OSCE.

Interview and Focus Group Setup

After the OSCE, all students received an email inviting them to sign up for interviews and focus groups, which remained open for 1 week. Participation was voluntary and not restricted to those who had used the VR simulation. To uphold ethical guidelines and prevent any sense of obligation, all students were encouraged to share their experiences regardless of VR participation. The interviews and focus groups lasted approximately 45 to 60 minutes. Copies of the interview and focus group questionnaires are available in Multimedia Appendices 1 and 2, respectively. Interview and focus group data were summarized using an interpretative thematic analysis based on an approach developed by Burnard [34], the interview process guide by Kvale [35], and the general approach by Maykut and Morehouse [36].

Participant responses were grouped into various categories as recurring trends emerged. Analysis of field notes and transcripts was processed across 10 stages partitioned into 3 phases: phase 1, consisting of 4 stages in which the data were filtered and categorized; phase 2, consisting of 3 stages in which data were condensed, organized, and integrated; and phase 3, in which understanding and meaning were derived from the data. This structured, 10-step and open coding process primarily aligns with the approach by Burnard [34], whereas the constant comparison method aligns with the approach by Maykut and Morehouse [36] and the emphasis on interpretation and participant perspectives by Kvale [35]. Table 1 provides more information on the interview and focus group analysis process.

In keeping with recommendations by Maykut and Morehouse [36], the research assistants maintained reflexive journals to document their assumptions and any shifts in interpretation. In total, 2 researchers independently reviewed the transcripts, reconciling coding differences through iterative discussion, thereby supporting confirmability [34]. Dependability was strengthened by maintaining an audit trail of analytic decisions, as outlined by Maykut and Morehouse [36].

XSL•FO
RenderX

Table 1. Interview and focus group analysis process.

Phase and stage	Process
Phase 1: filtering and categorizing	
1	• Notes are written immediately after the interview. • Memos are written about ways of categorizing data.
2	• Each transcription is read to gain a better understanding of some of the larger themes. • Quick notes are made (eg, terms and definitions).
3	• Each transcript is read again, and detailed notes on general themes and major categories are recorded in table format.
4	• Open coding [34]: categories are created as needed to describe what is in the text. • Transcriptions are read again, and as many headings and themes as possible are recorded. • Fillers are deleted.
Phase 2: condensing, organizing, and integrating	
5	• Constant comparison method [36]: data segments are continuously compared to emerging categories. • Data meanings are compared to each new unit and existing units before grouping. • Data are condensed, organized, and integrated. • Headings and themes are narrowed down, and repetitive ones are removed.
6	• Headings and themes are grouped under higher-order headings.
7	• Transcriptions are reviewed again using the categories and higher-order headings.
Phase 3: understanding and meaning	
8	• Discovery pages are created with categories listed and themes grouped under them. • Short narrations briefly describing the primary points are created. • Comparisons are made between interviews.
9	• Narratives are shared with participants or within the research team to refine the data.
10	• Narratives are linked with what is found in the literature and how it relates to the research questions.

Secondary Outcome Measures

Perceived Preparedness Inventory

The aim of this investigation was to analyze individual components of student perceived preparedness, including self-confidence, awareness, knowledge, understanding, and skill level in accordance with OT learning objectives. The Perceived Preparedness Inventory survey was used to assess these specific components, allowing for the observation of which variables improved the most throughout the term. This web-based survey was adapted from the physiotherapy study by Weeks and Horan [37] and the nursing study by Massey et al [11] and used a 5-item, 5-point Likert scale to determine the variables that showed the strongest improvements in student perceived preparedness within their OT coursework. The survey's wording was intended to probe students' perceived preparedness variables for OSCEs in a general sense and did not address the impact of immersive VR. A copy of this survey is available in Multimedia Appendix 3.

OSCE Scores

OSCE score comparisons between the NoVR and YesVR groups were conducted to determine the effects of the VR simulation on student performance. Specifically, comparisons were made between the 2 groups' history taking and cognitive assessment OSCE scores, as well as their total course grades. During the OSCE, each student was required to complete 2 stations followed by a written component. The first station required students to interact with a standardized patient (actor) who was afflicted with a random physical ailment, making up the OSCE's history taking and physical check component. This random physical complaint (scenario) would be an acute injury in a

XSL•FO
RenderX

random peripheral anatomical location. The second station required students to interact with a different standardized patient representing a client with a mild cognitive disability. After these 2 stations were completed, the students would complete a written component requiring them to apply their clinical findings, demonstrate critical thinking, and clarify their clinical reasoning related to the scenario.

Simulation Design

Overview

The VR simulation in this investigation included the following components:

- A virtual environment depicting a health sciences clinic rendered using the Unity game engine software (Unity Technologies). The environment allowed the user to select from 1 of 2 modules: History Taking or the Saint Louis University Mental Status (SLUMS) exam cognitive assessment [38].
- 3 virtual avatars—the first appeared in the History Taking module as a virtual standardized patient who responded to a user's questions, the second appeared in the cognitive assessment module as a virtual standardized patient who responded to a user question from the SLUMS cognitive assessment, and the third was a virtual exam evaluator who observed the user and wrote notes into a clipboard during each module.
- Speech recognition and response software using Azure Cognitive Services (Microsoft Corp). The process pipeline includes user speech to text, open AI for language processing and generation of the virtual avatar's text response, and the avatar's text converted from text to speech.
- Meta Quest 2 headsets that ran the VR software, uploaded using SideQuest (Khronos Group). These headsets were portable and free of cables. The headset could detect a user's hand gestures without the use of controllers. The headset was also equipped with a microphone to detect user speech for communicating with the virtual avatars.

Virtual Health Sciences Clinic

This investigation rebuilt the VR simulation based on the system mentioned in the study by Concannon et al [1] using the same interdisciplinary team members' expertise in computer science, physical therapy, communication and science disorders, rehabilitation medicine, and OT. Student and researcher feedback from the aforementioned study was implemented to further improve the system used in this investigation. The virtual clinic in this investigation had the same clinical office layout as that used in the previous study, yet this simulation had graphical, functional, and quality-of-life improvements. The virtual clinic's tools, tables, and plinth layout were arranged to closely match those used in the actual OSCE room. This study's virtual clinic had a main door with a menu. This menu allowed users to select the virtual module they wanted to attempt: History Taking or the SLUMS exam. The user could do this by using hand gestures to push buttons on the door, which the headset could detect. Upon selecting an option, the door would open to a hall where the user could walk to the examination room to meet the standardized patient.

Depending on the module, the virtual clinic had a different clipboard for the user to interact with and read from. The History Taking clipboard contained preliminary notes about the virtual patient, similar to what the student would receive before interacting with their real-world OSCE standardized patient. The virtual patient in this module had a random physical complaint (scenario), such as an acute injury in a random peripheral anatomical location. Once the user entered the virtual room, a buzzer sounded to signal the start of the virtual OSCE, and a miniature timer on a desk began counting down from 8 minutes. The user could ask the virtual patient questions in natural language, with the user's voice being detected by the headset's microphone to convert the question from speech to text. The virtual patient would respond to health questions, including those about physical, social, psychological, spiritual, emotional, workplace, and daily living health domains. Figure 2 provides a sample screenshot of the History Taking module.

For the SLUMS module, the clipboard contained the questions and scoring rubric from the SLUMS cognitive assessment. A copy of the SLUMS cognitive assessment and scoring rubric has been included in Multimedia Appendix 4. The virtual patient in the SLUMS module would attempt to solve each test question asked by the student, with a random 25% chance of error. For example, one question required the student to request the virtual patient to draw the hands of a clock inside a circle, showing a time of 10:50. The virtual patient would answer appropriately, conducting the motion of drawing the clock on a piece of paper to display an illustration of a clock with hour and minute hands indicating a time of 10 minutes to 11. There was a 25% random chance that the virtual patient would make a mistake while attempting this task, drawing the clock with the hour and minute hands in switched positions or the incorrect time altogether. The SLUMS module's miniature timer began counting down from 15 minutes once the module started as the user entered the virtual room.

Real-world OSCEs are typically station oriented, attempting to replicate clinical scenarios with as much detail as possible. They usually feature standardized patients who represent clinical patients in health-related situations. A typical OSCE will grade a student's performance across an array of evaluation categories, including (the percentages represent mean scoring components) [39] history taking (30%-40%), physical examination (20%-40%), patient education (0%-10%), physician-patient interaction (20%-40%), and note practice (5%).

XSL•FO
RenderX

Figure 2. Screenshot of the virtual History Taking module.



Speech Recognition Software

The virtual standardized patients' text responses were generated using a generative pretrained transformer (GPT; GPT-3.5 [OpenAI]), which was fine-tuned (ie, trained) on recorded live student and patient actor interactions. GPT models use a neural network that has undergone pretraining using a substantial literature database. When a GPT creates text, it uses learned information to choose the most likely subsequent word based on the previous words in a sequence, aligning with the word probability distribution learned from its training data. This procedure is reiterated for each following word in the sequence, leading to the formation of a naturally flowing sentence. For this investigation, the avatar training was conducted using a transfer learning technique on the base model using real-world student and standardized patient text files that were collected from transcribed recordings. The simulation's GPT-3.5–based model was deployed on the Azure Machine Learning Cloud service (Microsoft Corp) for real-time inference. Virtual avatar actions (behaviors) were generated in text form by GPT-3.5 and then linked to appropriate animations, allowing them to respond to user requests (eg, if a user asked the virtual patient the following— "Can you raise both arms up into the air?"—the virtual avatar would respond accordingly). GPT-3.5 selected appropriate avatar action and verbal responses deployed through the Azure Machine Learning Cloud service using custom training and deployment scripts. In other words, GPT-3.5 acted as the actual *language process and avatar action–deciding* engine. Azure Machine Learning Cloud functioned as a bridge, relaying information between the VR headset and the GPT. The virtual patient could provide basic responses to convergent questions. For instance, if a user said the following—"Can you tell me more?"—the virtual patient could respond with related information based on the previous question or simply reply, "That's all I know."

Statistical Analysis

Statistical significance was evaluated at α=.05, and a 2-sided *P* value of ≤.05 was considered to be statistically significant. Comparisons between the NoVR and YesVR groups were conducted using independent *t* tests (2-tailed), which compared STAI, TAI, and OSCE scores between the groups. Repeated *t* tests were used to compare the YesVR group's perceived preparedness results before and after the VR sessions. The Cohen *d* effect size was checked for each *t* test between the groups.

Results

Overview

A total of 60 students participated in the quantitative portion of this study: 32 (53%) in the NoVR group and 28 (47%) in the YesVR group. A total of 25 students participated in the qualitative portion of this study: 16 (64%) in the interviews and 9 (36%) in the focus groups. Most YesVR participants (26/28, 93%) used the VR simulation for both modules, which typically meant 8 minutes of the History Taking module and an additional 15 minutes with the SLUMS module, in addition to some participants retrying one or both of the modules. The mean VR simulation time spent by the YesVR group was 24.11 (SD 8.00) minutes per participant. Table 2 provides the participant data. The main statistical analysis results are provided in Table 3. Perceived-preparation results for the YesVR group are presented in Table 4.

XSL•FO
RenderX

Table 2. Participant data.

Group	Participants, n (%)
Quantitative portion (n=60)[a]	
NoVR[b] participants[c]	32 (53)
YesVR[d] participants[e]	28 (47)
Qualitative portion (n=25)	
Interview participants	16 (64)
Focus group participants	9 (36)

[a]Mean age 24.53 (SD 2.64) years.
[b]NoVR: participants not exposed to the virtual reality simulation.
[c]Mean age 24.52 (SD 2.42) years.
[d]YesVR: participants exposed to the virtual reality simulation.
[e]Mean age 24.54 (SD 2.93) years.

Table 3. Statistical analysis results.

Test	NoVR[a], mean (SD)	YesVR[b], mean (SD)	*t* test (*df*)	*P* value	Cohen *d*
STAI[c] (form Y-1) score	55.00 (10.18)	43.04 (13.19)	3.96 (58)	<.001	1.02
TAI[d] score	41.81 (14.36)	43.50 (13.13)	−0.47 (58)	.64	0.12
OSCE[e] score: history taking	86.38 (6.53)	87.25 (7.64)	0.48 (58)	.63	0.12
OSCE[e] score: cognitive assessment	88.47 (8.04)	90.07 (5.88)	0.87 (58)	.39	0.23
OSCE[e] score: course total	90.03 (4.26)	90.25 (4.92)	0.19 (58)	.85	0.05

[a]NoVR: participants not exposed to the virtual reality simulation.
[b]YesVR: participants exposed to the virtual reality simulation.
[c]STAI: State-Trait Anxiety Inventory. Measure taken 3 days before the Objective structured clinical examination.
[d]TAI: Test Anxiety Inventory. Measure taken 3 days before the objective structured clinical examination.
[e]OSCE: objective structured clinical examination.

Table 4. Perceived preparation analysis results.

Preparedness domain	Pre-VR[a], mean (SD)	Post-VR[b], mean (SD)	*t* test (*df*)	*P* value	Cohen *d*
Perceived preparedness: confidence	2.71 (0.94)	3.46 (1.00)	2.68 (27)[c]	.01	0.77
Perceived preparedness: awareness	3.18 (1.16)	3.89 (0.86)	2.68 (27)[c]	.01	0.70
Perceived preparedness: knowledge	2.82 (0.82)	3.61 (1.03)	3.16 (27)[c]	.004	0.85
Perceived preparedness: understanding	3.32 (0.90)	3.93 (0.98)	2.62 (27)[c]	.01	0.65
Perceived preparedness: skill	3.00 (0.98)	4.04 (0.96)	4.16 (27)[c]	<.001	1.07

[a]PreVR: student scores 3 weeks before the objective structured clinical examination, before their virtual reality training session.
[b]PostVR: student scores 3 days before the objective structured clinical examination, after their virtual reality training session.
[c]Participants who completed the perceived preparedness surveys were in the YesVR group only; thus, df were changed accordingly.

Primary Outcome Measures

State Anxiety Scores

There was a significant difference in state anxiety scores between the groups, with NoVR showing higher anxiety scores (mean 55.00, SD 10.18) than YesVR (mean 43.04, SD 13.19; t_{58}=3.96; *P*<.001; Cohen *d*=1.02). The mean difference was 11.96 units (95% CI 5.92-18.01).

Test Anxiety Scores

There was no significant difference in test anxiety scores between the groups, with NoVR showing similar anxiety scores (mean 41.81, SD 14.36) to those of YesVR (mean 43.50, SD 13.13; t_{58}=−0.47; *P*=.64; Cohen *d*=0.12). The mean difference was −1.69 units (95% CI −8.84 to 5.46).

XSL•FO
RenderX

Student Perspectives and Experience

Overview

A total of 42% (25/60) of the students participated in the interviews and focus groups. Each of the 8 interview sessions had 2 participants interviewed separately. The first focus group session had 5 participants, whereas the second focus group session had 4 participants. Participants remarked that the actual exam was not as difficult or challenging as they thought it would be. Participants also remarked that time management was the most stressful part of the OSCE. VR was cited as being useful in helping with student orientation regarding the exam procedure while allowing students to fail and work through difficulties in a low-stakes environment. Regarding areas of improvement, students stated that they wanted improved feedback from the VR simulation and the ability to practice components of the OSCE physical check. Table 5 provides a summary of the interview and focus group themes.

Table 5. Interview and focus group theme analysis summary.

Themes	Subthemes
Major themes	
Experience and background	• Previous clinical exam experience • Personality traits (talkative vs reserved)
Need for feedback and support	• Instructor and peer feedback • Laboratory size (number of students) challenges
Fear of the unknown or unexpected	• Anxiety about uncertain scenarios • Unexpected events during the OSCE[a]
Self-consciousness and insecurities	• Comparing oneself to peers • Heightened anxiety when observed and fear of failure
Knowledge of the exam environment	• Importance of knowing the exam room ahead of time • Useful for the VR[b] exam setting to match the exam room layout
Practice should mimic the actual exam	• Formal note-taking aligned with the OSCE format • Graded mock OSCE sessions to reduce anxiety
Student grade mindset	• Viewing the OSCE as “just another test” vs a major assessment • Impact of learning vs performance focus
Time constraint	• Rushed feeling during the written component • Differences between class practice time and OSCE format
Peer influences	• Social atmosphere in group practice as exams drew near • Peer influence on VR uptake (encouraging or discouraging)
Minor themes	
Framing	• Impact of how the faculty present the OSCE • Mentally reframing the exam (game-like vs high stakes)
Pace of practice laboratories	• Laboratories perceived as too fast • Inadequate time to practice new skills
Inconsistencies	• Contradictory instructions in coursework • VR avatar responses that do not align with questions
OSCE not realistic	• Delays in standardized patient responses • Exam setting differs from real-world clinical practice

[a]OSCE: objective structured clinical examination.

[b]VR: virtual reality.

Major Themes

Experience and Background

Most students indicated that previous experience and background were factors influencing OSCE performance. Students said that having experience with clinical exams would lessen exam stress. As one student said, “I think the reason that the OSCE is quite stressful is because many of us have not done a practical exam like this.” Other students said that their comfort with the OSCE was influenced by their previous education,

XSL•FO
RenderX

work experience, or personality. For instance, students said that "having previous knowledge or possessing a degree in anatomy helped [us] prepare for the physical assessment more than students without." There were students who were more confident with the history taking because of "the previous work experiences and interactions that [they] have had." Regarding personality, one student believed that she did well in the history taking because she was naturally "a talkative, sociable person," whereas another student felt that she would struggle because she was "not hugely a people person." A student remarked that people with gaming experience may help them appreciate the VR simulation.

Need for Feedback and Support

Another consistent theme was the mention that any practice sessions in which students received feedback were especially helpful, such as the practice history-taking sessions with the senior students and the mock OSCEs in which instructors graded students using the marking rubric. In class, some students felt that the class sizes made it more difficult to receive feedback:

> *...to get feedback was not always easy because there is only one instructor to how many students?*

Due to the lack of immediate feedback, a few students expressed that it may be beneficial to be paired with a laboratory partner who has more experience and knowledge on physical assessments. This way, novices could receive quality feedback when needed as opposed to practicing with someone else who is not as familiar with the content. Overall, students said that having access to timely feedback and support played a main role in their sense of preparedness. Regarding the VR simulation, a focus group unanimously agreed (with nodding and affirmative gestures) that the VR modules provided feedback in an indirect manner:

> *VR made us be more aware to ask simpler questions in the OSCE, because the VR would glitch-up if we talked too much.*

When probed about this further, students said that the glitches would often result in the virtual patient saying an off-topic remark, such as "I like to play ice-hockey!" One student said that she found the VR to be stress relieving because the glitches were funny and made her laugh. Most VR students recommended that the simulation be upgraded to provide immediate feedback.

Fear of the Unknown or Unexpected

Another theme that emerged involved students mentioning feeling anxious and nervous about the OSCE because they did not know what to expect. One of the main sources of stress was not knowing which scenario they were going to be assigned ahead of time. As one student said, "I dreaded the scenario selection because you didn't know what you were going to get." Other students felt stressed because they were unfamiliar with the clinical exam process, such as not knowing the test environment or procedure. Notably, several students had unexpected events happen during their OSCEs, which contributed to their sense of stress. For example, one student's client was an instructor she knew, whereas another student said that "my client kept looking at the evaluator."

Self-Consciousness and Insecurities

Another theme that emerged involved students' feelings of self-consciousness, insecurity, and stress, especially when they compared themselves to their classmates. These students admitted to forming comparisons without regard to study progress, knowledge, and competency. Some students felt intimidated when they saw that their classmates had started preparing for the OSCE before them:

> *...it made me anxious to know how early the people were starting* [to study]

Others felt incompetent and doubted their ability to pass the test. Notably, for one participant, the stress from comparing herself to her peers was even greater than the actual OSCE. As a subtheme, other participants said that their self-consciousness manifested whenever they felt fear of being observed and the fear of failure. A few students feared being watched because they associated those situations with judgment (ie, being evaluated). These students worried that they would be deemed incompetent. Some students had fears of failing the OSCE and what others would think of them. One student described the following:

> *I didn't want to fail because I didn't want to look incompetent to my peers, more so than me actually caring for my own sake.*

Another student made a similar remark:

> *...no one wanted to be known as the person who failed it.*

Knowledge of the Exam Environment

Another consistent theme was the importance of having knowledge of the OSCE room before the exam. These students believed that it would be helpful to be shown the exam room ahead of time, whether through a tour or through practice sessions in it. Specifically one student stated they would be "more thrown off by the foreign environment." Another student made a similar remark, stating that the move from a classroom (practice) to a clinical (exam) setting was intimidating. As one student described, the differences between the room where she conducted her practice interview with second-year students and the actual exam room initially "threw me off." In total, 8% (2/25) of the students said that they had difficulty operating the plinths or locating tools in the actual OSCE room. One student said that it would be useful to see the environment ahead of time so that she could visualize the room and rehearse. Several students recommended that simulated exam rooms mimic the actual test environment as closely as possible. Regarding the VR simulation, many students said that VR helped them visualize what the actual OSCE would look like. One student remarked the following:

> *I found it helpful. [It was] my first VR experience. I got to see what the layout of the room would be like.*

Practice Should Mimic the Actual Exam

Another major theme that emerged related to practicing for the OSCEs. The students recommended that their practice should closely mimic the actual exam as much as possible. With regard

XSL•FO
RenderX

to the standardized patients, some students believed that practicing with classmates was less useful because "there is an expectation there where if I give [a classmate] a little prompt, they should be doing what I asked. Whereas if you do it with a patient, they don't know what you are talking about unless you give them full instructions." One student remarked that there were few male students in the course, which resulted in increased feelings of nervousness during the OSCE because her practice was only on female students. Several students remarked that they did not have enough practice with formal documentation of clinical notes in class. They further mentioned that note-taking was "different" between in-class and exam sessions:

> *In the OSCE, [we] have 30 minutes for two notes and three questions.*

Some students remarked that practice notes, done in laboratories, did not realistically fulfill the expectations of the OSCE. Students found practice sessions useful when instructors would grade them based on the rubric. One of the students remarked that, after she did a practice OSCE in front of the class and was marked on it, she felt more relieved because "you realize there's room to make mistakes." Regarding the VR simulation, one student pointed out that *VR gave us different ways to ask questions.* Another student remarked the following:

> *I think it would be helpful to reiterate the fact that VR is a different way to practice for the OSCE.*

Student Grade Mindset

Students' mindsets regarding the OSCE impacted their perceived anxiety and response. Some students mentally *equalized the weight* of the OSCE by saying the following:

> *I think anything [worth] over 10% as the same in my world. You have to treat it just like any other test.*

These thought processes were admitted by students to potentially reduce their anxiety about the OSCE. In contrast, others viewed it more as a learning opportunity:

> *I've viewed it just as a learning experience. I don't think I really look at [the OSCE] as the 50% or maybe the 30% for human systems or whatever.*

Some students agreed that viewing the OSCE as only a learning experience felt less distressing to them, saying the following:

> *...if I did something poor enough to actually fail, then I hope I would get corrected and learn how to do it right.*

In contrast, one student said that her peers were "...all super high achieving, wanting to be the best kind of student," which led to her feeling scared about the possibility of failing. Regarding the VR simulation, one student remarked the following:

> *I would try any practice method presented to me. I think VR helped me chill out a little. I think in the future, in general, I'll try to remember this.*

Although the students' mindsets impacted their perceived anxiety, when they focused on the grades attached to the OSCE, they admitted to having potentially greater anxiety levels than those who viewed the OSCE as a learning activity only. Students admitted that the VR simulation may have helped them reduce their anxiety as it functioned as a stand-alone learning activity in this manner.

Time Constraint

A common theme reported by students was admitting to being under pressure to complete the OSCE on time, which felt rushed. In particular, many students found the written component to be the most fast-paced. For instance, one participant said the following:

> *I ran of time, when I was answering the questions, so I knew my answers sucked...[I] had two minutes to answer all three questions so I just wrote whatever came into my mind.*

A significant number of students felt that the time allotted was not enough given what they were expected to complete. One student asked the following:

> *...how can we write two notes and answer the questions in twenty minutes?*

The students also mentioned that the time on the OSCE was not similar to that in the practice sessions they had received in class. Notably, a participant expressed the following:

> *...in class, we have twenty minutes to write just a SOAP note and you are still feeling rushed. Whereas in the OSCE, you have about fifteen-ish minutes to do each.*

While one student mentioned running out of time during the physical check, most remarks on time constraints were about the written test.

Peer Influences

A theme emerged regarding the perceived change in the social atmosphere as the appointed time of the OSCE drew near, with statements such as the following:

> *...the vibe of the class nearing the OSCE made it more stressful.*

One student said the following:

> *...being in the group setting while everyone was practicing, you got that very nervous environment going on and everyone's asking questions about how you do this and that. I thought that this was spreading fear in our class.*

One student said that peer pressure was more anxiety inducing than the upcoming OSCE itself. She remarked the following:

> *...it wasn't even the OSCE [causing anxiety], because as long as you are practicing and you prepared yourself, I felt it was okay. But it was the pressure from everyone else to be like we've got to do this, this, and this.*

Another student said that peer perception added to anxiety, leading to some students avoiding their peers altogether. Despite mentions of negative peer influences, students reported that their classmates were also helpful in offering assurances and knowledge. Students were influenced by their peers to attempt

XSL•FO
RenderX

or refuse the VR simulation based on how effective they believed it was. For example, one student said the following:

> *Friends that did VR were suggesting people to not take VR. After hearing [about] other people's experiences, some decided to back out.*

Minor Themes

Framing

An additional theme emerged indicating how the OSCE could have been framed differently by the faculty. Students suggested that the instructors could have introduced the OSCE in a less anxiety-provoking way. Students mentioned that bringing up the OSCE on the first day of class in the academic year for first-year students caused them stress. One student mentioned the following during a focus group (acting in panic to represent their mood at the time): "What is this?" and "How is this half of my mark?" Students suggested that the faculty could make the OSCE less intimidating by changing the wording to make exams sound less important (eg, calling a midterm a quiz) and avoiding the use of acronyms. In total, 8% (2/25) of the students said that they mentally reframed the OSCE, lessening their anxiety over it. Regarding the VR simulation, one student pointed out the following:

> *...for some people, seeing the OSCE as a video-game or something fun to engage with, rather than all this pressure...it could create a shift in some people.*

Pace of Practice Laboratories

Students also felt that the practice laboratories were too fast, which failed to allow students adequate time to learn and practice the content:

> *The speed of the content [coursework] was a problem.*

Students in both interviews and focus groups mentioned that the material in the laboratory was taught too fast, which made it harder for them to learn and remember. These students said that they did not feel as if they had adequate time for both partners to practice a task, leaving some students confused after the laboratories. One student remarked the following:

> *...labs in general always felt kind of rushed, so when we practiced, we'd only get fifteen minutes and then, okay, we are starting [to learn new content] again.*

Inconsistencies

Other concerns raised by students that added to their stress and anxiety were being given inconsistent information. According to these students, there were inconsistencies in both course delivery and exam marking. While they acknowledged that such inconsistencies were eventually cleared up, one felt *stressed and confused by the ordeal*. One student added the following:

> *...having to re-learn a ROM or MMT movement, due to being told incorrect information, was highly inefficient.*

Regarding the VR simulation, inconsistencies involving the virtual patient's responses were also reported:

> *I asked the VR patient if he was a father and he responded by saying "not really." And then as time went on, I figured out he had two children, so yeah, he was a father.*

OSCE Not Realistic

Another theme that emerged was that students felt that, while the actual OSCE was designed to represent a real-world clinical exam, it did not accurately reflect everyday practice. During the physical check on a standardized patient, one student thought that looking for a single impaired movement was unnatural because rarely would an injury only affect 1 movement in real life. Students also noted a "lag time between asking a question and waiting for the patient to reply." Another 8% (2/25) of the students commented on the time limit for the interview:

> *[the] timeframe threw me off a bit—it seemed so inorganic. Like, it won't be like that in real life, probably.*

Regarding the VR simulation, one user said the following:

> *I think the simulation did not accurately reflect what it was going to be in the [actual] OSCE. Even if it did, every OSCE is different, so it may not help so much for the future.*

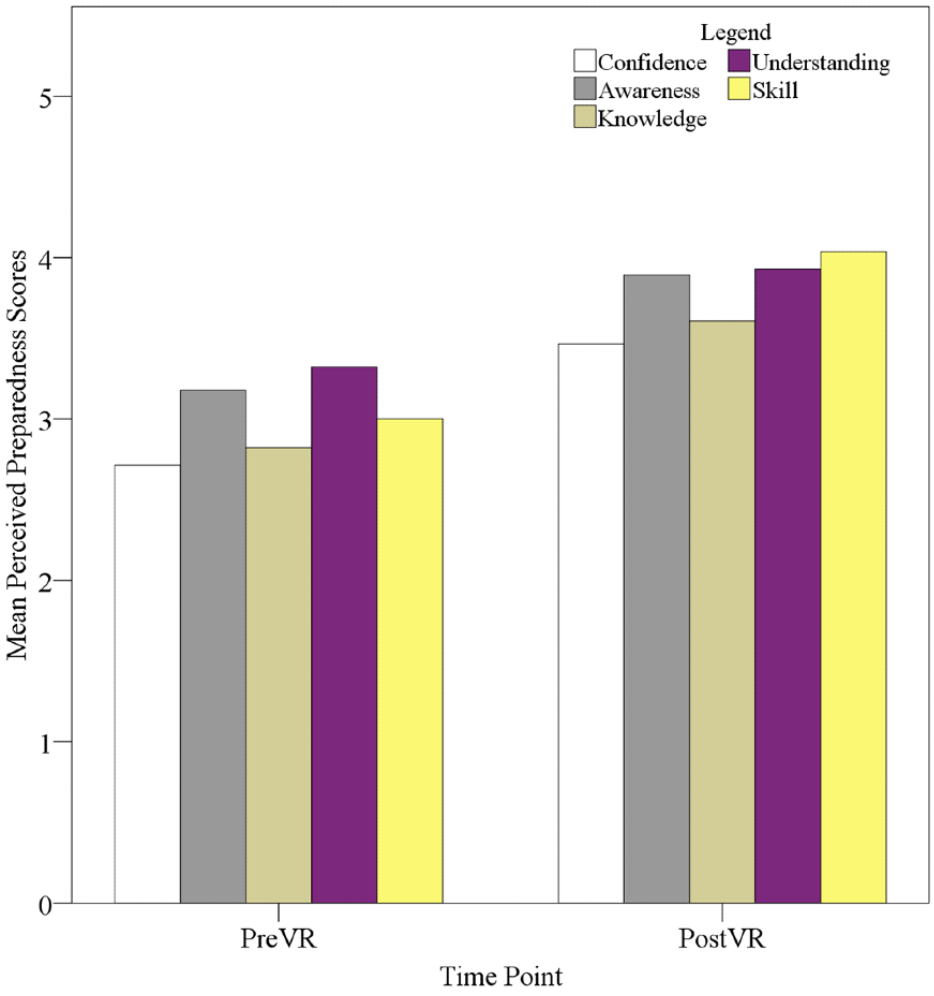
Secondary Outcome Measures

Perceived Preparedness Inventory

Figure 3 shows student Perceived Preparedness Inventory scores for student confidence, awareness, knowledge, understanding, and skill levels. There was a significant increase in all variables from before to after VR. It should be noted that the presentation of scores in this report for perceived preparedness was arranged with higher values representing greater mean levels of student-reported ability, unlike the traditional scoring of this instrument, which represents lower scores as being better. There was a significant difference in perceived confidence between mean scores before VR, which showed the lowest values (mean 2.71, SD 0.94) out of all variables, and mean scores after VR (mean 3.46, SD 1.00; t_{27}=2.68; *P*=.01; Cohen *d*=0.77), which were also the lowest of all variables at that time point. The mean difference was 0.75 units (95% CI 0.18-1.32). Perceived awareness showed a significant difference from before VR (mean 3.18, SD 1.16) to after VR (mean 3.89, SD 0.86; t_{27}=2.68; *P*=.01; Cohen *d*=0.70), with the mean difference being 0.71 units (95% CI 0.17-1.26). Perceived knowledge scores showed a significant mean increase from before VR (mean 2.82, SD 0.82) to after VR (mean 3.61, SD 1.03; t_{27}=3.16; *P*=.004; Cohen *d*=0.85), with the mean difference being 0.79 units (95% CI 0.28-1.30). Of all the variables, understanding showed the highest mean scores before VR (mean 3.32, SD 0.90) and the second-highest mean scores after VR (mean 3.93, SD 98; t_{27}=2.62; *P*=.01; Cohen *d*=0.65). The mean difference was 0.61 units (95% CI 0.13-1.08). Finally, student-perceived skill showed the greatest significant increase from before VR (mean 3.00, SD 0.98) to after VR (mean 4.04, SD 0.96; t_{27}=4.16; *P*<.001; Cohen *d*=1.07) and also showed the highest mean scores at that time point. The mean difference was 1.04 units (95% CI 0.52-1.55).

XSL•FO
RenderX

Figure 3. Student perceived preparedness scores. PreVR: measure taken 3 weeks before virtual reality simulation; PostVR: measure taken after virtual reality simulation.



OSCE Scores

There was no significant difference in mean OSCE scores between the NoVR group (history taking: mean 86.38, SD 6.53; cognitive assessment: mean 88.47, SD 8.04; course total: mean 90.03, SD 4.26) and the YesVR group (history taking: mean 87.25, SD 7.64; cognitive assessment: mean 90.07, SD 5.88; course total: mean 90.25, SD 4.92; history taking: t_{58}=0.48, *P*=.63, and Cohen *d*=0.12; cognitive assessment: t_{58}=0.87, *P*=.39, and Cohen *d*=0.23; course total: t_{58}=0.19, *P*=.85, and Cohen *d*=0.05). However, it is worth noting that the YesVR group had higher mean OSCE scores across all scoring categories, although the differences were not statistically significant when compared to the NoVR group.

Discussion

Principal Findings

The integration of qualitative and quantitative findings provided a more comprehensive understanding of students' experiences. While STAI, TAI, and perceived preparedness scores quantified student anxiety and preparedness levels, the interviews and focus groups offered insights into underlying factors. The theme of *knowledge of the exam environment* suggested that VR simulation may have familiarized students with the OSCE room, helping reduce uncertainty about the exam setting—one of the major themes linked to anxiety. While this study was not a randomized controlled trial, the mixed methods approach allowed for a deeper exploration of how students perceived the role of VR in their exam readiness, highlighting its potential as a preparatory tool.

The data showed that students who used the VR simulation had lower state anxiety scores compared to those who did not, but without a controlled study design, a causal relationship cannot be confirmed. Student interviews provided additional context that may help explain this observed difference. Intolerance to uncertainty is a significant risk factor for increasing anxiety symptoms [40]. However, *knowledge of the exam environment* was a recurring theme among the students, who mentioned that seeing and experiencing the exam room in advance was useful. Students reported rethinking their study strategies and adjusting their expectations accordingly, although it is unclear whether these changes were directly due to the VR experience. It could be that the small moment spent contemplating the exam setting outside of VR was enough to initiate their expectation adjustments. Regarding recurrent self-focused worry, *self-consciousness and insecurities* and *peer influences* were additional themes mentioned by students during the interviews, stating that their worries were caused by thoughts of inadequate

XSL•FO
RenderX

performances while being watched. When practicing in real-world social situations, the students were aware of their potential mistakes and criticisms becoming known by peers. Student remarks such as "I didn't want to fail because I didn't want to look incompetent to my peers..." reflected this. While it was mentioned that having peer feedback provided assurances and important ideas for learning, some students said that they felt anxious when peer influences instilled a "we have to do this, this and this" mentality. Having the option to practice in a private VR simulation allowed students to make mistakes with no perceived consequences, potentially allowing them to hone their practice strategies and make mistakes without worry. The VR simulation may have offered a break from demanding peer objectives as they had a learning environment in which they could set the pace for themselves. Depending on how well they thought the VR simulation worked, students were persuaded by their peers to either participate or avoid the self-serve use of VR. The VR simulation was announced solely as a tool for potentially decreasing student anxiety—it was not advertised to raise students' test scores. Some students may have perceived the VR system to be a less valuable source of learning after being informed that it was not designed to boost their grades. Examining student OSCE scores for this experiment revealed no differences between students who used VR and those who did not. The relationship between learning technology resources and student achievement is poorly understood and challenging to evaluate. The VR simulation would need to be significantly improved if the students' only goal was to use it to raise their OSCE scores. Although there was a choice between 2 separate modules in the VR simulation, the system was not particularly thorough. Students could pose basic questions or try conducting a cognitive assessment on a virtual patient, but there was no feedback or scoring system. Such a system should be grounded in behavioral learning principles, include a wide array of patient scenarios, and provide scoring and reinforcement cues to help students understand how to improve their OSCE performance.

Regarding the handling of associated symptoms of anxiety leading to overanalyzing thought processes, the student themes of *feedback and support* and *time constraint* showed that students reported adjusting their questioning strategies while using the VR simulation, suggesting a possible link between VR sessions and changes in how students posed questions. The VR system glitched if students spoke with excessive dialogue. However, students noted that simplifying their questions to short, concise phrases seemed to boost their attentional focus. Some suggested that this helped them avoid overanalyzing and feeling anxious, although whether these outcomes resulted specifically from VR remains unclear. Greater anxiety levels are associated with overactivation of orienting processes, resulting in decreased capacity for attentional regulation [41]. However, the factors and directionality between manifested anxiety symptoms and attention are largely unknown [42]. Regarding the handling of rumination thought processes, themes of *framing* and *student grade mindset* emerged from the interviews and focus groups. When students adopt a positive and learnable growth mindset, there are perceptual changes in student aptitude, reducing anxiety about the course material [43]. The students in this investigation spoke of mentally changing the severity of the OSCE by saying either "the OSCE is only a learning experience" or "just pretend it's a game," which is different from learning growth mindsets. This shift in framing may have freed up students' attentional resources, enabling them to focus more effectively on the immediate task rather than on future outcomes, potentially reducing rumination associated with test anxiety. However, the impact of a student's *gaming mindset* on test anxiety cannot be conclusively determined from this study.

The TAI was not found to be different between the NoVR and YesVR groups. This suggests a similarity in general test anxiety traits between the groups. The TAI was created as a tool to assess test anxiety as a *situation-specific personality trait* according to the manual by Spielberger et al [33]. The study by Concannon et al [1] found this exact outcome between the NoVR and YesVR groups throughout multiple time points across the student curriculum. Consistent with trait test-anxiety theory [5], students with higher trait test-anxiety scores exhibit a stable, cross-context anxiety response that surfaces in nearly all evaluative situations. In a report by Sargolzaei et al [44], a 21-minute, 360° instructional VR video for heart surgery students was able to reduce trait anxiety compared to a control group [44]. For the reduction of trait anxiety in internship settings, immersive VR interventions are recommended to be supplemented with instructional audio commentary and augmented reality notes [44]. For the reduction of trait anxiety in other anxiety-related disorders such as social anxiety, meta-analysis reports recommend using sophisticated VR exposure therapy across 4 to 14 sessions, each lasting 60 to 90 minutes [45].

Between 3 weeks and 3 days before the OSCE, all variables of perceived preparedness among YesVR students increased. A similar trend was reported in the study by Massey et al [11], which used video OSCE exemplars as blended learning materials for nursing students. They reported that a reduction in student stress can be attained by improving preparation and understanding of assessment expectations [11,46]. Of all the perceived preparedness variables, student-perceived confidence levels were found to be the lowest at both the beginning and the end of the OSCE preparation timeline despite showing a significant improvement (P=.01). Previous reports have stated that generation Y OT students (those born in the 1980s) exhibited overconfidence behaviors that reduced their openness to feedback from educators [47]. However, our investigation found student self-reported levels of confidence to be the lowest of all perceived preparedness variables, with *feedback and support* being a recurring theme mentioned by students to be important for their confidence development. This discrepancy could be due to a couple of factors. First, the students in our investigation reported that the large class sizes made it difficult to receive feedback. Increased class sizes and teaching workloads in OT means that new technology is recommended to be implemented in student curriculums with around-the-clock access [48-51]. This technology is recommended to provide feedback specific to each student, filling gaps in their knowledge based on performances in evaluative situations. Second, this investigation's Perceived Preparedness Inventory measured OT student confidence of a specific type—perceived confidence in performance in an upcoming OSCE. The specificity of this

XSL•FO
RenderX

confidence type is crucial. For example, a report by DeCleene Huber et al [52] found that increased clinical experience in OT students made an impact on how confident they felt when using evidence-based practice as future practitioners. Without this experience, occupational therapists would be less likely to perform evidence-based practice in clinical environments. However, the report by DeCleene Huber et al [52] also found that OT students showed less confidence when evidence-based practice skills were used in association with different tasks, such as statistical procedures and tests. Student confidence, knowledge, and skills have shown to be improved through motivational interviews and instruction [53]. Motivational interviews contain technical components (ie, open questioning, affirmation, reflective listening, and summarizing aspects) and relational components (ie, collaboration, evocation, autonomy support, and acceptance) [53]. One issue with these motivational interviewing sessions is that they typically have limited availability. However, immersive VR modules that offer simulated patient encounters could support interactive, experiential motivational interviews around the clock, which may enhance student perceived preparedness.

For OT student-perceived understanding, the *American Journal of Occupational Therapy* (2021) created a curricular design framework that recommended using formative assessment to measure student understanding while providing course adjustments and feedback throughout course progression [54]. These formative assessments can be in the form of simple, 1-minute reflections with *nonjudgmental* characteristics, allowing students to correct mistakes [55]. Activity resources reportedly improve student awareness levels, with technology playing a role in the form of web-based polls, student-uploaded pictures, videos, and comments on discussion forums [54]. The Perceived Preparedness Inventory was intended to probe respondents' knowledge, understanding, awareness, skill level, and confidence surrounding OSCEs. The impact of our investigation's immersive VR simulation was not specified in the Perceived Preparedness Inventory questions. Regardless of whether the students used the immersive VR simulation, it is possible that they could still experience higher perceived preparedness for their upcoming OSCE. The study by Concannon et al [1] found that student academic self-efficacy levels continuously increased throughout their first-year OT coursework regardless of whether the students had used VR simulation. Thus, it was beneficial to understand the degree to which each perceived preparedness parameter improved in a general sense.

Future Recommendations

New cutting-edge components are available to enhance our investigation's VR simulation and its general capabilities. However, adding new features based on student feedback from this investigation may enhance a student's sense of immersion. The GPT AI can be upgraded to the latest version (ie, GPT-4o [omni] at the time of this writing) and trained on future recorded real-time interactions between student and patient actors. This would likely improve virtual patient response continuity for specific scenarios during the History Taking module. Regarding the reported lag time between student questions and virtual patient responses, using GPT-4o may reduce delays, although some models with simpler reasoning capabilities can generate replies more quickly. If used with a portable immersive VR headset such as the VIVE XR Elite, the simulation would be accessible around the clock in a portable manner with improved sound definition. Other technical improvements such as making the avatars more realistic (eg, rendered from Unreal Engine MetaHumans) while improving the animations of their movements may further support students' sense of immersion. A greater variety of selectable avatars, including those from different cultures and backgrounds, might broaden student perspectives of real-world clientele while promoting diversity and inclusion.

Students were concerned about their training mistakes being known by their peers. It would be important for future VR learning environments to use disclaimers, informing users that their performances will not be recorded or used for assessment. From an ethical standpoint, if the VR simulation uses student interactions with virtual patients to train the AI system, this should be clearly informed to the student, and the ability to opt out should be offered. Overall, students said that their sense of preparedness for the OSCE was affected by their ability to access timely feedback and support during their clinical skill practice. This investigation found that, when students questioned the virtual patients during the History Taking module, it was beneficial to know when such questions became excessively worded. GPT is capable of providing condensed suggestions for such questions, which could be used to inform the student. Formative learning aspects of the simulation could also be improved to include AI assessment of interactions with the virtual patient, allowing the system to provide student feedback on the completeness of questioning procedures and evaluation of their bedside manner.

Programming virtual environments can be a costly and time-consuming process. However, if a module allowed the VR simulation to have its testing environments adjusted by the users themselves (ie, allow for positional adjustment of apparatuses such as desks and plinths), this would result in virtual test settings better representing those experienced in the real world. This would adhere to student preferences, who desire the practice environment to be similar to that of the actual OSCE.

Strengths and Limitations

This investigation did not use a double-blind, randomized controlled trial; thus, causation of immersive VR's ability to reduce student anxiety cannot be inferred. Conducting a randomized controlled trial could have been challenging as it may have disrupted the student study routines while potentially segregating them from their study partners. In addition, this investigation could not match student responses to demographic factors such as sex. However, this study was able to test students from the same academic year, potentially reducing group differences due to student admission requirements.

The overall response rate was relatively low. The respondents may not represent all enrolled students as those who chose to participate in the VR simulation might have been more open minded, performance focused, or generally eager to try new study methods, potentially limiting the broader applicability of these findings. Despite these limitations, the interview and focus

XSL•FO
RenderX

group data offered meaningful context for the quantitative data, supplemented by student recommendations to understand their perceptions of immersive VR's role in first-year health sciences coursework. Our qualitative aim was not to produce generalizable findings but, rather, to gain a deeper understanding and insight into student perspectives. We used focus groups and interviews to explore and generate themes; participants reiterated similar thematic categories, suggesting the data were approaching saturation; however, we do not claim full saturation, as additional participants or new cohorts might still reveal additional themes.

Conclusions

Immersive VR simulation may help first-year OT students lower their state anxiety when preparing for clinical practical exams. Students reported feeling more oriented to the exam setting and found that VR sessions allowed them to practice questioning strategies in a low-stakes environment. One key theme was the need for additional practice on new exam components and student performance feedback. Future enhancements might incorporate automated scoring or simplified interview prompts for the students. Overall, VR could provide a safe environment for practice and feedback, potentially easing students' anxiety in clinical test settings.

Acknowledgments

The authors wish to acknowledge the contributions of the Rehabilitation Robotics Laboratory at the Edmonton Clinic Health Academy, which developed the objective structured clinical examination simulation as described in this paper. This study was supported by the Teaching and Learning Enhancement Fund, University of Alberta.

Data Availability

The relevant quantitative datasets (State-Trait Anxiety Inventory, Test Anxiety Inventory, and perceived preparedness scores) generated or analyzed during this study are available in Multimedia Appendices 5, 6, and 7, respectively.

Although objective structured clinical examination score data were collected and assessed, releasing the full dataset could compromise student identity by making it possible to match specific scores to individuals. Therefore, these data are not publicly available but can be obtained from the corresponding author upon reasonable request.

The qualitative interview and focus group datasets are not publicly available because they contain researcher identifiers (eg, usernames and personal anecdotes) and fully deidentifying them for an open repository is unfeasible. However, key excerpts and themes have been included in the manuscript, and the underlying data are available from the corresponding author upon reasonable request.

Authors' Contributions

BC and SE coconceptualized this investigation. SE contributed to the vision design of the clinical practical simulation. BC outlined the psychometric tool selection, data intake, and analysis process. BC led the manuscript writing process. SE contributed to the writing process and revisions. Both authors approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview guide.
[DOCX File , 18 KB - games_v13i1e69428_app1.docx]

Multimedia Appendix 2
Focus group guide.
[DOCX File , 18 KB - games_v13i1e69428_app2.docx]

Multimedia Appendix 3
Perceived Preparedness Inventory.
[DOCX File , 13 KB - games_v13i1e69428_app3.docx]

Multimedia Appendix 4
Saint Louis University Mental Status Exam.
[PDF File (Adobe PDF File), 80 KB - games_v13i1e69428_app4.pdf]

XSL•FO
RenderX

Multimedia Appendix 5
State-Trait Anxiety Inventory raw data.
[XLSX File (Microsoft Excel File), 9 KB - games_v13i1e69428_app5.xlsx]

Multimedia Appendix 6
Test Anxiety Inventory raw data.
[XLSX File (Microsoft Excel File), 10 KB - games_v13i1e69428_app6.xlsx]

Multimedia Appendix 7
Perceived preparedness raw data.
[XLSX File (Microsoft Excel File), 10 KB - games_v13i1e69428_app7.xlsx]

References

1. Concannon BJ, Esmail S, Roduta Roberts M. Immersive virtual reality for the reduction of state anxiety in clinical interview exams: prospective cohort study. JMIR Serious Games 2020 Jul 09;8(3):e18313 [FREE Full text] [doi: 10.2196/18313] [Medline: 32673223]
2. Perrotta G. Anxiety disorders: definitions, contexts, neural correlates and strategic therapy. J Neur Neurosci 2019;6(1):042 [FREE Full text]
3. Spielberger C, Gorsuch R, Lushene R, Vagg P, Jacobs G. State-trait anxiety inventory for Adults™ (forms Y1 and Y2). Mind Garden. 2015. URL: https://www.mindgarden.com/145-state-trait-anxiety-inventory-for-adults [accessed 2024-11-29]
4. Hong E, Karstensson L. Antecedents of state test anxiety. Contemp Educ Psychol 2002 Apr;27(2):348-367. [doi: 10.1006/ceps.2001.1095]
5. Woldeab D, Brothen T. 21st Century assessment: online proctoring, test anxiety, and student performance. Int J E Learn Distance Educ 2019;34(1):1 [FREE Full text]
6. Sommer M, Arendasy ME. Further evidence for the deficit account of the test anxiety–test performance relationship from a high-stakes admission testing setting. Intelligence 2015 Nov;53:72-80. [doi: 10.1016/j.intell.2015.08.007]
7. Roos AL, Goetz T, Voracek M, Krannich M, Bieg M, Jarrell A, et al. Test anxiety and physiological arousal: a systematic review and meta-analysis. Educ Psychol Rev 2020 Aug 16;33(2):579-618. [doi: 10.1007/S10648-020-09543-z]
8. Conijn R, Kleingeld A, Matzat U, Snijders C. The fear of big brother: the potential negative side‐effects of proctored exams. Comput Assist Learn 2022 Feb 08;38(6):1521-1534. [doi: 10.1111/jcal.12651]
9. DordiNejad FG, Hakimi H, Ashouri M, Dehghani M, Zeinali Z, Daghighi MS, et al. On the relationship between test anxiety and academic performance. Procedia Soc Behav Sci 2011;15:3774-3778. [doi: 10.1016/j.sbspro.2011.04.372]
10. Brand HS, Schoonheim-Klein M. Is the OSCE more stressful? Examination anxiety and its consequences in different assessment methods in dental education. Eur J Dent Educ 2009 Aug 15;13(3):147-153. [doi: 10.1111/j.1600-0579.2008.00554.x] [Medline: 19630933]
11. Massey D, Byrne J, Higgins N, Weeks B, Shuker M, Coyne E, et al. Enhancing OSCE preparedness with video exemplars in undergraduate nursing students. A mixed method study. Nurse Educ Today 2017 Jul;54:56-61 [FREE Full text] [doi: 10.1016/j.nedt.2017.02.024] [Medline: 28477564]
12. Pascoe MC, Hetrick SE, Parker AG. The impact of stress on students in secondary school and higher education. Int J Adolesc Youth 2019 Apr 11;25(1):104-112. [doi: 10.1080/02673843.2019.1596823]
13. Vanstone DM, Hicks RE. Transitioning to university: coping styles as mediators between adaptive-maladaptive perfectionism and test anxiety. Pers Individ Dif 2019 Apr;141:68-75. [doi: 10.1016/j.paid.2018.12.026]
14. Shangraw AM, Silvers J, Warholak T, Vadiei N. Prevalence of anxiety and depressive symptoms among pharmacy students. Am J Pharm Educ 2021 Feb;85(2):8166 [FREE Full text] [doi: 10.5688/ajpe8166] [Medline: 34283739]
15. Arria AM, O'Grady KE, Caldeira KM, Vincent KB, Wilcox HC, Wish ED. Suicide ideation among college students: a multivariate analysis. Arch Suicide Res 2009;13(3):230-246 [FREE Full text] [doi: 10.1080/13811110903044351] [Medline: 19590997]
16. Eisenberg D, Golberstein E, Gollust SE. Help-seeking and access to mental health care in a university student population. Med Care 2007 Jul;45(7):594-601. [doi: 10.1097/MLR.0b013e31803bb4c1] [Medline: 17571007]
17. Scheffler RM, Qazi H, Vora N. Anxiety disorder on college campuses: the new epidemic. Berkeley Public Policy. 2019. URL: https://gspp.berkeley.edu/assets/uploads/page/Anxiety_Disorder_on_College_Campuses_UCB_Study_FINAL.pdf [accessed 2024-11-29]
18. Cynthia IN. Exam anxiety, academic burnout, and cheating behavior an empirical study of undergraduates in enugu state. J Adv Res Soc Sci Humanit 2022;8(5):11-16. [doi: 10.53555/nnssh.v8i5.1267]
19. Dugas MJ, Gagnon F, Ladouceur R, Freeston MH. Generalized anxiety disorder: a preliminary test of a conceptual model. Behav Res Ther 1998 Feb;36(2):215-226. [doi: 10.1016/s0005-7967(97)00070-3] [Medline: 9613027]

XSL•FO
RenderX

20. Huntley C, Young B, Tudur Smith C, Jha V, Fisher P. Testing times: the association of intolerance of uncertainty and metacognitive beliefs to test anxiety in college students. BMC Psychol 2022 Jan 05;10(1):6 [FREE Full text] [doi: 10.1186/s40359-021-00710-7] [Medline: 34986890]
21. Freitas JR, Velosa VH, Abreu LT, Jardim RL, Santos JA, Peres B, et al. Virtual reality exposure treatment in phobias: a systematic review. Psychiatr Q 2021 Dec 26;92(4):1685-1710. [doi: 10.1007/s11126-021-09935-6] [Medline: 34173160]
22. Bouchard S, Dumoulin S, Robillard G, Guitard T, Klinger É, Forget H, et al. Virtual reality compared with exposure in the treatment of social anxiety disorder: a three-arm randomised controlled trial. Br J Psychiatry 2017 Apr;210(4):276-283. [doi: 10.1192/bjp.bp.116.184234] [Medline: 27979818]
23. Cant R, Cooper S, Sussex R, Bogossian F. What's in a name? Clarifying the nomenclature of virtual simulation. Clin Simul Nurs 2019 Feb;27:26-30. [doi: 10.1016/j.ecns.2018.11.003]
24. Kardong-Edgren S, Farra SL, Alinier G, Young HM. A call to unify definitions of virtual reality. Clin Simul Nurs 2019 Jun;31:28-34. [doi: 10.1016/j.ecns.2019.02.006]
25. Concannon BJ, Esmail S, Roduta Roberts M. Head-mounted display virtual reality in post-secondary education and skill training. Front Educ 2019 Aug 14;4:33 [FREE Full text] [doi: 10.3389/feduc.2019.00080]
26. Yip BC, Man DW. Virtual reality-based prospective memory training program for people with acquired brain injury. NeuroRehabilitation 2013 Feb 28;32(1):103-115. [doi: 10.3233/NRE-130827] [Medline: 23422463]
27. Grealy MA, Johnson DA, Rushton SK. Improving cognitive function after brain injury: the use of exercise and virtual reality. Arch Phys Med Rehabil 1999 Jun;80(6):661-667. [doi: 10.1016/s0003-9993(99)90169-7] [Medline: 10378492]
28. De Luca R, Maggio MG, Maresca G, Latella D, Cannavò A, Sciarrone F, et al. Improving cognitive function after traumatic brain injury: a clinical trial on the potential use of the semi-immersive virtual reality. Behav Neurol 2019;2019:9268179 [FREE Full text] [doi: 10.1155/2019/9268179] [Medline: 31481980]
29. Man DW, Poon WS, Lam C. The effectiveness of artificial intelligent 3-D virtual reality vocational problem-solving training in enhancing employment opportunities for people with traumatic brain injury. Brain Inj 2013;27(9):1016-1025. [doi: 10.3109/02699052.2013.794969] [Medline: 23662639]
30. Kyaw BM, Saxena N, Posadzki P, Vseteckova J, Nikolaou CK, George PP, et al. Virtual reality for health professions education: systematic review and meta-analysis by the digital health education collaboration. J Med Internet Res 2019 Jan 22;21(1):e12959 [FREE Full text] [doi: 10.2196/12959] [Medline: 30668519]
31. Donnelly MR, Reinberg R, Ito KL, Saldana D, Neureither M, Schmiesing A, et al. Virtual reality for the treatment of anxiety disorders: a scoping review. Am J Occup Ther 2021 Nov 01;75(6):7506205040 [FREE Full text] [doi: 10.5014/ajot.2021.046169] [Medline: 34817595]
32. Bennett S, Rodger S, Fitzgerald C, Gibson L. Simulation in occupational therapy curricula: a literature review. Aust Occup Ther J 2017 Aug 05;64(4):314-327. [doi: 10.1111/1440-1630.12372] [Medline: 28378428]
33. Spielberger C, Gonzalez H, Taylor C, Anton E, Algaze B, Ross G, et al. Test anxiety inventory. Mind Garden. URL: https://www.mindgarden.com/150-test-anxiety-inventory [accessed 2024-11-29]
34. Burnard P. A method of analysing interview transcripts in qualitative research. Nurse Educ Today 1991 Dec;11(6):461-466. [doi: 10.1016/0260-6917(91)90009-y] [Medline: 1775125]
35. Kvale S. Doing Interviews. Thousand Oaks, CA: Sage Publications; 2012.
36. Maykut P, Morehouse R. Beginning Qualitative Research: A Philosophical and Practical Guide. London, UK: Routledge; 1994.
37. Weeks BK, Horan SA. A video-based learning activity is effective for preparing physiotherapy students for practical examinations. Physiotherapy 2013 Dec;99(4):292-297. [doi: 10.1016/j.physio.2013.02.002] [Medline: 23806841]
38. SLU mental status exam. Saint Louis University. URL: https://www.slu.edu/medicine/internal-medicine/geriatric-medicine/aging-successfully/assessment-tools/mental-status-exam.php [accessed 2024-11-29]
39. Park WB, Kang SH, Lee YS, Myung SJ. Does objective structured clinical examinations score reflect the clinical reasoning ability of medical students? Am J Med Sci 2015 Jul;350(1):64-67 [FREE Full text] [doi: 10.1097/MAJ.0000000000000420] [Medline: 25647834]
40. Carleton RN. The intolerance of uncertainty construct in the context of anxiety disorders: theoretical and practical perspectives. Expert Rev Neurother 2012 Aug;12(8):937-947. [doi: 10.1586/ern.12.82] [Medline: 23002938]
41. Fernández-Castillo A, Caurcel MJ. State test-anxiety, selective attention and concentration in university students. Int J Psychol 2015 Aug 08;50(4):265-271. [doi: 10.1002/ijop.12092] [Medline: 25104475]
42. Putwain DW, Langdale HC, Woods KA, Nicholson LJ. Developing and piloting a dot-probe measure of attentional bias for test anxiety. Learn Individ Differ 2011 Aug;21(4):478-482. [doi: 10.1016/j.lindif.2011.02.002]
43. Smith TF, Capuzzi G. Using a mindset intervention to reduce anxiety in the statistics classroom. Psychol Learn Teach 2019 Apr 01;18(3):326-336. [doi: 10.1177/1475725719836641]
44. Sargolzaei F, Omid A, Mirmohammad-Sadeghi M, Ghadami A. The effects of virtual-augmented reality training on anxiety among operating room students attending coronary artery bypass graft surgery. Nurs Midwifery Stud 2021;10(4):229. [doi: 10.4103/nms.nms_95_20]

XSL•FO
RenderX

45. Carl E, Stein AT, Levihn-Coon A, Pogue JR, Rothbaum B, Emmelkamp P, et al. Virtual reality exposure therapy for anxiety and related disorders: a meta-analysis of randomized controlled trials. J Anxiety Disord 2019 Jan;61:27-36. [doi: 10.1016/j.janxdis.2018.08.003] [Medline: 30287083]
46. Stunden A, Halcomb E, Jefferies D. Tools to reduce first year nursing students' anxiety levels prior to undergoing objective structured clinical assessment (OSCA) and how this impacts on the student's experience of their first clinical placement. Nurse Educ Today 2015 Sep;35(9):987-991. [doi: 10.1016/j.nedt.2015.04.014] [Medline: 26004437]
47. Hills C, Ryan S, Smith DR, Warren-Forward H. The impact of 'Generation Y' occupational therapy students on practice education. Aust Occup Ther J 2012 Apr 28;59(2):156-163. [doi: 10.1111/j.1440-1630.2011.00984.x] [Medline: 22448996]
48. Foy C. Identifying barriers and pathways to success for new occupational therapy faculty members: a pilot survey. Occup Ther Health Care 2017 Oct;31(4):329-340. [doi: 10.1080/07380577.2017.1354269] [Medline: 28836884]
49. Beane-Katner L. Developing the next generation of faculty: taking a learning community approach. J Cent Teach Learn 2013;5:22 [FREE Full text]
50. Benson RT, Trower CA. Data, leadership, and catalyzing culture change. Change Mag High Learn 2012 Jun 29;44(4):27-34. [doi: 10.1080/00091383.2012.691862]
51. Serrat N, Rubio A. Coming from outside the academy. Values and 2.0 culture in higher education. Interact Learn Environ 2012 Jun;20(3):293-308. [doi: 10.1080/10494820.2011.641684]
52. DeCleene Huber KE, Nichols A, Bowman K, Hershberger J, Marquis J, Murphy T, et al. The correlation between confidence and knowledge of evidence-based practice among occupational therapy students. Open J Occup Ther 2015 Jan 01;3(1):497. [doi: 10.15453/2168-6408.1142]
53. Fortune J, Breckon J, Norris M, Eva G, Frater T. Motivational interviewing training for physiotherapy and occupational therapy students: effect on confidence, knowledge and skills. Patient Educ Couns 2019 Apr;102(4):694-700 [FREE Full text] [doi: 10.1016/j.pec.2018.11.014] [Medline: 30482468]
54. American Occupational Therapy Association. Occupational therapy curriculum design framework. Am J Occup Ther 2021;75((Supplement_3)):7513420020. [doi: 10.5014/ajot.2021.75s3008]
55. Wood D. Formative assessment. In: Swanwick T, Forrest K, O'Brien BC, editors. Understanding Medical Education: Evidence, Theory, and Practice. 3rd edition. Hoboken, NJ: John Wiley & Sons; 2018:361-373.

Abbreviations

AI: artificial intelligence
GPT: generative pretrained transformer
OSCE: objective structured clinical examination
OT: occupational therapy
SLUMS: Saint Louis University Mental Status
STAI: State-Trait Anxiety Inventory
TAI: Test Anxiety Inventory
VR: virtual reality

Edited by A Coristine; submitted 16.12.24; peer-reviewed by A Graf, X Cheng; comments to author 05.02.25; revised version received 10.03.25; accepted 25.03.25; published 30.04.25.

Please cite as:
Esmail S, Concannon B
Immersive Virtual Reality and AI (Generative Pretrained Transformer) to Enhance Student Preparedness for Objective Structured Clinical Examinations: Mixed Methods Study
JMIR Serious Games 2025;13:e69428
URL: https://games.jmir.org/2025/1/e69428
doi:10.2196/69428
PMID:

©Shaniff Esmail, Brendan Concannon. Originally published in JMIR Serious Games (https://games.jmir.org), 30.04.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Children's Improvement After Language and Rhythm Training With the Digital Medical Device Poppins for Dyslexia: Single-Arm Intervention Study

Charline Grossard[1,2], PhD; Melanie Descamps[2], PhD; Hugues Pellerin[1], MA; François Vonthron[2], MA; David Cohen[1], Prof Dr Med

[1]Department of Child and Adolescent Psychiatry, Pitié-Salpêtrière Hospital, Paris, France

[2]Poppins, Paris, France

Corresponding Author:
Charline Grossard, PhD
Department of Child and Adolescent Psychiatry
Pitié-Salpêtrière Hospital
47-83 boulevard de l'Hôpital
Paris, 75013
France
Phone: 33 142162383
Email: charline.grossard@aphp.fr

Abstract

Background: Specific learning disorder in reading (SLD reading), commonly named dyslexia, is a neurodevelopmental condition affecting reading. Current best practice recommendations for SLD reading emphasize the necessity of including graphophonological interventions. The serious game Mila-Learn, which is based on rhythm training, showed promising results in a prior randomized trial. However, it lacked a component of graphophonological training.

Objective: This study aimed to evaluate the effectiveness of Poppins, a new digital medical device that combines rhythm-based and graphophonological training for improving reading and phonological skills in children with SLD reading. We also explored its performance against Mila-Learn, the earlier version based on rhythm training only.

Methods: A single-arm study without an active control group was conducted with 38 children (aged 7-11 years) diagnosed with SLD reading. The participants completed an 8-week training program with Poppins (five 20-minute sessions per week). Pre- and posttraining assessments measured reading accuracy and speed, phoneme deletion, and phonological discrimination. Statistical analysis included pre- and postcomparisons (primary analysis) and comparisons with children's improvement from a previous randomized controlled trial of Mila-Learn, an earlier version of the device (exploratory analysis).

Results: The participants demonstrated significant improvements in reading accuracy (+11.46 words correctly read; $P<.001$), reading speed (+10.26 words read; $P<.001$), and phoneme deletion (+2.87 points; $P<.001$). No significant change was observed in reading comprehension for younger participants (grades 2-3; $P=.09$), although improvements were noted in older children (grades 4-5, $P=.03$). Exploratory analysis comparing children's improvements with Mila-Learn and Poppins revealed similar gains in reading accuracy and speed but revealed superior improvement in phonological skills for the Poppins group, with a moderate effect size according to the benchmarks by Cohen (Cohen $d=0.48$, $P=.02$).

Conclusions: Poppins is an effective and safe tool for enhancing reading and phonological skills in children with SLD reading. By integrating rhythm-based and graphophonological exercises, the device aligns with best practice recommendations for curative intervention. Future research should explore its long-term effects and medicoeconomic impact and compare outcomes with those of conventional therapy, as serious games provide an engaging, scalable method for delivering such interventions.

Trial Registration: ClinicalTrials.gov NCT06596980; https://clinicaltrials.gov/study/NCT06596980

(JMIR Serious Games 2025;13:e76435) doi:10.2196/76435

KEYWORDS

serious game; medical device; specific learning disorder; reading; rhythm

XSL•FO
RenderX

Introduction

Specific learning disorder in reading (SLD reading), commonly named dyslexia, is a neurodevelopmental condition characterized by impairments in reading, writing, and phonological processing. SLD reading affects 5% to 10% of the population, with a relatively high prevalence in males. SLD reading is associated with deficits in phonological awareness, rapid naming, and auditory-verbal memory, which impair the ability to decode written language and recognize words fluently [1,2]. The recommended treatment always includes graphophonological interventions. Indeed, among the leading theories explaining SLD reading, the phonological deficit hypothesis posits that impaired representation, storage, or access to phonemes—the basic sound units of language—underlies reading difficulties [3].

A related theory, the temporal processing hypothesis, suggests that difficulties in perceiving rhythmic and prosodic features of speech, such as syllable stress and amplitude modulations, further exacerbate phonological and reading challenges [4,5]. Recent neurophysiological studies using electroencephalography have shown that children with SLD reading exhibit atypical low-frequency cortical encoding of speech, particularly in the delta and theta bands (1-8 Hz), which are critical for processing rhythmic and temporal aspects of language [6]. These findings align with genetic studies showing moderate but significant correlations between rhythm perception, language skills, and SLD reading, supporting the Atypical Rhythm Risk Hypothesis [7].

Rhythmic and musical interventions have emerged as promising tools for addressing these deficits. A scoping review examined the use of music interventions to improve reading skills in children with SLD reading [8]. The authors identified 12 studies that focused on how music interventions target auditory processing, phonological processing, and temporal processing. Rhythmic activities are the most common musical element used in interventions, primarily aimed at improving reading accuracy. These activities were designed to enhance phonological awareness, syllable segmentation, and word recognition. Music and rhythm engage neural networks similar to those involved in language processing, enhancing phonological awareness, auditory memory, and reading accuracy [9,10]. For example, rhythmic training improves the synchronization of neural oscillations with auditory stimuli, which is crucial for phonological segmentation and temporal prediction in speech [11,12]. These interventions are particularly effective because they target both the underlying cognitive deficits and the motivational barriers often faced by children with SLD reading.

Serious games, which combine educational objectives with engaging gameplay, offer a novel and scalable approach to SLD reading intervention [13]. These digital tools provide intensive, repetitive training in phonological and graphophonological skills, which are essential for reading development [14]. By incorporating adaptive difficulty levels, immediate feedback, and rewards, serious games maintain high levels of engagement and motivation, which are critical for children with SLD reading who may otherwise avoid conventional literacy activities [15,16]. Furthermore, these games can be accessed remotely, addressing disparities in access to specialized resources and enabling frequent, home-based practice [13,17]. To be effective, interventions for children with SLD reading must be intensive (5 times per week). However, achieving this intensity is not feasible in conventional in-person therapy, where speech therapists are typically limited to scheduling 1 to 2 sessions per week [18]. Therefore, using serious games appears to be a solution to support high-intensity training.

With this in mind, we developed Mila-Learn, a digital medical device in the form of a serious game designed for rhythmic training to improve reading skills in children aged 7 years to 11 years with SLD reading [17]. It includes a series of music-based games focusing on rhythm repetition, synchronization, and rhythm completion, which are embedded in a story where the child (here, the player) is the hero. Between October 2021 and March 2023, we conducted a randomized placebo-controlled trial with 154 children with SLD reading who played with Mila-Learn or with a placebo game for 2 months. The results revealed improvements in reading accuracy and speed, supporting the ability of Mila-Learn to improve reading in children with SLD reading [19].

However, recommendations for best practices in the treatment of written language disorders in children [1] include 3 different modes of intervention for patients with SLD reading: (1) curative treatment targets underlying cognitive deficits (eg, phonological or visual-attentional deficits) and the various processes involved in written word identification (eg, graphophonological conversion, phonological decoding, orthographic memory, orthographic recoding), (2) adaptive treatment aiming to strengthen the reader's natural compensatory strategies (eg, development of lexical orthographic memory), and (3) compensatory treatment to reduce written language disorders by replacing deficient cognitive functions (eg, using digital aids). These intervention modes are not mutually exclusive and should be combined or alternated according to a treatment plan tailored to each patient. However, whenever possible, curative treatment should be the first intervention used with the patient. Curative treatment needs to combine phonological training with graphophonological training, as phonological training alone results in limited improvement in reading skills [1].

The publication of these recommendations led to modifications in the Mila-Learn game to better address the various aspects of curative treatment. In addition to the rhythmic games, we computed and added games focusing on written word identification. The new medical device, named Poppins, thus combines rhythmic tasks from Mila-Learn with new written language training tasks. In this article, we first describe the language games added to Poppins. We describe a single-arm study in which we focus on the impact of Poppins on the reading and phonological skills of children with SLD reading. As an exploratory analysis, we then compared the effects of Poppins with those of the medical device Mila-Learn [19] on the reading and phonological skills of children with SLD reading.

XSL•FO
RenderX

Methods

Device Description

Poppins is an app available on tablets and smartphones (iOS and Android) in which the child is led to carry out different activities divided into 2 categories: language activities and rhythm activities. Game data are stored anonymously on secure servers that comply with medical data protection regulations.

Overall, Poppins contains a number of different short activities, allowing for a variety of activities within a single session. An activity lasts an average of 1.40 minutes, to not demand too much attention from the child [20,21]. This structure of short activities ensures a fluid, coherent, and continuous experience for the user. To offer an experience suitable for as many children as possible (from 7 years to 11 years old, gamers and nongamers, with a wide range of tastes), the visual and narrative aesthetics of Poppins are aligned with mainstream video game standards such as Rayman and Mario games, games for all audiences.

Rhythmic activities (Figure 1) consist of rhythmic activities that were previously described [17]. To quickly summarize, all the tasks were designed to work on rhythm, as rhythm appears to be directly related to reading skills, whereas melody is not [22]. However, each task requires the mobilization of other skills, such as attention, inhibition, working memory, and motor skills, which are also often impaired in children with SLD reading [23].

Figure 1. Screenshots of rhythm tasks in the Poppins medical device: (A) In the Temple of Music, the child must recreate the rhythm of a melody using predefined rhythmic blocks; (B) in the King Song, the child must sing the displayed syllables in rhythm; (C) in the Beat Jumper, the child must make their character jump in rhythm by shaking the tablet; and (D) in Pop'n Run, the child must keep the rhythm by tapping the screen with their finger.



In March 2022, Collège Français d'Orthophonie (French College of Speech Therapy [CFO]) [1] published its recommendations for best practices in written language. These best practice recommendations were developed using the formal consensus recommendation method described in the scientific rationale published by the Haute Autorité de Santé [24] and include a section on the management of SLD reading. These recommendations reiterate the principle of the phonological hypothesis and the deficit in temporal processing of auditory signals. The approach of rhythm-based intervention is presented and supported by recommendations, which advise a "curative" intervention as a first-line treatment, meaning that it aims to rehabilitate underlying processes (such as rhythm processing) in the context of SLD reading.

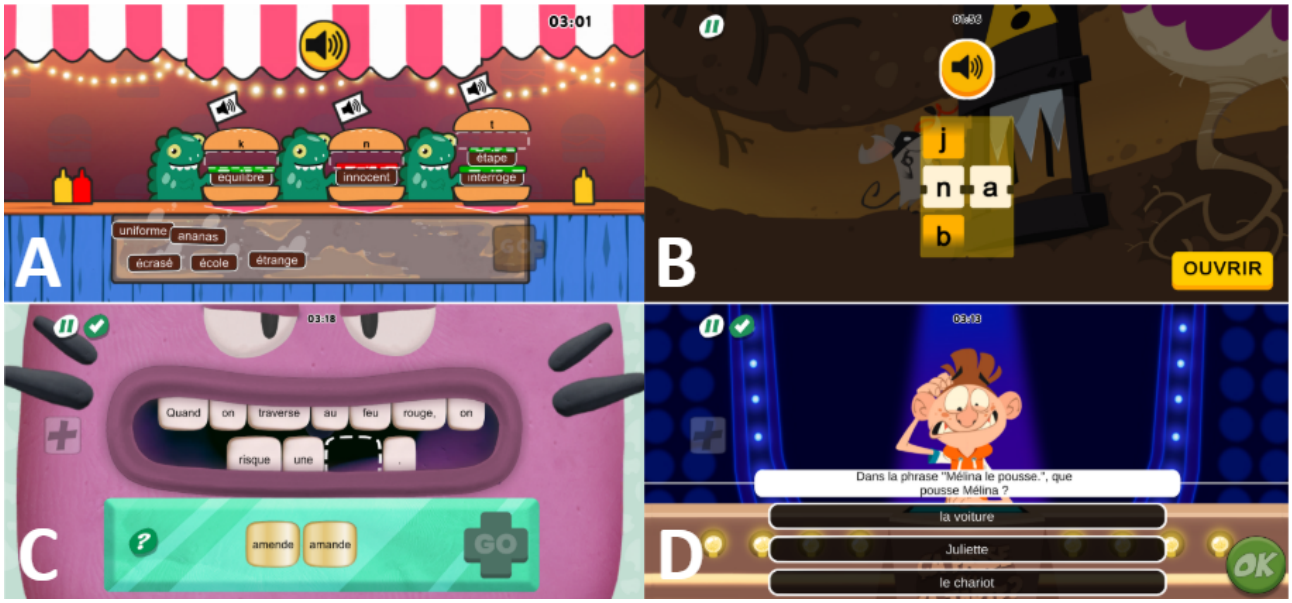
However, this curative intervention includes 2 components: the rehabilitation of underlying processes and direct work on graphophonological conversion (ie, the link between graphemes [letters] and phonemes [sounds]). The CFO's recommendations [1] also provide guidelines for implementing graphophonological conversion training. Following the publication of these recommendations, the game Mila-Learn was revised to include exercises targeting graphophonological conversion, which we will refer to as the language games module, in addition to the rhythmics tasks already developed. The 20-minute training time is evenly divided between written language exercises and rhythmic tasks. Parents are advised to aim for 5 play sessions per week. In the game, specific rewards appear starting from the 3rd session to encourage the child to complete at least 3 sessions per week, with additional rewards unlocked upon reaching the fifth session. The child's play time was monitored directly by the app. Once the child had played for 20 minutes, the game automatically stopped. The play time data were transmitted directly to Poppins' company using the internet. Parents were encouraged to ensure that the tablet was connected to the internet after each session that was played by the child. If the child did not play for a week, notifications were sent by the app: one on Monday to remind the child to play and another on Saturday to encourage play over the weekend. In addition, parents received an email on Wednesday, reminding them that their child should play Poppins.

All written language games (Figure 2) are designed to present the child with the grapheme linked to its phoneme and consist of real words selected on the basis of their frequency (using the Manulex-infra database [25]), orthographic complexity (verified

XSL•FO
RenderX

using the Eole database [26]), and length. Words are presented to the child on the basis of their infralexical characteristics. The difficulty level is continuously adapted to the child on the basis of their responses in the game: Depending on the number of errors in each exercise, the complexity level of the words is adjusted, and the child's progress in the game is tailored. The games are designed to maximize errorless learning, and immediate feedback is provided if the child makes a mistake. For example, in the word search game, the target words are displayed to the child, minimizing the risk of spelling errors.

Figure 2. Screenshots of language tasks in Poppins medical device: (A) In Dino-Burger, the child must categorize words on the basis of the sounds they contain; (B) in Veggie Crok, the child must reconstruct a syllable or a word; (C) in Supersmile, the child must choose the correct word to complete the sentence; and (D) in Ca. passe ou ça tarte (meaning "Pass or Smash"), the child must answer a question to demonstrate sentence comprehension.

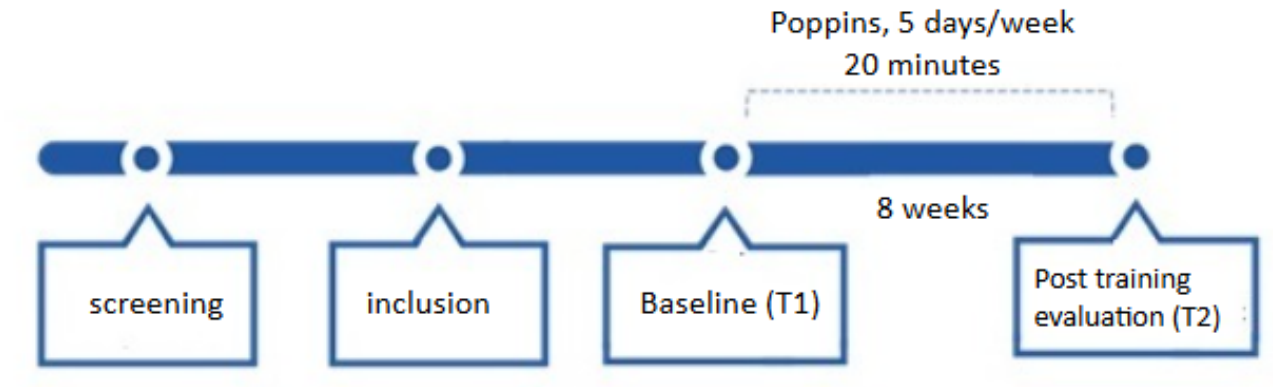


The written language games are divided into 4 areas: one focused on phonology ("I listen"), one focused on reading ("I read"), one focused on transcription ("I write"), and the last focused on comprehension ("I understand"). These areas always work in pairs to limit the number of different tasks (Recommendation 2.8) [1]. Additionally, these areas alternate on the basis of the child's progress, allowing for the coordination of reading tasks with transcription tasks (Recommendation 2.13) [1].

Study Protocol

The study design is a single-arm trial with no active control group examining the effect of the Poppins digital medical device for children with SLD reading (Figure 3). Each participant included in the study followed the training protocol, which consists of 5 training sessions per week on Poppins. Each session lasts approximately 20 minutes. After an assessment of their reading and phonological skills, all the children trained for 8 weeks with the Poppins medical device. Their skills were reassessed after the training period.

Figure 3. Description of the protocol for a single-arm study involving dyslexic children aged 7 years to 11 years.



Participants

We included children aged between 7 years and 11 years with a diagnosis of SLD reading according to the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-V) criteria, as confirmed by the recruiting center's physicians. They had to be native French speakers or have been schooled in France for at least 3 years. The children were not allowed to

XSL•FO
RenderX

have previously played Mila-Learn or Poppins. Children with autism, intellectual disabilities, or epilepsy were also excluded.

The sample size determination was based on the main objective and performed in G.Power 3.1.9.7.

The null (H0) and alternative (H1) hypotheses are defined as follows: H0: μD=0 and H1: μD≠0, where D=EVAL2MT2-EVAL2MT1 (defined earlier as the primary end point).

A 1-sample bilateral *t* test, assuming a dz effect size of 0.5, a 5% type I error, and 80% power, determined that the study needs 34 participants. To account for a 10% dropout rate (34/0.9), 38 participants were needed.

Outcomes

The primary variable was reading ability as assessed using the Eval2M task, which involves reading a list of words within a 2-minute time frame and produces 2 scores (speed and accuracy; Evaleo 6-15) [27].

The secondary variables included phonological skills, measured through a phoneme deletion task (BALE) [28] and a phonemic discrimination task (Evaleo 6-15) [27]. Additionally, reading skills were assessed in terms of speed and accuracy using the nonsensical text "Evalouette" [27], as well as in terms of speed, accuracy, and comprehension using texts from Evaleo 6-15 [27]. We also assessed parental stress with the shortened version of the Parenting Stress Index [29] and children's quality of life with the quality-of-life questionnaire of the PedsQL 4.0 Generic Core Scales child report [30] 8 weeks after the start of the training. The PedsQL 4.0 Generic Core Scales is a widely used and validated 23-item instrument measuring health-related quality of life in children and adolescents. For this study, the child self-report version was used. The Parenting Stress Index-Short Form is a 36-item parent self-report questionnaire. It efficiently measures parental stress across 3 domains: Parental Distress, Parent-Child Dysfunctional Interaction, and Difficult Child. Finally, all adverse events reported by the participants were recorded to evaluate the safety of Poppins.

Ethical Considerations

The study protocol was approved by the local ethics committee (Comité de protection des personnes Ile de France; approval number 24.00890.000301) and national regulatory agencies (Agence nationale de sécurité du médicament et des produits de santé and Commission nationale de l'informatique et des libertés; approval number 2024-A00432-45).

Prior to providing consent, the patient and their parents received a complete and comprehensive explanation of the study, including the study rationale, the procedures, the benefits and risks, that participation was voluntary, and that the participant could withdraw from the study at any time without any negative consequences.

Written informed consent was obtained from the participant or legal guardian in accordance with local practice and regulations prior to any study assessment or test being conducted. The parents provided their signature at the end of the consent form, and the delegated site team member countersigned the form. Consent was obtained via a wet-ink signature. A copy of the fully executed informed consent form was provided to the participants and their parents (either a paper version or electronically sent via email), and a copy was retained by the site in a secure area with restricted access.

Given the patient's condition (specific learning disability with reading or written expression deficit) and his or her age, thus potentially being unable to provide written consent, the child provided verbal assent to participate in the trial. The investigator read the content of the consent form during the information visit, clearly stating that the child was free to decline his or her participation at any moment. The consent form was adapted to an appropriate language to facilitate the child's understanding. Once the child verbally agreed to participate in the trial, the investigator signed the consent form on their behalf and registered it in the electronic case report form.

Information and inclusion visits could be carried out remotely or in person at the discretion of the parents or the investigator if he or she considered it necessary to assess that the child was not subjected to undue pressure.

Identifiable participant details (confidential details, including name and date of birth) were held in a database separated from the research database after receiving participants' consent. The research database never holds personal identifiable information. Only clinical sites had access to this information.

The study was registered in ClinicalTrials.gov retrospectively on September 11, 2024, after it started on June 28, 2024.

Statistical Analysis

Statistical analyses were carried out using R Version 4.2.2. First, we aimed to assess the delta before and after training with Poppins. In line with our predefined statistical analysis plan, the evaluation of absolute changes in skills between the baseline assessment (T1) and final assessment (T2) was conducted using a mixed linear regression with the following formula: score ~ time (T2 vs T1) + (1|Subject), with a *t* test of the time effect (lme4 package). The participant was used as a random intercept to account for individual variability. *P* values and 95% CIs were obtained using bootstrap resampling (boot package; 10,000 replications).

In a preceding clinical trial [19], 8 weeks of training with Mila-Learn was compared with 8 weeks of training with a placebo game. Our second exploratory analysis aimed to compare the scores of the children exposed to Poppins (this study) with the scores of the children exposed to the former version of the digital medical device Mila-Learn (previous study). To compare the results, the former version of the digital medical device will be compared when the results are significantly different from those of the placebo in search of superiority and difference descriptions. The comparison of change scores (T2-T1) between groups was performed using the Welch *t* test. Missing data were handled using a complete-case analysis.

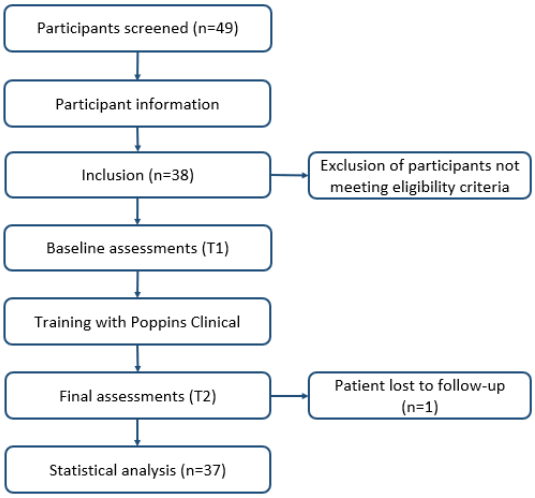
XSL•FO
RenderX

Results

Study Follow-Up

The study started in June 2024 and ended in February 2025, and 49 patients were screened for this study. The inclusion criteria were not met by 11 patients. A total of 38 patients gave their informed written consent to participate in the study and were subsequently enrolled (see Figure 4). One patient was lost to follow-up. Consequently, 37 patients were included in the analysis. All patients followed the training during the study.

Figure 4. Flowchart of patient recruitment and study follow-up.



Participants

The characteristics of the participants enrolled are described in Table 1.

Table 1. Description of the participants recruited for the study on the basis of their grade level.

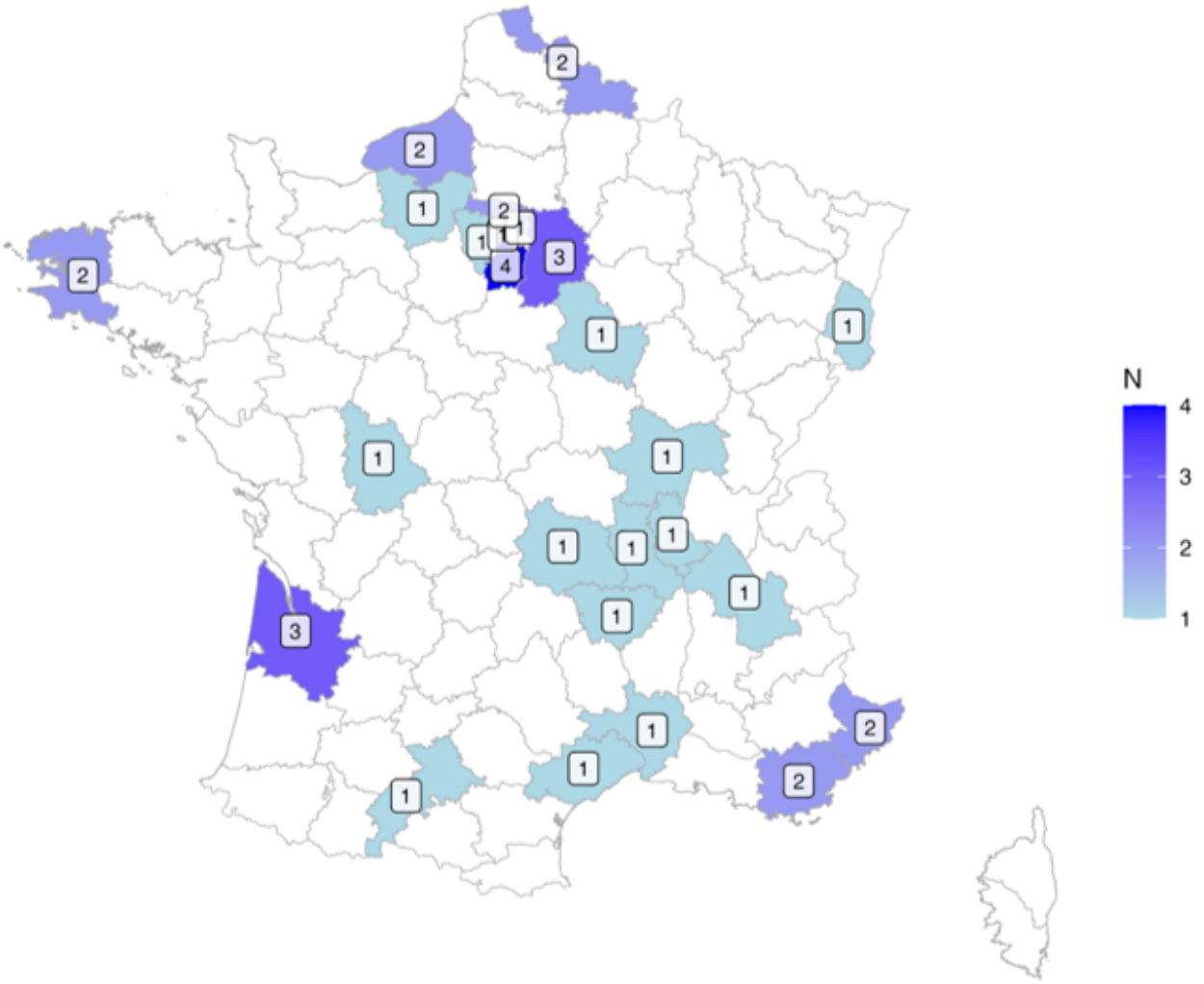
Characteristics	2nd grade (n=6)	3rd grade (n=13)	4th grade (n=14)	5th grade (n=5)
Age (years), mean (SD)	8.21 (0.47)	8.94 (0.42)	9.91 (0.62)	10.88 (0.46)
Sex, n				
Female	1	8	4	0
Male	5	5	10	5
Speech therapy, n (%)	2 (33)	11 (85)	9 (64)	2 (40)
ADHD[a], n (%)	4 (67)	4 (31)	3 (21)	2 (40)
Language disorder, n (%)	0 (0)	1 (8)	0 (0)	0 (0)
Dyspraxia, n (%)	1 (17)	2 (15)	2 (14)	0 (0)
Dyscalculia, n (%)	1 (17)	3 (23)	1 (7)	0 (0)

[a]ADHD: attention-deficit/hyperactivity disorder.

Among the 38 participants, 20 had a medical history. The highest level of education attained by the parents was controlled. Among the 38 families, none did not have a high school diploma (Bac), 18 (47%) had an education level ranging from Bac to Bac +3, and 20 (53%) had an education level equal to or higher than Bac +4. The mean total time played by the children was 653 (SD 202) minutes. The distribution of children in France is shown in Figure 5. No adverse events were reported during the study.

XSL•FO
RenderX

Figure 5. Geographical distribution of children included in the study across France.



Outcomes

The primary end point was absolute improvement in reading skills, specifically through the lens of word reading capability, using the EVAL2M subtest from EVALEO 6-15 [27] after an 8-week training regimen (T2). The secondary end point was improvement in reading skills in terms of accuracy, speed and comprehension, and phonological skills (phoneme deletion and phonological discrimination). The scores at T1 and T2 are presented in Table 2.

XSL•FO
RenderX

Table 2. Comparison of participants' scores on the proposed tasks at pretest (T1) and posttest (T2).

Tasks	T2 (n=37), mean (SD)	T1 (n=38), mean (SD)	Difference (95% CI)	*P* value
Words read in 2 minutes (EVAL2M)				
Words correctly read	101.49 (33.86)	88.39 (34.00)	11.46 (8.75 to 14.25)	<.001
Words read	110.46 (32.43)	98.50 (32.54)	10.26 (6.77 to 13.85)	<.001
Phoneme deletion (BALE)				
Total score	16.58 (3.38)	13.54 (4.51)	2.87 (1.64 to 4.09)	<.001
Total time (seconds)	205.19 (101.37)	213.14 (102.56)	–5.37 (–31.97 to 21)	.69
Text with no meaning (Evalouette)				
Words read	105.59 (42.94)	93.21 (38.76)	10.95 (4.87 to 17.19)	<.001
Words correctly read	94.49 (42.80)	79.76 (38.92)	13.25 (7.78 to 18.86)	<.001
Phonological discrimination (Evaleo 6-15)				
Total score	21.84 (3.95)	20.21 (3.86)	1.61 (0.2 to 3.06)	.03
Text comprehension 4th and 5th grades				
Number of reading errors	5.50 (5.73)	4.50 (4.88)	1.00 (–0.72 to 2.67)	.25
Reading time (seconds)	97.76 (41.51)	85.29 (37.46)	12.46 (4.96 to 19.76)	.001
Comprehension score	15.45 (1.50)	4.75 (1.12)	0.70 (0.06 to 1.32)	.03

The results indicate that participants correctly read significantly more words at T2 than at T1 (*P*<.001). On average, children read 11.46 more words correctly at T2 than at T1. The participants using Poppins also showed improvements in reading speed, with more words read at T2. The difference in mean improvement (+10.26 words read between T1 and T2) reached statistical significance (*P*<.001).

Children significantly improved their performance between T1 and T2 for reading speed and fluency (*P*<.001) and phonological tasks. No significant improvement between T1 and T2 was found in the parental stress questionnaire (*P*=.15) or in the child's quality of life questionnaire (*P*=.92).

Text comprehension was assessed using the paragraph reading task from EVALEO 6-15 [27]. This assessment varies depending on grade level, as the texts provided differ between the 2nd and 3rd grades and the 4th and 5th grades. The results were therefore analyzed by grade level, 2nd/3rd vs 4th/5th. No significant improvement in reading time (*P*=.66), the number of reading errors (*P*=.052), or the comprehension score (*P*=.86) was found in the text comprehension task for the 2nd/3rd graders. A significant improvement in the comprehension score (*P*=.03) was found for the 4th/5th graders, as the reading time increased (*P*=.001).

Finally, we conducted an exploratory analysis comparing the group who played Poppins and the group who played Mila-Learn in a previous study [19] in terms of absolute improvement in the number of words read and number of words correctly read using the subtest EVAL2M from the Evaleo 6-15 [27] and in terms of absolute improvement in the score of the phoneme deletion task from the BALE [28] after an 8-week training regimen (T2). The change in each group was quantified as the difference between the posttraining mean score at T2 and the pretraining baseline mean score. The results are summarized in Table 3.

Table 3. Improvement in scores between posttest (T2) and pretest (T1) for children using Poppins compared with the improvement in scores for children using Mila-Learn from the previous study.

Tasks	Poppins (N=38), mean (SD)	Mila-Learn (N=75), mean (SD)	Difference (95% CI)	Cohen *d*	*P* value
Number of words correctly read (EVAL2M)					
T2-T1	11.41 (8.44)	10.63 (12.51)	0.77 (–3.31 to 4.85)	0.07	.70
T1	88.39 (34.00)	89.32 (30.39)	–0.92 (–14.05 to 12.20)	–0.03	.90
Number of words read (EVAL2M)					
T2-T1	10.16 (10.86)	11.25 (13.43)	–1.09 (–5.89 to 3.72)	–0.09	.70
T1	98.50 (32.54)	102.72 (27.07)	–4.22 (–16.54 to 8.10)	–0.15	.50
Phoneme deletion score (BALE)					
T2-T1	2.75 (3.67)	0.90 (3.94)	1.85 (0.31 to 3.40)	0.48	.02
T1	13.54 (4.51)	12.38 (5.22)	1.17 (–0.75 to 3.08)	0.2	.20

XSL•FO
RenderX

The progression of children's scores was similar between Mila-Learn and Poppins for speed (number of words read) and reading fluency (number of words read correctly). However, the phonological skills of the children in the Poppins game group were better than those of the children in the Mila-Learn game group.

Discussion

Principal Findings

In this study, we evaluated the impact of 8 weeks of training with Poppins on the reading performance of children aged 7 years to 11 years with SLD reading. We also assessed the phonological skills of the children and compared the results of the group who used the medical app Poppins with those of a group who used a former version of the digital medical device (named Mila-Learn) in a prior study. SLDs manifest as marked deficits in reading fluency and accuracy and are typically characterized by slow, effortful reading and substantial inaccuracies [2]. Reading accuracy precedes the development of reading speed, with significant improvements in speed occurring once a foundational accuracy level is established [31]. Therefore, both reading accuracy and speed are crucial metrics for assessing reading proficiency.

First, Poppins appears to be a safe device, as no adverse events were reported during the study. No device malfunction could lead to a serious adverse event, and no serious breaches were reported.

We explored clinical performance by assessing children after 8 weeks of training with Poppins. With respect to reading accuracy, the children improved their scores in nearly all reading tasks after 8 weeks of training with Poppins. Children showed improvement in the primary end point, with the number of correctly read words increasing significantly (P<.001), by 11 words. For comparison, children in the placebo group in the previous study [19] improved their reading by 5 words at the second evaluation, 8 weeks later. This improvement was also observed in other reading tasks, such as reading nonsense text (Evalouette; P<.001). Word reading is a measure used in clinical and research settings, serving as a proxy for reading proficiency in day-to-day life.

The impact of Poppins on reading speed has also been investigated. Reading fluency is recognized as a key marker of SLD reading across different languages and is a persistent symptom of SLD reading, even when basic decoding skills have been acquired [32]. In our study, reading speed was reported by counting the number of words read, which improved at T2 (P<.001).

Phonological awareness is an important capacity when exploring reading. Indeed, phonological deficit is one of the most common causes of SLD reading [3,33] and results, among others, in phonological awareness difficulties. In our study, children's scores improved on both the discrimination and the phoneme deletion tasks (P=.03 and P<.001, respectively). The children did not show significant improvement in speed in the deletion task (P=.69). Assessing phonological skills in children with SLD reading typically involves evaluating both accuracy and speed, as these dimensions provide a comprehensive understanding of their phonological processing abilities. Many children with SLD reading exhibit deficits in phonological processing speed, meaning that they process phonological information more slowly than their peers do [34]. Speed is often assessed using tasks such as rapid automated naming, which measures how quickly a child can name familiar items (eg, letters, colors, or objects). In future research, it could be valuable to include such tasks in the assessment battery to better capture improvements in phonological processing speed.

We evaluated the reading comprehension of children after an 8-week intervention with Poppins. The assessment was conducted using the paragraph comprehension task of EVALEO 6-15 [27]. This task involves 2 texts to avoid retest effects: one for the initial evaluation and another for the follow-up evaluation. The texts are adapted to the children's grade level, with children in 2nd and 3rd grades reading shorter texts consisting of only a few sentences, whereas children in 4th and 5th grades read a full text composed of 3 paragraphs, each containing 5 to 6 sentences. Reading comprehension was assessed by asking questions about the text. The results revealed no significant effect of the training on the comprehension scores of children in the 2nd and 3rd grades. However, the comprehension scores of older children in grades 4 and 5 improved, as evidenced by the analyses (P=.03). The observed difference in results between 2nd/3rd grade and 4th/5th grade students likely stems from the texts used. The texts for 4th/5th grade students are significantly longer than those for 2nd/3rd grade students, leading to increased working memory load. Additionally, the comprehension of 4th/5th grade students is assessed with a greater number of questions than that of 2nd/3rd grade students (7 questions vs 4 questions), allowing for a more detailed evaluation of their text comprehension abilities. We also observed that children read more slowly at T2 than at T1 (+12.46 seconds; P=.01). This suggests that children may have slowed their reading speed to ensure better comprehension of the text, which is a common and effective strategy among children with SLD reading [35].

Comparison With the Former Version Mila-Learn

Finally, we compared the evolution of reading accuracy and reading speed scores obtained from the 2-minute word reading task (EVAL2M) between children who used Poppins and those who used Mila-Learn. No significant differences were found between the 2 groups, suggesting that the new version of the device, Poppins, remains effective. We also examined the impact of Poppins on phonological skills, as measured by the phoneme deletion task, compared with the impact of Mila-Learn. Children who used Poppins improved their scores by 1.85 points more than those who used Mila-Learn did (P=.02). In the previous Poppins-01 study, only children in the 2nd/3rd grades demonstrated significantly greater improvement in the Mila-Learn group than in the placebo group. These results support the hypothesis that Poppins enhances children's phonological skills more effectively than the previous version, Mila-Learn. Indeed, several studies have independently demonstrated the benefits of rhythmic training, as well as phonemic and graphophonological training, in improving phonological skills in children with dyslexia [36,37]. These

XSL•FO
RenderX

different approaches have shown substantial effects, which are generally comparable across interventions. Thomson et al [36] suggested that combining rhythmic training with graphophonological training may be particularly effective for enhancing phonological skills in children with dyslexia.

Among the study population, 9 children played with Poppins for less than 500 minutes, representing 27% of the population, whereas 36% of the children in the previous study using the Mila-Learn game did not reach the 500-minute play threshold. The median playing time was 752.50 minutes, with the majority of the children playing for more than 500 minutes (Q1=524.00 minutes, Q3=809.50 minutes). The mean global playing time was 653 minutes, which is comparable to the playing time reported in a previous study with Mila-Learn (global playing time=660 minutes). This suggests that the addition of language exercises did not negatively impact adherence to the treatment.

Maintaining treatment adherence makes Poppins a relevant tool for health care professionals to enhance patient care by enabling home-based training. Moreover, serious games offer affordable and accessible solutions, providing support when health care services are limited or in the context of remote follow-up. Treatment adherence is further supported by the game's engaging and playful elements [38].

Limitations

Despite these promising results, our study has several major limitations. First, we did not include a control group that followed their treatment as usual, which makes it difficult to attribute the observed improvements to the use of Poppins rather than to natural progression for the majority of our results. To address this, we used data from a previous study [19] as a comparison, but this approach may affect the validity of the comparison. To minimize bias, we conducted recruitment in the same way for both studies and used the same inclusion criteria, except for the speech therapy sessions. Effectively, children in our current study were allowed to continue their usual treatment. Among the 38 recruited children, 24 (63%) were receiving speech therapy. In contrast, the children in the previous Poppins-01 study did not have speech therapy sessions. This difference may have introduced bias, as children in the Poppins group could have received additional training than those in the Mila-Learn group did, and we did not monitor the exact number of speech therapy sessions attended by the children during their participation in the study. However, since a large portion of the children were recruited between June and July, many did not regularly attend speech therapy during this period, as July and August are holiday months in France. Finally, the groups also differed in terms of age and gender distribution, with a greater percentage of children in 5th grade (CM2) in the Mila-Learn group (23/75, 30% for Mila-Learn vs 5/38, 13% for Poppins) and a greater percentage of girls who played Mila-Learn (39/75, 52% for Mila-Learn vs 13/38, 34% for Poppins) [19].

Conclusion

In conclusion, despite several limitations, the study revealed that Poppins appears to be a promising tool, with participants showing significant improvements in reading fluency as well as phonological skills. The improvements in reading skills were comparable to those seen with the previous version of the device, Mila-Learn, but Poppins might have a greater impact on phonological skills. The improvement in reading seems to have lightened the load on working memory, allowing children to better understand long texts. Moreover, no adverse events were reported, and the device was found to be safe and well tolerated.

Acknowledgments

We would like to thank the family and the participants. Our thanks also go to Sara Cadoni and Clemence Faulcon, who participated in this work.

Data Availability

The data sets generated or analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

Conceptualization – CG, MD, DC, FV
Statistical plan – MD, HP
Writing - original draft – CG
Writing - review & editing – MD, DC, FV, HP

Conflicts of Interest

FV is the CEO of Poppins, the company that developed the game Poppins. MD and CG are employees at Poppins. DC reports a relationship with bMotion Technologies that includes equity or stocks.

References

1. Les Recommandations de Bonne Pratique d'Évaluation, de Prévention et de Remédiation des troubles du langage écrit chez l'enfant et l'adulte. Collège Français d'Orthophonie. 2022 Mar 31. URL: https://www.college-francais-orthophonie.fr/recommandations-de-bonne-pratique-en-langage-ecrit/ [accessed 2025-07-07]

XSL•FO
RenderX

2. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, Text Revision (DSM-5-TR). Washington, DC: American Psychiatric Association Publishing; 2022.
3. Peterson RL, Pennington BF. Developmental dyslexia. The Lancet 2012 May;379(9830):1997-2007. [doi: 10.1016/s0140-6736(12)60198-6]
4. Huss M, Verney JP, Fosker T, Mead N, Goswami U. Music, rhythm, rise time perception and developmental dyslexia: perception of musical meter predicts reading and phonology. Cortex 2011 Jun;47(6):674-689. [doi: 10.1016/j.cortex.2010.07.010] [Medline: 20843509]
5. Goswami U. Sensory theories of developmental dyslexia: three challenges for research. Nat Rev Neurosci 2015 Jan 5;16(1):43-54. [doi: 10.1038/nrn3836] [Medline: 25370786]
6. Araújo J, Simons BD, Peter V, Mandke K, Kalashnikova M, Macfarlane A, et al. Atypical low-frequency cortical encoding of speech identifies children with developmental dyslexia. Front Hum Neurosci 2024 Jun 7;18:1403677 [FREE Full text] [doi: 10.3389/fnhum.2024.1403677] [Medline: 38911229]
7. Alagöz G, Eising E, Mekki Y, Bignardi G, Fontanillas P, 23andMe Research Team, et al. The shared genetic architecture and evolution of human language and musical rhythm. Nat Hum Behav 2025 Feb 21;9(2):376-390. [doi: 10.1038/s41562-024-02051-y] [Medline: 39572686]
8. Lewis J, Kim SJ. Scoping review of music interventions aimed at improving reading skills in children with specific learning disorders in reading. J Music Ther 2024 Sep 12;61(3):218-243. [doi: 10.1093/jmt/thae010] [Medline: 38683748]
9. Habib M, Lardy C, Desiles T, Commeiras C, Chobert J, Besson M. Music and dyslexia: a new musical training method to improve reading and related disorders. Front Psychol 2016 Jan 22;7:26 [FREE Full text] [doi: 10.3389/fpsyg.2016.00026] [Medline: 26834689]
10. Flaugnacco E, Lopez L, Terribili C, Montico M, Zoia S, Schön D. Music training increases phonological awareness and reading skills in developmental dyslexia: a randomized control trial. PLoS One 2015;10(9):e0138715 [FREE Full text] [doi: 10.1371/journal.pone.0138715] [Medline: 26407242]
11. Bégel V, Dalla Bella S, Devignes Q, Vandenbergue M, Lemaître MP, Dellacherie D. Rhythm as an independent determinant of developmental dyslexia. Dev Psychol 2022 Feb;58(2):339-358. [doi: 10.1037/dev0001293] [Medline: 35007111]
12. Goswami U. Language acquisition and speech rhythm patterns: an auditory neuroscience perspective. R Soc Open Sci 2022 Jul;9(7):211855 [FREE Full text] [doi: 10.1098/rsos.211855] [Medline: 35911192]
13. Ren X, Wu Q, Cui N, Zhao J, Bi H. Effectiveness of digital game-based trainings in children with neurodevelopmental disorders: a meta-analysis. Res Dev Disabil 2023 Feb;133:104418. [doi: 10.1016/j.ridd.2022.104418] [Medline: 36603312]
14. Dalla Bella S. Rhythmic serious games as an inclusive tool for music-based interventions. Ann N Y Acad Sci 2022 Nov 17;1517(1):15-24. [doi: 10.1111/nyas.14878] [Medline: 35976673]
15. Khaleghi A, Aghaei Z, Behnamghader M. Developing two game-based interventions for dyslexia therapeutic interventions using gamification and serious games approaches entertainment computing journal. Entertainment Computing 2022 May;42:100482. [doi: 10.1016/j.entcom.2022.100482]
16. Vasalou A, Khaled R, Holmes W, Gooch D. Digital games-based learning for children with dyslexia: a social constructivist perspective on engagement and learning during group game-play. Computers & Education 2017 Nov;114:175-192. [doi: 10.1016/j.compedu.2017.06.009]
17. Vonthron F, Yuen A, Pellerin H, Cohen D, Grossard C. A serious game to train rhythmic abilities in children with dyslexia: feasibility and usability study. JMIR Serious Games 2024 Jan 11;12:e42733 [FREE Full text] [doi: 10.2196/42733] [Medline: 37830510]
18. Helloin M. Restitution et analyse au questionnaire : Prise en soin des troubles du langage écrit. L'Orthophoniste 2019(392):15-23 [FREE Full text]
19. Descamps M, Grossard C, Pellerin H, Lechevalier C, Xavier J, Matos J, et al. Rhythm training improves word-reading in children with dyslexia. Sci Rep 2025 May 21;15(1):17631 [FREE Full text] [doi: 10.1038/s41598-025-02485-y] [Medline: 40399331]
20. Carlier S, Van der Paelt S, Ongenae F, De Backere F, De Turck F. Empowering children with ASD and their parents: design of a serious game for anxiety and stress reduction. Sensors (Basel) 2020 Feb 11;20(4):966 [FREE Full text] [doi: 10.3390/s20040966] [Medline: 32054025]
21. Yusoff A, Crowder R, Gilbert L. Validation of Serious Games Attributes Using the Technology Acceptance Model. 2010 Presented at: Second International Conference on Games and Virtual Worlds for Serious Applications; March 25-26, 2010; Braga, Portugal. [doi: 10.1109/vs-games.2010.7]
22. Sousa J, Martins M, Torres N, Castro SL, Silva S. Rhythm but not melody processing helps reading via phonological awareness and phonological memory. Sci Rep 2022 Aug 02;12(1):13224 [FREE Full text] [doi: 10.1038/s41598-022-15596-7] [Medline: 35918357]
23. Lonergan A, Doyle C, Cassidy C, MacSweeney Mahon S, Roche RA, Boran L, et al. A meta-analysis of executive functioning in dyslexia with consideration of the impact of comorbid ADHD. Journal of Cognitive Psychology 2019 Oct 03;31(7):725-749. [doi: 10.1080/20445911.2019.1669609]
24. Haute ADS. Recommandations par consensus formalisé (RCF). Saint-Denis La Plaine: HAS 2015:1-40 [FREE Full text]

XSL•FO
RenderX

25. Peereman R, Lété B, Sprenger-Charolles L. Manulex-infra: distributional characteristics of grapheme-phoneme mappings, and infralexical and lexical units in child-directed written material. Behav Res Methods 2007 Aug;39(3):579-589. [doi: 10.3758/bf03193029] [Medline: 17958171]
26. Pothier B, Pothier P. Échelle d'acquisition de l'orthographe lexicale. Retz. 2003. URL: https://eole.editions-retz.com/ [accessed 2025-07-08]
27. Launay L, Roustit J, Maeder C, Touzin M. Evaluation du langage écrit et du langage oral 6-15 ans. Evaléo 6-15. 2018. URL: https://www.evaleo6-15.fr/ [accessed 2025-07-08]
28. Jacquier-Roux M, Lequette C, Pouget G, Valdois S, Zorman M. BALE: batterie analytique du langage écrit. Groupe Cogni-Sciences, Laboratoire de Psychologie et NeuroCognition. 2010. URL: https://www1.ac-grenoble.fr/media/14951/download [accessed 2025-07-08]
29. EuroQol Group. EuroQol--a new facility for the measurement of health-related quality of life. Health Policy 1990 Dec;16(3):199-208. [doi: 10.1016/0168-8510(90)90421-9] [Medline: 10109801]
30. Varni JW, Seid M, Kurtin PS. Pediatric Quality of Life Inventory 4.0 Generic Core Scales. APA PsycTests. 2001. URL: https://psycnet.apa.org/doiLanding?doi=10.1037%2Ft54487-000 [accessed 2025-07-08]
31. Juul H, Poulsen M, Elbro C. Separating speed from accuracy in beginning reading development. Journal of Educational Psychology 2014 Nov;106(4):1096-1106 [FREE Full text] [doi: 10.1037/a0037100]
32. Carroll J, Holden C, Kirby P, Thompson PA, Snowling MJ. Towards a consensus on dyslexia: findings from a Delphi study. OSF; preprint available online on May 2, 2024 ;3. [doi: 10.31219/osf.io/tb8mp]
33. Share DL. Common misconceptions about the phonological deficit theory of dyslexia. Brain Sci 2021 Nov 14;11(11):1510 [FREE Full text] [doi: 10.3390/brainsci11111510] [Medline: 34827508]
34. da Silva PB, Engel de Abreu PMJ, Laurence PG, Nico, Simi LGV, Tomás RC, et al. Rapid automatized naming and explicit phonological processing in children with developmental dyslexia: a study with Portuguese-speaking children in Brazil. Front Psychol 2020 May 27;11:928 [FREE Full text] [doi: 10.3389/fpsyg.2020.00928] [Medline: 32528363]
35. Walczyk JJ, Griffith‐Ross DA. How important is reading skill fluency for comprehension? The Reading Teacher 2011 Nov 09;60(6):560-569. [doi: 10.1598/rt.60.6.6]
36. Thomson JM, Leong V, Goswami U. Auditory processing interventions and developmental dyslexia: a comparison of phonemic and rhythmic approaches. Read Writ 2012 Feb 15;26(2):139-161. [doi: 10.1007/s11145-012-9359-6]
37. Bhide A, Power A, Goswami U. A rhythmic musical intervention for poor readers: a comparison of efficacy with a letter‐based intervention. Mind Brain and Education 2013 May 17;7(2):113-123. [doi: 10.1111/mbe.12016]
38. Dereje D, Lamba D, Abessa TG, Kenea C, Ramari C, Osama M, et al. Unlocking the potential of serious games for rehabilitation in low and middle-income countries: addressing potential and current limitations. Front Digit Health 2025;7:1505717 [FREE Full text] [doi: 10.3389/fdgth.2025.1505717] [Medline: 39957725]

Abbreviations

Bac: high school diploma
CFO: Collège Français d'Orthophonie
DSM-V: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
SLD-reading: specific learning disorder with reading impairment

Edited by A Coristine; submitted 23.04.25; peer-reviewed by M Adnan, A Giannakopoulou, C Curatoli; comments to author 10.06.25; revised version received 24.06.25; accepted 01.07.25; published 01.08.25.

Please cite as:
Grossard C, Descamps M, Pellerin H, Vonthron F, Cohen D
Children's Improvement After Language and Rhythm Training With the Digital Medical Device Poppins for Dyslexia: Single-Arm Intervention Study
JMIR Serious Games 2025;13:e76435
URL: https://games.jmir.org/2025/1/e76435
doi:10.2196/76435
PMID:

©Charline Grossard, Melanie Descamps, Hugues Pellerin, François Vonthron, David Cohen. Originally published in JMIR Serious Games (https://games.jmir.org), 01.08.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic

XSL•FO
RenderX

information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Empowering Children With Down Syndrome by Enhancing Emergency Preparedness Through Serious Games: Quasi-Experimental Study With a Between-Group Design

Samaa M Shohieb[1*], PhD; Suzan Hassan Bakhit[2*], PhD; Ensaf Mohammed[2*], PhD; Abdelghafar M Elhady[3*], PhD

[1]Faculty of Computers and Information, Mansoura University, Mansoura, Egypt

[2]Department of Computer Engineering and Information, College of Engineering-Wadi Addawasir, Prince Sattam Bin Abdulaziz University, Reyad, Saudi Arabia

[3]General Administration of the University, Mansoura University, Mansoura, Egypt

[*]all authors contributed equally

Corresponding Author:
Samaa M Shohieb, PhD
Faculty of Computers and Information
Mansoura University
1 EL Gomhoria Street
Mansoura, 35511
Egypt
Phone: 20 01142558595
Email: sm_shohieb@mans.edu.eg

Abstract

Background: Children with Down syndrome (DS) often experience cognitive and adaptive challenges that affect their ability to acquire and retain critical life skills, including those needed for effective response during emergencies. Traditional training methods used to prepare children for crises are frequently static, noninteractive, and insufficiently tailored to the unique learning profiles of children with DS. These limitations contribute to reduced engagement, poor knowledge retention, and inadequate real-world preparedness. Recent advancements in game-based learning, particularly serious games, have demonstrated potential for enhancing education and skill development among individuals with cognitive impairments.

Objective: This study aimed to design, implement, and evaluate Risk Resist, an adaptive serious game developed to improve emergency preparedness in children with DS. The game incorporates a dynamic difficulty adjustment algorithm that personalizes the learning experience by dynamically modifying game difficulty based on real-time behavioral performance metrics. The study also assessed whether this adaptive game-based learning approach leads to superior learning gains and engagement compared to conventional teacher-led training.

Methods: A quasi-experimental, between-group design was used with 18 children diagnosed with DS, aged 8 to 12 years. Participants were randomly assigned to either an experimental group (n=9), which played Risk Resist, or a control group (n=9), which received traditional instruction on emergency scenarios. Learning outcomes were assessed using pre- and postintervention knowledge tests composed of 5 emergency-related questions. Engagement levels were measured through a structured 5-point Likert scale questionnaire completed by observing teachers. The game used a machine learning–driven dynamic difficulty adjustment model, specifically a Random Forest Regressor, which adjusted difficulty in response to individual performance indicators such as success rate, response time, and behavioral patterns during gameplay.

Results: The experimental group achieved significantly higher learning gains (mean 3.6, SD 0.5) than the control group (mean 2.0, SD 0.5; $P<.001$). Engagement levels were also significantly greater in the game-based group (mean 4.54, SD 0.4) compared to the control group (mean 4.01, SD 0.32; $P=.002$). A strong positive correlation was identified between engagement and learning gain ($r=0.85$; $P<.001$), indicating that higher engagement contributed directly to improved knowledge acquisition.

Conclusions: The results support the effectiveness of Risk Resist in enhancing both engagement and learning outcomes for children with DS. The integration of adaptive difficulty algorithms provides a personalized, responsive learning experience, positioning serious games as a viable and impactful tool for emergency preparedness training in special education settings.

(JMIR Serious Games 2025;13:e73690) doi:10.2196/73690

XSL•FO
RenderX

KEYWORDS

crisis management; Down syndrome; dynamic difficulty adjustment; game-based learning; game-based rehabilitation; machine learning; serious games

Introduction

Children with Down syndrome (DS) represent a significant portion of the global population, with approximately 1 in every 700 children born each year affected by this chromosomal abnormality [1]. As the most common chromosomal condition, DS presents unique challenges in cognitive development and skill acquisition, necessitating specialized educational strategies to optimize learning outcomes [2].

Problem Statement and the Significance of the Study

Children with DS face significant cognitive and adaptive challenges that affect their ability to respond to emergencies effectively. Difficulties in processing complex instructions, recalling learned responses, and making rapid safety decisions increase their vulnerability during crises [3]. Many struggle to interpret danger cues, follow emergency protocols, and apply prior knowledge under stressful conditions, making traditional preparedness training methods insufficient [4].

Despite efforts to teach crisis response, conventional teacher-led instruction often fails to engage children with DS effectively, as these methods lack interactive and adaptive learning mechanisms [5]. The limitations of static training approaches reduce knowledge retention and do not accommodate the individualized learning needs of children with DS. Consequently, children may remain underprepared for real-world emergencies due to the absence of engagement-driven learning models [6,7].

To address these gaps, this study introduces Risk Resist, a game-based learning (GBL) tool designed to provide personalized, adaptive emergency training through interactive gameplay and real-time feedback. In addition to addressing the immediate learning needs of children with DS, the development of adaptive serious games like Risk Resist has broader implications for educational policy and curriculum design in special education. By providing personalized, engaging, and scalable learning tools, serious games can complement traditional instructional methods and support inclusive education mandates. This aligns with growing international calls for leveraging technology to enhance educational accessibility [8], equity, and lifelong learning opportunities for individuals with cognitive disabilities. By incorporating dynamic difficulty adjustment (DDA) and engagement-driven mechanics, Risk Resist enhances learning retention and supports cognitive accessibility for children with DS, ensuring a safe and effective learning experience.

Emergency preparedness is essential for ensuring the safety and well-being of individuals in crises. However, children with DS face unique cognitive challenges that impair their ability to process emergency information, recall procedures, and react appropriately under stress. Traditional training methods, such as teacher-led instruction or passive learning approaches, fail to engage these children effectively and lack the adaptive mechanisms necessary for personalized crisis management training. These limitations hinder learning retention, reduce motivation, and fail to equip children with practical, real-world emergency response skills [3,4,6].

GBL has demonstrated significant benefits for engagement, knowledge retention, and skill acquisition, particularly for individuals with cognitive disabilities [7]. However, conventional static GBL designs do not automatically adjust difficulty based on individual learning progression, making them less effective for children with DS, whose attention span, cognitive processing speed, and learning abilities vary significantly.

To address these challenges, this study introduces Risk Resist, an adaptive serious game that incorporates DDA to personalize learning and optimize engagement levels. DDA dynamically modifies challenge levels based on player performance metrics, ensuring that learning remains accessible, motivating, and appropriately challenging. By using machine learning techniques, the game adjusts difficulty in real time, preventing frustration from overly difficult scenarios while ensuring continued cognitive engagement and learning progression.

The integration of adaptive mechanics in serious gaming for crisis preparedness represents an innovative, evidence-based approach to enhancing emergency response skills in children with DS. Risk Resist bridges the gap between static training methods and real-world applicability, ensuring effective knowledge transfer while maintaining high levels of engagement and accessibility.

Literature Review

Overview

To identify relevant studies on serious games designed for children with DS, a literature search was conducted using multiple academic databases, including PubMed, Scopus, IEEE Xplore, and Web of Science. The search included publications up to 2024, using combinations of keywords such as “serious games,” “Down syndrome,” “educational games,” “cognitive development,” and “assistive technologies.” Studies were included if they (1) focused on the development, evaluation, or application of serious games for individuals with DS; (2) addressed cognitive, motor, language, or daily living skill development; and (3) were peer-reviewed journal papers or conference proceedings. Exclusion criteria included studies not specifically targeting individuals with DS, nonpeer-reviewed literature, and papers lacking empirical data or detailed descriptions of the game design. This selection process ensured that the reviewed literature was directly relevant to the target population and aligned with the scope of this research.

Serious Games for DS

The use of serious games has been extensively explored as a means to provide individuals with DS essential daily concepts and training for various life situations. For instance, “Junk-Food

XSL•FO
RenderX

Destroyer," a video game developed by Hatzigiannakoglou [9], is designed to help teenagers with DS understand the importance of healthy eating. Similarly, "EnCity," developed specifically for individuals with DS by Bourazeri et al [10], includes a variety of real-life skills such as meal preparation and bill payment, encouraging players to engage without fear of failure.

Brandão et al [11] proposed a video game aimed at aiding preschool children with DS in enhancing their cognitive development. They found that controlling a mouse promoted interaction between the game and the player. However, Macedo [12] noted that while players with DS enjoyed the game, they also experienced boredom and frustration in response to game stimuli. AKA New Media Inc developed online games to teach DS individuals aged 14 to 20 years to use the internet [13]. The authors suggested modifying these games to enhance engagement through auditory reinforcement, animated feedback, and immediate positive support.

González-Ferreras et al [14] conducted usability testing to demonstrate that DS games should adapt to players' skills, incorporate age-appropriate graphics, provide clear instructions, and offer positive feedback. Research on the physical effects of video games on individuals with DS has shown that Wii games can improve sensorimotor skills [15], balance [16], and coordination [17]. Berg et al [17] observed significant improvements in a child's balance and agility after 8 weeks of playing Nintendo Wii, although there were no bimanual coordination improvements [18]. Wuang et al [15] found that Nintendo Wii games motivated players to continue learning, attributing improvements to increased brain plasticity during gaming.

Whalen et al [19] used "TeachTown" to improve language and memory in 4-year-old children with DS, finding significant performance increases after 8 weeks. Lopez-Basterretxea et al [20] provided 3 money-related iPad apps to children with DS, revealing positive interactions and success in counting and recognizing money. Shalash et al [21] designed "No Limit," an interactive game developed using Unity 3D and Kinect 360, to enhance the intellectual and physical abilities of children with DS, showing promising results during testing. The literature review of serious games for individuals with DS is displayed in Table 1, which includes "Research future work or gap" that summarizes the limitations, open issues, or suggested directions for future work identified in each referenced study. Including this column helps to highlight existing gaps in the literature that our own game seeks to address, such as limited real-life emergency preparedness interventions specifically tailored for children with DS.

XSL•FO
RenderX

Table 1. Serious games for children with Down syndrome (DS).

Author (year)	Game name	Purpose	Technique	Findings	Research future work or gap
Hatzigiannakoglou (2015) [9]	Junk-Food Destroyer	Teach teenagers with DS the importance of healthy eating	First-person shooting	Enhanced understanding of healthy eating	The game has not been applied to real users
Bourazeri et al (2017) [10]	EnCity	Teach real-life skills (eg, meal preparation and bill payment) to individuals with DS	Virtual reality serious game	Covered a broad range of daily skills	Continuous engagement with the end users (young people with DS)
Brandão et al (2010) [11]	Jecripe	Aid preschool children with DS in cognitive development	2D computer game	Promoted interaction via mouse control	Observe more users playing the game
Macedo (2015) [12]	Jecripe (evaluation)	Evaluate the previously proposed game	2D computer game	Proved the game's efficiency for children with DS	Apply to more users
Kirijian et al (2007) [13]	Web Fun Central	Teach individuals with DS to use the internet	Online learning tools for DS	Individuals with DS use Web Fun Central across North America	Using the system for different cultural users
González-Ferreras et al (2017) [14]	N/A[a]	Improve voice skills, particularly prosody-related people with DS	Video game	Good level of acceptance by professionals	Include languages other than Spanish and redesign activities to train vocabulary, syntax, and phonetics
Wuang et al (2011) [15]	Wii game (not identified)	Improve motor skills using Nintendo Wii in children with DS aged 10-13 years	Virtual reality or Wii-Fit game	Significant improvements in balance	Apply to more users
Rahman and Rahman (2010) [16]	Wii game (not identified)	Measure the balance of children with DS	Virtual reality or Wii-Fit game	High improvement of equilibrium in the study group	N/A
Berg et al (2012) [17]	Wii game	Improve a child's stability and postural control	Wii game	Need improvements in postural control and motor skills	N/A
Whalen et al (2006) [19]	TeachTown	Improve language and memory in children	Video computer game	Significant performance increases	Apply to more users
Lopez-Basterretxea et al (2014) [20]	iPad game	Count and recognize money for children with DS	OSI[b] games on iPads	Positive interactions, successful money counting, and recognition	N/A
Shalash et al (2018) [21]	No Limit	Enhance motor and cognitive skills for young people	3D Unity game	Promising intellectual and physical enhancement	Add a wider range of coins and notes and design real-life activities like shopping at supermarkets

[a]N/A: not available.

[b]OSI: Open Systems Interconnection.

Crisis and Emergency Management Games

Several crisis and emergency management games have been developed to teach individuals how to handle various disaster scenarios. "StopDisasters" by the International Strategy for Disaster Reduction engages players in activities to manage a specified budget and mitigate potential disaster consequences. Players can choose specific natural disasters such as floods, tsunamis, wildfires, hurricanes, and earthquakes, aiming to allocate resources efficiently to reduce community vulnerability [22].

"Crisis of Nations" [23] is an international civics simulation, where players assume the role of a country and must collaborate

XSL•FO
RenderX

with 3 other nations to resolve issues. Effective resource use is a key to winning the game, and a time limit ensures that the game does not continue indefinitely.

Babu et al [24] discussed learning from disasters through collaborative game-based disaster management training. In a multiplayer game, players face a simulated catastrophe with an incident management structure focusing on cooperation and communication. Students and academic members from Amrita University tested the game, showing that the proposed frameworks successfully provided fundamental postdisaster management knowledge to participants.

Shohieb [25] and Doenyas and Shohieb [26] proposed the "Crises and Disaster Management Game", piloted by 12 students at a government primary school in Mansoura, Egypt. The results revealed that the Crises and Disaster Management Game efficiently provided basic crisis management knowledge to players.

Although these crisis management games have been beneficial for adults and young people, they are not always suitable for children with DS [27,28]. Additionally, these games are typically in English, which can be a barrier for those from Arabic cultures [29]. Consequently, "Risk Resist" was developed specifically for children with DS and features an English or Arabic interface.

Objectives

The objectives of this study are 2-fold. This study introduces Risk Resist, an adaptive serious game designed to enhance emergency preparedness in children with DS. The primary aim is to evaluate the effectiveness of the game in improving both learning outcomes and engagement levels compared to traditional teacher-led training. Specifically, the study tests 2 hypotheses: (hypothesis 1) that children who use Risk Resist will demonstrate significantly greater learning gains than those receiving conventional instruction and (hypothesis 2) that Risk Resist will elicit significantly higher engagement levels. These hypotheses are examined through a controlled experimental design involving pre- and postintervention assessments and teacher-reported engagement metrics.

Methods

Overview

Effective learning for children with DS relies heavily on visual and interactive elements, as research has consistently demonstrated that illustrated content enhances comprehension and retention [30]. Educational approaches that integrate photos, animations, and interactive experiences have been shown to significantly improve learning engagement for individuals with DS, making game-based methodologies an ideal choice for emergency preparedness training.

Recognizing these cognitive needs, Risk Resist was specifically designed to ensure that children with DS could absorb and apply emergency response strategies effectively. Studies highlight that many children with DS struggle with basic life skills [6], which underscores the importance of reinforcing training concepts through adaptive gameplay mechanics. This principle aligns with Freina et al [31], who found that learning transfer—the ability to apply acquired knowledge across different environments—is significantly enhanced through immersive and interactive learning experiences.

Risk Resist Game Design Process and Its Rationale

Concept Development and General Description of Risk Resist

Risk Resist is a GBL tool designed to teach crisis and emergency management in a safe, controlled environment, preventing real-world risks to children with DS. Through engaging gameplay, players can practice methodical crisis response, such as handling indoor fires, in a way that reinforces learning without exposure to danger. The game's systemic framework allows for broad application across different types of emergencies, spanning both human-caused and ecological crises. The game is structured into two main categories: (1) human-caused crises includes electrical fires, first aid emergencies (such as fainting and bleeding), and road-crossing safety; and (2) natural catastrophes teaches essential survival strategies during disasters like earthquakes.

Learning Needs and Cognitive Accessibility

To ensure accessibility for children with DS, Risk Resist incorporates specialist recommendations and evidence-based learning theories [32-36]. The design specifically integrates:

- Visual learning: Vivid graphics, animations, and interactive storytelling engage players and reinforce emergency concepts.
- Repetition and feedback: Scenarios provide repeated exposure and immediate positive reinforcement, enhancing skill retention.
- Simplified instructions: Clear guidance, intuitive controls, and step-by-step instructions ensure accessibility and comprehension for children with DS.

Game Mechanics and Adaptive Learning

The game's homepage emphasizes peer support, fostering collaboration and engagement among students. The mechanics implement ongoing trials, ensuring players grasp key emergency responses through adaptive learning techniques essential for children with cognitive difficulties [7,32]. These features are seamlessly integrated within the Galaxy Shop, making them readily available for use within gameplay. The Galaxy Shop is a virtual store integrated within the game environment, where children can redeem points or rewards earned through gameplay for various customization options. These include avatar personalization, acquisition of accessories, and unlocking of bonus content, all of which are designed to maintain children's interest and enhance their intrinsic motivation through gamification principles such as reward systems, autonomy, and personalization (Figure 1).

- Constructivist learning theory: Active learning where individuals acquire knowledge through experiential engagement rather than passive instruction. This aligns with Risk Resist's interactive gameplay, which enables children to practice crisis management in a simulated environment, reinforcing problem-solving and decision-making skills [37,38].

XSL•FO
RenderX

- Skill transfer model: The concept of learning transfer refers to the ability to apply knowledge gained in one environment to real-world situations. Research indicates that learning transfer improves when activities mimic real-world challenges [30,31]. To promote skill transfer, Risk Resist integrates real-life scenarios, including fire emergencies, first-aid response, and road-crossing safety, ensuring that children develop applicable emergency skills beyond the gaming environment.
- Reinforcement theory: The game incorporates positive reinforcement by providing instant feedback, reward mechanisms, and gradual difficulty progression. According to Skinner's reinforcement theory [39], positive reinforcement (praise, progress tracking, and visual rewards) enhances motivation and learning retention.
- Prototype creation and iterative design are described in detail in Multimedia Appendix 1.

Figure 1. The Galaxy Shop.



Technical Implementation

Risk Resist is built using the Unity 3D engine, ensuring smooth performance and an optimized user experience. The user interface is available with instructions provided in Arabic, catering to Arabic-speaking children with DS and enhancing inclusivity. By combining educational principles, cognitive accessibility, and dynamic engagement strategies, Risk Resist stands out as an effective and innovative tool for enhancing emergency preparedness in children with DS, supporting both their learning and personal development.

DDA Algorithm

Overview

One significant issue in many traditional games is the static difficulty of each level [40,41]. Often, game designers predefine the difficulty level without accounting for the actual performance of each player. This static approach can lead to levels that are either too challenging or too easy for some players. To address this issue, a technique known as DDA [42-45] is used.

DDA automatically adjusts scenarios, parameters, and game behaviors in real time based on the player's skill level. This ensures that the game maintains an optimal balance between challenge and engagement, avoiding boredom (when tasks are too easy) or frustration (when tasks are too difficult). By comparing the player's current performance to an ideal or reference performance (such as the average first-time performance on a level), the game dynamically modifies difficulty settings.

For children with DS, the DDA algorithm incorporates additional behavioral data points, such as response times, success rates, and specific patterns of behavior. These factors help tailor the game to the child's unique skill level, ensuring an experience that is neither overwhelming nor trivial.

Steps for Implementing the DDA Algorithm Using a Random Forest Regressor

To simplify understanding, the implementation steps are outlined as follows:

XSL•FO
RenderX

Step I: Initialization

The process of initialization is as follows:

- Reference baseline: The algorithm starts with a predefined baseline performance measure (eg, the average number of attempts taken by a typical player to complete a level).
- Adjustment rate: An adjustment rate is defined to control how quickly the difficulty level changes based on the player's performance.
- Initial difficulty: The initial difficulty level is set at 0.5, representing a medium difficulty on a scale from 0.0=easiest to 1.0=hardest.

Step II: Performance Calculation

The player's performance is calculated using factors such as:

- Number of attempts to complete the task: how many tries the player needs to succeed.
- Time taken to complete the task: longer times suggest increased difficulty.
- Success rate (percentage of successful attempts): tracks completion accuracy and failure patterns.
- Average response time: indicates reaction speed and cognitive engagement.

Difficulty level adjustment:

- The difficulty level is updated by comparing the player's performance to the reference baseline.
- It is normalized based on the time taken, with lower difficulty levels indicating easier tasks.

Ease level is defined as the complement of the difficulty level and reflects how easy the game feels for the player.

Step III: Premodel Difficulty Adjustment (Before Machine Learning Predictions)

Since the machine learning model (Random Forest Regressor) requires data accumulation before making accurate predictions, the system uses simplified difficulty adjustment logic in the early stages: if the player struggles (low success rate and long response times), the difficulty decreases to enhance accessibility; if the player performs well (high success rate and fast response times), the difficulty increases to provide a greater challenge; and difficulty is constrained within 0.0 to 1.0, ensuring balanced play.

This fallback mechanism prevents extreme difficulty spikes or drops before enough data have been collected for machine learning–based adjustments.

Step IV: Data Collection and Model Training

Once the player completes at least 5 gameplay interactions, the collected performance data are stored to train the Random Forest Regressor, which identifies patterns in player performance, predicts an optimal difficulty level for upcoming gameplay, and gradually replaces the initial simplified logic with machine learning–driven difficulty optimization.

Step V: Machine Learning–Based Difficulty Adjustment

Once the model is trained, difficulty modifications become more precise, ensuring real-time adaptation based on deeper behavioral analysis. The algorithm predicts the ideal difficulty based on player history, refines the adjustment process using progressively more performance data, and maintains engagement levels while preventing frustration and boredom.

Justification for Premodel Adjustment Logic

Justification for premodel adjustment logic is present in Textbox 1.

Textbox 1. Justification for premodel adjustment logic.

- Before the Random Forest Regressor gathers enough data for predictions, the system applies a simplified manual difficulty adjustment mechanism to ensure that players experience appropriate challenge levels from the start.
- This logic determines whether difficulty should increase or decrease based on immediate performance metrics such as number of attempts, success rate, and response time.
- If fewer than 5 gameplay interactions are available, difficulty changes based on predefined rules, preventing extreme difficulty spikes before the machine learning model starts making predictions.
- Once the player completes at least 5 gameplay interactions, sufficient performance data have been collected, allowing the Random Forest Regressor to train.
- The machine learning model replaces manual logic, making predictions based on the child's past performance trends.
- Unlike early-stage adjustments, these predictions continuously refine difficulty adaptation, ensuring optimal challenge levels by detecting deeper behavioral patterns.

Justification for Incorporating Behavioral Data Points

Overview

To personalize and optimize the gaming experience for children with DS, we incorporate additional behavioral data points such as response times, success rates, and specific patterns of behavior. These data points are critical for several reasons (Textbox 2).

XSL•FO
RenderX

Textbox 2. Behavioral data points.

Response times

- Significance: Response time is a valuable indicator of a player's processing speed and reaction ability. For children with Down syndrome, who may have varying cognitive and motor skills, measuring response times helps in understanding how quickly they can process and act on game stimuli [46].
- Use: By analyzing response times, the dynamic difficulty adjustment algorithm can dynamically adjust the game pace to match the player's capabilities, ensuring that tasks are neither too fast nor too slow, thus maintaining an optimal challenge.

Success rates

- Significance: The success rate reflects the player's ability to complete tasks accurately. It provides insight into how well the player understands and executes game objectives [47].
- Use: Monitoring success rates allows the algorithm to identify when a player is consistently succeeding or struggling. This information helps in adjusting the difficulty level to provide a balanced experience that fosters learning and skill development without causing frustration or disengagement.

Specific patterns of behavior

- Significance: Specific patterns of behavior, such as repeated mistakes or preferred strategies, offer a deeper understanding of the player's learning style and problem-solving approach [48].
- Use: By identifying these patterns, the dynamic difficulty adjustment algorithm can tailor game scenarios to align with the player's learning preferences, providing personalized support and encouragement. This ensures that the game remains engaging and educationally effective.

Implementation and Impact

These behavioral data points are used to dynamically adjust the game's difficulty in real time. The use of a Random Forest Regressor, a sophisticated machine learning model, allows the algorithm to predict the optimal difficulty level based on the collected performance data. By continuously adapting the challenge to the player's skill level, the DDA algorithm ensures that the game remains appropriately challenging and engaging. This approach prevents boredom from tasks that are too easy and frustration from tasks that are too difficult, ultimately enhancing the learning and development of children with DS.

The Random Forest Regressor was trained using performance data from gameplay sessions. The input features included average response time, success rate, number of attempts, and error frequency. The target variable was the difficulty score, a continuous value between 0 (easy) and 1 (hard), which represented the ideal challenge level for the next session. A total of 162 valid session logs were collected from the 9 children in the experimental group. To evaluate model performance, we applied 5-fold cross-validation. The following regression metrics were used: mean absolute error (MAE), root-mean-square error (RMSE), and R^2 score.

The pseudocode snippet in Textbox 3 demonstrates how difficulty adjustments occur before the Random Forest Regressor is fully trained. The DDA algorithm is presented in Textbox 4. Figure 2 represents the implementation of the DDA algorithm, illustrating how player performance metrics, such as attempts, success rate, and response time, are processed to dynamically adjust game difficulty. The algorithm ensures an optimized balance between challenge and engagement through real-time adaptation.

XSL•FO
RenderX

Textbox 3. Pseudocode representation of early-stage difficulty adjustments before machine learning predictions take effect.

```
BEGIN DDA Algorithm
//Step I: Initialization
// Set the baseline performance reference
baseline_performance ← average_attempts_to_complete_level
adjustment_rate ← predefined_adjustment_rate
current_difficulty ← 0.5 // Medium difficulty (range: 0.0 to 1.0)
//Step II: Performance Calculation
WHILE game_running DO
// Calculate player performance based on key metrics
player_attempts ← get_number_of_attempts()
time_taken ← get_task_completion_time()
success_rate ← calculate_success_rate(player_attempts)
response_time ← calculate_average_response_time()
// Update difficulty level based on player performance
performance_measure ← player_attempts - baseline_performance
normalized_difficulty ← normalize(performance_measure, time_taken)
current_difficulty ← normalized_difficulty
// Compute ease level (complement of difficulty)
ease_level ← 1.0 - current_difficulty
END WHILE
//Step III: Data Collection
WHILE game_running DO
// Collect and store player performance data
data_collected ← {player_attempts, time_taken, success_rate, response_time, current_difficulty}
store(data_collected)
END WHILE
//Step IV: Model Training
IF collected_data_count > 5 THEN
// Train Random Forest Regressor using collected data
model ← RandomForestRegressor()
model.train(collected_data, difficulty_levels)
END IF
//Step V: Difficulty Adjustment
IF model.is_trained THEN
// Predict difficulty using trained model
predicted_difficulty ← model.predict(current_player_data)
current_difficulty ← clamp(predicted_difficulty, 0.0, 1.0)
ELSE
// Use fallback logic for adjusting difficulty
IF current_difficulty > 0.5 THEN
current_difficulty ← current_difficulty - adjustment_rate // Make easier
ELSE
current_difficulty ← current_difficulty + adjustment_rate // Make harder
```


XSL•FO
RenderX

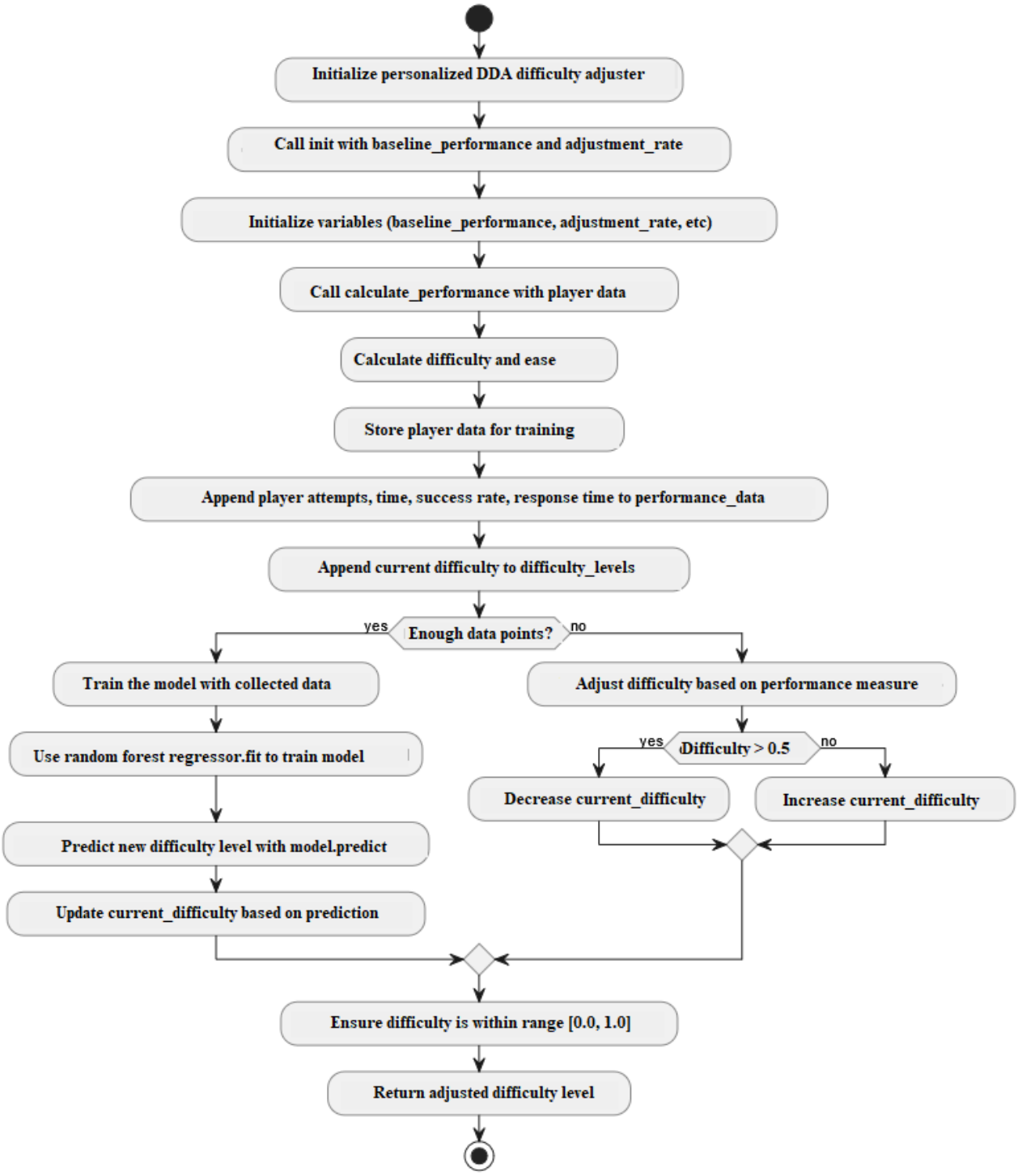
```
END IF
current_difficulty ← clamp(current_difficulty, 0.0, 1.0)
END IF
//Step VI: Implementation
// Use scikit-learn's RandomForestRegressor for model implementation
model.initialize()
model.calculate_performance_measures()
model.train(collected_data)
model.adjust_difficulty(current_difficulty)
END DDA Algorithm
```

Textbox 4. Pseudocode representation of the dynamic difficulty adjustment algorithm.

```
# Initialize difficulty level at a moderate starting point (0.5)
difficulty_level = 0.5
min_difficulty = 0.0
max_difficulty = 1.0
# Function to manually adjust difficulty before ML predictions start
def pre_model_difficulty_adjustment(success_rate, avg_response_time):
global difficulty_level
# Check if the player is struggling
if success_rate < 0.4 or avg_response_time > threshold_time:
difficulty_level -= 0.1 # Decrease difficulty to make tasks easier
elif success_rate > 0.8 and avg_response_time < threshold_time:
difficulty_level += 0.1 # Increase difficulty to make tasks more challenging
# Ensure difficulty remains within valid bounds
difficulty_level = max(min_difficulty, min(difficulty_level, max_difficulty))
return difficulty_level
# Example usage before machine learning model is trained
new_difficulty = pre_model_difficulty_adjustment(player_success_rate, player_avg_response_time)
```


XSL•FO
RenderX

Figure 2. The DDA algorithm flow diagram. DDA: dynamic difficulty adjustment.



Experimental Design

To assess the effectiveness of Risk Resist, a controlled experimental design was used, comparing GBL with traditional human-based training. The study aimed to evaluate learning gains and engagement outcomes, structured around the following hypotheses:

- Null hypothesis (hypothesis 0): No significant difference in learning outcomes or engagement between Risk Resist and traditional human-based training.
- Hypothesis 1 (learning gain hypothesis): Children who play Risk Resist will demonstrate significantly higher learning gains compared to those receiving traditional human-based training.
- Hypothesis 2 (engagement level hypothesis): Children who play Risk Resist will exhibit significantly higher engagement levels compared to those receiving traditional human-based training.

A between-group experimental design was used, with participants divided equally into treatment and control groups (n=9 each). The control group received traditional teacher-led

XSL•FO
RenderX

training for the same emergency scenarios presented in the game, while the experimental group engaged in all Risk Resist gameplay sessions. The main group used laptops to engage with Risk Resist, while the control group underwent teacher-led training. The sample consisted of 18 children with DS, aged 8 to 12 years, recruited from special needs facilities and schools. Selection criteria emphasized functional living skills, such as the ability to use the toilet and recognize objects, with the severity of DS being a more critical criterion than age.

The Baseline Equivalence Verification

To ensure the internal validity of the study and account for the relatively small sample size (n=18), a baseline equivalence test was conducted to verify that the experimental (Risk Resist) and control (traditional training) groups did not differ significantly in their initial emergency preparedness knowledge before the intervention. An independent samples 2-tailed *t* test was performed on the pretest scores collected from both groups. Testing was conducted individually in classrooms over 2 weeks. Teachers participated in the evaluation phase by completing observation-based questionnaires designed to measure learning outcomes and engagement levels.

The study used various experimental materials, including pretest and posttest surveys, participant performance assessments, and parent consent forms. Testing was conducted individually in classrooms over a 2-week period. Teachers participated in the evaluation phase by completing observation-based questionnaires designed to measure learning outcomes and engagement levels.

Procedures for Administering the Learning Test

Pretest and Posttest Overview

Participants were given identical pretests and posttests comprising 5 questions covering emergency scenarios (Multimedia Appendices 2 and 3). These tests aimed to measure learning gains and were presented in Arabic for accessibility.

Role of Teacher Help

Teachers were available to assist participants in understanding the test questions. Their assistance was limited to rephrasing or explaining the questions in simpler terms without providing any answers. This ensured that participants understood the instructions while maintaining the integrity of the test.

Scoring Methodology

To evaluate engagement levels, a Mann-Whitney *U* test was conducted due to the ordinal nature of the data. Learning gains were calculated using the difference between the number of correct answers in the posttest and pretest:

Learning gain=posttest score–pretest score **(1)**

Method for Calculating Engagement Scores

A 5-point Likert scale questionnaire evaluated engagement factors like attention, motivation, and effort during gameplay or training. Teachers completed the questionnaire based on their observations of participants. The total engagement score for each participant was derived by summing their responses across all items. Scores were normalized to allow consistent comparison between groups.

Validation of Instruments and Testing

Learning Test Validation

Participants were asked 5 questions regarding crisis and emergency scenarios during both the pretest and posttest phases. Learning gains were calculated as the difference between correct answers in the pretest and posttest, as shown in equation 2:

Learning gain=correct answers in the posttest–correct answers in the pretest **(2)**

The questions asked of the children are presented in Textbox 5.

Textbox 5. The questions asked for each child in the pre- and posttests (learning gain).

Questions
1. What will you do if there is a fire caused by electricity? What is the difference between a regular fire and a fire caused by electricity?
2. Can you give quick help to anyone who faints in front of you?
3. What can you do if you or the other one in front of you has bleeding because of an injury?
4. Can you cross the road? What do you do while crossing the road?
5. If there is an earthquake, what will you do? (with a quick explanation what is the meaning of earthquake)

Engagement Questionnaire Validation

Pilot testing with a small sample of children ensured the reliability and suitability of the game for its target demographic. To assess the effectiveness of the game, learning outcomes were compared between the 2 groups by analyzing the average learning gains of participants in each condition. Engagement levels were measured using a 5-point Likert scale questionnaire [40], evaluating factors related to control and enjoyment [41].

After the session, teachers were provided hard copies of the questionnaire and instructed to complete them based on their observations of children's engagement and learning behaviors. An engagement questionnaire presented to the teachers upon their observations of participants (Table 2). All questions were originally presented in Arabic and translated into English for clarity and readability. Engagement levels were assessed through the Likert scale questionnaire, measuring factors such as attention, effort, motivation, and enjoyment.

XSL•FO
RenderX

Table 2. Engagement questionnaire presented to the teachers upon their observations of participants.

	Question	1=very poor	2=poor	3=fair	4=good	5=very good
1.	To what extent did the game or training hold the child's attention?					
2.	How much effort did the child put into playing the game or getting the training?					
3.	Does the game or training motivate the child to learn?					
4.	The child was motivated to get involved in the training or game.					
5.	I believe the game or training has improved the child's understanding of the covered topics.					
6.	To what extent the child was tired after playing the game?					
7.	How well does the game maintain the child's interest over time?					
8.	Was the total duration of the game or training satisfactory to understand the required scenarios?					
9.	Would the child like to play the game or get training again?					
10.	How well do you think the child performed in the game or training?					

Correlation of Learning Gain and Engagement

To assess the relationship between learning gains and engagement, a Pearson correlation coefficient (r) was calculated using the participants' scores in both variables using the formula:



where x_i is the learning gain for each participant,  is the mean of learning gains, y_i is the engagement score for each participant, and  is the mean of engagement scores.

Ethical Considerations

This study involving human participants was conducted following the ethical standards of the institutional review board at the Faculty of Computers and Information, Mansoura University, Egypt (protocol 202110016). Parents of all participating children provided written informed consent before enrollment, authorizing both participation in the study and the subsequent use of data for research and publication purposes. Participation was entirely voluntary, and no monetary or material compensation was offered to participants or their families. To protect participant privacy and confidentiality, all collected data were anonymized and deidentified prior to analysis; no personally identifiable information was stored or shared. Additionally, no identifiable images of participants are included in the manuscript or supplementary materials. The study was conducted following ethical standards outlined in the Declaration of Helsinki.

Results

Learning Gain

The results demonstrated a significant difference, with the experimental group (mean 3.6, SD 0.5) outperforming the control group (mean 2.0, SD 0.5). The test statistic (t_{16}=5.35), with a P value <.001, confirms the effectiveness of Risk Resist in enhancing emergency preparedness.

- The null hypothesis (hypothesis 0) was rejected.
- The alternative hypothesis (hypothesis 1) was supported, demonstrating that Risk Resist significantly improves learning compared to traditional training.

The Baseline Equivalence Verification

The results revealed no statistically significant difference between the experimental group (mean 1.78, SD 0.67) and the control group (mean 1.67, SD 0.71; t_{16}=0.30; P=.77).

Engagement Test

To evaluate engagement levels, a Mann-Whitney U test was conducted due to the ordinal nature of the data. The experimental Risk Resist group (mean 4.54, SD 0.4) had significantly higher engagement than the control group (mean 4.01, SD 0.32). The test statistic (U=12), with a P value of .002, confirms that Risk Resist enhances engagement more effectively than traditional methods.

- The null hypothesis (hypothesis 0) was rejected.
- The alternative hypothesis (hypothesis 2) was supported, confirming that Risk Resist significantly increases engagement compared to traditional training.

This result indicates that Risk Resist is more engaging for children with DS, leading to higher levels of motivation and involvement.

The results indicate that participants who underwent the Risk Resist game-based training achieved significantly higher outcomes in both learning gain and engagement compared to those in traditional human-based training. Specifically, the mean learning gain for Risk Resist was 3.6 (SD 0.5), which was substantially higher than the mean of 2.0 (SD 0.5) for the traditional method, with a highly significant difference (P<.001). Engagement levels were also greater in the game-based condition, with a mean of 4.54 (SD 0.4), compared to 4.01 (SD 0.32) for traditional training, and this difference was statistically significant (P=.002). Overall, the data demonstrate that Risk Resist not only enhanced participants' knowledge acquisition but also fostered higher engagement during the learning process.

XSL•FO
RenderX

Table 3 represents learning gain and engagement correlation (game-based training vs human-based training scenarios).

Table 3. Learning gain and engagement correlation (game-based training vs human-based training scenarios).

Measure	Engagement level		Learning gain	
	Risk Resist	Human-based	Risk Resist	Human-based
All participants, scores spread	3.8, 4.0, 4.1, 4.3, 4.4, 4.6, 4.8, 5.0	3.7, 3.8, 3.9, 4.0, 4.1, 4.2, 4.3, 4.4	2.8, 3.0, 3.1, 3.3, 3.4, 3.6, 3.8, 4.0	2.0, 2.1, 2.3, 2.5, 2.6, 2.8, 3.0
Mean (SD)	~4.6 (~0.3)	~4.0 (~0.4)	~3.7 (~0.35)	~2.0 (~0.45)
SE	~0.1	~0.15	~0.1	~0.2

Model Evaluation

To assess the effectiveness of the machine learning–based DDA system embedded in Risk Resist, we conducted a comprehensive evaluation of the Random Forest Regressor responsible for adjusting game difficulty in response to player performance. The evaluation focused on 3 key areas: predictive accuracy, convergence stability, and adaptive usability.

Model Performance and Predictive Accuracy

The model was trained on a dataset comprising 162 valid gameplay session logs collected from 9 children in the experimental group. Input features included success rate (% of correct responses), average response time (in seconds), number of failed attempts, and task completion time.

The target variable was a normalized difficulty score ranging from 0.0 to 1.0, representing the ideal challenge level for the player's next session.

Using 5-fold cross-validation, the model achieved the following average performance metrics: MAE=0.062, RMSE=0.089, and coefficient of determination (R^2)=0.81. These results indicate that the model demonstrated high predictive accuracy and stability across varied input conditions.

Feature Importance

A feature importance analysis of the trained model revealed that the most influential predictors were success rate (43%), average response time (31%), number of failed attempts (21%), and task completion time (5%). This indicates that the model's decision logic was grounded in core behavioral and cognitive indicators, supporting its educational alignment.

Convergence and Stability

To evaluate the minimum data required for reliable performance, we analyzed how model accuracy improved with increased session data per participant. The model exhibited stable convergence after approximately 5-7 gameplay sessions: R^2=0.49 after 2 sessions, R^2=0.78 after 5 sessions, and R^2=0.81 after 7 sessions. This pattern suggests the model reaches effective learning capacity within a manageable number of short sessions, suitable for special education settings.

Simulation-Based Usability Testing

To further validate the adaptive behavior of the model, we simulated 3 hypothetical player profiles—novice, intermediate, and advanced—based on predefined performance inputs. The model assigned difficulty scores are present in Table 4.

The difficulty levels scaled appropriately with performance, supporting the model's real-time personalization capabilities.

Table 4. The model assigned difficulty scores.

Profile	Success rate (%)	Response time (seconds)	Predicted difficulty
Novice	40	6.8	0.28
Intermediate	70	3.5	0.54
Advanced	90	1.8	0.81

Case Examples From Participants

Two representative participants from the experimental group further illustrate the model's effectiveness:

- Participant 7 began with low accuracy and long response times. The model initially reduced difficulty, resulting in improved performance. As the child's scores stabilized, difficulty levels were gradually increased.
- Participant 2 showed consistently high success rates and fast response times. The model responded by increasing difficulty levels early in gameplay, which were sustained throughout the sessions.

These examples demonstrate the model's responsiveness to individual learning trajectories and its ability to dynamically modulate challenge without manual intervention.

A full breakdown of feature importance scores, simulation results, and participant case examples is provided in Multimedia Appendix 4.

Correlation of Learning Gain and Engagement

The correlation analysis revealed a strong positive relationship between learning gains and engagement levels (r=0.85; P<.001). This indicates that higher engagement levels were associated with greater learning gains. The experimental Risk Resist group, which exhibited higher engagement levels, also demonstrated

XSL•FO
RenderX

significantly better learning outcomes. In contrast, the control group displayed lower engagement levels and correspondingly lower learning gains.

Discussion

Principal Findings

Advancing skill acquisition through technological platforms is essential for improving education accessibility for children with DS, who often face high cognitive demands. Numerous studies have explored the use of serious games to support learning and development, demonstrating their effectiveness in engaging learners through interactive environments. Games such as "Junk-Food Destroyer" [9] and "EnCity" [10] have successfully helped individuals with DS acquire essential life skills, including meal preparation and financial management. These digital tools have proven beneficial by enabling repeated practice without the fear of failure, fostering confidence, and skill mastery.

Despite the success of these games in daily life training, emergency preparedness education for children with DS remains an underdeveloped area. Traditional human-based training methods, though widely used, often lack adaptability and fail to account for individual learning progressions and engagement levels. Accessibility barriers within special education, along with the burdens placed on caregivers, highlight the need for adaptive, personalized learning approaches. To address these limitations, this study introduces Risk Resist, an adaptive serious game designed to improve emergency preparedness among children with DS. The game leverages a personalized DDA algorithm, ensuring that the challenge level adapts to each child's abilities and engagement patterns.

The results provide strong evidence that GBL surpasses traditional training methods in both learning acquisition and engagement. Findings from the study confirm the rejection of the null hypothesis (hypothesis 0), which proposed no difference in learning outcomes or engagement levels. In contrast, results strongly support the first hypothesis (hypothesis 1), demonstrating that Risk Resist significantly enhances learning outcomes compared to traditional training methods.

Similarly, engagement levels among participants revealed a clear advantage in favor of Risk Resist. Traditional instruction, despite its structured approach, did not maintain consistent engagement, particularly for children who require stimulating and interactive learning environments. Statistical analyses led to the rejection of the second null hypothesis (hypothesis 2), confirming that children using Risk Resist were significantly more engaged than those undergoing conventional training. These findings emphasize the interactive nature of adaptive GBL as a key factor in sustaining motivation and involvement, reinforcing the benefits of personalized gaming experiences in crisis preparedness training.

A critical component of these outcomes is the personalized DDA algorithm, which dynamically modifies difficulty levels based on real-time behavioral metrics such as response times, success rates, and task completion patterns. By continuously adjusting challenges to align with individual capabilities, the algorithm prevents boredom (from tasks that are too easy) and frustration (from tasks that are too difficult), ensuring consistent engagement and optimal learning progression. This approach tailors emergency preparedness education to the specific learning needs of children with DS, bridging gaps left by static, noninteractive training methods. This confirms that both groups were comparable at baseline and that any observed differences in postintervention outcomes are likely attributable to the instructional method rather than pre-existing disparities. The study used various experimental materials, including pretest and posttest surveys, participant performance assessments, and parent consent forms.

Risk Resist stands out as the first game dedicated to emergency preparedness for individuals with DS and the first in this field to incorporate an adaptive DDA framework. By personalizing difficulty in real time, the game ensures effective knowledge retention, skill development, and engagement, addressing the shortcomings of traditional education models.

The trained Random Forest Regressor achieved an MAE of 0.062, an RMSE of 0.089, and an R^2 score of 0.81, indicating good model fit and reliable performance. The model showed stable convergence after approximately 5-7 gameplay sessions per participant, consistent with design expectations. Since the model predicts continuous difficulty values, classification metrics such as area under the receiver operating characteristic curve, precision, and recall were not applicable.

Correlation of Learning Gain and Engagement

Participants in the experimental group (Risk Resist), who exhibited higher engagement levels, also demonstrated greater learning improvements compared to the control group trained through traditional methods. The control group, which showed lower engagement, exhibited limited learning progress, reinforcing the importance of interactive and adaptive learning environments in emergency preparedness training for children with DS.

Implications and Integration of Risk Resist Into Educational Frameworks

The findings of this study suggest multiple applications for Risk Resist within both educational and special-needs learning environments. Risk Resist can be incorporated into existing special education curricula as a complementary tool that reinforces theoretical emergency training with interactive, practical experiences [6]. To support effective implementation, professional development sessions and educator training workshops should be introduced to equip teachers with the skills necessary to integrate the game into classroom instruction [31]. The broader use of serious games such as Risk Resist also fosters inclusive and engaging learning environments, accommodating diverse cognitive profiles and aligning with established adaptive education frameworks [32,33]. From a policy and accessibility standpoint, decision makers are encouraged to prioritize the adoption of adaptive serious games in special education programs. This includes allocating appropriate funding and resources to support digital innovation, technology-driven learning, and the equitable delivery of educational content for students with cognitive disabilities [14]. Collectively, these initiatives confirm that serious games represent a powerful tool

XSL•FO
RenderX

for enabling children with DS to develop essential life skills in a controlled, structured, and motivating environment.

Game Performance in Real-Time Learning Contexts

In evaluating the game's real-time performance, several technical factors were considered to ensure a smooth and effective learning experience for children with DS. The system demonstrated minimal latency during gameplay, with an average response time of less than 150 milliseconds, ensuring that feedback and adaptive difficulty adjustments were delivered promptly. Loading times for each game level were kept under 3 seconds to maintain engagement and avoid frustration. Error handling mechanisms were implemented to address unexpected user inputs or hardware interruptions, allowing the game to resume seamlessly without significant disruption. Moreover, the game was tested on various hardware platforms with differing specifications to assess adaptability and performance consistency. The results indicated stable functionality across devices, which is critical for ensuring accessibility in diverse real-world settings such as homes, schools, and therapy centers. These performance considerations contribute to maintaining the user's cognitive focus and engagement, which are essential for effective skill acquisition and learning transfer in this population.

Comparison With Similar Studies

Numerous serious games have been developed to support individuals with DS, targeting various domains such as cognitive development, social skills, and physical coordination. For example, EnCity [10] was designed to teach real-life activities like meal preparation and bill payment, promoting independence in everyday tasks. TeachTown [19] focused on improving language and memory, while No Limit [21] aimed to enhance motor and cognitive skills using interactive 3D gameplay. While these games demonstrated positive outcomes in their respective domains, none directly addressed emergency preparedness, a critical and time-sensitive skill area for individuals with cognitive disabilities.

Moreover, Risk Resist is distinct in its integration of a personalized DDA algorithm, which enables real-time tailoring of difficulty levels based on individual behavioral metrics such as success rates and response times. This adaptive feature ensures an optimized balance between challenge and engagement—an innovation not found in the previously mentioned games. By combining emergency response training with adaptive machine learning techniques, Risk Resist introduces a novel and comprehensive framework for empowering children with DS in crises. This dual focus on both content innovation and technological advancement positions Risk Resist as a pioneering contribution within the field of serious games for special education, as shown in Table 5.

Table 5. Comparison of Risk Resist with similar serious games for individuals with Down syndrome.

Game name	Target skills	Emergency preparedness	Adaptive difficulty	Platform
EnCity [10]	Daily living skills	No	No	Virtual reality
TeachTown [19]	Language and memory	No	No	Desktop
No Limit [21]	Motor and cognitive enhancement	No	No	3D, Kinect
Risk Resist	Emergency response and cognition	Yes	Yes (DDA[a] algorithm)	Unity (Arabic or English)

[a]DDA: dynamic difficulty adjustment.

Limitations

This study has several limitations. First, the sample size (n=18) is relatively small, which limits the statistical power and generalizability of the findings. Additionally, all participants were recruited from a specific geographic region, and the study was limited to children aged 8 to 12 years, potentially narrowing the applicability of results across broader populations or age ranges. Some technical limitations were also noted, such as delayed keypress registration, which may have influenced task performance for children with motor skill challenges.

Future Directions

Future research should focus on replicating this study with a larger and more diverse population across different cultural and linguistic contexts. It would also be beneficial to conduct longitudinal studies to evaluate long-term retention of emergency skills and assess the transferability of knowledge to real-life situations. Further refinement of the game's usability and expansion to cover a broader range of life skills (eg, health management and social interaction) are recommended to enhance educational value.

Recommendations

While the present findings are preliminary, they indicate the potential of adaptive serious games in special education. Education authorities and curriculum developers may consider pilot-testing similar tools in special-needs classrooms. However, any policy-level implementation should be based on larger-scale evidence and stakeholder engagement. Training workshops for teachers and caregivers could support integration into structured learning environments, contingent on further validation.

Conclusions

The integration of serious games into educational systems for children with DS is an increasingly important research area, offering interactive and engaging platforms to enhance learning experiences. This study introduced Risk Resist, a novel serious game designed to improve emergency preparedness skills for children with DS. The game leverages a personalized DDA algorithm, ensuring that the learning experience is tailored to each child's cognitive abilities and engagement level.

A key contributor to these improved outcomes was the DDA algorithm, which dynamically adjusted game difficulty based

XSL•FO
RenderX

on behavioral metrics, such as response times, success rates, and learning progression. This personalization ensured that each child remained challenged at an optimal level, thereby enhancing both learning outcomes and engagement in crisis preparedness training.

Acknowledgments

The authors would like to thank Prince Sattam Bin Abdulaziz University for funding this project. This study is supported via funding from Prince Sattam Bin Abdulaziz University project number (PSAU/2025/R/1446).

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Authors' Contributions

All authors were equally responsible for the analysis and modeling, as well as the preparation and presentation of this manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Game design.
[DOCX File , 1671 KB - games_v13i1e73690_app1.docx]

Multimedia Appendix 2
The questionnaires in Arabic language.
[DOCX File , 13 KB - games_v13i1e73690_app2.docx]

Multimedia Appendix 3
The questionnaires translated in English language.
[DOCX File , 12 KB - games_v13i1e73690_app3.docx]

Multimedia Appendix 4
Model evaluation and usability validation.
[DOCX File , 22 KB - games_v13i1e73690_app4.docx]

References

1. Facts, myths, & truths about Down syndrome. National Down Syndrome Society (NDSS). URL: http://ndss.org/myths-truths [accessed 2025-09-27]
2. What is the Down's syndrome association? Down Syndrome Association (DSA). URL: https://www.downs-syndrome.org.uk/about-dsa/ [accessed 2025-09-27]
3. Oelwein PL. Teaching Reading to Children with Down Syndrome: A Guide for Parents and Teachers. Bethesda, MD: Woodbine House; 1995:1-371.
4. Kokol P, Vošner HB, Završnik J, Vermeulen J, Shohieb S, Peinemann F. Serious game-based intervention for children with developmental disabilities. Curr Pediatr Rev 2020;16(1):26-32. [doi: 10.2174/1573396315666190808115238] [Medline: 31393252]
5. Henke S, Widera A, Hellingrath B. Evaluation-driven disaster management exercises: a collaborative toolkit. 2023 Presented at: Proceedings of the 20th International ISCRAM Conference; May 28, 2023; Omaha, NE, United States p. 45-52 URL: https://tinyurl.com/5n853twd [doi: 10.59297/PSIX3279]
6. Faragher R, Brown RI. Numeracy for adults with Down syndrome: it's a matter of quality of life. J Intellect Disabil Res 2005;49(Pt 10):761-765. [doi: 10.1111/j.1365-2788.2005.00747.x] [Medline: 16162123]
7. Hanus MD, Fox J. Assessing the effects of gamification in the classroom: a longitudinal study on intrinsic motivation, social comparison, satisfaction, effort, and academic performance. Comput Educ 2015;80:152-161. [doi: 10.1016/j.compedu.2014.08.019]
8. Reimagining our futures together: a new social contract for education. United Nations Educational, Scientific and Cultural Organization. 2021. URL: https://unesdoc.unesco.org/ark:/48223/pf0000379707 [accessed 2025-09-10]

XSL•FO
RenderX

9. Hatzigiannakoglou P. Junk-food destroyer: helping adolescents with Down syndrome to understand healthy eating through serious game. 2015 Presented at: 7th International Conference on Games and Virtual Worlds for Serious Applications (VS-Games); September 16-18, 2015; Skovde, Sweden p. 1-5. [doi: 10.1109/vs-games.2015.7295778]
10. Bourazeri A, Bellamy-Wood T, Arnab S. EnCity: a serious game for empowering young people with Down's syndrome. 2017 Presented at: IEEE 5th International Conference on Serious Games and Applications for Health (SeGAH); April 2-4, 2017; Perth, Australia p. 1-6. [doi: 10.1109/segah.2017.7939267]
11. Brandão A, Trevisan DG, Brandão L, Moreira B, Nascimento G, Vasconcelos CN. Semiotic inspection of a game for children with Down syndrome. 2010 Presented at: Brazilian Symposium on Games and Digital Entertainment; November 8-10, 2010; Florianopolis, Brazil p. 199-210. [doi: 10.1109/sbgames.2010.24]
12. Macedo E. Base Nacional Comum para Currículos: direitos de aprendizagem e desenvolvimento para quem? [Common National Curricula Base: learning and development rights for whom?]. Educ Soc 2015;36(133):891-908. [doi: 10.1590/es0101-73302015155700]
13. Kirijian A, Myers M, Charland S. Web fun central: online learning tools for individuals with Down syndrome. In: Universal Usability: Designing Computer Interfaces for Diverse User Populations. Hoboken, NJ: Wiley; 2007:195-230.
14. González-Ferreras C, Escudero-Mancebo D, Corrales-Astorgano M, Aguilar-Cuevas L, Flores-Lucas V. Engaging adolescents with Down syndrome in an educational video game. Int J Hum-Comput Interact 2017;33(9):693-712. [doi: 10.1080/10447318.2017.1278895]
15. Wuang Y, Chiang C, Su C, Wang C. Effectiveness of virtual reality using Wii gaming technology in children with Down syndrome. Res Dev Disabil 2011;32(1):312-321. [doi: 10.1016/j.ridd.2010.10.002] [Medline: 21071171]
16. Rahman SA, Rahman A. Efficacy of virtual reality-based therapy on balance in children with Down syndrome. World Appl Sci J 2010;10(3):254-261 [FREE Full text]
17. Berg P, Becker T, Martian A, Primrose KD, Wingen J. Motor control outcomes following Nintendo Wii use by a child with Down syndrome. Pediatr Phys Ther 2012;24(1):78-84. [doi: 10.1097/PEP.0b013e31823e05e6] [Medline: 22207475]
18. Spanò M, Mercuri E, Randò T, Pantò T, Gagliano A, Henderson S, et al. Motor and perceptual-motor competence in children with Down syndrome: variation in performance with age. Eur J Paediatr Neurol 1999;3(1):7-13. [doi: 10.1053/ejpn.1999.0173] [Medline: 10727186]
19. Whalen C, Liden L, Ingersoll B, Dallaire E, Liden S. Behavioral improvements associated with computer-assisted instruction for children with developmental disabilities. J Speech Lang Pathol Appl Behav Anal 2006;1(1):11-26. [doi: 10.1037/h0100182]
20. Lopez-Basterretxea A, Mendez-Zorrilla A, Garcia-Zapirain B. A telemonitoring tool based on serious games addressing money management skills for people with intellectual disability. Int J Environ Res Public Health 2014;11(3):2361-2380 [FREE Full text] [doi: 10.3390/ijerph110302361] [Medline: 24573223]
21. Shalash WM, AlTamimi S, Abdu E, Barom A. No limit: a down syndrome children's educational game. 2018 Presented at: IEEE Games, Entertainment, Media Conference (GEM); August 15-17, 2018; Galway, Ireland p. 352-358. [doi: 10.1109/gem.2018.8516519]
22. Play and learn to stop disasters. United Nations Office for Disaster Risk Reduction. 2022. URL: https://www.undrr.org/children-and-youth/stop-disasters-game [accessed 2025-09-27]
23. Crisis of Nations by iCivics. 2022. URL: https://www.icivics.org/games/crisis-nations [accessed 2025-09-27]
24. Babu SK, McLain ML, Bijlani K, Jayakrishnan R, Bhavani RR. Collaborative game-based learning of post-disaster management: serious game on incident management frameworks for post-disaster management. 2016 Presented at: IEEE Eighth International Conference on Technology for Education (T4E); December 2-4, 2016; Mumbai, India p. 80-87. [doi: 10.1109/t4e.2016.024]
25. Shohieb SM. CDMG: crises and disaster management game (video game to teach Arab children how to handle emergencies and crises). 2018 Presented at: ICSCA '18: Proceedings of the 2018 7th International Conference on Software and Computer Applications; February 8-10, 2018; Kuantan, Malaysia p. 364-368. [doi: 10.1145/3185089.3185147]
26. Doenyas C, Shohieb SM. Leveraging technology for the wellbeing of individuals with autism spectrum disorder and their families during Covid-19. Front Psychiatry 2021;12:566809. [doi: 10.3389/fpsyt.2021.566809] [Medline: 34262486]
27. Juntunen E, Kalla C, Widera A, Hellingrath B. Digitalization potentials and limitations of cash-based assistance. Int J Disaster Risk Reduct 2023;97:104005. [doi: 10.1016/j.ijdrr.2023.104005]
28. Grirk S. Computer games for learning crisis management skills. Learning 4 Security (L4S) Project. 2024. URL: https://cordis.europa.eu/article/id/91288-computer-games-for-learning-crisis-management-skills [accessed 2025-09-10]
29. Ignjatovi D, Karagiannis V, Chettakattu A, Havlik D, Neubauer G. Crisis management in the era of the IoT, edge computing, and LLMs. 2024 Presented at: 11th International Conference on Internet of Things: Systems, Management and Security (IOTSMS); September 2-5, 2024; Malmö, Sweden p. 45-52. [doi: 10.1109/iotsms62296.2024.10710254]
30. Ortega‐Tudela JM, Gómez‐Ariza CJ. Computer‐assisted teaching and mathematical learning in Down syndrome children. J Comput Assist Learn 2006;22(4):298-307. [doi: 10.1111/j.1365-2729.2006.00179.x]
31. Freina L, Busi M, Canessa A, Caponetto I, Ott M. Learning to cope with street dangers: an interactive environment for intellectually impaired. 2014 Presented at: EDULEARN14 Conference; July 8-9, 2014; Barcelona, Spain p. 972-980 URL: https://iris.cnr.it/handle/20.500.14243/251861

XSL•FO
RenderX

32. Kim B, Park H, Baek Y. Not just fun, but serious strategies: using meta-cognitive strategies in game-based learning. Comput Educ 2009;52(4):800-810. [doi: 10.1016/j.compedu.2008.12.004]
33. Feng J, Lazar J, Kumin L, Ozok A. Computer usage by children with Down syndrome. ACM Trans Access Comput 2010;2(3):1-44. [doi: 10.1145/1714458.1714460]
34. Faragher R, Clarke B. Educating Learners with Down Syndrome: Research, Theory, and Practice. London: Routledge; 2013.
35. Guidelines for inclusive education. National Down Syndrome Society (NDSS). 2021. URL: https://ndss.org/Guidelines_for_Inclusive_Education.pdf [accessed 2025-09-10]
36. Recommendations for the education of learners with Down syndrome. Down Syndrome International. 2021. URL: https://ds-int.org/recommendations-for-the-education-of-learners-with-down-syndrome/ [accessed 2025-09-10]
37. Vygotsky LS. Mind in Society: The Development of Higher Psychological Processes. Cambridge: Harvard University Press; 1978.
38. Piaget J. Part I: cognitive development in children: Piaget development and learning. J Res Sci Teach 2006;2(3):176-186. [doi: 10.1002/tea.3660020306]
39. Skinner BF. Science and Human Behavior. New York, NY: Macmillan; Jun 1953.
40. Pearce J, Ainley M, Howard S. The ebb and flow of online learning. Comput Hum Behav 2005;21(5):745-771. [doi: 10.1016/s0747-5632(04)00036-6]
41. Salah J, Abdennadher S, Sabty C, Abdelrahman Y. Super Alpha: Arabic alphabet learning serious game for children with learning disabilities. 2016 Presented at: Proceedings of the Joint International Conference on Serious Games; November 7-8, 2024; New York City, NY, United States p. 104-115. [doi: 10.1007/978-3-319-45841-0_9]
42. Sepulveda GK, Besoain F, Barriga NA. Exploring dynamic difficulty adjustment in video games. 2019 Presented at: IEEE CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON); November 13-27, 2019; Valparaiso, Chile p. 1-6. [doi: 10.1109/chilecon47746.2019.8988068]
43. Eid MM, El-Kenawy ESM, Ibrahim A. A binary sine cosine-modified whale optimization algorithm for feature selection. 2021 Presented at: National Computing Colleges Conference (NCCC); March 27-28, 2021; Taif, Saudi Arabia p. 1-6. [doi: 10.1109/nccc49330.2021.9428794]
44. El-Kenawy EM, Mirjalili S, Ghoneim SSM, Eid MM, El-Said M, Khan ZS, et al. Advanced ensemble model for solar radiation forecasting using sine cosine algorithm and Newton's laws. IEEE Access 2021;9:115750-115765. [doi: 10.1109/access.2021.3106233]
45. Shohieb SM, Doenyas C, Elhady AM. Dynamic difficulty adjustment technique-based mobile vocabulary learning game for children with autism spectrum disorder. Entertain Comput 2022;42:100495. [doi: 10.1016/j.entcom.2022.100495]
46. Mortazavi F, Moradi H, Vahabie A. Dynamic difficulty adjustment approaches in video games: a systematic literature review. Multimed Tools Appl 2024;83(35):83227-83274. [doi: 10.1007/s11042-024-18768-x]
47. Rahimi M, Moradi H, Vahabie AH, Kebriaei H. Continuous reinforcement learning-based dynamic difficulty adjustment in a visual working memory game. ArXiv Preprint posted online on August 24, 2023. [doi: 10.48550/arXiv.2308.12726]
48. Fisher N, Kulshreshth AK. Exploring dynamic difficulty adjustment methods for video games. Virtual Worlds 2024;3(2):230-255. [doi: 10.3390/virtualworlds3020012]

Abbreviations

DDA: dynamic difficulty adjustment
DS: Down syndrome
GBL: game-based learning
MAE: mean absolute error
RMSE: root-mean-square error

Edited by A Coristine; submitted 15.03.25; peer-reviewed by A Sharaf, T Mentler, ST Carvalho, G Fan; comments to author 14.05.25; revised version received 26.06.25; accepted 30.08.25; published 17.10.25.

Please cite as:
Shohieb SM, Bakhit SH, Mohammed E, M Elhady A
Empowering Children With Down Syndrome by Enhancing Emergency Preparedness Through Serious Games: Quasi-Experimental Study With a Between-Group Design
JMIR Serious Games 2025;13:e73690
URL: https://games.jmir.org/2025/1/e73690
doi:10.2196/73690
PMID:

XSL•FO
RenderX

©Samaa M Shohieb, Suzan Hassan Bakhit, Ensaf Mohammed, Abdelghafar M Elhady. Originally published in JMIR Serious Games (https://games.jmir.org), 17.10.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

An Integrated Model for Circular Waste Management Using the Internet of Things, Semantic Web, and Gamification (Circonomy): Case Study in Indonesia

Vitri Tundjungsari[1], MSc, PhD; Bambang Purnomosidi Dwi Putranto[2], MMSI, PhD; Muhamad Bahrul Ulum[3], MKom; Nizirwan Anwar[3], MT

[1]Department of Master in Computer Science, Faculty of Computer Science, Universitas Esa Unggul, Jl Arjuna Utara 9, Kebon Jeruk, Jakarta, Indonesia

[2]Department of Information Technology, Universitas Teknologi Digital Indonesia, Yogyakarta, Indonesia

[3]Department of Informatics, Faculty of Computer Science, Universitas Esa Unggul, Jakarta, Indonesia

Corresponding Author:
Vitri Tundjungsari, MSc, PhD
Department of Master in Computer Science, Faculty of Computer Science, Universitas Esa Unggul, Jl Arjuna Utara 9, Kebon Jeruk, Jakarta, Indonesia

Abstract

Background: The problem of how to deal with waste is a global issue all countries face. Like many developing countries, Indonesia has inadequate infrastructure to process the extremely high volume of waste produced throughout the country and minimal public participation in proper waste management. Although the Indonesian government regulates waste banks as a community-based waste management solution, there is a lack of integrated technological innovations to support waste banks. This study fills the gap by developing Circonomy, a model combining Internet of Things, gamification, and semantic web technologies to advance community-based circular waste management.

Objective: The aim of this study is to develop Circonomy as a circular waste model that integrates an Internet of Things–based smart bin, semantic web, and gamification as an innovative technological solution.

Methods: We identified the problem by observing Indonesian waste banks at 3 locations in Jakarta and Yogyakarta to define and design Circonomy. The Circonomy prototype was developed using the Design Science Research Methodology and evaluated based on technical performance and user experience. The technical performance has three indicators: bin capacity accuracy with a minimum of 80% precision, bin lid response time <5 seconds for a minimum of 80% of trials, and data transmission success rate of at least 80%. The user experience metric has two indicators: a minimum of 80% reporting high usability and ease of use, and at least 80% of users reporting that they feel more motivated using the prototype than the traditional waste bank. We selected 10 random participants aged 18-60 years to perform a user experience evaluation of our prototype.

Results: The Circonomy prototype demonstrated sound and stable performances related to technical performance and user experience. Circonomy achieved at least 80% technical performance accuracy, comparable to industry standards. The accuracy problem lies in the placement of the ultrasonic sensor. The waste should be placed directly under the ultrasonic sensor to ensure the bin's capacity measurement accuracy. The user experience testing results from 10 participants indicated that Circonomy has excellent user engagement, and 100% felt motivated by gamification and 80% found the mobile app easy to use.

Conclusions: The testing results showed that Circonomy has acceptable performances for early-stage prototyping, with at least an 80% accuracy rate in technical performance and user experience. This ensures that Circonomy operates effectively in real-world conditions while remaining cost-efficient and scalable. For future development, Circonomy will prioritize enhancing the accuracy and reliability of sensor-based occupancy detection through improved sensor placement, the integration of multiple sensors, and an exploration of alternative technologies for regions with limited IT resources. In addition, more gamification features such as challenges and quizzes should be added to improve user experience and motivation.

(JMIR Serious Games 2025;13:e66781) doi:10.2196/66781

KEYWORDS

IoT; gamification; semantic web; waste management; circular economy; smart bin; Circonomy; semantic; game; gamified; Indonesia; reuse; reduce; recycle; 3Rs; prototype; circular waste; Internet of Things; DSRM

XSL•FO
RenderX

Introduction

Background

The waste problem is complex and presents significant environmental, social, and economic challenges worldwide. It is characterized by increasing waste production, inadequate waste management systems, and the ecological impacts of waste disposal. Rapid urbanization, population growth, and industrialization, especially in developing countries, are mitigating this problem. Some of the root causes of the global waste problem are capitalist consumption patterns [1], ineffective waste management systems [2,3], and hazardous waste disposal [4,5]. Developing countries such as Indonesia face persistent challenges, including limited infrastructure, inefficient waste handling processes, low community participation, and inadequate awareness of proper waste sorting and disposal processes [2,3,5]. In contrast, developed countries, such as those in Europe [6] (including Germany [7] and Spain [8]), have successfully embedded responsible waste management practices in the culture of their communities through effective incentives.

Based on data collected from 341 districts and cities throughout Indonesia in 2023 released by the National Information System for Waste Management (Sistem Informasi Pengelolaan Sampah Nasional) and the Minister of Environment and Forestry [9], Indonesia generates 37,811,293.71 tons of waste annually, of which only 62.54% can be managed and 13.64% can be reduced per year. The data show that Indonesia urgently needs strategies and innovations to improve waste reduction, handling, and management.

We identified several potential solutions and opportunities to address the waste problem:

1. Integrated waste management systems: Integrated waste management systems, comprising reduce, reuse, recycle (3R) strategies, can significantly mitigate waste-related issues. These systems require coordinated policies, financial support, and infrastructure development [3,10].
2. Innovative waste disposal technologies: Most existing studies typically investigate the use of Internet of Things (IoT)–based smart bins [11-15], integration of IoT and artificial intelligence algorithms [16], gamification to encourage recycling behaviors [17,18], and semantic web technologies for enhanced data integration [19]. Microbial waste utilization and mobile waste processing plants are also used to enhance waste management efficiency and reduce the environmental impact [20].
3. Community-based initiatives: In Indonesia, community-based initiatives have been started and promoted by the government with the Regulation by the Ministry of Environment and Forestry Number 14 of 2021 concerning Waste Management at Waste Banks [21], which emphasizes community-driven, integrated, and circular economy–focused waste management solutions. Nikmah et al [20] discussed the success story of waste banks, which uses an eco-enzyme and demonstrates the potential of community-based solutions to transform organic waste into valuable products and promote environmental sustainability. Eka et al [22] discussed community involvement in waste management through waste banks, which are run by the community to enhance waste segregation, handling, and management. Effendi et al [23] developed a desktop-based waste bank information system as a tool to support waste bank activities in Indonesia.

From a review of the literature, we found that there are some research gaps:

1. Lack of integrated technological solutions: Some innovations that are often found in the literature to support waste management are IoT-based smart bins, gamification, and semantic web. These innovations have generally been applied separately rather than integrated into a comprehensive circular waste management model. Our proposed model addresses this gap by integrating these 3 elements into a cohesive system designed specifically for community-based waste management operations.
2. Limited technology adoption in community-based circular waste systems: Previous waste management solutions often focused on municipal- or city-level waste management and neglected community-managed operations such as waste banks, which are particularly prevalent in Indonesia. Most studies rarely address how technology can support local waste banks as a community initiative along with automated processes, incentivization, and participation. There are some studies related to technological support waste bank management systems within Indonesia. However, none of them propose a comprehensive approach combining automated waste bank innovations and increasing community motivation to take action in circular waste management. Our proposed model deals not only with how technology and innovation can be used for circular waste management but also with how to engage the community and deliver social impact to all stakeholders. We address the gap by using technology and the concepts of IoT, gamification, and semantic web to provide real-time semantic interconnections between stakeholders and waste data and enhance data-driven decision-making at the community level.
3. Limited use of gamification in circular waste management systems, especially in Indonesia: Indonesia has a regulation about waste banks (Number 14 of 2021 concerning Waste Management at Waste Banks [21]), emphasizing community-driven, integrated, and circular economy–focused waste management solutions. Our proposed model explicitly targets this gap by automating local operational processes such as waste weighing and recording integrated with points, leaderboards, and rewards schema using a gamification approach and technology-driven automatic process, with real-time feedback, explicitly tailored for community-based settings. We also address the gap by encouraging recycling behaviors and integrating it with the circular economy.

In short, Circonomy addresses these critical gaps by integrating IoT, semantic web, and structured gamification into a unified system tailored for community-managed waste management. This significantly improves operational efficiency and enhances

XSL•FO
RenderX

stakeholder participation within the circular economy framework.

Objectives

This research aims to address the waste problem in Indonesia by introducing Circonomy, an innovative circular waste management model. This research makes the following primary contributions to the literature and practice of circular waste management:

1. Integration of technologies: It combines IoT-based smart bins, semantic web technology, and gamification into a unified, community-centered system, uniquely enhancing operational efficiency and community engagement simultaneously.
2. Automation of waste bank processes: It aligns with the applicable Indonesian regulation, addressing limitations of manual processes (eg, inaccurate weighing, data entry errors). Real-time IoT sensor data and semantic web integration optimize waste management and decision-making.
3. Structured gamification approach: This approach implements incentive-based gamification (points, leaderboards, rewards) to motivate sustainable 3R behaviors and foster continuous community participation.

The specific objectives targeted by this research are as follows:

1. Integrated technological solution: develop a comprehensive, technology-driven circular waste management model that automates traditional waste bank operations (addressing gaps 1 and 3).
2. Real-time data collection and accuracy: develop an IoT technology that facilitates accurate real-time data collection (addressing gaps 1 and 2).
3. Semantic web integration: enhance operational and decision-making efficiency by enabling real-time stakeholder interactions and optimizing waste treatment processes (addressing gaps 1 and 2).
4. Enhanced community engagement through gamification: encourage sustainable behaviors through structured gamification incentives explicitly tailored for local communities (addressing gaps 1 and 3).
5. Scalability and adaptability: provide a flexible model that can be deployed across various socioeconomic contexts, including resource-constrained environments (addressing gaps 1 and 2).

Methods

Overview

The Design Science Research Methodology (DSRM) is used in this research. It consists of six main activities [24]: (1) problem identification, (2) solution proposal, (3) prototyping, (4) evaluation, (5) statement of learning, and (6) communication. Figure 1 shows the DSRM stages and activities.

Figure 1. The stages and activities of the DSRM (adopted from [24]). DSRM consists of 6 main stages: problem identification, solution proposal, prototyping, evaluation, statement of learning, and communication and dissemination. DSRM: Design Science Research Methodology.



Our decision to use DSRM aligns with our objectives involving integrated waste management, innovative circular waste technology, and community-driven initiatives. DSRM has been used widely in several smart city innovation projects. For example, Effendi et al [23] developed a desktop-based waste bank information system in Indonesia. Henaien et al [24] applied DSRM to develop smart waste management systems, further validating its efficacy in creating impactful solutions for environmental management. Similarly, Lapão et al [25] implemented the Smart Public Health City framework in Europe using DSRM. Duque [26] successfully used DSRM for IoT-based smart city innovations, demonstrating how the methodology effectively addresses urban technological challenges. Moreover, DSRM explicitly supports iterative development, rigorous evaluation, and clear dissemination. Duque [26] mentioned in his research that DSRM is a methodology that allows for the production of innovative prototypes and can be used to solve an organization's problems.

Problem Identification

At this stage, our research project investigated the waste problem faced by Indonesia and how current waste management–related regulations propose to solve it. The Indonesian government regulation as agreed in the Regulation of the Ministry of Environment and Forestry Number 14 of 2021 concerning Waste Management at Waste Banks [21] states that: "(a) Waste management should be comprehensive and integrated, applying circular economy principles collaboratively by government and society to achieve economic, health, and environmental benefits."

"Waste management is as intended in point (a) can be done synergistically through the waste bank." A waste bank is defined as a facility for managing waste with the 3R principle, as well as educational facilities, behavioral changes in waste management, and the implementation of a circular economy, which is formed and managed by communities, business entities, and local governments.

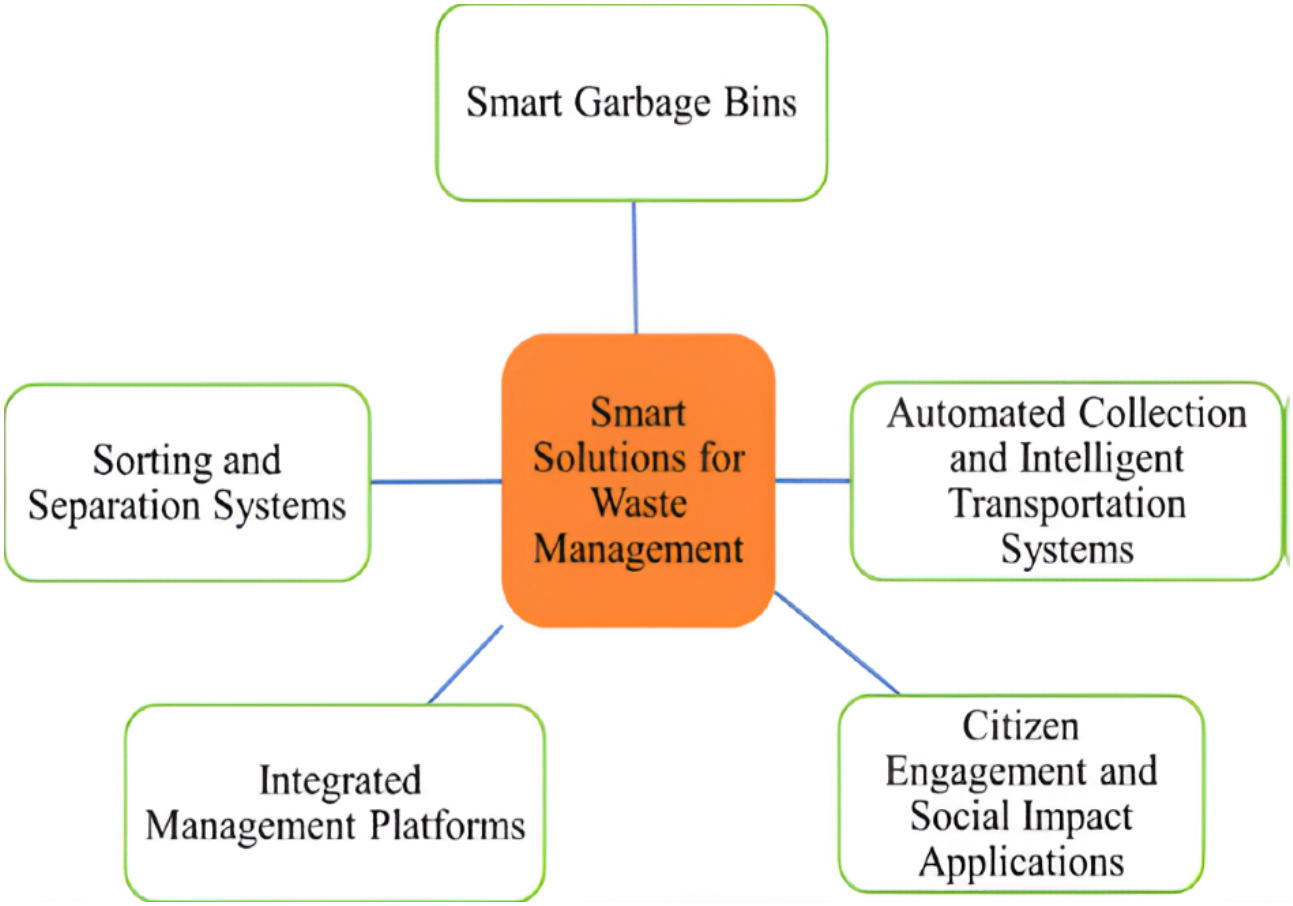
XSL•FO
RenderX

To identify the problem, we observed 3 waste banks located in 2 different cities in Indonesia (Jakarta and Yogyakarta).

Solution Proposal

Henaien et al [24] mentioned that there are five smart city solid waste management system classifications: (1) smart garbage bins, (2) sorting and separation systems, (3) integrated management platforms, (4) automated collection and intelligent transportation systems, and (5) citizen engagement and social impact applications. This research is a pilot project aiming to develop a first-stage prototype based on the Circonomy model. In developing the first stage of our prototype, we only focused on the design and development of (1) smart garbage bins, (2) citizen engagement mechanisms, and (3) integrated management platforms regarding circular waste management. This is in accordance with the objectives we want to achieve, as we discussed in the Introduction. This targeted focus was chosen due to the particular relevance and effectiveness of these 3 areas in addressing the identified waste management challenges in Indonesia and related countries. Furthermore, those 3 focuses align with Indonesia's regulatory framework, particularly the Regulation of the Minister of Environment and Forestry Number 14 of 2021 concerning Waste Management at Waste Banks, which emphasizes community-driven, integrated, and circular economy–focused waste management solutions. In the development of the second-stage prototype, the features of (1) sorting and separation functions and (2) automated collection and intelligent transportation systems will be added, resulting in a comprehensive integrated circular waste management system. Figure 2 shows the smart city solid waste management classifications [24].

Figure 2. Classification of smart city solid waste management systems [24]. There are 5 classifications: smart garbage bins; sorting and separation systems; automated collection and intelligent transportation systems; integrated management platforms; and citizen engagement and social impact applications.



Prototyping

This phase is crucial as it involves creating prototypes that demonstrate the practical application of our proposed hardware and software solutions. Our prototype, a smart bin, serves as an automated waste collection and weighing tool in our model. We tested the prototype-based Circonomy model using a combination of technical tests, user experience feedback, and preliminary impact projections. The primary objective was to assess the model's automation capabilities, data transmission reliability, and user engagement via gamification incentives. Controlled experiments were conducted in multiple scenarios, focusing on (1) automated bin operations (lid opening/closing based on user interaction); (2) automated weighing operations (ensuring precise tracking of deposited waste); (3) bin capacity estimation (evaluating ultrasonic sensor effectiveness); (4) data transmission and processing (assessing server communication and response times); and (5) gamification effectiveness (measuring user engagement based on points, rewards, and leaderboard interactions).

XSL•FO
RenderX

Evaluation and Statement

To carry out the testing process, we needed to determine the workflow scenario of our proposed model. The workflow is described as follows:

1. Installation of Circonomy app: The user downloads the Circonomy mobile app.
2. User registration and authentication: Users register and log in via the Android app. User data are stored on the server.
3. Connecting to smart bin: The user selects the bin by scanning the QR code and then sends a command to open the lid.
4. Bin lid control: Based on commands from the app, the ESP32 system-on-a-chip microcontroller controls the servo motor to open or close the bin lid. The lid will close automatically after the waste is thrown into it.
5. Collecting waste data: After the bin is loaded, the load cell and ultrasonic sensor measure the weight and capacity of the bin, and the data are sent to the ESP32. The bin's capacity data are calculated by measuring the distance from the bottom of the bin to the lid, and then converting it to a percentage using an ultrasonic sensor. For example, if there is plastic bottle waste with a height of 10 centimeters thrown away in a bin with a depth of 20 centimeters, then the bin capacity is 50%. The LED will show the weight and bin capacity after the measurement.
6. Sending data to server: The ESP32 sends waste weight and bin capacity data to the server via the REST application programming interface (API). The waste's weight is measured using a load cell sensor, while the bin's capacity is calculated using an ultrasonic sensor. The server processes the data received, converts waste weight into points, and updates the leaderboard and history data.
7. Feedback to user: The mobile app displays the points earned on the user profile and the leaderboard and updates the bin's status on the web-based application.

To measure the testing results, we used the following evaluation metrics:

1. Technical performance indicators consisted of three metrics: (1) bin capacity accuracy (minimum 80% precision); (2) bin lid response time (<5 seconds in at least 80% of trials); and (3) data transmission success rate (minimum 80%).
2. User experience metrics were gathered from 10 users, with two targets: (1) minimum 80% reporting high usability and ease of use and (2) at least 80% of users feeling more motivated using the prototype compared to the traditional waste bank. We selected 10 random users aged 18 to 60 years to be included in prototype user testing. Cazanas et al [27] mentioned in their research that 10 users are enough to achieve 90% problem discovery on an app's usability.

Ethical Considerations

This research involved the development and evaluation of the Circonomy model, which was designed for community-driven circular economy initiatives. Ethical considerations were diligently addressed throughout the study. The study protocol underwent an ethical review process by the Esa Unggul University Ethical Board, which determined that the research qualified for an exemption due to minimal risk and the use of anonymized secondary data. Participant privacy and confidentiality were strictly maintained through anonymization protocols. No personal identifiable information was collected or disclosed and no compensation was provided to participants.

Results

Problem Identification

We investigated how waste banks run by doing a case study in some provinces in Indonesia (Daerah Khusus Jakarta and Daerah Istimewa Yogyakarta). Here are some findings obtained from our observation of three waste banks:

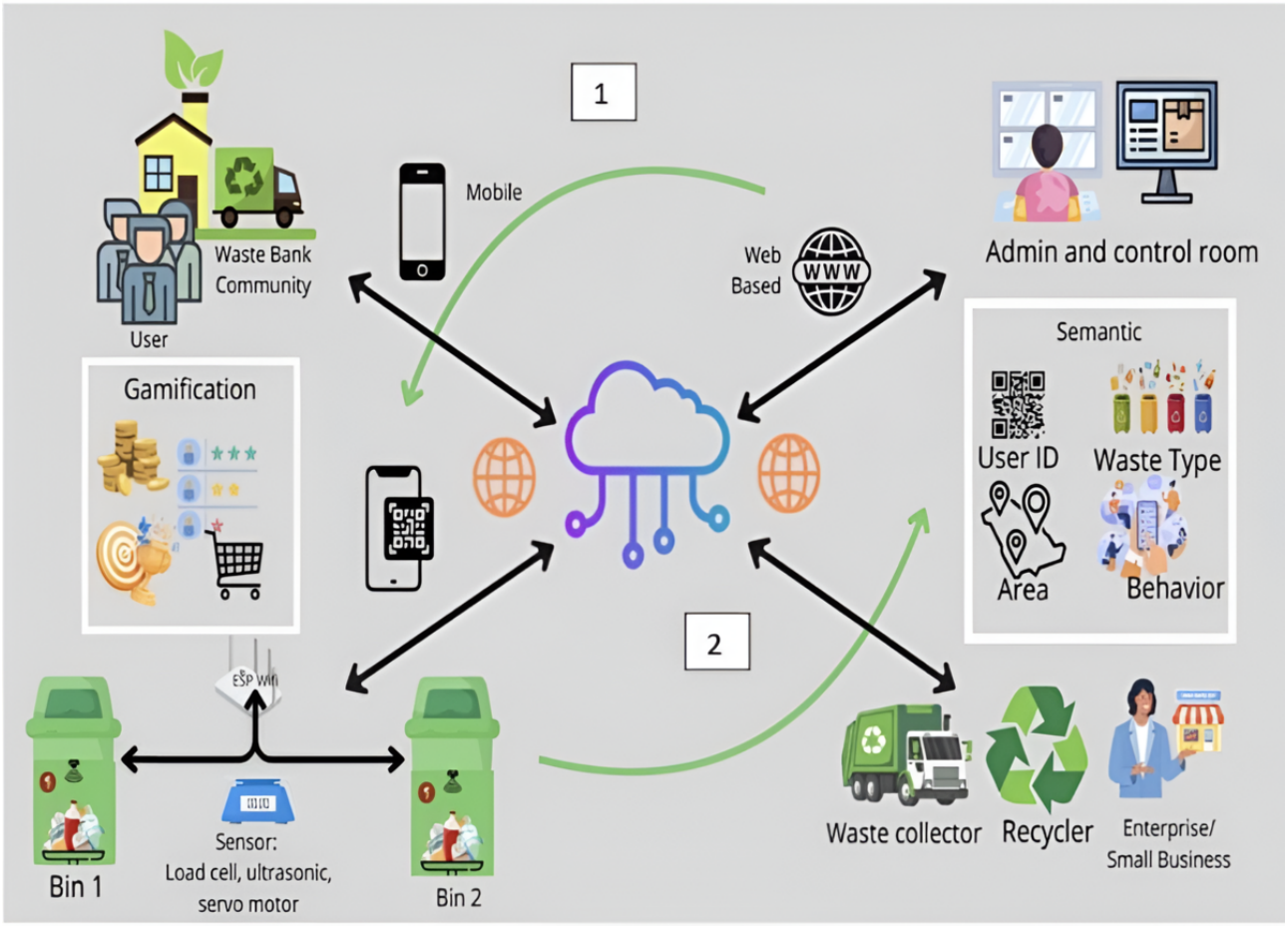
1. Waste banks have common workflows, which is in line with previous research [28]. The workflows of the waste banks in Indonesia consist of several activities: (1) waste separation and segregation, (2) waste deposit, (3) waste weighing, (4) waste recording and ledgers, and (5) waste collection for further treatment.
2. There are some common problems faced by waste banks, such as (1) inefficiency of waste bank management, especially when handling the waste due to uncertain working hours; (2) lost and incorrect waste recording documents; (3) lack of staff at the waste bank; (4) slow processes in waste weighing and no automated record (manually handwritten) in the waste book ledger; (5) people are often reluctant to come to the waste bank because the process and surrounding environment are not hygienic; and (6) people are not interested in supporting the waste bank because there are minimal incentives and rewards.

Solution Proposal: Circonomy

Our proposed model, Circonomy, is inspired by the circular economy concept. The circular economy contains different elements required for the transformation from a linear to a circular system, which includes: (1) design for sustainability, (2) cleaner production, (3) reverse logistics, (4) industrial ecology, (5) energy efficiency, (6) waste management, (7) product life extension, and (8) product as service [29]. Circonomy focuses on the waste management element, which consists of two subelements: (1) waste treatment and disposal and (2) reduce, reuse, and recycle. The model is presented in Figure 3.

XSL•FO
RenderX

Figure 3. The proposed Circonomy model. Circonomy has two main workflows: (1) waste treatment and disposal, in which the user (individual or waste bank community) puts the garbage in the smart bins using the mobile app, which is enhanced by gamification features; and (2) garbage collection by a waste collector, recycler, enterprise, or small business and monitoring by an administrator. This workflow is supported by the semantic web concept to enhance reduce, reuse, and recycle behavior.



As we can see from the model, there are cycles described by the numbers 1 and 2. Workflow 1 is the process involved in the waste and treatment disposal subelement, while workflow 2 is engaged in the 3R subelement. There are three main actors in our proposed model: (1) the community and users of the waste bank (the waste disposers), (2) the system administrator or operator, and (3) the waste collector, recycler, enterprise, or small business. The model has three main components: (1) smart bins, (2) a mobile app, and (3) a web-based application.

Gamification and the semantic web approach are used in our model to enhance user engagement in circular waste management by having these features:

1. User profiles and points: Users are categorized into residents and businesses. Users can create user profiles linked to waste collection and recycling activities.
2. Points and leaderboard: Users receive points for correctly placing waste in the smart bin, reducing waste, and participating in community cleanups. The points gathered by each user are displayed on the leaderboard in rank order.
3. Rewards: Points can be redeemed for rewards, for example, to get discounts at local eco-friendly stores or from other third parties.
4. Cross-platform data sharing: Waste data from municipalities, manufacturers, and recycling centers can be standardized and shared across platforms, providing better visibility into product life cycles. This enables real-time insights into waste flows and statuses to decide on the best disposal or recycling options based on local conditions.

Prototyping

To design the smart garbage bin in our model, we separated it into hardware and software components, as follows:

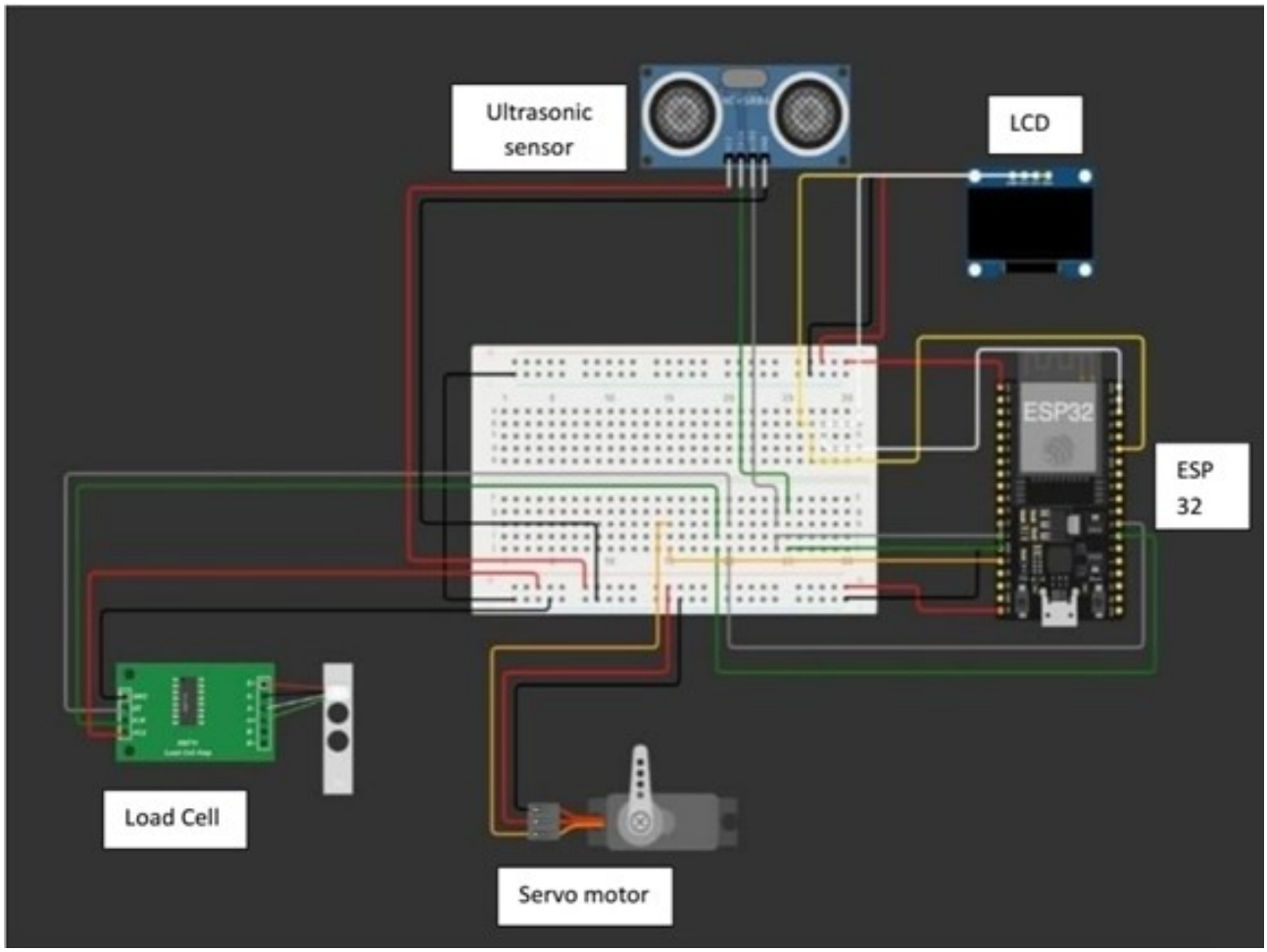
1. Hardware components (IoT). These components form the smart garbage bin by sending waste bin data to the software component. The hardware component consists of the following: (1) ESP32, which acts as the main controller of the IoT system. ESP32 will integrate and control other modules, such as load cells, ultrasonic sensors, and servo motors. Additionally, the ESP32 will communicate with the REST API server to send and receive data. (2) A load cell, which measures the weight of waste placed in the bin. These data are sent to the ESP32 for further processing, as well as to the ultrasonic sensor, servo motor, and LED. (3) The ultrasonic sensor measures the distance between the bin and the sensor to determine the level of bin fullness. These data are converted into a percentage of occupied capacity. (4) A servo motor is used to open and close the bin lid based on commands from the Android app sent via the REST API server to the ESP32. (5) The LED displays

XSL•FO
RenderX

the information on the waste bin. Figure 4 shows the IoT components in our prototype.

2. The software components in this model include a server and API, an Android app as an interface to the user, and a web-based application as an interface to the administrator. We used the Node.js/Express REST API server to handle communication between the ESP32 and the Android app. Functions included (1) user management (registration, authentication, and user data management); (2) smart bin management (registration and monitoring of each connected bin); (3) reward and challenge management (manage rewards and challenges in which users can participate); (4) leaderboard management (save and display rankings based on points from collected trash); (5) history data (stores waste weight data from each connected bin); and (6) mobile apps and web-based applications with features, for example, connecting to a smart bin so that the user can select a bin to dispose of trash, waste disposal history so that the user can view the weight of the waste automatically after he/she puts the waste in the bin, monitoring and control so that the application sends commands to the server to open the bin lid and presents the waste's weight and the bin's capacity on the LED, points and rewards systems so that users can view their points and redeem rewards, and a leaderboard to display user rankings based on collected waste points. Figure 5 presents the smart bin prototype, which was developed based on the Circonomy model.

Figure 4. The Internet of Things component in Circonomy. This component consists of the following: (1) an ultrasonic sensor to measure the bin's capacity; (2) a load cell to measure the waste's weight; (3) a servo motor to open or close the bin's lid automatically; (4) an ESP32 to serve as microcontroller, which sends and receives sensor data over a WiFi connection; (5) an LED to display the waste's weight and bin's capacity data.



XSL•FO
RenderX

Figure 5. The prototype of the Circonomy smart bin. The prototype connects hardware and software through the node.js/Express REST API server to handle communication between the ESP32 in the smart bin and the Android app. ESP32 serves as a microcontroller to communicate with other modules (such as the load cells, ultrasonic sensors, and servo motors) using the REST API server to send and receive data. API: application programming interface.



The Circonomy model incorporates a gamification system designed to increase user engagement, motivation, and long-term participation in community-based waste management activities. Gamification elements are embedded into mobile apps and web-based applications directly connected to the IoT-enabled smart bin system. The gamification framework aims to encourage sustainable waste behaviors, particularly the 3R practices, through the following mechanisms:

1. Points system: The point system is triggered every time a user deposits waste into the smart bin. The system automatically records the type of waste (eg, plastic, organic, paper) and the weight of the waste is measured via IoT-connected load cell sensors. Based on these inputs, the system calculates points for each user. Different waste types may earn different point values to encourage the sorting of high-priority recyclables. The heavier and more frequent the deposits, the more points a user can collect and accumulate.
2. Rewards: Users can exchange accumulated points for various rewards, such as vouchers (for groceries, services, or local products), eco-friendly goods (such as reusable bags or compost kits), or community recognition (through badges or certificates). Figure 6 shows the interface of the Circonomy mobile app with the points and rewards feature.
3. Leaderboards and challenges: The Circonomy mobile app features a leaderboard to visualize users' point rank from the highest to the lowest. This fosters healthy competition and public recognition. Periodic challenges and events are also introduced to keep engagement high, such as the top recycler of the month, community recycling drives with bonus points, and group challenges where groups compete to achieve collective waste reduction goals. Figure 7 shows the interface of the Circonomy mobile app with the rewards and leaderboard feature.
4. Incentive schema: The gamification system is designed around a structured incentive schema to maintain ongoing motivation, including immediate feedback via notifications after deposits, accumulated rewards over time, and social incentives and badges as participation status in the community.

XSL•FO
RenderX

Figure 6. The interface of the Circonomy mobile app. The left figure is the interface of the mobile app used to open and close the bin's lid. The middle figure is the interface showing the smart bin history (positive values show that waste was placed in the bin and negative values indicate that waste was taken out from the bin). The right figure presents the interface of gamification features (reward and leaderboard) in the mobile app as a reward mechanism to enhance user engagement.

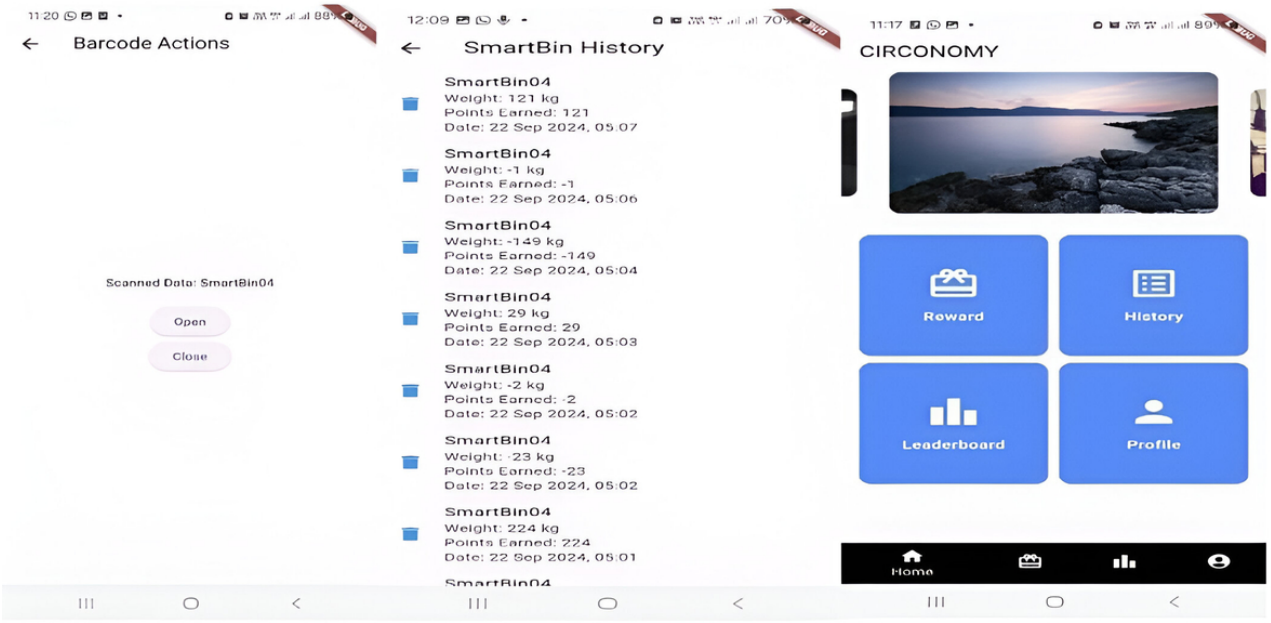
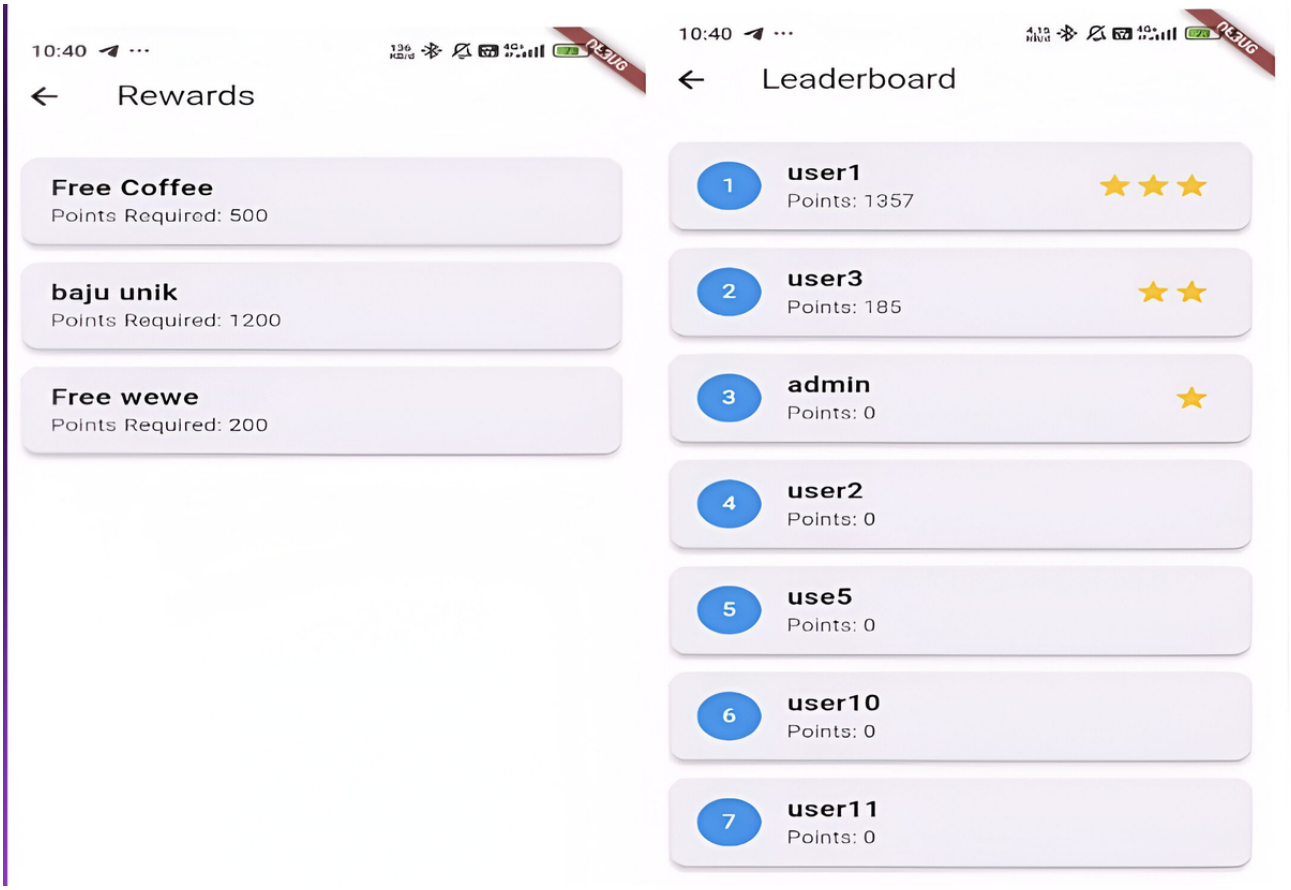


Figure 7. The interface of the Circonomy mobile app: rewards and leaderboard. The interface shows the leaderboard based on the points gained from using Circonomy (right side); these points can be exchanged for rewards such as free coffee, clothes, or others (left side).



Evaluation and Statement

To measure the testing results, we determined some evaluation metrics: (1) technical performance indicators and (2) user experience metrics. Technical performance indicators had three metrics: (1) weighing accuracy with at least 80% precision, with minor discrepancies due to sensor positioning; (2) bin lid response time <5 seconds in 90% of trials; and (3) data transmission with a minimum 80% success rate, affected by

XSL•FO
RenderX

network reliability and processing speed. User experience metrics were gathered from 10 users (aged 18-60 years), with two targets: 80% reported high usability and ease of use and 100% found the gamification incentives motivating for waste disposal, with the leaderboard and reward features receiving positive engagement feedback. We selected 10 random users aged 18 to 60 years to perform usability testing on our prototype. The results of the testing are summarized in Table 1.

Table . Prototype testing results based on technical performance and user experience.

Category and process		Result	Metrics
Technical performance indicator			
	Installation of Circonomy mobile app	All users find that the process is easy	100% installation success (10/10 tests)
	User registration and authentication	All users find that the process is easy	100% registration success (10/10 tests)
	Connecting to smart bin by scanning a QR code using the mobile app	All users find that the process is easy	100% successful connection (10/10 tests)
	Open and close the bin's lid automatically	The result depends on the position of the servo motor, which sometimes moves unexpectedly (still needs further improvements to the design)	90% response time <5 seconds (9/10 tests)
	Collecting waste data and showing weight and bin capacity data on an LED screen	The result depends on the viewpoint of the ultrasonic sensor toward the waste's position	Ultrasonic accuracy of 80% (8/10 tests). See Table 2 for more detailed measurements.
	Sending data to the server and updating waste history data that have been collected for each user and converting weight to points	The successful rate of data transmission depends on process number 5 and internet speed	80% data transmission success (8/10)
	Feedback to the user by displaying points earned, leaderboard, and the status of the bin on the mobile app and web application	The successful rate of data transmission depends on process number 6 and internet speed	80% data accuracy updating points and leaderboard success (8/10)
User experience			
	Usability and ease of use	Most users strongly agree that the prototype has excellent usability	80% (Likert scale 1 - 5). See Table 3 for more detailed testing results.
	Gamification incentives are motivating	All users are completely in agreement that Circonomy enhanced motivation	100% (Likert scale 1 - 5). See Table 3 for more detailed testing results.

XSL•FO
RenderX

Table . Numeric results of the bin's capacity (sample of 10 trials). Net bin dimensions (15 cm width × 20 cm length × 20 cm depth) with a total volume of 6000 cm^3 or equal to 6000 mL. Displayed bin capacity on the LED is measured from the distance from the top of the waste to the ultrasonic sensor divided by depth of the bin. The ultrasonic sensor is placed under the bin's lid.

Trial number	Waste type and size	Displayed bin capacity (the waste position should be observed carefully)	Actual bin capacity (%)	Result
1	1 plastic bottle with a height of 20 cm	0% (the bottle is placed under the ultrasonic sensor)	0% available	Accurate
2	1 box with a height of 10 cm	50% (the box is placed under the ultrasonic sensor)	50% available	Accurate
3	1 box with a height of 5 cm	75% (the box is placed under the ultrasonic sensor)	75% available	Accurate
4	1 plastic bottle with a height of 14 cm	70% (the plastic bottle is in a fallen position)	30% available	Not accurate
5	1 plastic bag contains food waste with a height of 5 cm	75% (the plastic bag is placed under the ultrasonic sensor)	75% available	Accurate
6	9 plastic bottle with a height of 6 cm	70% (the bottles are placed and distributed evenly inside the bin)	70% available	Accurate
7	1 box with a height of 15 cm	25% (the box is placed under the ultrasonic sensor)	25% available	Accurate
8	1 plastic cup with a height of 10 cm	100% (the plastic bottle is placed not directly under the ultrasonic sensor)	50% available	Not accurate
9	1 box with a height of 20 cm	0% (the box is placed under the ultrasonic sensor)	0% available	Accurate
10	1 box with a height of 4 cm	80% (the box is placed under the ultrasonic sensor)	80% available	Accurate

Table . User experience testing results in terms of (1) usability and ease of use and (2) the gamification incentive scheme. Ten respondents rated their opinion on provided questions using a Likert scale (1 to 5).

Topics	1	2	3	4	5	Means	Interpretation
Overall reaction to the system: ease of use and features			2	6	2	4	80% of users agree that Circonomy is easy to use (mean score 4/5)
Enhance motivation with gamification (points, rewards, leaderboard)					10	5	100% of users agree that gamification enhances motivation to use Circonomy (mean score 5/5)

Discussion

Preliminary Findings

The Circonomy model successfully integrates IoT, gamification, and semantic web technologies into a circular waste management system, addressing critical gaps in traditional waste bank operations. Our prototype testing demonstrated at least 80% accuracy in waste weighing, bin capacity estimation, and data transmission, aligning with industry standards. User engagement was notably high, with 100% of participants reporting increased motivation due to gamification features, and 80% finding the mobile app easy to use. The structured incentives, including leaderboards and reward systems, encouraged active participation, distinguishing Circonomy from conventional waste management models. Furthermore, the system's economic feasibility highlights the potential for public-private partnerships to enhance its sustainability. These findings indicate that Circonomy is an effective and scalable solution for optimizing community-based waste management and promoting circular economy principles.

Model Evaluation

Our model is proven to perform well, with an 80% accuracy threshold. The 80% accuracy threshold for weighing precision,

XSL•FO
RenderX

data transmission, and overall system performance is justifiable and acceptable for the following reasons.

Acceptable Performance for Early-Stage Prototyping

The Circonomy model is in its prototype and testing phase. At this stage, achieving 100% accuracy is unrealistic due to system constraints such as sensor limitations, environmental factors, and calibration challenges. Most early-stage IoT-based waste management systems report similar performance levels, with accuracy typically between 70% and 85% during initial testing phases [24,26]. An 80% accuracy rate ensures the system functions reliably under real-world conditions while allowing room for iterative improvements in future versions.

Sensor and Environmental Constraints

There are several sensor and environmental constraints that need to be taken into account, including the following:

- Weight measurement: The 80% weighing accuracy accounts for minor errors caused by uneven waste placement, bin tilting, or variations in material density (eg, lightweight plastics vs heavy glass).
- Ultrasonic bin capacity detection: How waste settles inside the bin influences sensor accuracy, which may create slight variations in readings.
- Data transmission: The 80% data success rate mainly depends on internet speed and server response times, which can vary due to network congestion or IoT device latency.

Comparison With Industry Standards

Similar IoT-based smart waste bin studies have reported performance accuracy in the 75% - 90% range under varying conditions [25], while for waste sorting and collection models, an accuracy threshold of 80% is typically considered operationally efficient in smart city applications [6].

Balancing Accuracy With Cost and Feasibility

Achieving more than 80% accuracy would require higher-end sensors and more expensive technology (eg, industrial-grade load cells, lidar-based bin occupancy detection), which would increase system costs and reduce scalability for community-level waste banks.

Thus, we can conclude that the current 80% accuracy rate ensures that Circonomy operates effectively in real-world conditions while remaining cost-efficient and scalable. However, we still require incremental refinements and continue to enhance accuracy performance over time.

Challenges and Limitations

Scalability and Adaptability

Circonomy is designed with IoT, sensor components, and mobile apps that require sound and stable internet connection access, however, we also provide flexible schemas to enable Circonomy to operate in regions with limited IoT infrastructure and intermittent internet connectivity. These adaptations allow the circular economy to remain operational across diverse socioeconomic contexts, from urban centers with full digital infrastructure to rural areas with limited technological access.

The technology flexibility ensures that the benefits of the circular economy and incentivized waste management can be extended to underresourced communities, fostering inclusive participation regardless of technological limitations. To ensure accessibility in such areas, the model incorporates several adaptive strategies:

1. Offline functionality. The mobile app and web application support offline data entry. Waste transactions can be recorded locally on devices and synchronized with the central system once a stable internet connection is available.
2. QR code validation. In low-connectivity areas, QR codes are used to identify users and link waste deposits to their profiles without needing constant real-time communication with the server.
3. Delayed synchronization. Data collected during offline periods are stored locally and batch-uploaded during periods of connectivity, ensuring no loss of transactional data.
4. In other resource-constrained areas with limited availability of IoT, we can exchange IoT with simpler technology, such as manual weighing systems integrated with mobile/web data logging.

Sustainability Strategies

To ensure long-term sustainability across diverse socioeconomic contexts, Circonomy incorporates several adaptive strategies that address financial, social, and operational challenges, making the model resilient and scalable in various environments. Community ownership and empowerment should be encouraged through Circonomy to foster a sense of responsibility, strengthen social cohesion, and promote long-term participation, for example, by transferring ownership and daily operational responsibilities to local waste bank members.

Collaboration between local governments, private companies, and nongovernmental organizations is also required to ensure shared investment in infrastructure, rewards, and educational campaigns. These partnerships reduce financial burdens placed on communities and provide technical support to maintain system functionality. Flexible incentive models also ensure sustainability by enabling the incentive scheme to be adapted to fit local needs—such as financial incentives, social recognition, vouchers, or access to community services—ensuring relevance and attractiveness to diverse user groups. Circonomy is also built to operate at various scales, from small communities to large urban areas, and it can gradually integrate additional technologies and features as resources grow.

Local resource utilization (materials and services for system maintenance, hardware replacement, and community events) is also important as a sustainability strategy. This not only reduces costs but also stimulates local economies.

Continuous community education is crucial to sustaining participation. Regular workshops, digital campaigns, and school programs promote an understanding of the circular economy and reinforce the value of waste sorting and recycling. Last but not least, aligning Circonomy with national and local regulations is crucial to ensuring legal support and access to public funding programs that can assist in long-term sustainability.

XSL•FO
RenderX

Alignment With Global Initiatives

Circonomy aligns directly with several global sustainability frameworks, notably the United Nations Sustainable Development Goals (SDGs), providing concrete contributions toward achieving global sustainability targets: (1) SDG 11 (Sustainable Cities and Communities): Circonomy facilitates sustainable urban environments by optimizing local waste management infrastructure through smart, community-centered approaches. Improving waste segregation accuracy and operational efficiency contributes to building sustainable and resilient urban communities. (2) SDG 12 (Responsible Consumption and Production): By integrating IoT-driven real-time monitoring, gamification-based incentives, and semantic web data management, Circonomy promotes responsible consumption habits and waste reduction among citizens. Its structured incentive framework directly encourages the adoption of recycling and reusing behaviors, aligning individual actions with broader circular economy principles. (3) SDG 13 (Combat Climate and Its Effect): Circonomy promotes 3R principles, preventing organic waste from being mixed with recyclables. This reduces emissions from landfills, a significant contributor to climate change.

Comparison With Other Global Waste Management Systems

After having several comprehensive discussions, we determined that it is important to compare and analyze how Circonomy can be applied not only as a potential solution for the local and national context but also for the global context. We conducted a structured comparison of traditional waste banks (Indonesia), Circonomy (Indonesia), Reciclos (Spain) [8], and Germany's Deposit Refund Scheme [7] in terms of their mechanisms, incentives, technology used, gamification, impact, scalability, challenges, alignment with global waste management, and regulatory framework. Table 4 shows how the Circonomy features compare to these other waste management systems.

XSL•FO
RenderX

Table . Comparison table of Circonomy with the traditional waste bank in Indonesia, Reciclos (Spain), and Deposit Refund Scheme (Germany).

Feature	Traditional waste banks (Indonesia)	Circonomy (Indonesia)	Reciclos (Spain) [8]	Deposit Refund Scheme (Germany) [7]
Mechanism	Users bring recyclable materials to a Waste Bank, which weighs and records their contributions manually.	Uses IoT-enabled smart bins to monitor waste disposal; integrates gamification with points and leaderboards.	Uses a mobile app and QR codes to track recycling activity and reward participation with digital tokens.	A deposit fee is added to beverage purchases, refunded upon returning the container.
Incentives	Users receive money in exchange for recyclable materials, but prices fluctuate based on market demand.	Points-based system for waste disposal, with rewards for users and businesses.	Lottery-based system where users can exchange digital tokens for prizes or public benefits.	Financial incentive (deposit refund) for each returned container.
Technology used	Mostly manual processes, with some using basic digital weighing systems for record-keeping.	IoT-enabled smart bins, mobile app, and semantic web integration for waste tracking.	Mobile web app, blockchain-based smart contracts, and digital token system.	Simple barcode scanning system at return points, integrated with retailers.
Gamification	No structured gamification, but some waste banks provide community recognition programs.	Yes, with leaderboards, rewards for individuals and businesses, and interactive challenges.	Yes, lottery-based incentive program with digital tokens.	No direct gamification, but monetary rewards encourage participation.
Impact on recycling behavior	Encourages community participation in recycling but relies on voluntary participation and fluctuating material prices.	Encourages responsible waste disposal through automation and competition.	Increases recycling rates through incentive-based engagement.	Ensures high return rates for beverage containers, reducing littering.
Scalability	Limited scalability; depends on local community participation and funding.	Designed for urban environments with IoT infrastructure but faces challenges in areas with limited internet access.	Easily scalable due to mobile-based participation; requires local government support.	Highly scalable, already implemented nationwide with success.
Challenges	Dependent on fluctuating recyclable material prices; often lacks proper funding and long-term sustainability.	Infrastructure dependency (IoT and internet connectivity); sensor inaccuracies; needs more extensive gamification features.	User engagement depends on prize incentives; blockchain-based reward system requires public trust.	Administrative and logistical costs; effectiveness relies on high participation and proper waste sorting.
Alignment with circular economy	Moderate, as it promotes recycling but lacks systemic optimization and technology integration.	Strong, integrates smart technology for waste monitoring and promotes long-term behavior change.	Strong, supports a digitalized circular economy by incentivizing proper waste sorting and recycling.	Strong, creates a closed-loop system for beverage containers, ensuring material reuse.
Regulatory framework	Supported by the Indonesian government regulation about waste banks.	Supported by the Indonesian government regulation about waste banks but there is a lack of financial support from government and private sector partnerships.	Supported by the Spanish government and private sector partnerships (Ecoembes).	Mandated by German law; highly structured under Extended Producer Responsibility.

The table comparison illustrates that Circonomy bridges the gap between traditional and technology-driven waste management by integrating IoT, gamification, and the semantic web. Unlike traditional waste banks in Indonesia, which rely on manual tracking, Circonomy automates waste monitoring using IoT sensors, ensuring real-time tracking of bin capacity, waste weight, and user participation. Compared to Reciclos in Spain and Germany's Deposit Refund Scheme, which focus mainly on plastic and beverage containers, Circonomy supports multimaterial waste sorting (plastic, organic, paper, metal, and glass) and assigns dynamic reward values to encourage responsible disposal. Gamification features such as leaderboards and digital incentives enable Circonomy to engage communities more effectively than existing systems, making waste management interactive and rewarding. Circonomy provides scalability and accessibility by using low-cost IoT components, making it more adaptable to regions with limited infrastructure than Germany's Deposit Refund Scheme, which requires expensive reverse vending machines.

Conclusion and Future Works

Circonomy has been implemented and evaluated from technical and user experience aspects. As an initial prototype, Circonomy has at least an 80% accuracy rate to ensure the model works reliably under real-world conditions while allowing room for incremental improvements and further interdisciplinary collaboration. Circonomy has potential as a solution encouraging circular waste management using technological innovation and socioenvironmental intervention, as it is aligned with research in [30]. Circonomy is grounded in 3 primary fields of study: computer science and engineering, social sciences, and

XSL•FO
RenderX

environmental management. It serves as a potential circular waste solution not only in local and national contexts but also in a global context to address sustainable waste management challenges by moving toward a robust circular economy.

Future development of the Circonomy model will prioritize enhancing the accuracy and reliability of sensor-based occupancy detection through improved sensor placement, multiple sensor integration, and exploring alternative technologies for regions with limited IT resources. Additional features will be introduced to the gamification component, including interactive challenges and quizzes to educate users, strengthen community involvement, and reinforce long-term sustainable waste management behaviors. Further integration of semantic web technologies will optimize data management and facilitate better decision-making by effectively linking waste-related data, user actions, and recycling processes across digital platforms.

Acknowledgments

The authors would like to express their gratitude to the Ministry of Education and Culture, Higher Directorate General (DIKTI), the Republic of Indonesia, for providing a grant for this research under the National Fundamental Research Grant scheme.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Authors' Contributions

VT (lead author) and BPDP were involved in conceptualization and formal analysis. VT, BPDP, and MBU were involved in data extraction and curation. MBU and NA were involved in validation and visualization. VT and BPDP supervised the project implementation, while MBU and NA handled the project administration. VT wrote the original draft of the manuscript. All authors provided feedback on the manuscript and approved the final version.

Conflicts of Interest

None declared.

References

1. Gregson N. The Waste of the World: Consumption, Economies and the Making of the Global Waste Problem: Bristol University Press; 2023. [doi: 10.56687/9781529232462]
2. Zavolokina L, Diesperova N. RETRACTED: On the issue of building an integrated waste management system: world practice and its application in the Russian Federation. E3S Web Conf 2024;549:03001. [doi: 10.1051/e3sconf/202454903001]
3. Kumari T. Waste management practices in the developing nations: challenges and opportunities. In: Waste Management and Resource Recycling in the Developing World: Science Direct; 2022. [doi: 10.1016/B978-0-323-90463-6.00017-8]
4. Walters R, Fuentes Loureiro MA. Waste crime and the global transference of hazardous substances: a southern green perspective. Crit Crim 2020 Sep;28(3):463-480. [doi: 10.1007/s10612-020-09522-4]
5. Parvez SM, Hasan SS, Knibbs LD, et al. Ecological burden of e-waste in Bangladesh-an assessment to measure the exposure to e-waste and associated health outcomes: protocol for a cross-sectional study. JMIR Res Protoc 2022 Aug 16;11(8):e38201. [doi: 10.2196/38201] [Medline: 35972788]
6. Laureti L, Costantiello A, Anobile F, Leogrande A, Magazzino C. Waste management and innovation: insights from Europe. Recycling 2024;9(5):82. [doi: 10.3390/recycling9050082]
7. Rhein S, Sträter KF. Intended and unintended effects of statutory deposit return schemes for single-use plastic bottles: lessons learned from the German experience. GAIA - Ecological Perspectives for Science and Society 2021 Dec 16;30(4):250-256. [doi: 10.14512/gaia.30.4.8]
8. Gibovic D, Bikfalvi A. Incentives for plastic recycling: how to engage citizens in active collection. Empirical evidence from Spain. Recycling 2021;6(2):29. [doi: 10.3390/recycling6020029]
9. Sistem Informasi Pengelolaan Sampah Nasional (SIPSN) [Website in Indonesian]. URL: https://sipsn.menlhk.go.id/sipsn/ [accessed 2024-09-11]
10. Masud H, Mourshed M, Hossain MS. Generation of waste: problem to possible solution in developing and underdeveloped nations. In: Waste Management and Resource Recycling in the Developing World: Elsevier; 2022:21-59. [doi: 10.1016/B978-0-323-90463-6.00021-X]
11. Singhvi RK, Lohar RL, Kumar A, Sharma R, Sharma LD, Saraswat RK. IoT basedsmart waste management system: India prospective. Presented at: 2019 4th International Conference on Internet of Things; Apr 18-19, 2019; Ghaziabad, India. [doi: 10.1109/IoT-SIU.2019.8777698]
12. Tandon R, Bansal A. IoT-enabled smart waste management: leveraging the power of IoT for sustainable solutions. Presented at: 2023 3rd International Conference on Innovative Sustainable Computational Technologies (CISCT); Sep 8-9, 2023; Dehradun, India. [doi: 10.1109/CISCT57197.2023.10351210]

XSL•FO
RenderX

13. Zebua AB, Husni MFA, Naufal M, Irviantina S. IoT-based smart bin using smell, weight, and height sensors. Presented at: 2022 Seventh International Conference on Informatics and Computing (ICIC); Dec 8-9, 2022; Denpasar, Bali, Indonesia p. 1-5. [doi: 10.1109/ICIC56845.2022.10006894]
14. Filho RH, de Brito WA, Paiva MG, Cavaignac Klemensov PA, de Oliveira LS. A sustainable recycling ecosystem scheme based on IoT to promote social benefits. Procedia Comput Sci 2024;238:344-351. [doi: 10.1016/j.procs.2024.06.034]
15. Saha S, Chaki R. IoT based smart waste management system in aspect of COVID-19. Journal of Open Innovation: Technology, Market, and Complexity 2023 Jun;9(2):100048. [doi: 10.1016/j.joitmc.2023.100048]
16. Pant G, Rawat P, Kathuria S, Gehlot A, Rathor N. Waste management system with technological intervention: advantages and challenges. Presented at: 2023 4th International Conference on Electronics and Sustainable Communication Systems (ICESC); Jul 6-8, 2023; Coimbatore, India. [doi: 10.1109/ICESC57686.2023.10193334]
17. Ouariachi T, Li CY, Elving WJL. Gamification approaches for education and engagement on pro-environmental behaviors: searching for best practices. Sustainability 2020;12(11):4565. [doi: 10.3390/su12114565]
18. Helmefalk M, Rosenlund J. Make waste fun again! A gamification approach to recycling. Presented at: Interactivity, Game Creation, Design, Learning, and Innovation: 8th EAI International Conference, ArtsIT 2019, and 4th EAI International Conference, DLI 2019; Nov 6-8, 2019; Aalborg, Denmark p. 415-426. [doi: 10.1007/978-3-030-53294-9_30]
19. Motevallian M, Esfar-E-Alam AM, Taherkordi A, Abbasi G. Semantic modeling of waste dataflow for automating circular economy systems. Presented at: 2024 20th International Conference on Distributed Computing in Smart Systems and the Internet of Things (DCOSS-IoT); Apr 29 to May 1, 2024; Abu Dhabi, United Arab Emirates p. 677-684. [doi: 10.1109/DCOSS-IoT61029.2024.00105]
20. Nikmah A, Kurniawan D, Putri A, Sahdi J. Pengolahan Sampah Rumah Tangga Menjadi Eco Enzim Yang Dapat Dimanfaatkan Masyarakat Desa Arul Putih [Article in Indonesian]. JPPMI 2023;2(5):38-42 [FREE Full text] [doi: 10.55542/jppmi.v2i5.857]
21. Peraturan Menteri Lingkungan Hidup dan Kehutanan Republik Indonesia No 14/ 2021 tentang Pengelolaan Sampah pada Bank Sampah [Website in Indonesian]. URL: https://peraturan.bpk.go.id/Details/233754/permen-lhk-no-14-tahun-2021 [accessed 2024-09-11]
22. Eka O, Regina C, Rasha A, Agnes S, Nur S, Nuraisyah A. Waste bank in Indonesia: problem and opportunities. In: Proceedings of the Business Innovation and Engineering Conference (BIEC 2022). [doi: 10.2991/978-94-6463-144-9_27]
23. Effendi D, Noviansyah B, Lestary L, Mauluddin S, Nurhayati S. Implementation of waste bank information system to manage waste transaction data from the community. JESTEC 2023;18(6):2835-2846.
24. Henaien A, Ben Elhadj H, Chaari Fourati L. A sustainable smart IoT-based solid waste management system. Future Generation Computer Systems 2024 Aug;157:587-602. [doi: 10.1016/j.future.2024.03.056]
25. Lapão LV, Correia JC, Jevtic M. Public health framework for smart cities within the comprehensive approach to sustainability in Europe: case study of diabetes. Sustainability 2023;15(5):4269. [doi: 10.3390/su15054269]
26. Duque J. The IoT to smart cities - a design science research approach. Procedia Comput Sci 2023;219:279-285. [doi: 10.1016/j.procs.2023.01.291]
27. Cazanas A, AdS M, Parra E. Estimating sample size for usability testing. In: Enfoque UTE 2017:172-185. [doi: 10.29019/enfoqueute.v8n1.126]
28. Prasetyo ME, Rosita YD, Ristono J. Pengembangan website pengelolaan manajemen bank sampah di bank sampah rosella kota mojokerto [Website in Indonesian]. Jurnal Bank Sampah. 2019. URL: http://repository.unim.ac.id/259/2/JURNAL%20BANK%20SAMPAH.pdf [accessed 2025-04-18]
29. Srikanth PK, Visvanathan C. Chapter 5: circular economy for plastic waste management. In: Marine Plastic Abatement: IWA Publishing; 2023. [doi: 10.2166/9781789063202_0185]
30. Rønn C, Wieland A, Lehrer C, et al. Circular business model for digital health solutions: protocol for a scoping review. JMIR Res Protoc 2023 Nov 24;12:e47874. [doi: 10.2196/47874] [Medline: 37999949]

Abbreviations

3R: reduce, reuse, recycle
API: application programming interface
DSRM: Design Science Research Methodology
IoT: Internet of Things
SDG: Sustainable Development Goal

XSL•FO
RenderX

Edited by A Coristine; submitted 24.09.24; peer-reviewed by E Toki, H Soesanto, M Karthik, R Gorantla; revised version received 29.03.25; accepted 29.03.25; published 06.05.25.

Please cite as:
Tundjungsari V, Putranto BPD, Ulum MB, Anwar N
An Integrated Model for Circular Waste Management Using the Internet of Things, Semantic Web, and Gamification (Circonomy): Case Study in Indonesia
JMIR Serious Games 2025;13:e66781
URL: https://games.jmir.org/2025/1/e66781
doi:10.2196/66781

© Vitri Tundjungsari, Bambang Purnomosidi Dwi Putranto, Muhamad Bahrul Ulum, Nizirwan Anwar. Originally published in JMIR Serious Games (https://games.jmir.org), 6.5.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Bridging Needs and Expectations of Individuals With Physical Disabilities and Community Services Stakeholders for the Cocreation of an Adapted Physical Activity Platform in Virtual Reality: Qualitative Study

Aurélien Ramos[1,2], BSc; Maxence Boisvert[2], BSc; Elodie Traverse[2], PhD; Danielle Levac[3,4], PhD; Martin Lemay[3,5], PhD; Marika Demers[4,6], PhD; Martine Bordeleau[7,8], PhD; Sarah-Maude Ruest[1,2]; Roxanne Périnet-Lacroix[9], BSc; Krista L Best[1,2], PhD; Maxime T Robert[1,2], PhD

[1]School of Rehabilitation Sciences, Université Laval, Québec, QC, Canada

[2]Centre interdisciplinaire de recherche en réadaptation et en intégration sociale, Research Center, 525 Bd Wilfrid-Hamel, Québec, QC, Canada

[3]Centre hospitalier universitaire Sainte-Justine, Research Center, Montreal, QC, Canada

[4]School of Rehabilitation, Faculty of Medicine, Université de Montréal, Montréal, QC, Canada

[5]Department of Physical Activity Sciences, Université du Québec a Montreal, Montréal, QC, Canada

[6]Division of Biokinesiology and Physical Therapy, University of Southern California, Los Angeles, CA, United States

[7]Centre intégré universitaire de santé et de services sociaux de l'Estrie, Centre hospitalier universitaire de Sherbrooke, Sherbrooke, QC, Canada

[8]Faculty of Medicine and Health Sciences, Université de Sherbrooke, Sherbrooke, QC, Canada

[9]Adaptavie, Community Organization, Québec, QC, Canada

Corresponding Author:
Maxime T Robert, PhD
School of Rehabilitation Sciences, Université Laval, Québec, QC, Canada

Abstract

Background: Physical activity supports the health and well-being of individuals with physical disabilities. Despite the significance of engaging in physical activity, barriers faced by individuals with disabilities, such as limited access to adapted facilities and lack of transportation, can restrict their participation. Community organizations play a role in addressing these challenges, but virtual reality (VR) also offers a way to diversify adapted activities. In some situations, VR can help overcome the resource limitations of organizations by providing accessible, engaging, and highly personalized options for physical activity.

Objective: The aim of this study was to explore the needs and expectations of individuals with physical disabilities and their interventionists for the use of a VR physical activity platform in a community organization.

Methods: A descriptive qualitative study was conducted using semistructured interviews with individuals with physical disabilities and their interventionists, all associated with a nonprofit organization promoting physical activity among people with disabilities. Data were analyzed using an inductive thematic approach.

Results: In total, 15 participants, including 8 people with physical disabilities and 7 interventionists, were interviewed. Through this discussion, we gained insights into the everyday challenges faced by individuals with disabilities and identified priorities for community organizations. Subsequently, we discussed key considerations for using VR, including adapting activities, the possibility of fostering a more positive perception of physical activity, and harnessing the potential of VR to improve access to physical activity. We also discussed the importance of supporting personal goals and creating inclusive experiences while recognizing challenges such as technical difficulties and connectivity issues.

Conclusions: By exploring the needs and expectations regarding VR technology from individuals with physical disabilities and their interventionists, this study provided essential insights for integrating immersive and nonimmersive VR into community organizations, informing next steps for the design of adapted physical activities in VR.

(JMIR Serious Games 2025;13:e59704) doi:10.2196/59704

KEYWORDS

virtual reality; physical activity; community organizations; accessibility; physical disability

XSL•FO
RenderX

Introduction

Physical activity is a cornerstone of health and well-being, yet individuals with physical disabilities face barriers in access to physical activity opportunities compared to their nondisabled peers [1]. According to Statistics Canada, of the 6.2 million people living with at least one disability in Canada, more than half do not engage in any leisure-time physical activity, and less than 15% are considered physically active [2]. Promoting access to physical activity among individuals with physical disabilities is important to prevent comorbidities [3], reduce the risk factors associated with sedentary behaviors [4], and increase participation in sports, social, and community activities [5].

Access to physical activity opportunities poses a significant challenge for individuals with physical disabilities. These barriers are multifaceted, stemming from both personal and environmental factors that impede access to physical activity. On the environmental front, difficulties are mainly due to a lack of facilities and equipment adapted to the needs of people with physical disabilities [1,5-8], inadequate means of transport, and a lack of professionals trained to work with this population [6]. These problems are particularly pronounced in rural areas, where resources are often more limited. Personal factors among people with physical disabilities, such as fear of failure, negative self-perception, or negative perception of the practice environment [1,7,9], play an important role in access to physical activities. This gap in accessible physical activity underscores the importance of community services in supporting continued engagement and overcoming these barriers.

Community services can be defined as associations that aim to improve the well-being of a community by addressing its specific needs, often through volunteer-driven initiatives and social support programs. For the purpose of this research, community services include recreational centers with adapted social and physical activities, accessible gyms with a wide heterogenous clinical profile, sports, social clubs, as well as healthy support programs. Community services can greatly influence one's engagement in physical activity, particularly after discharge from rehabilitation services. Community-based approaches for promoting physical activity take into account the dynamic interplay of social, organizational, cultural, socioeconomic, environmental, and policy influences on an individual's functioning [10]. Consequently, community-based interventions, defined as interventions led by therapists to improve various health-related domains (motor, cognitive, social, etc), have emerged as essential strategies for promoting engagement in physical activity [11]. Despite evidence of their positive impact on preventing secondary conditions and on improving overall health and function in people with disabilities, developing tailored programs to meet individual needs remains challenging. Adapted and accessible physical activities are essential for all individuals, regardless of health status or ability level [8,12,13]. A greater understanding of options for accessible physical activities tailored to individuals with physical disabilities is essential.

Virtual reality (VR), defined as the use of computer-generated interactive simulations, comprises a spectrum of technologies varying in type and technical intricacies, spanning from fully immersive (head-mounted displays [HMDs]) to nonimmersive (Nintendo Wii, Nintendo Switch, PlayStation, Xbox, etc) technologies [14]. VR allows users to engage in physical activities at different levels of immersion in digital landscapes where they can participate in games and tasks that are both appealing and adaptive. The growing adoption of VR, highlighted by its global market valuation of US $59.96 billion in 2022 [15], underscores its potential in various domains, including rehabilitation. One of the primary reasons VR is considered an accessible option for physical activity is its ability to create stimulating, interactive, and multimodal environments tailored to individual needs. Advantages of both nonimmersive and immersive VR interaction include promoting motor learning, motivation, adapting activities to users' needs, and developing a sense of confidence [13,14,16-18]. Over the past decade, numerous studies, both in children and adults with physical disabilities, have highlighted the effectiveness of immersive VR for improving locomotion [18], postural control [19], upper limb functions [20], and physiological outcomes [21]. Moreover, VR offers significant potential for improving cognitive functions [22], such as attention and perception [23]. VR is recognized for sustaining motivation during sessions, thereby enhancing engagement levels and adherence to both therapeutic exercises and physical activities [24,25]. In this way, the effectiveness of VR in enhancing the accessibility of physical activity for people with disabilities has been increasingly acknowledged [26,27].

The practical application and integration of this technology within community settings, such as the accessibility of hardware to individuals that may not be familiar with this technology, remain unexplored. Indeed, the use of these technologies in community settings requires careful consideration. The transition from laboratory use to actual practice in community settings can raise challenges such as safety, space organization, training requirements, and ease of use. The aim of this study was to identify the needs and expectations of people with physical disabilities and their interventionists for the use of a VR physical activity platform in community organizations.

Methods

Study Design

A descriptive qualitative study was conducted using semistructured interviews.

Participants and Recruitment

Individuals with physical disabilities and their interventionists (students or professionals in health-related domains such as physical therapists and kinesiologists), who were associated with a respite camp managed by the nonprofit organization (Adaptavie [28]), were recruited using a convenience sampling approach. No screening process was conducted, as our goal was to capture the broad spectrum of experiences of individuals engaging with the community organization, regardless of their specific disability. To be included in this study, individuals with physical disabilities and interventionists needed to have an active involvement in the summer services for a minimum of 2 weeks and written and oral French proficiency. Participants needed to be able to follow simple instructions.

XSL•FO
RenderX

Setting

The interviews took place in an elementary school that was being used by the community partner (Adaptavie) to offer respite camps. The overarching goal of Adaptavie, a nonprofit community-based organization, is to facilitate physical activity and encourage healthy lifestyles for people with disabilities, with the aim to improve autonomy, socialization, well-being, and global health for this community. Aligned with Adaptavie's goals, the study was integrated into the respite camps that were offered in the summer and during weekends.

Procedure

Participants completed a **s**ociodemographic questionnaire including age and gender. For individuals with physical disabilities, additional questions included the medical diagnosis or type of disability, occupation, assistive technology use, and the frequency of practice of physical activity weekly. For interventionists, information about professional background and educational trajectory, years of experience working with individuals with disabilities, and years of experience within Adaptavie was collected.

The research team developed 2 interview guides—one for individuals with physical disabilities and one for interventionists—through an iterative process (repeatedly refining with input from all stakeholders). These guides were designed to explore and understand the challenges of everyday life and access to physical activities for individuals with physical disabilities, as well as the interventionists' perceptions of various types of adapted physical activities at Adaptavie. The guides also examined previous experiences with VR and perspectives on using VR to promote physical activity, considering participants' viewpoints, expectations, needs, and preferences. Each technical term was accompanied by its definition to ensure clarity for all participants. The questions were then validated by 3 pilot participants (an occupational therapist, a kinesiologist, and an individual with a physical disability). This led to a content validation of both interview guides [29]. The interviews started with introductions and a presentation of the study objectives, followed by a period of discussion. The discussion was structured around 20 open-ended questions covering the following topics: (1) participant's daily activities, (2) interventionist's perception of community services, (3) prior VR experiences, and (4) perspectives on the use of VR during community services. Themes 1 and 2 allow us to understand the realities of individuals with physical disabilities and the workings of the community organization in order to understand if VR could address their specific needs. Theme 3 provides a nuanced perspective, enabling us to refine our understanding of the participants' and interventionists' views based on their previous experiences with VR, as well as their reactions to and feelings about these technologies. A PowerPoint presentation was used to illustrate the definition and types of VR (presenting various immersive and nonimmersive VR devices such as the Wii, the Kinect, VR headsets, etc).

The semistructured interviews were conducted in person during the summer respite services by MB and ET, both trained in qualitative data collection and interviewing skills. The data collection took place over 1 week, with interviews video recorded and averaging 50 minutes in duration.

Data Analysis

Descriptive statistics were calculated for the sociodemographic data. Video files were converted to audio files and transcribed by MB and ET or through TranscribeMe [30] (a web-based transcription service). All transcripts were verified by MB to ensure accuracy. A thematic inductive approach was used to analyze the interviews [31]. The codebook was initially written by MB and ET following several verbatim readings. Subsequently, the codebook underwent continuous iterative updates throughout the analysis, adapting to emerging information from the interviews. Line-by-line coding of the interviews was conducted using NVivo software (version 1.7.1; Lumivero). To ensure coding consensus and ensure rigor and quality of the data, 20% of the interviews were independently coded by 2 research team members, MB and ET. The remaining interviews were independently coded by MB. The final versions were then reviewed and approved by the research team. Following the coding process, unique categories and themes were discerned, drawing upon the extensive expertise of the research team in the use of VR in individuals with physical disabilities and qualitative studies conducted within community organizations. The identified themes were then presented to the research team for thorough discussion and interpretation. In addition, to ensure all discussed topics were included, we reviewed the themes with 2 interventionists. Given that the interviews were conducted in French, the quotes presented in this article have been translated into English and verified by bilingual team members. To enhance the rigor of this study, we employed the Standards for Reporting Qualitative Research [32].

Ethical Considerations

Ethical approval for this study was obtained from the CIUSSS Capitale-Nationale research ethics committee (2021-2064, 2023-2569). All participants provided informed consent and had the right to decline answering our questions or to stop the interview at any time if they wished. All collected data were anonymized and stored on the university's server, with access restricted to the researchers involved in the project to ensure data confidentiality. A financial compensation of CAD $50 (US $35.30) was provided to all participants to cover their travel expenses.

Results

User Statistics

A total of 15 participants, including 8 individuals with physical disabilities aged between 18 and 28 (mean 24, SD 4.28) years and 7 interventionists aged between 19 and 31 (mean 24.87, SD 4.97) years, were interviewed. Tables 1 and 2 present the characteristics of the individuals with physical disabilities and those of the interventionists.

XSL•FO
RenderX

Table . Sociodemographic information of participants with physical disabilities (n=7).

Characteristic	Values
Age (years), mean (SD)	24 (4.28)
Gender (male), n (%)	6 (85.71)
Medical diagnosis, n (%)	
Cerebral palsy	4 (57.14)
Duchenne muscular dystrophy	1 (14.29)
Congenital muscular dystrophy	1 (14.29)
Stroke aftereffects	1 (14.29)
Occupation, n (%)	
Student	3 (42.86)
Worker	1 (14.29)
Worker and student	2 (28.57)
N/A[a]	1 (14.29)
Technical aids, n (%)	
Motorized wheelchair	4 (57.14)
Walking cane	1 (14.29)
Wheelchair	1 (14.29)
N/A	1 (14.29)
Number of physical activities per week, n (%)	
4 times a week	2 (28.57)
2 times a week	1 (14.29)
0 times a week	4 (57.14)

[a]N/A: not applicable.

XSL•FO
RenderX

Table . Sociodemographic information of interventionists working in community organization (n=8).

Characteristic	Values
Age (years), mean (SD)	24.87 (4.97)
Gender, n (%)	
Female	6 (75)
Male	1 (12.5)
N/A[a]	1 (12.5)
Level of education, n (%)	
Bachelor's degree in a health-related field	5 (62.5)
College diploma in a health-related field	2 (25)
Others	1 (12.5)
Role in the association, n (%)	
Coordinator	1 (12.5)
Interventionist	3 (37.5)
Kinesiologist	1 (12.5)
Social interventionist consultor	1 (12.5)
Special education instructor	1 (12.5)
Team leader	1 (12.5)
Years of experience within the organization, n (%)	
10 years	1 (12.5)
2 years	2 (25)
1 year	5 (62.5)

[a]N/A: not applicable.

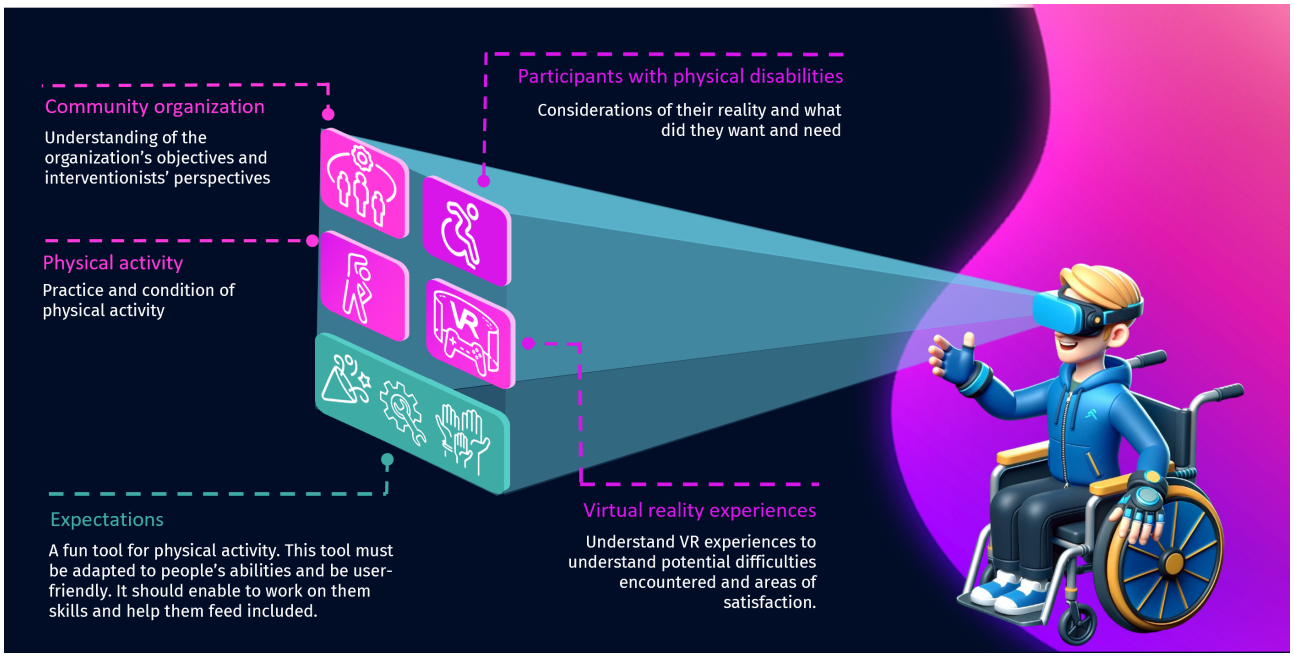
Interviews

The qualitative analysis revealed 4 themes reflecting important elements to consider when implementing VR-based physical activities in community settings: (1) daily activities of individuals with physical disabilities, (2) interventionists' perception of community services, (3) prior VR experiences, and (4) perspectives on the use of VR during community services.

The first theme related exclusively to individuals with physical disabilities, while the second theme was relevant only to interventionists. The first 2 themes provided a better understanding of the lifestyles of the participants with physical disabilities, as well as the interventionist's perceptions of the community services provided by the nonprofit organization. Themes 3 and 4 were applicable to interventionists and individuals with physical disabilities. Theme 3 delved into participants' prior knowledge and experiences with VR, offering insights into their level of comfort and understanding of the technology. Theme 4, on the other hand, outlined essential considerations that must be addressed before establishing a VR platform within community services. These results are summarized in Figure 1.

XSL•FO
RenderX

Figure 1. Considerations for implementing virtual reality within community organizations.



Theme 1: Daily Activities of Individuals With Physical Disabilities

All participants reported that their free time was primarily spent on passive activities, including watching television, browsing social networks, or reading. Activities of daily living reported by all participants were tasks that required the use of their upper limbs, such as cooking, dressing, reaching and grasping objects, or brushing teeth. However, 3 participants mentioned engaging in different active physical activities weekly, with neighborhood walks being the most common among all other reported activities. These physical activities were mostly conducted in services provided by a local nonprofit community organization throughout the year. Participants also expressed interest in participating in 18 physical activities, including soccer, boccia, and swimming, while completing the sociodemographic questionnaire.

Participants expressed that their motor limitations were the primary limiting factor in carrying out activities of daily living:

> *Yes, there are quite a few activities that I find difficult to do. Because of my condition....When things are too heavy, or I have to throw something, or hold them in my hands….*

The participants also highlighted the lack of adaptation in their environment:

> *For sure, cooking is a bit harder because my kitchen isn't adapted at all.*

As for the practice of physical activity, several obstacles were highlighted by all participants. Most participants reported the challenge of accessing the adapted infrastructure, such as specialized gyms, which presents a significant challenge for individuals with disabilities. For example, one participant expressed frustrations over the lack of availability and accessible suitable facilities for daily practice:

> *I'd like to be able to exercise at home, but space is really limited and it's hard to get into a gym. This lack of space is a particular problem when it comes to playing games like boccia. So, I'd like to be able to practice this activity virtually.*

Some participants reported a decrease in rehabilitation services once they reached adulthood, which in turn limited their accessibility of adapted services:

> *I always had followed up when I was a child and at a certain point nothing happened, it was as if by magic I was cured.*

The lack of adapted equipment was also expressed. Moreover, all participants emphasized feeling neglected in access to rehabilitation services, mainly because they perceived individuals with temporary injuries are given priority to maximize recovery:

> *I find that physiotherapy prioritizes those who have accidents, because we can almost bring the person back to independence. But for us who have always had a disability....It's as if we've been left on the side.*

Despite the availability of adapted physical activity programs, the public's lack of awareness of these programs restricts opportunities for participation.

> *Not many disabled people know about all the programs available to them.*

Theme 2: Interventionist's Perception of Community Services

The interventionists highlighted the diverse yet essential goals of the community service organizations. While these objectives vary, they all revolve around promoting the well-being of individuals. Most interventionists emphasized the significance of social integration within the community services to encourage social interactions and improve the social skills of the participants.

XSL•FO
RenderX

> *Above all, we aim to promote social interaction…*
>
> *It's a real opportunity for our members to be in a group and work on their social skills.*

Another objective reported by most interventionists of the community services is the regular practice of (adapted) physical activity. The definition of physical activity varied significantly among interventionists, ranging from simple movement to a more detailed description of adapted physical activity. However, they all emphasized the idea of fun and entertainment in their respective descriptions of physical activity. The participant's level of physical activity also differed from one interventionist to another. For example, while a few noted a moderate level, most interventionists reported that most participants had a low level of physical activity. Importantly, the perceived level of physical activity varied significantly among participants.

Some interventionists perceived community services as a means to promote healthy lifestyles for individuals with physical disabilities, emphasizing their role in enhancing their autonomy. Interventionists also noted that regardless of the primary objectives of community services, these programs offer rare opportunities for individuals with physical disabilities to be engaged in a wide variety of adapted activities. They also highlighted that community services also act as a respite service for participants' families.

Theme 3: Prior VR Experiences

Both individuals with physical disabilities and interventionists reported various levels of familiarity with the general use of VR, encompassing both immersive and nonimmersive experiences. Specifically, their familiarity with nonimmersive VR experiences ranged from 0 to over 12 years.

The systems used to experience VR, primarily for entertainment and recreation, exhibited significant diversity. Half of the individuals with physical disabilities and interventionists had experienced immersive VR devices such as the Oculus. While just under half of the interviewees mentioned experiencing immersive VR through a headset, the majority had engaged with nonimmersive VR via various video game consoles, with the Wii being notably prevalent.

Nevertheless, few individuals with physical disabilities who had prior experience of immersive VR reported side effects during prolonged use, such as dizziness and feelings of disorientation, particularly when engaging in simulations of amusement park rides:

> *In fact, it was a roller coaster simulation, and you had to be seated, because the environment moves. So, the body doesn't necessarily understand everything that's going to happen....It's probably mainly in the duration of exposure that, at some point, things can happen. But for me, it was like that, at the beginning, it was fine, but at some point, you go: Ouh!*

Regarding the installation of VR systems, the majority of interventionists expressed ease in using these systems:

> *It's quite simple, I think. I'm definitely not the most tech-savvy person, but despite that, I still think it's fairly simple.*

The sentiment was echoed by other interventionists, emphasizing the increasing simplicity of technology overall:

> *I'm still amazed every day at how complex it used to be, and now it's become so accessible.*

Both individuals with physical disabilities and interventionists also reported factors influencing the ease of using VR systems, including the importance of concise and precise instructions as well as a period of familiarization. Several interventionists reported the simplicity of VR games, particularly the limited number of buttons, which is crucial in reducing complexity:

> *There are things that are really simple, with not many buttons, and then there are games where you can get lost in 50 million buttons to use with lots of different levels, and all that.*

Some interventionists acknowledged that the complexity of using VR systems may vary depending on the participant's clinical characteristics:

> *And I'm still pretty sure that if there are people with physical disabilities, but who are used to playing with these kinds of consoles, they'll be able to adapt. But maybe for people who play a little less, it can be complicated.*

Most individuals with physical disabilities reported the need for assistance when putting on a VR headset. The difficulty in using the systems varied based on specific games, equipment, and the clinical profile. In fact, many participants reported facing challenges with the controllers:

> *I find the controller is not designed for people like us who have a bit of difficulty.*

Despite these technical difficulties, some participants found adaptive solutions:

> *But instead of holding the controller upright, I hold it horizontally.*

Theme 4: Perspectives on the Use of VR During Community Services

Advantages of Using VR Within Community Settings

Interventionists noted that VR provides numerous possibilities to improve access to physical activities and rehabilitation.

> *I really think it would be great in the sense that it would open up a lot of possibilities that we don't necessarily have right now. It's like having access to more material. Especially in camps for people with physical disabilities, accessibility isn't always easy.*

Interventionists highlighted the potential to assist individuals in reaching personal objectives, such as improving overall health, enhancing cognitive skills, and maximizing autonomy. Specifically, interventionists pointed out a significant benefit of VR, which involves incorporating movements needed for daily activities into a diverse array of engaging VR experiences.

> *Some [individuals with physical disabilities] find it difficult to cook, to chop vegetables, because they have lost dexterity or other skills. So, I think that by doing it in the form of a virtual game, they could*

XSL•FO
RenderX

> *develop these abilities, but gradually, and by doing it virtually, they can't really get hurt. It would allow them to practice, and maybe by doing it virtually, they would improve, and they could transfer it back into their daily life.*

Participants expressed a shared perspective, highlighting that VR could allow regular practice of specific movement skills essential for daily activities, all while avoiding the perception of engaging in a formal exercise:

> *to help me be more mobile and less, to have less trouble picking things up… without it seeming like an exercise.*

Facilitator

Interventionists emphasized that a crucial factor for the success of VR lies in creating experiences that inspire movement and designing games that are tailored to accommodate a range of user abilities:

> *VR needs to be tailored to their level, with various levels, different ways of doing things, and different settings....I just see it as a wonderful opportunity to do things beyond what they have the chance to do here.*

Participants, echoing this need, requested a variety of gameplay options (eg, 1- or 2-handed gameplay, ergonomic controller, etc) that cater to a range of physical capabilities:

> *I expect it to be accessible...that I feel included.*

The VR platform should feature an intuitive and user-friendly interface.

> *It would be great if it were easy to use, if it worked. That there aren't too many unforeseen events. Of course, you can't control it, but to say to everyone, 'We're going to do this,' and then it doesn't work. It's kind of disappointing...I'm thinking mainly of testing. Not installing someone, and then realizing that they can't do it. It's always a bit disappointing.*

Creating a positive experience is key, as another interventionist noted:

> *I think one of the most important objectives would be to reconcile the relation they have with physical activity, because some of them associate it with their rehabilitation, or with things that aren't necessarily pleasant. Really make the experience positive, then, if need be, reconcile the relationship with physical activity.*

Additionally, the importance of having a dedicated support contact for assistance was highlighted. This need is complementary to the request for comprehensive training materials, such as detailed user guides and instructional videos, to effectively familiarize both interventionists and participants with the platform's functionalities. The interventionists also emphasized the importance of a safe, practical, and user-friendly interface. They noted the need for thorough supervision of participants during equipment setup and gameplay, enabling them to observe and, if necessary, intervene.

Social interaction within the games is highlighted for fostering a sense of community, friendship, and enjoyment. Both participants and interventionists recognized the value in scenarios where players can interact, compete, or cooperate within various virtual environments.

Participants and interventionists also showed a keen interest in diverse virtual settings, covering a spectrum from modified popular commercial games to those specifically designed for physical activity. Desirable features include customizable sound settings, motivational reward systems, the ability to create and save user profiles, and improved movement sensitivity in games that cater to individuals with severe limitations.

Barriers and Fears

The main obstacle identified by interventionists is the heterogeneity of participants' limitations.

> *One barrier might be their limitations. It's hard to predict for each limitation, what you can do.*

Lack of financial and human resources also emerged as major obstacles for interventionists.

> *Ensuring we have the right resources for implementation is key. This means having ample financial backing for a diverse range of games. Plus, we can't overlook the importance of human resources, especially the number of interventionists involved.*

Technological issues, such as malfunctioning equipment and poor internet connectivity, were also a concern.

Expected Benefits

The expected benefits of using VR are numerous, ranging from encouraging people to move while having fun, to providing access to new activities. According to an interventionist, it can enable them to move and gain confidence in their abilities:

> *allow them to discover new possibilities that they might not have thought they could do....Doing things that they wouldn't necessarily do in their everyday life, and it could give them confidence in their abilities.*

The benefits for participants vary widely. Some participants wish to engage in activities that they cannot do in the real world:

> *Experiences that I can't do because of my disability, but to be able to do them, thanks to this, might be nice.*

Others really want to maintain their skills:

> *I know that playing a game won't help me regain my mobility, but at least I'll be able to keep it for longer.*

Some hope to become more equal with able-bodied individuals:

> *Maybe become a little more equal with everyone else. In a virtual world, we can become equal again.*

Discussion

Principal Results

This study addressed the needs and expectations of using VR within community settings, as perceived by individuals with

XSL•FO
RenderX

physical disabilities and interventionists. Our study revealed 4 essential themes for the implementation of VR within community settings, including participants' daily activities, stakeholders' perceptions of community services, previous VR experiences, and perspectives on the use of VR within these services.

Both individuals with physical disabilities and interventionists were enthusiastic about the possibility of having personalized VR-based interventions. In other words, the adaptability and accessibility of VR platforms were identified as crucial factors for tailoring interventions to the specific needs of individuals with physical disabilities or the organization objective. Such adaptability allows for the creation of unique experiences that are tailored to the individual's capabilities and preferences. Ultimately, there is a strong belief in the potential of VR to improve access to physical activity, promote social integration, and foster autonomy among people with disabilities.

One of the main challenges underscored by both individuals with physical disabilities and interventionists involves maintaining the gains obtained through rehabilitation while aligning with the individual's objectives. VR technologies offer a promising solution to engage in physical activity, serving as either a supplementary tool to rehabilitation or as a post-rehabilitation intervention [26]. One participant expressed concern about immersive VR, since it can isolate the user. This concern was of importance for this participant, as their personal objective was to alleviate loneliness and ultimately enhance social participation. Although adapting VR games to facilitate social engagement presents challenges, a recent study indicates that immersive VR can foster connections within a virtual environment, enabling individuals to support each other's progress and thereby enriching the social aspect of rehabilitation [33].

Both individuals with physical disabilities and interventionists emphasized the importance of tailoring VR games to the clinical profiles and individuals' goals and abilities to maximize their benefits. One approach to achieve this is by integrating motor learning principles, which involve manipulating factors such as task difficulty, repetition, motivation, and engagement [34]. Motor learning principles encompass a set of processes grounded in neuroplasticity principles associated with practice or experience, leading to relatively permanent changes in motor skills. While the primary objective of community services may not always focus on improving motor functions, our results highlighted the demand for adapted games and the aspiration to foster various skills among individuals with physical disabilities. Consequently, the development of fully customizable games could meet the expectations of both participants and interventionists by allowing all necessary elements to be adapted, added, or removed to enhance the user experience.

Crafting customizable games presents a unique blend of challenges and innovative solutions, aimed at bridging the gap between entertainment and the rehabilitative process. At the core of this initiative is the necessity to design VR games and applications that suit a diverse spectrum of interests and abilities, all the while prioritizing rehabilitation objectives. Based on the findings of this study, the goal is to design experiences that allow participants to engage in the development of various skills that resemble leisure activities more so than therapeutic interventions. However, this ambition encounters several obstacles, notably in usability and the transition from development to user experience, a transition that is often expensive and time-consuming, limiting the extent of testing and refinement needed to ensure the applications are user-friendly. This issue is particularly pronounced in rehabilitation-focused VR applications, as they cater to a more specialized audience compared to mainstream gaming. To overcome these obstacles, a structured approach to VR content development is essential, one that balances functionality with affordability. Tools like StellarX [35], which advocate for a "drag and drop" method to create VR environments without the need for coding, represent promising steps towards more accessible VR content creation.

It is important to ascertain the suitability of different game types in addressing participants' expectations. Immersive VR has been shown to enhance the sense of presence, making virtual experiences more engaging and appealing. This is supported by advancements in VR technology, which continue to become more sophisticated and financially accessible. Nonetheless, challenges such as cybersickness with symptoms like nausea and dizziness affect many users, especially during high-motion experiences. Cybersickness is notably more prevalent in immersive VR than nonimmersive VR, with around 65.2% of users affected and 23.9% experiencing severe discomfort [36]. So this is important to think about strategy to decrease this rate as "rotation snapping and translation snapping" [37] or "foveated blur" [38]. In addition, the design of HMDs requires careful consideration to ensure positive user experiences. Studies have highlighted the advantages of HMD-VR, including elevated immersion and emotional responses, which can significantly increase satisfaction and creative freedom in gameplay [39-41].

Implementing VR in community organizations presents significant challenges, especially regarding the financial and human resources required for effective deployment. As noted by participants, it is crucial to adopt affordable technologies tailored to community organizations, ideally standalone systems that do not require additional computers. This approach aims to streamline the deployment process and ensure accessibility for community organizations with limited resources.

Finally, to effectively use VR technologies for adapted physical activity, it is crucial to understand how the technological capabilities of VR align with the specific needs of both practitioners and interventionists. This customization process involves ensuring the VR equipment is comfortable to use, providing clear and concise instructions, and offering prompt assistance to resolve any issues encountered during usage. The goal is to create an environment where professionals and individuals with physical disabilities have the skills and knowledge to use VR for rehabilitation and leisure activities. The implementation of knowledge dissemination tools, such as introductory coaching sessions, user manuals, and instructional videos, can play a significant role in achieving this goal [42].

XSL•FO
RenderX

Limitations

One of the primary limitations of this study is the sample size. Despite our efforts to encompass a wide range of participants, the relatively small number of people involved may not fully represent the diverse populations served by community services. Another limitation of this study is the diversity of backgrounds, expertise, and varying levels of experience in VR of the interventionists who participated in this study. The heterogeneity of their academic and professional backgrounds and their varying degrees of familiarity with VR could influence their understanding of the implementation of an adapted physical activity program for people with physical disabilities.

In addition, it is important to note that while this study focused on specific types of community services, its findings may not be universally applicable across all community organizations. The diversity in operational structures, target populations, and services among community centers suggests that the considerations about implementing these kinds of interventions might vary. The findings of this study apply primarily to individuals with similar disabilities and may not extend to populations with varying medical diagnoses or a broader range of disabilities.

Conclusions

The integration of VR to provide accessible physical activity opportunities in community services offers promising opportunities but also presents significant challenges. Tailoring VR experiences to individual needs, addressing the barriers to physical activity participation, and ensuring continuity of care from childhood to adulthood are critical areas that need attention. Community organizations play a vital role in facilitating physical activity and social integration for individuals with disabilities. Future research should focus on developing standardized, accessible, and engaging VR interventions that can be easily implemented in community settings, with appropriate resources and training for interventionists. Such an approach could provide an alternative form of physical activity designed to improve the quality of life and independence of individuals with disabilities.

Acknowledgments

This project was funded by the Ministry of Economy and Innovation (GQ133218). We would like to thank the partner organization Adaptavie for their collaboration.

Authors' Contributions

MTR, KLB, MD, and ML conceptualized the study and acquired the funding. MTR and MB designed the research question. ET and MB collected and analyzed the data. AR and MB verified the analysis. AR wrote the paper. MTR supervised the research. All authors participated in drafting and critically reviewing the manuscript.

Conflicts of Interest

None declared.

References

1. Rimmer JH, Riley B, Wang E, Rauworth A, Jurkowski J. Physical activity participation among persons with disabilities: barriers and facilitators. Am J Prev Med 2004 Jun;26(5):419-425. [doi: 10.1016/j.amepre.2004.02.002] [Medline: 15165658]
2. L'emploi du temps des personnes ayant une incapacité au Canada [Article in French]. Statistique Canada. URL: https://www150.statcan.gc.ca/n1/pub/89-654-x/89-654-x2022001-fra.htm [accessed 2025-01-09]
3. Downs M, MacDermid J, Connelly D, McDougall J. Current experiences and future expectations for physical activity participation: perspectives of young people with physical disabilities and their rehabilitation clinicians. Dev Neurorehabil 2021 Feb;24(2):73-84. [doi: 10.1080/17518423.2020.1771787] [Medline: 32496839]
4. Lee IM, Shiroma EJ, Lobelo F, et al. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. Lancet 2012 Jul 21;380(9838):219-229. [doi: 10.1016/S0140-6736(12)61031-9] [Medline: 22818936]
5. Kissow AM. Participation in physical activity and the everyday life of people with physical disabilities: a review of the literature. Scand J Disabil Res 2015 Apr 3;17(2):144-166. [doi: 10.1080/15017419.2013.787369]
6. Calder A, Sole G, Mulligan H. The accessibility of fitness centers for people with disabilities: a systematic review. Disabil Health J 2018 Oct;11(4):525-536. [doi: 10.1016/j.dhjo.2018.04.002] [Medline: 29936142]
7. Martin JJ. Benefits and barriers to physical activity for individuals with disabilities: a social-relational model of disability perspective. Disabil Rehabil 2013;35(24):2030-2037. [doi: 10.3109/09638288.2013.802377] [Medline: 23781907]
8. Rimmer JH, Padalabalanarayanan S, Malone LA, Mehta T. Fitness facilities still lack accessibility for people with disabilities. Disabil Health J 2017 Apr;10(2):214-221. [doi: 10.1016/j.dhjo.2016.12.011] [Medline: 28143707]
9. Martin Ginis KA, Ma JK, Latimer-Cheung AE, Rimmer JH. A systematic review of review articles addressing factors related to physical activity participation among children and adults with physical disabilities. Health Psychol Rev 2016 Dec;10(4):478-494. [doi: 10.1080/17437199.2016.1198240] [Medline: 27265062]

XSL•FO
RenderX

10. Sallis JF, Owen N, Fisher EB. Ecological models of health behaviour. In: Health Behavior and Health Education: Theory, Research, and Practice: John Wiley & Sons; 2015:465-486.
11. Kahn EB, Ramsey LT, Brownson RC, et al. The effectiveness of interventions to increase physical activity. A systematic review. Am J Prev Med 2002 May;22(4 Suppl):73-107. [doi: 10.1016/s0749-3797(02)00434-8] [Medline: 11985936]
12. Christensen KM, Holt JM, Wilson JF. Effects of perceived neighborhood characteristics and use of community facilities on physical activity of adults with and without disabilities. Prev Chronic Dis 2010 Aug;7(5):A105. [Medline: 20712932]
13. Demers M, Martinie O, Winstein C, Robert MT. Active video games and low-cost virtual reality: an ideal therapeutic modality for children with physical disabilities during a global pandemic. Front Neurol 2020;11:601898. [doi: 10.3389/fneur.2020.601898] [Medline: 33381077]
14. Demers M, Fung K, Subramanian SK, Lemay M, Robert MT. Integration of motor learning principles into virtual reality interventions for individuals with cerebral palsy: systematic review. JMIR Serious Games 2021 Apr 7;9(2):e23822. [doi: 10.2196/23822] [Medline: 33825690]
15. Virtual reality (VR) market size and share report, 2030. Grand View Research. URL: https://www.grandviewresearch.com/industry-analysis/virtual-reality-vr-market [accessed 2025-01-09]
16. Heller T, Ying Gs GS, Rimmer JH, Marks BA. Determinants of exercise in adults with cerebral palsy. Public Health Nurs 2002;19(3):223-231. [doi: 10.1046/j.0737-1209.2002.19311.x] [Medline: 11967109]
17. Robert MT, Levin MF. Validation of reaching in a virtual environment in typically developing children and children with mild unilateral cerebral palsy. Dev Med Child Neurol 2018 Apr;60(4):382-390. [doi: 10.1111/dmcn.13688] [Medline: 29427357]
18. Winter C, Kern F, Gall D, Latoschik ME, Pauli P, Käthner I. Immersive virtual reality during gait rehabilitation increases walking speed and motivation: a usability evaluation with healthy participants and patients with multiple sclerosis and stroke. J Neuroeng Rehabil 2021 Apr 22;18(1):68. [doi: 10.1186/s12984-021-00848-w] [Medline: 33888148]
19. Goffredo M, Pagliari C, Turolla A, et al. Non-immersive virtual reality telerehabilitation system improves postural balance in people with chronic neurological diseases. J Clin Med 2023 Apr 28;12(9):3178. [doi: 10.3390/jcm12093178] [Medline: 37176618]
20. Shahmoradi L, Almasi S, Ahmadi H, et al. Virtual reality games for rehabilitation of upper extremities in stroke patients. J Bodyw Mov Ther 2021 Apr;26:113-122. [doi: 10.1016/j.jbmt.2020.10.006] [Medline: 33992230]
21. Qian J, McDonough DJ, Gao Z. The effectiveness of virtual reality exercise on individual's physiological, psychological and rehabilitative outcomes: a systematic review. Int J Environ Res Public Health 2020 Jun 10;17(11):4133. [doi: 10.3390/ijerph17114133] [Medline: 32531906]
22. Georgiev DD, Georgieva I, Gong Z, Nanjappan V, Georgiev GV. Virtual reality for neurorehabilitation and cognitive enhancement. Brain Sci 2021 Feb 11;11(2):221. [doi: 10.3390/brainsci11020221] [Medline: 33670277]
23. Mura G, Carta MG, Sancassiani F, Machado S, Prosperini L. Active exergames to improve cognitive functioning in neurological disabilities: a systematic review and meta-analysis. Eur J Phys Rehabil Med 2018 Jun;54(3):450-462. [doi: 10.23736/S1973-9087.17.04680-9] [Medline: 29072042]
24. Laver KE, Lange B, George S, Deutsch JE, Saposnik G, Crotty M. Virtual reality for stroke rehabilitation. Cochrane Database Syst Rev 2017 Nov 20;11(11):CD008349. [doi: 10.1002/14651858.CD008349.pub4] [Medline: 29156493]
25. Levac D, Rivard L, Missiuna C. Defining the active ingredients of interactive computer play interventions for children with neuromotor impairments: a scoping review. Res Dev Disabil 2012;33(1):214-223. [doi: 10.1016/j.ridd.2011.09.007] [Medline: 22093667]
26. Zhou S. Using virtual reality to promote physical activity. J Software Eng Appl 2020;13(11):312-326. [doi: 10.4236/jsea.2020.1311021]
27. Woessner MN, Tacey A, Levinger-Limor A, Parker AG, Levinger P, Levinger I. The evolution of technology and physical inactivity: the good, the bad, and the way forward. Front Public Health 2021;9:655491. [doi: 10.3389/fpubh.2021.655491] [Medline: 34123989]
28. Adaptavie. URL: https://www.adaptavie.org/ [accessed 2025-01-09]
29. Sánchez-Guardiola Paredes C, Aguaded Ramírez EM, Rodríguez-Sabiote C. Content validation of a semi-structured interview to analyze the management of suffering. Int J Environ Res Public Health 2021 Oct 29;18(21):11393. [doi: 10.3390/ijerph182111393] [Medline: 34769919]
30. TranscribeMe: fast & accurate human transcription services. TranscribeMe. 2024. URL: https://www.transcribeme.com/ [accessed 2025-01-09]
31. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
32. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: a synthesis of recommendations. Acad Med 2014 Sep;89(9):1245-1251. [doi: 10.1097/ACM.0000000000000388] [Medline: 24979285]
33. Kenyon K, Kinakh V, Harrison J. Social virtual reality helps to reduce feelings of loneliness and social anxiety during the Covid-19 pandemic. Sci Rep 2023 Nov 7;13(1):19282. [doi: 10.1038/s41598-023-46494-1] [Medline: 37935718]
34. Kleim JA, Jones TA. Principles of experience-dependent neural plasticity: implications for rehabilitation after brain damage. J Speech Lang Hear Res 2008 Feb;51(1):S225-S239. [doi: 10.1044/1092-4388(2008/018)] [Medline: 18230848]

XSL•FO
RenderX

35. StellarX par OVA | Platforme sans code pour les orginasations | IA, XR, VR, AR [Article in French]. StellarX. URL: https://fr.stellarx.ai/ [accessed 2025-01-09]
36. Garrido LE, Frías-Hiciano M, Moreno-Jiménez M, et al. Focusing on cybersickness: pervasiveness, latent trajectories, susceptibility, and effects on the virtual reality experience. Virtual Real 2022;26(4):1347-1371. [doi: 10.1007/s10055-022-00636-4] [Medline: 35250349]
37. Farmani Y, Teather RJ. Evaluating discrete viewpoint control to reduce cybersickness in virtual reality. Virtual Real 2020 Dec;24(4):645-664. [doi: 10.1007/s10055-020-00425-x]
38. Hussain R, Chessa M, Solari F. Mitigating cybersickness in virtual reality systems through foveated depth-of-field blur. Sensors (Basel) 2021 Jun 10;21(12):4006. [doi: 10.3390/s21124006] [Medline: 34200616]
39. Hunicke R, LeBlanc M, Zubek R. MDA: a formal approach to game design and game research. Association for the Advancement of Artificial Intelligence. URL: https://aaai.org/papers/ws04-04-001-mda-a-formal-approach-to-game-design-and-game-research/ [accessed 2025-01-09]
40. Slater M, Sanchez-Vives MV. Enhancing our lives with immersive virtual reality. Front Robot AI 2016 Dec 19;3. [doi: 10.3389/frobt.2016.00074]
41. Tao G, Garrett B, Taverner T, Cordingley E, Sun C. Immersive virtual reality health games: a narrative review of game design. J Neuroeng Rehabil 2021 Feb 11;18(1):31. [doi: 10.1186/s12984-020-00801-3] [Medline: 33573684]
42. Glegg SMN, Levac DE. Barriers, facilitators and interventions to support virtual reality implementation in rehabilitation: a scoping review. PM R 2018 Nov;10(11):1237-1251. [doi: 10.1016/j.pmrj.2018.07.004] [Medline: 30503231]

Abbreviations

HMD: head-mounted display
VR: virtual reality

Edited by K Erenli, T Leung; submitted 21.04.24; peer-reviewed by F Altinay, L Bruton, M Spewak; revised version received 05.11.24; accepted 20.11.24; published 17.01.25.

Please cite as:
Ramos A, Boisvert M, Traverse E, Levac D, Lemay M, Demers M, Bordeleau M, Ruest SM, Périnet-Lacroix R, Best KL, Robert MT Bridging Needs and Expectations of Individuals With Physical Disabilities and Community Services Stakeholders for the Cocreation of an Adapted Physical Activity Platform in Virtual Reality: Qualitative Study
JMIR Serious Games 2025;13:e59704
URL: https://games.jmir.org/2025/1/e59704
doi:10.2196/59704

© Aurélien Ramos, Maxence Boisvert, Elodie Traverse, Danielle Levac, Martin Lemay, Marika Demers, Martine Bordeleau, Sarah-Maude Ruest, Roxanne Périnet-Lacroix, Krista L Best, Maxime T Robert. Originally published in JMIR Serious Games (https://games.jmir.org), 17.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Feasibility and Usability of an Artificial Intelligence—Powered Gamification Intervention for Enhancing Physical Activity Among College Students: Quasi-Experimental Study

Yanan Gao[1,2], PhD; Jinxi Zhang[2], PhD; Zhonghui He[3], PhD; Zhixiong Zhou[2], PhD

[1]School of Physical Education and Sport Science, Fujian Normal University, Fuzhou, China

[2]Institute of Artificial Intelligence in Sports, Capital University of Physical Education and Sports, 11 Beisanhuan West Road, Beijing, China

[3]The Department of Physical Education, Peking University, Beijing, China

Corresponding Author:
Zhixiong Zhou, PhD
Institute of Artificial Intelligence in Sports, Capital University of Physical Education and Sports, 11 Beisanhuan West Road, Beijing, China

Abstract

Background: Physical activity (PA) is vital for physical and mental health, but many college students fail to meet recommended levels. Artificial intelligence (AI)-powered gamification interventions through mobile app have the potential to improve PA levels among Chinese college students.

Objective: This study aimed to assess the feasibility and usability of an AI-powered gamification intervention.

Methods: A quasi-experimental study spanning 2 months was conducted on a sample of college students aged 18 to 25 years old from 18 universities in Beijing. PA data were recorded using the ShouTi Fitness app, and participant engagement was evaluated through surveys. User satisfaction was gauged through the System Usability Scale, while the intervention's feasibility was assessed through Spearman rank correlation analysis, Mann-Whitney tests, and additional descriptive analyses.

Results: As of July 2023, we enrolled 456 college students. In total, 18,073 PA sessions were recorded, with men completing 8068 sessions and women completing 10,055 sessions. The average PA intensity was 7 metabolic equivalent of energy (MET)s per session. Most participants preferred afternoon sessions and favored short-duration sessions, with men averaging 66 seconds per session and women 42 seconds. The System Usability Scale score for the intervention based on app is 65.2. Users responded positively to the integration of AI and gamification elements, including personalized recommendations, action recognition, smart grouping, dynamic management, collaborative, and competition. Specifically, 341 users (75%) found the AI features very interesting, 365 (80%) were motivated by the gamification elements, 364 (80%) reported that the intervention supported their fitness goals, and 365 (80%) considered the intervention reliable. A significant positive correlation was observed between the duration of individual PA and intervention duration for men (ρ=0.510, P<.001), although the correlation was weaker for women (ρ=0.258, P=.046). However, the frequency of PA declined after 35 days.

Conclusions: This study provides pioneering evidence of the feasibility and usability of the AI-powered gamification intervention. While adherence was successfully demonstrated, further studies or interventions are needed to directly assess the impact on PA levels and focus on optimizing long-term adherence strategies and evaluating health outcomes.

(JMIR Serious Games 2025;13:e65498) doi:10.2196/65498

KEYWORDS

physical activity; gamification; artificial intelligence; digital health; digital intervention; feasibility study

Introduction

Background

Physical activity (PA) is crucial for maintaining both physical and mental health. It plays a significant role in preventing chronic diseases, improving cardiovascular health, and building strength and endurance [1]. For college students, regular PA is especially important as it can alleviate academic stress, enhance cognitive function, and contribute to overall well-being [2]. Despite these benefits, many college students adopt sedentary lifestyles, often due to increased academic pressures, social activities, and time constraints, and time constraints following their transition to college life [3]. This global trend of sedentary behavior poses a significant public health challenge. In China, national surveys reveal that many students fail to meet the recommended 150 minutes of moderate PA per week [4]. A 2021 national survey further highlighted that 12.94% of students from "Double First-Class" universities failed physical fitness standards, while 14.73% were classified as overweight and 5.5%

XSL•FO
RenderX

as obese [5]. Addressing these issues is crucial for promoting students' health and well-being.

Various interventions have been developed to address these concerns, including traditional methods [6-8], digital approaches [9-11], and gamification strategies [11-13]. Traditional interventions typically include structured PA programs, workshops, and physical education classes. While these methods have shown positive outcomes in increasing PA levels and improving metrics like cardiovascular fitness and body composition [14], they often struggle with maintaining long-term participant engagement [15]. Digital interventions, which include mobile apps, web-based platforms, and wearable devices, have become increasingly popular. These tools offer features such as activity tracking, reminders, goal setting, and personalized feedback, which enhance motivation and adherence [16]. However, many existing PA apps and wearables are not optimized for efficiency or comprehensive assessment, and they often fail to fully engage users over time [17].

Gamification interventions, which incorporate game design elements like points, levels, badges, and leaderboards into nongame contexts, have emerged as an effective strategy to increase PA participation [18]. These interventions aim to make PA more enjoyable and engaging, thereby enhancing motivation and participation. Popular among young people, gamification has been shown to increase PA frequency, intensity, and participation, particularly when multiple game elements are used simultaneously [19,20]. However, more research is needed to identify the optimal combination of gamification elements for promoting PA. The field of gamification in PA interventions is still developing, and further advancements are required [21].

The integration of artificial intelligence (AI) into digital interventions holds significant promise. AI-powered solutions can personalize user experiences, enhance engagement, and improve adherence by analyzing user data (eg, activity levels, preferences, and behavior patterns) to tailor content, set realistic goals, and provide feedback [21,22]. Despite these benefits, research on combining AI and gamification in PA interventions is still in its infancy. Most studies in this field focus on machine learning techniques and are limited in scope and quality [23]. A recent scoping review identified AI-powered digital solutions, such as recommendation systems and conversational agents, as key components in PA interventions, but few studies have explored the combination of AI with gamification elements [24]. In addition, most existing research focuses on adults with chronic conditions [25] or children [26], leaving a gap in research targeting college students, a population that could benefit greatly from innovative PA interventions.

Objective

This study aimed to evaluate the feasibility and usability of an AI-powered gamification intervention based on the ShouTi Fitness app to improve PA levels among Chinese college students. By assessing user experience, engagement, and the intervention's impact on PA frequency and intensity, this study seeks to provide valuable insights into the potential of integrating AI and gamification to promote healthier lifestyles in this population. Through this investigation, we aim to fill existing gaps in the literature by offering a personalized and engaging solution for increasing PA among college students.

Methods

Study Design

This study employed a quasi-experimental design without a control group to assess the feasibility and usability of an AI-powered gamification intervention that includes personalized recommendations, team collaboration and competition, reward systems, and challenge mechanisms.

Participants

Participants for the study were recruited through advertisements for the "Beijing Student AI Physical Fitness Competition", which were posted on the official websites of the 18 participating universities and on the ShouTi Fitness app. The recruitment process lasted for 1 month. Prospective participants expressed initial interest by responding to these advertisements. Eligibility was then determined based on their age, enrollment in universities in Beijing and their residence within university premises. Eligible individuals received informed consent forms that outlined the study's purpose, potential risks, and benefits. Consent was obtained both in person and through email, and participants provided their contact information for follow-up purposes. Participants were excluded from the study if they had missing or abnormal PA data recorded in the app, ensuring the reliability and accuracy of the collected data. The final sample comprised 456 college students (Figure 1), aged 18 to 25 years old, with 239 men and 217 women.

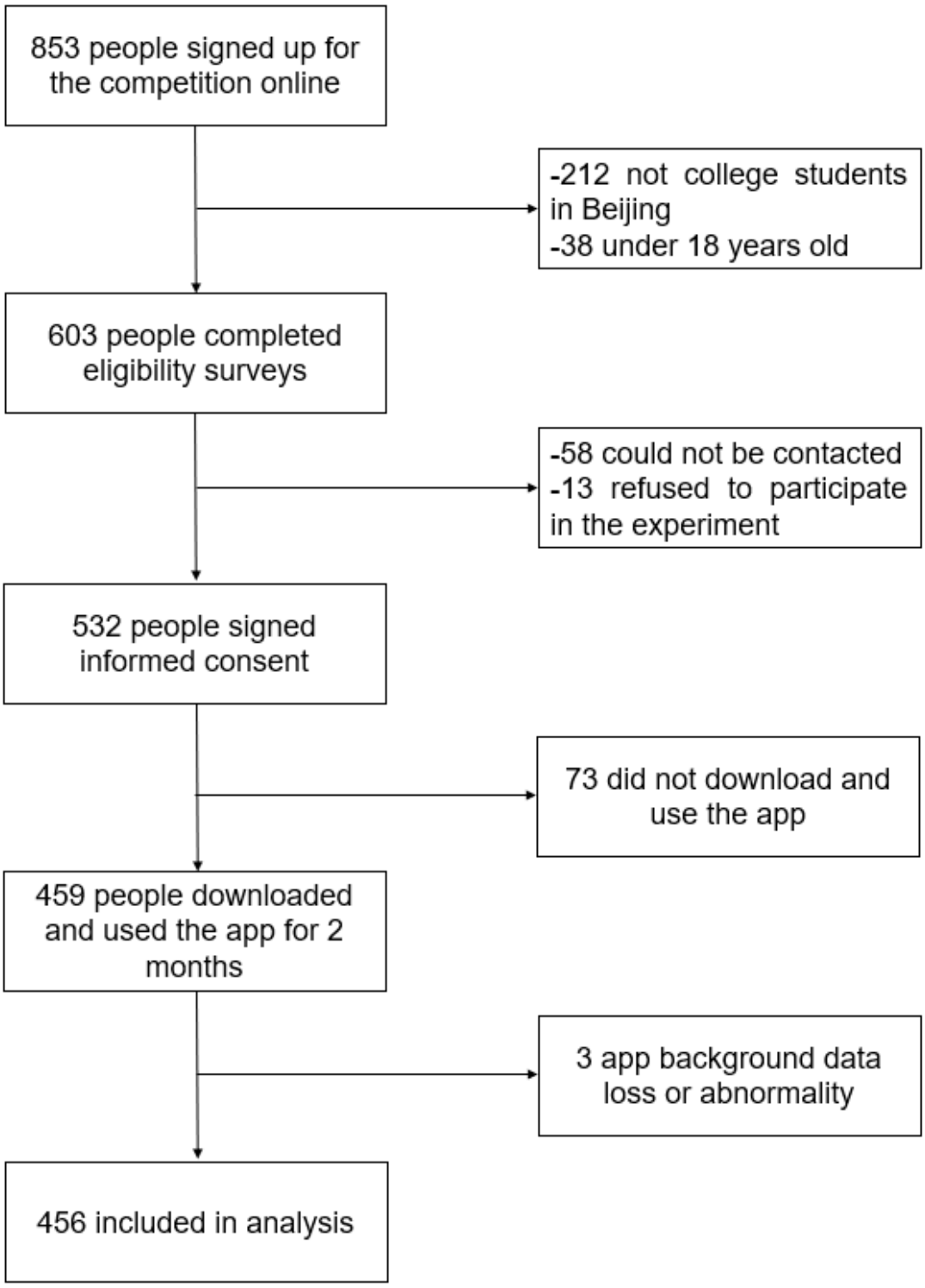
Eligible individuals received informed consent forms that outlined the study's purpose, potential risks, and benefits. Consent was obtained both in person and via email, and participants provided their contact information for follow-up purposes.

Participants were excluded from the study if they had missing or abnormal physical activity (PA) data recorded in the app, ensuring the reliability and accuracy of the collected data.

The final sample comprised 456 college students (Figure 1), aged 18 to 25 years old, with 239 men and 217 women.

XSL•FO
RenderX

Figure 1. Participant flowchart.



Procedures

Participants who agreed to take part in the study provided written consent. Upon completing these initial steps, participants were instructed to download and register the ShouTi Fitness app. After successful registration, participants selected their university and joined the "Beijing Student AI Physical Fitness Competition," where they were randomly assigned to teams based on their university affiliation.

The data collection phase of the study began on May 1, 2023, and concluded on June 30, 2023, spanning a total of 2 months. During this period, participants engaged in an AI-powered gamification intervention using the ShouTi Fitness app. Participants had the autonomy to manually start and stop the app. Throughout the study, participants seamlessly balanced their daily academic responsibilities and personal commitments while using the app. This flexibility allowed them to engage in

XSL•FO
RenderX

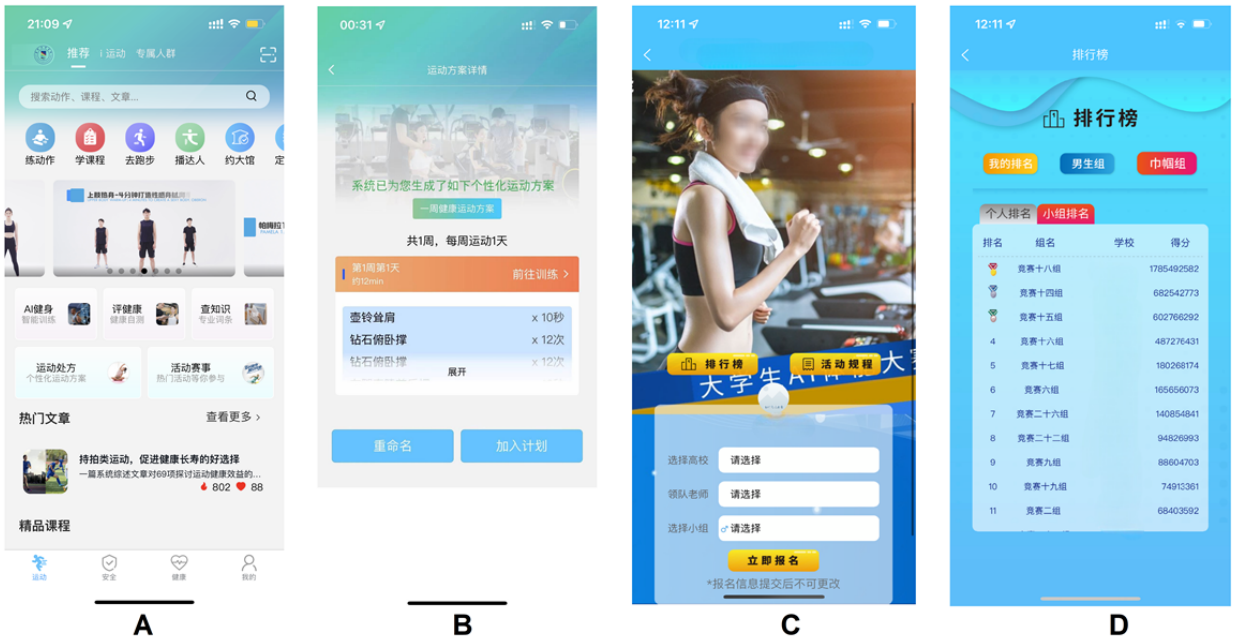
PA at times that best fit their schedules, supporting a more adaptable approach to increasing PA.

Intervention

ShouTi Fitness App Overview

ShouTi Fitnessis a free app available for both Android and iOS devices, (Figure 2A), designed to serve as part of the AI-powered gamification intervention. The app leverages AI to assess exercise capabilities and generate personalized exercise prescription that align with users' fitness levels and goals (Figure 2B). The app provides the latest fitness trends, tips, and equipment reviews, and includes tools to track exercise frequency, intensity and duration. In addition, ShouTi Fitness has a built-in exercise risk screening feature to help users identify potential risks and a community interaction function that allows users to connect, share experiences, and motivate each other. All data are securely managed to protect user privacy. The ShouTi Fitness app was developed using principles from Self-Determination Theory, Goal-Setting Theory, and Behavior Change Techniques, with a focus on intrinsic motivation, personalized goals, and sustained engagement. Gamification elements such as team collaboration, dynamic goals, and self-monitoring were designed to meet users' psychological needs and promote long-term PA.

Figure 2. ShouTi fitness interface screenshot.



AI-Powered Gamification Intervention

The ShouTi Fitness app integrates a variety of AI- and gamification-powered features designed to optimize PA engagement, enhance motivation, and personalize the user experience. Below are the key components of the AI-powered gamification intervention:

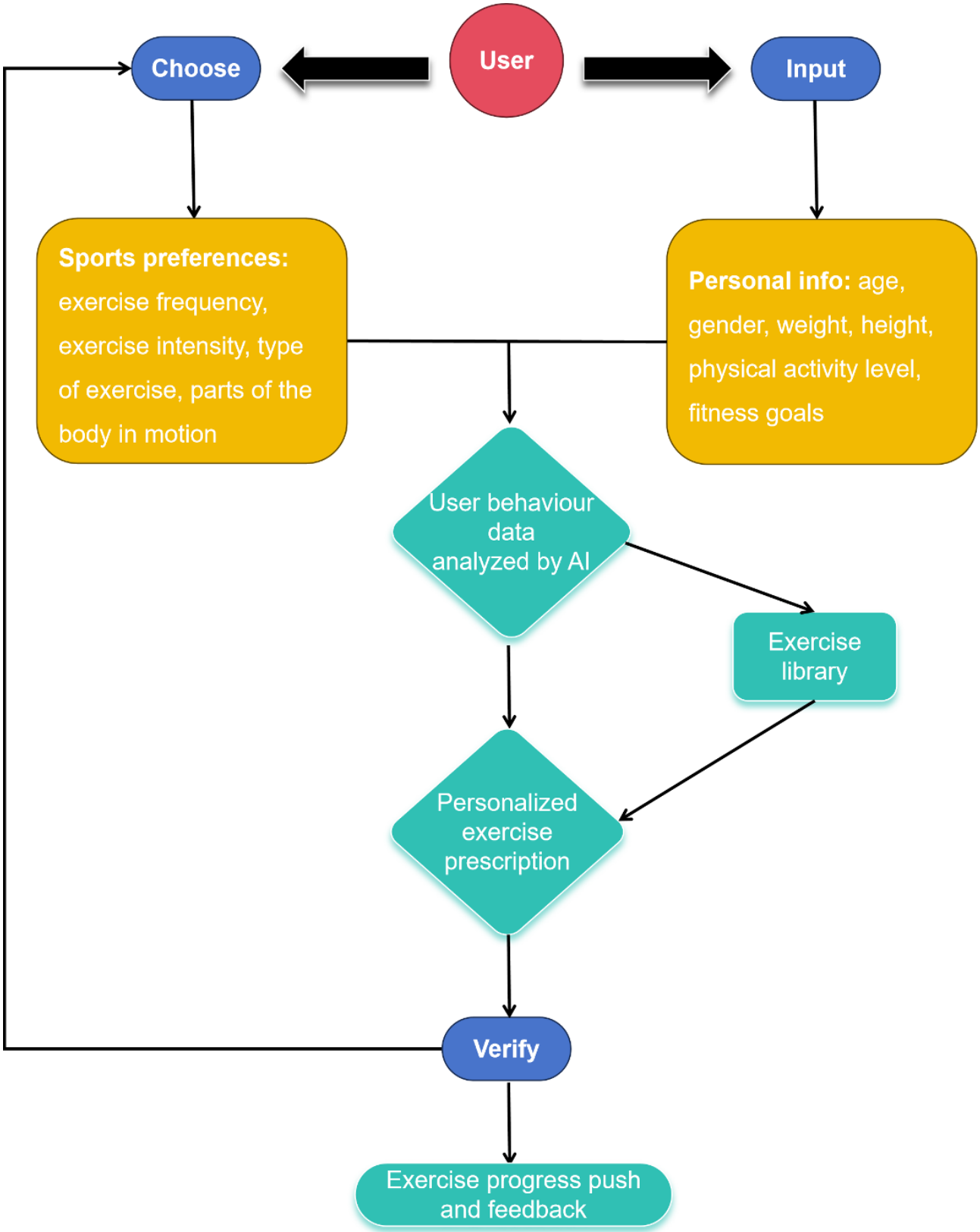
Personalized Exercise Recommendations

The main purpose of personalized exercise recommendations is to ensure that users are engaging in targeted and goal-oriented exercise that is effective in meeting their fitness objectives. By receiving daily personalized exercise recommendations, participants are more likely to stay motivated and maintain consistency (Figure 2B). This personalized approach promotes the achievement of individual fitness goals while simultaneously contributing to the overall progress of their teams. As each user completes their personalized workout plan, the team's collective performance improves, which directly influences their rankings and engagement in the gamification intervention.

The app's AI recommendation system analyses user data, including activity levels, preferences, and behavior patterns, to create highly personalized exercise prescriptions (Figure 3). The exercise library in the ShouTi Fitness app features a diverse selection of activities targeting various body parts, ranging from moderate to vigorous PA (3 - 10 metabolic equivalent of energy [METs), with an average MET value of 7 METs (Multimedia Appendix 1).

XSL•FO
RenderX

Figure 3. Personalized exercise prescription development. AI: artificial intelligence.

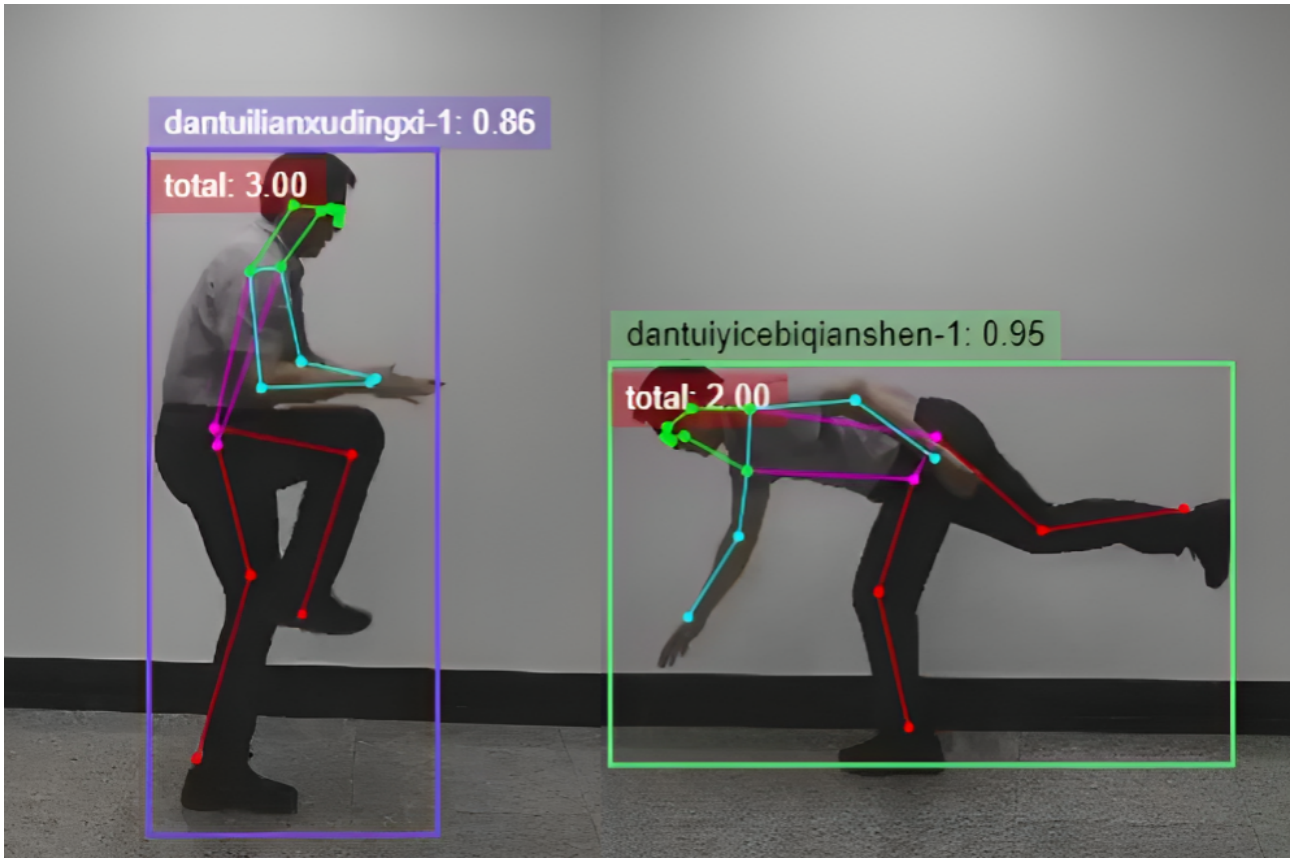


Action Recognition

The ShouTi Fitness app uses AI technology to provide real-time feedback and movement recognition (Figure 4). Its aim is to enhance the interactivity and accuracy of users' exercise routines. The app functions by using motion recognition algorithms to monitor the user's movements and posture throughout workouts, offering immediate feedback. This prompt correction assists users in adjusting their form in real time, thereby improving the effectiveness of their workouts and reducing the risk of performing exercises incorrectly, which could otherwise lead to injury.

XSL•FO
RenderX

Figure 4. Artificial Intelligence motion recognition and monitoring.



Smart Grouping

AI plays a central role in creating balanced, optimized teams for the gamification. Using a combination of random assignment and a matching algorithm, participants are grouped into teams of 5 (Figure 2C). The AI analyzes a range of user characteristics, including age, gender, height, weight, PA level, and fitness goals, to create teams that are comparable in terms of physical capacity and fitness objectives. The algorithm uses a K-means clustering method to optimize team composition by minimizing intra-team differences based on users' characteristics, ensuring that the competition remains fair and engaging for all participants:

$$Dij=\sum k=1n(xik-xjk)2$$

where Dij is the distance between participants i and j based on their characteristics.

Dynamic Management

The main function of dynamic management is to reset weekly rankings, match participants with new competitors, and dynamically adjust the competitive environment to keep the challenge fresh and motivating. At the end of each week, the AI calculates a team's total score by aggregating the individual performances of all 5 members:

$$\text{Total Score} = \sum i=15(\text{PA Frequency } i + \text{PA Duration } i)$$

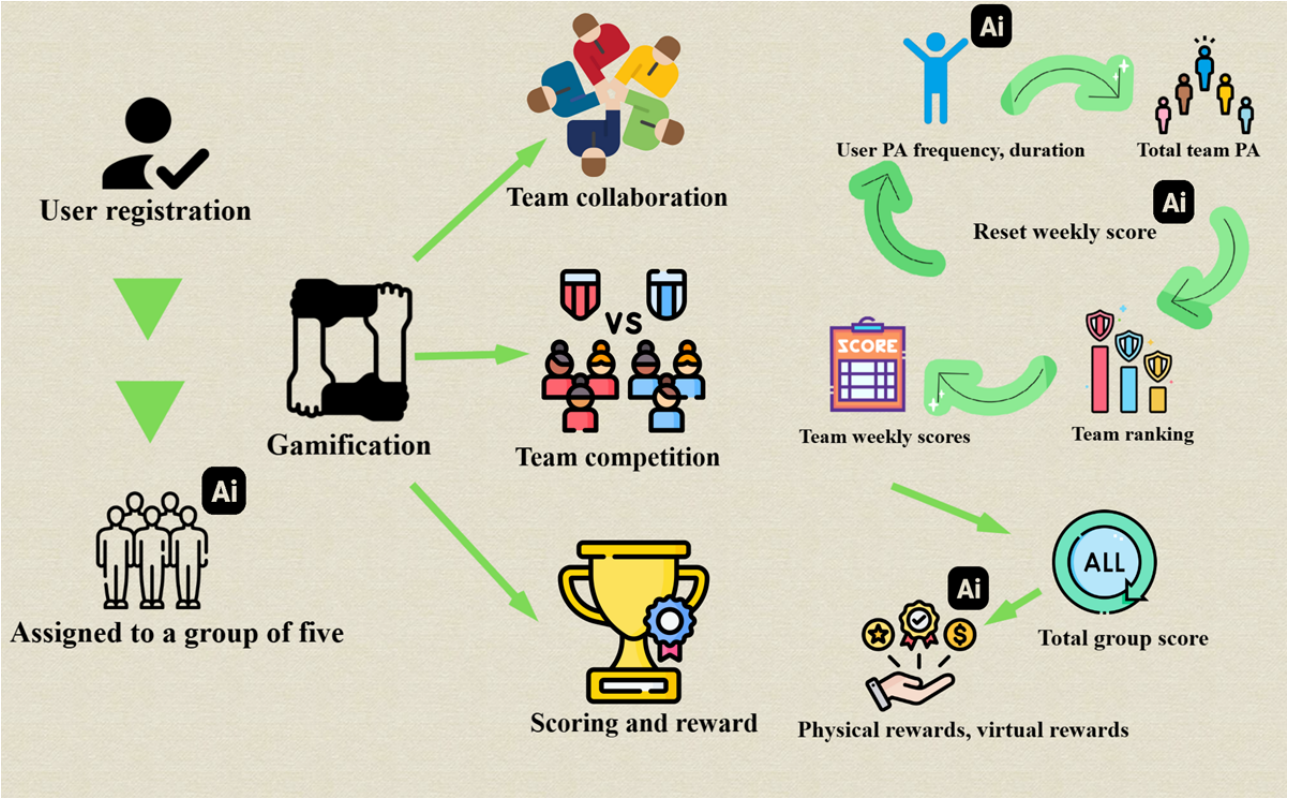
By resetting weekly scores and rematching competitors, it ensures a fair and balanced competitive environment that encourages consistent effort, sustained engagement, and long-term success.

Team Collaboration and Competition

Collaboration and competition are central to the gamification experience (Figure 5). Teams are not only created with balanced composition, but also engaged in friendly competition through weekly rankings. At the end of each week, the AI assigns points to the top-performing teams based on their accumulated scores, with the highest-ranking teams earning rewards (Figure 2D). The competition fosters a sense of community, as users work together within teams to achieve collective fitness goals. The reward system includes both virtual achievements, such as badges and unlockable content, and tangible physical prizes, further incentivizing engagement and consistent effort across the app's user base.

XSL•FO
RenderX

Figure 5. Artificial intelligence—powered gamification development. Note: The artificial intelligence functions involved in the diagram are: artificial intelligence-based smart grouping; artificial intelligence-based personalized exercise recommendation; artificial intelligence-based visual action recognition and feedback; artificial intelligence-based gamification intervention dynamic management. AI: artificial intelligence; PA: physical activity.



Measures

Demographic Characteristics

At baseline, participants completed an online form within the app to provide their demographic information. This included age, gender, school, residence, height, weight, PA levels, and fitness goals.

Objective Physical Activity Measurement

PA was recorded using the ShouTi Fitness app, which logged the start time, end time, and duration of each PA session with high precision, down to the nearest second. This detailed data collection ensured accurate tracking of participants' activity levels and allowed for precise measurement of their engagement with the exercise program. Descriptive statistics were used to summarize PA session counts and durations. Given the nonnormal distribution of the data, assessed using the D'Agostino test for normality, the median IQR were reported. The IQR is provided from the 25th to the 75th percentile. Percentages were used for categorical data. Scatter plots visualized the usage characteristics of ShouTi Fitness throughout the intervention period.

Usability and Acceptability of the AI-Powered Gamification Intervention

To evaluate the usability and acceptability of the AI-powered gamification intervention, participants completed an online survey within the app at follow-up. The survey included the System Usability Scale (SUS) and additional self-reported questions. The SUS, consisting of 10 items with responses ranging from strongly agree to strongly disagree, was scored on a 0 - 100 scale [27]. Scores were calculated by summing the converted values and multiplying by 2.5. Based on Bangor et al's [28] cutoffs, SUS scores were categorized as follows: excellent (85.58 - 100), good (72.75 - 85.57), OK (52.01 - 72.74), and poor (0 - 52.00). Additional questions assessed participants' views on the intervention's usefulness, their willingness to continue using it, and their likelihood of recommending it. Participants also provided feedback on technical issues and their engagement with the intervention's features, including the AI technology and gamification elements.

Feasibility and Adherence of the AI-Powered Gamification Intervention

The study primarily focused on assessing the feasibility and adherence to the AI-powered gamification intervention by analyzing the correlation between PA session duration and the overall intervention duration. While this measure provides insights into user engagement over time, further assessments are necessary to directly evaluate the intervention's effectiveness in increasing PA levels. The adherence effectiveness of the intervention was evaluated through Spearman rank correlation analysis, examining the relationship between the duration of individual PA sessions and the length of the intervention.

Characteristics of Participation: Gender Differences

To investigate gender differences in intervention participation characteristics, Mann-Whitney U tests were conducted on PA patterns for different genders. These analyses aimed to identify usage trends and preferences, providing insights for optimizing the intervention's features and user-engagement strategies.

XSL•FO
RenderX

Data Analysis

Statistical analyses were conducted using SPSS (version 26.0, IBMCorp) for data processing and GraphPad Prism for graphical representations. A significance level of $P<.05$ was used for all tests to determine statistical significance.

Ethical Considerations

Eligible individuals received informed consent forms that outlined the study's purpose, potential risks, and benefits. Consent was obtained both in person and through email, and participants provided their contact information for follow-up purposes. The study adhered to ethical guidelines, reviewed by the Academic Committee of Capital Institute of Physical Education (approval number: 2024A063), with informed consent obtained from all participants, who were fully briefed about the study and their rights. Participant confidentiality was strictly maintained throughout the research.

Results

Participant Activity Overview

This study involved 456 college students aged 18 to 25 years old, including 239 men (52.4%) and 217 women (47.6%). Over the 2-month study period, a total of 18,073 online PA sessions were recorded, all involving moderate or vigorous PA (3 - 10 METs), with men contributing 8068 sessions and women 10,055 sessions. On average, men participated in 15 PA sessions (median) over the 2 months, while women participated in 22 sessions. The average duration of a single PA session was 66 seconds for men and 42 seconds for women (Table 1).

Table . Participant characteristics.

	Total (n=456), median (IQR)	Men (n=239), median (IQR)	Women (n=217), median (IQR)
Sessions	18 (6-39.5)	15 (5-39)	22 (9-40.5)
Duration (seconds)	52 (22-111)	66 (33-139)	42 (13-93)

Usability and Acceptability of the Gamified AI App

Table 2 presents the usability and satisfaction results for the AI-powered gamification intervention. The average SUS score was 65.2, reflecting an "ok" usability rating. Of the 456 participants, 58 (13%) rated the experience as excellent, 197 (43%) rated it as good, 183 (40%) rated it as okay, and 18 (4%) rated it as poor, with the majority giving positive feedback. Regarding the intervention's effectiveness in helping users achieve their fitness goals, 205 out of 456 participants (45%) strongly agreed and 159 out of 456 participants (35%) agreed, while only 24 out of 456 participants (5%) disagreed, showing a general consensus on its effectiveness. Most participants were satisfied with the AI functions: 191 out of 456 participants (42%) were very satisfied, 173 out of 456 participants (38%) were satisfied, 68 out of 456 participants (15%) were neutral, and 24 out of 456 participants (5%) were dissatisfied, reflecting strong approval of the intervention's tailored plans.

XSL•FO
RenderX

Table . Participant ratings and satisfaction with the artificial intelligence—powered gamification intervention.

System Usability Scale score rating for intervention	Values, n (%)
System Usability Scale score for the intervention	
Excellent	58 (13%)
Good	197 (43%)
Ok	183 (40%)
Poor	18 (4%)
The intervention helped me achieve my fitness goals	
Strongly agree	205 (45%)
Agree	159 (35%)
Neutral	68 (15%)
Disagree	24 (5%)
Be satisfied with the artificial intelligence functions	
Strongly agree	191 (42%)
Agree	173 (38%)
Neutral	68 (15%)
Disagree	24 (5%)
Would continue to use the intervention in future	
Strongly agree	228 (50%)
Agree	137 (30%)
Neutral	68 (15%)
Disagree	23 (5%)
Would recommend the intervention to others	
Strongly agree	201 (44%)
Agree	164 (36%)
Neutral	68 (15%)
Disagree	23 (5%)
Rate the reliability of this intervention	
Excellent	160 (35%)
Good	205 (45%)
Ok	68 (15%)
Poor	23 (5%)
Be attracted by the artificial intelligence features (eg, personalized recommendations, action recognition, smart grouping, and dynamic management)	
Strongly agree	182 (40%)
Agree	159 (35%)
Neutral	68 (15%)
Disagree	45 (10%)
Be motivated by the gamification elements (eg, team collaboration, team competitions, and rewards)	
Strongly agree	219 (48%)
Agree	146 (32%)
Neutral	68 (15%)
Disagree	23 (5%)

XSL•FO
RenderX

Regarding future use, 228 out of 456 participants (50%) strongly agreed, and 137 out of 456 participants (30%) agreed that they would continue to use ShouTi Fitness for similar interventions, indicating good potential for continued use.The likelihood of recommending the intervention was also high, with 201 out of 456 participants (44%) strongly agreeing and 164 out of 456 participants (36%) agreeing, indicating users find it valuable. In terms of reliability, 160 out of 456 participants (35%) rated the intervention as excellent, 205 out of 456 participants (45%) rated it as good, 68 out of 456 participants (15%) rated it as okay.

The AI features, such as personalized recommendations, action recognition, smart grouping, and dynamic management, were appealing to 182 out of 456 participants (40%) who strongly agreed and 159 out of 456 participants (35%) who agreed. Gamification elements, such as team collaboration, team competitions and rewards, were well-received, with 219 out of 456 participants (48%) strongly agreeing and 146 out of 456 participants (32%) agreeing that these features were motivating, highlighting their effectiveness in boosting user engagement and motivation.

Feasibility and Adherence of the AI-Powered Gamification Intervention

The distribution of daily PA sessions among the 456 students throughout the intervention period is shown in Multimedia Appendix 2, with a breakdown by gender shown in Multimedia Appendix 3. The frequency of daily PA sessions varied over time, with an initial increase followed by a decrease.

A Mann-Kendall test was used to analyze the time series data and the resulting trend curve is shown in Multimedia Appendix 4. The analysis showed that participants increased their daily PA sessions during the first 30 days of the intervention. However, from day 35 a significant decrease in PA frequency was observed, which continued until the end of the intervention. This pattern was consistent for both male and female participants.

Table 3 presents the results of the Spearman rank correlation analysis, which examined the relationship between the duration of each PA session and the total duration of the intervention. The analysis revealed a significant positive correlation (ρ=0.464, P<.001), suggesting that participants tended to increase the duration of their individual PA sessions as the intervention progressed. This suggests that the intervention was effective in increasing PA session adherence over time.

Notably, the correlation was stronger for male participants (ρ=0.510, P<.001) compared with female participants (ρ=0.258, P=.046). This pattern supports the idea that the intervention was more effective in promoting sustained PA engagement among men.

Table . Time-series correlation of the duration of each physical activity session.

Physical activity sessions	Spearman coefficient	*P* value
Total (N=18,073)	0.464	<.001
Men (n=8068)	0.51	<.001
Women (n=10,055)	0.258	.046

Characteristics of Participation: Gender Differences

Multimedia Appendices 5 and 6 provide an insight into the usage patterns of the AI-powered gamification intervention among college students. The data show that men engaged in PA mainly in the dawn (23.4%), followed by the afternoon (21.8%), and the forenoon (20.3%). In contrast, women were most active in the afternoon (29.4%), followed by the forenoon (21.5%), and dawn (21.1%). Overall, the afternoon proved to be the peak time for PA, with minimal participation at dusk and in the evening (0.1% and 0.3% respectively). In addition, both men (94.%) and women (97.2%) showed a strong preference for short PA sessions of 1 - 10 minutes, with longer sessions less common. This suggests a clear preference for short PA sessions among students. A significant gender difference was observed in the duration of a single PA session (Z=29.585, P<.001). Men exhibited longer PA durations than women, indicating a significant difference in PA patterns between genders when using the intervention.

Discussion

Principal Findings

This study assessed the feasibility and usability of an AI-powered gamification intervention, designed to improve PA levels among Chinese college students. Results showed that the intervention was effective in engaging users, particularly through its gamification elements and AI features. While overall usability and satisfaction were high, the intervention's impact on PA frequency diminished after the first 30 days, suggesting the need for strategies to sustain long-term engagement. Notably, gender differences in activity patterns were also observed, underscoring the necessity for tailored interventions to address diverse user preferences.

Over the 2-month intervention period, the average PA intensity was 7 METs per session. Participants engaged in the intervention an average of 18 times, with men participating 15 times and women 22 times. Most participants preferred afternoon sessions and favored short-duration sessions, with men averaging 66 seconds per session and women 42 seconds.

AI-Powered Gamification Intervention

Gamification has long been recognized as an effective strategy for increasing user engagement and enhancing app effectiveness [29]. Previous research consistently shows that smartphone-based gamification can significantly boost PA participation rates by providing motivation through rewards, competition, and goal setting [20,30-33]. The integration of AI with gamification, as exemplified by the ShouTi Fitness app,

XSL•FO
RenderX

offers innovative opportunities to further enhance user motivation and engagement. In this study, the incorporation of AI (eg, personalized exercise recommendations, action recognition, smart grouping, and dynamic management) and gamification elements (eg, team competition, rewards, and leaderboards) played a pivotal role in motivating users.

The gamification elements, particularly team competition, rewards, and leaderboards, significantly boosted participants' accountability and fostered a sense of achievement [29]. Weekly team matches created both a competitive and collaborative atmosphere, where users could cooperate within their teams while striving to outperform other teams. This dynamic structure, supported by real-time team rankings, allowed users to track their performance and compare it with others, reinforcing the idea that individual contributions impacted the team's overall success. Rewards earned by reaching specific milestones, provided an extra layer of motivation, further promoting sustained engagement. During the first month, these gamification elements could contribute to longer PA session durations and increased session frequency, supporting the Achievement Goal Theory as a useful framework for understanding how gamification features drive motivation [34].

The combination of AI-powered personalization and gamification principles significantly contributed to the success of this intervention. Specifically, AI provides tailored feedback and challenges based on user data. This level of customization made the PA experience more relevant and engaging [35-37]. Users were more likely to maintain regular PA as the intervention continuously aligned with their fitness levels and preferences. Furthermore, AI-based visual simulation technology, which offers interactive content, holds great potential for further boosting engagement [38]. By providing an immersive and dynamic experience, this technology can sustain user interest and motivate continued participation. However, despite the positive reception of AI features, around 10% of participants reported that some recommendations felt repetitive or not sufficiently tailored to their fitness levels. This feedback highlights a critical area for improvement, refining the personalization algorithms to ensure that all recommendations remain fresh, relevant, and appropriately challenging.

Challenges in Long-Term Engagement

Despite the initial success, the observed decline in PA frequency after 35 days points to the challenges of maintaining long-term engagement in gamified AI interventions. This decline in activity may be due to several factors. First, the novelty of the intervention may have initially led to high engagement, but as this novelty wore off, users' interest likely declined [39]. In addition, the appeal of the intervention's challenges may fade as users become familiar with its features, while the absence of continuous incentives may reduce motivation. The competitive elements of gamification, while initially effective, can also lead to fatigue, especially when combined with privacy concerns or social overload [40]. The SUS score of 65.2 indicates that the app's overall usability fell short of expectations for a highly user-friendly application. Several factors may have contributed to this result. User feedback highlighted issues such as app stability and occasional malfunctions in AI features, which likely impacted the overall usability score. Technical disruptions can disrupt the seamless user experience and contribute to user dissatisfaction [41]. In addition, the app's interface and usability may not have fully aligned with the preferences of all users, especially those less familiar with technology-driven fitness interventions.

To combat this decline, regular updates to challenges and goals based on user feedback could reignite interest. Exploring innovative gamification techniques such as interactive storylines or augmented reality could help maintain engagement. In addition, regular reminders, special events, and seasonal challenges could help maintain user interest and enthusiasm over time. For the long-term success of the intervention, regular updates of the app are necessary to introduce new features and ensure compatibility with evolving platforms.

Strengths and Limitations

This study has several strengths.: It is the first to explore the integration of AI and gamification in a mobile health app to enhance PA, offering novel insights into how these technologies can boost PA among college students. With 456 participants from 18 universities in Beijing, the large sample size ensures diverse representation and improves the generalizability of the findings. The use of objective data to measure PA sessions ensures a reliable, bias-free assessment of user engagement, enhancing the study's validity. The detailed analysis of intervention patterns, including time of day, session duration, and gender differences, provides insights into user preferences that can inform future interventions.

However, this study also has limitations. The 2-month intervention period may not be sufficient to assess the long-term sustainability of PA behaviors. The decline in PA frequency after 35 days suggests that a longer intervention period might yield more insights into long-term effectiveness and user retention. While the AI-powered gamification effectively promoted initial engagement, the decline after day 35 highlights the need for strategies to maintain long-term adherence. Although the study indicates a positive trend in PA adherence, it does not assess the health effectiveness or long-term impact on physical health or fitness outcomes. Further research is needed to evaluate the broader health-related effectiveness of the intervention.

Conclusions

The AI-powered gamification intervention shows promise for improving PA levels among college students. While it demonstrates strong initial engagement and positive feedback, addressing challenges related to long-term engagement, app stability, and refinement of AI algorithms is essential to maximize its potential. Continued research into optimizing these features, along with a focus on long-term PA adherence, will be critical to improving the effectiveness of the intervention and ensuring sustained user engagement in future interventions.

XSL•FO
RenderX

Acknowledgments

The authors sincerely thank all co-authors and institutions for their valuable contributions to this research. This work was funded by the Emerging Interdisciplinary Platform for Medicine and Engineering in Sports (EIPMES).

This study was funded by the Emerging Interdisciplinary Platform for Medicine and Engineering in Sports (EIPMES) and the National Key Research and Development Program of China (Grant No. 2020YFC2006200).

Data Availability

The datasets generated during the current study are available from the corresponding author upon reasonable request.

Authors' Contributions

YG conceived and designed the study, performed the experiments, and wrote the manuscript. JZ analyzed the data and contributed to the experimental design. ZH provided critical revisions to the manuscript and assisted with data interpretation. ZZ supervised the research, secured funding, and reviewed the final manuscript. All authors read and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Exercise movement library.
[PDF File, 4569 KB - games_v13i1e65498_app1.pdf]

Multimedia Appendix 2
Distribution of daily physical activity frequency during the intervention.
[PNG File, 74 KB - games_v13i1e65498_app2.png]

Multimedia Appendix 3
Distribution of daily physical activity frequency by gender.
[PNG File, 138 KB - games_v13i1e65498_app3.png]

Multimedia Appendix 4
Trend plot of daily physical activity sessions and intervention.
[PNG File, 91 KB - games_v13i1e65498_app4.png]

Multimedia Appendix 5
Daily physical activity frequency patterns.
[PNG File, 29 KB - games_v13i1e65498_app5.png]

Multimedia Appendix 6
Duration of single physical activity session.
[PNG File, 45 KB - games_v13i1e65498_app6.png]

References

1. Warburton DER, Nicol CW, Bredin SSD. Health benefits of physical activity: the evidence. CMAJ 2006 Mar 14;174(6):801-809. [doi: 10.1503/cmaj.051351] [Medline: 16534088]
2. Herbert C, Meixner F, Wiebking C, Gilg V. Regular physical activity, short-term exercise, mental health, and well-being among university students: the results of an online and a laboratory study. Front Psychol 2020;11:509. [doi: 10.3389/fpsyg.2020.00509] [Medline: 32528333]
3. Lacaille LJ, Dauner KN, Krambeer RJ, Pedersen J. Psychosocial and environmental determinants of eating behaviors, physical activity, and weight change among college students: a qualitative analysis. J Am Coll Health 2011;59(6):531-538. [doi: 10.1080/07448481.2010.523855] [Medline: 21660808]
4. Zhang M, Ma Y, Xie X, et al. Trends in insufficient physical activity among adults in China 2010-18: a population-based study. Int J Behav Nutr Phys Act 2023 Jul 17;20(1):87. [doi: 10.1186/s12966-023-01470-w] [Medline: 37460936]
5. Jiang Q, Huang X, Wang Z, Dai X, Li R, Cui D. Regional differences of physical fitness and overweight and obesity prevalence among college students before and after COVID-19 pandemic since the “double first-class” initiative in China. Front Public Health 2024;11:1252270. [doi: 10.3389/fpubh.2023.1252270]

XSL•FO
RenderX

6. Worobetz A, Retief PJ, Loughran S, et al. A feasibility study of an exercise intervention to educate and promote health and WELL-being among medical students: the “MED-WELL” programme. BMC Med Educ 2020 Jun 3;20(1):183. [doi: 10.1186/s12909-020-02097-2] [Medline: 32493427]
7. Faro J, Wright JA, Hayman LL, Hastie M, Gona PN, Whiteley JA. Functional resistance training and affective response in female college-age students. Med Sci Sports Exerc 2019 Jun;51(6):1186-1194. [doi: 10.1249/MSS.0000000000001895] [Medline: 30640287]
8. Barfield J, Channell B, Pugh C, Tuck M, Pendel D. Format of basic instruction program resistance training classes: effect on fitness change in college students. Physical educator 2012;69:325 [FREE Full text]
9. Tulasiram B, Chandrasekaran B. Are smartphones better in guiding physical activity among sedentary young adults? A randomised controlled trial. Muscle Ligaments and Tendons J 2021;11(1):77. [doi: 10.32098/mltj.01.2021.10]
10. Al-Eisa E, Al-Rushud A, Alghadir A, et al. Effect of motivation by “Instagram” on adherence to physical activity among female college students. Biomed Res Int 2016;2016:1546013. [doi: 10.1155/2016/1546013] [Medline: 27034927]
11. St Quinton T, Morris B, Barwood MJ, Conner M. Promoting physical activity through text messages: the impact of attitude and goal priority messages. Health Psychol Behav Med 2021 Mar 1;9(1):165-181. [doi: 10.1080/21642850.2021.1891073] [Medline: 34104555]
12. Hamari J, Koivisto J. Social motivations to use gamification: an empirical study of gamifying exercise. 2013 Presented at: Proceedings of the 21st European Conference on Information Systems; Jun 5-8, 2013; Utrecht, The Netherlands p. 18-19 URL: http://aisel.aisnet.org/ecis2013_cr/105 [accessed 2024-12-25]
13. Mora-Gonzalez J, Pérez-López IJ, Esteban-Cornejo I, Delgado-Fernández M. A gamification-based intervention program that encourages physical activity improves cardiorespiratory fitness of college students: “The Matrix rEFvolution Program”. Int J Environ Res Public Health 2020 Jan 30;17(3):877. [doi: 10.3390/ijerph17030877] [Medline: 32019259]
14. Heath GW, Parra DC, Sarmiento OL, et al. Evidence-based intervention in physical activity: lessons from around the world. Lancet 2012 Jul 21;380(9838):272-281. [doi: 10.1016/S0140-6736(12)60816-2] [Medline: 22818939]
15. Müller-Riemenschneider F, Reinhold T, Nocon M, Willich SN. Long-term effectiveness of interventions promoting physical activity: A systematic review. Prev Med 2008 Oct;47(4):354-368. [doi: 10.1016/j.ypmed.2008.07.006]
16. Sanders JP, Loveday A, Pearson N, et al. Devices for self-monitoring sedentary time or physical activity: a scoping review. J Med Internet Res 2016 May 4;18(5):e90. [doi: 10.2196/jmir.5373] [Medline: 27145905]
17. McCallum C, Rooksby J, Gray CM. Evaluating the impact of physical activity apps and wearables: interdisciplinary review. JMIR Mhealth Uhealth 2018 Mar 23;6(3):e58. [doi: 10.2196/mhealth.9054] [Medline: 29572200]
18. Deterding S, Dixon D, Khaled R, Nacke L. From game design elements to gamefulness: defining" gamification. 2011 Presented at: Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments; Sep 28-30, 2011; Tampere, Finland p. 9-15. [doi: 10.1145/2181037.2181040]
19. Pickering K, Pringle A. Gamification for physical activity behaviour change. Perspect Public Health 2018 Nov;138(6):309-310. [doi: 10.1177/1757913918801447]
20. Xu L, Shi H, Shen M, et al. The effects of mHealth-based gamification interventions on participation in physical activity: systematic review. JMIR Mhealth Uhealth 2022 Feb 3;10(2):e27794. [doi: 10.2196/27794] [Medline: 35113034]
21. Hariri M, Stone R. Gamification in physical activity: state-of-the-art. IJACSA 2023;14(10). [doi: 10.14569/IJACSA.2023.01410105]
22. Fang J, Lee VC, Ji H, Wang H. Enhancing digital health services: A machine learning approach to personalized exercise goal setting. Digit Health 2024;10:20552076241233247. [doi: 10.1177/20552076241233247] [Medline: 38384365]
23. Tolks D, Schmidt JJ, Kuhn S. The role of AI in serious games and gamification for Health: scoping review. JMIR Serious Games 2024 Jan 15;12:e48258. [doi: 10.2196/48258] [Medline: 38224472]
24. Gabarron E, Larbi D, Rivera-Romero O, Denecke K. Human factors in AI-driven digital solutions for increasing physical activity: scoping review. JMIR Hum Factors 2024 Jul 3;11:e55964. [doi: 10.2196/55964] [Medline: 38959064]
25. Valladares-Rodriguez S, Fernández-Iglesias MJ, Anido-Rifón L, Facal D, Rivas-Costa C, Pérez-Rodríguez R. Touchscreen games to detect cognitive impairment in senior adults. A user-interaction pilot study. Int J Med Inform 2019 Jul;127:52-62. [doi: 10.1016/j.ijmedinf.2019.04.012] [Medline: 31128832]
26. Keshav NU, Vogt-Lowell K, Vahabzadeh A, Sahin NT. Digital attention-related augmented-reality game: significant correlation between student game performance and validated clinical measures of attention-deficit/hyperactivity disorder (ADHD). Children (Basel) 2019 May 28;6(6):72. [doi: 10.3390/children6060072] [Medline: 31142022]
27. Brooke J. SUS-A quick and dirty usability scale. In: Usability Evaluation in Industry 1996:189-194 URL: https://www.taylorfrancis.com/chapters/edit/10.1201/9781498710411-35 [accessed 2024-12-25]
28. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the System Usability Scale. Int J Hum Comput Interact 2008 Jul 29;24(6):574-594. [doi: 10.1080/10447310802205776]
29. Rehman U, Abbasi AZ, Ting DH, Hassan M, Khair N. Exploring the impact of gamified experiences on user engagement in fitness apps: a GAMEFULQUEST perspective. IEEE Trans Eng Manage 2023;71:3613-3628. [doi: 10.1109/TEM.2023.3347231]
30. Yang Y, Hu H, Koenigstorfer J. Effects of gamified smartphone applications on physical activity: a systematic review and meta-analysis. Am J Prev Med 2022 Apr;62(4):602-613. [doi: 10.1016/j.amepre.2021.10.005] [Medline: 34893387]

XSL•FO
RenderX

31. Xu L, Tong Q, Zhang X, et al. Smartphone-based gamification intervention to increase physical activity participation among patients with coronary heart disease: A randomized controlled trial. J Telemed Telecare 2024 Oct;30(9):1425-1436. [doi: 10.1177/1357633X221150943] [Medline: 36794484]
32. Morais Pereira Simões MDS, de Barros Gonze B, Leite Proença N, et al. Use of a smartphone app combined with gamification to increase the level of physical activity of adults and older adults: protocol of a sequential multiple assignment randomized trial. Trials 2019 Dec;20(1):1-8. [doi: 10.1186/s13063-019-3879-1]
33. Gonze BDB, Padovani RDC, Simoes MDS, et al. Use of a tsmartphone app to increase physical activity levels in insufficiently active adults: feasibility sequential multiple assignment randomized trial (SMART). JMIR Res Protoc 2020 Oct 23;9(10):e14322. [doi: 10.2196/14322] [Medline: 33094733]
34. Hamari J, Eranti V. Framework for designing and evaluating game achievements. 2011 Presented at: Proceedings of DiGRA 2011 Conference; Sep 14-17, 2011; Utrecht, Netherlands URL: https://dl.digra.org/index.php/dl/article/view/545/545 [accessed 2024-12-25] [doi: 10.26503/dl.v2011i1.545]
35. Lee JC, Lin R. The continuous usage of artificial intelligence (AI)-powered mobile fitness applications: the goal-setting theory perspective. IMDS 2023 May 29;123(6):1840-1860. [doi: 10.1108/IMDS-10-2022-0602]
36. Lee JC, Chen X. Exploring users' adoption intentions in the evolution of artificial intelligence mobile banking applications: the intelligent and anthropomorphic perspectives. IJBM 2022 Apr 25;40(4):631-658. [doi: 10.1108/IJBM-08-2021-0394]
37. Sjödin D, Parida V, Palmié M, Wincent J. How AI capabilities enable business model innovation: Scaling AI through co-evolutionary processes and feedback loops. J Bus Res 2021 Sep;134:574-587. [doi: 10.1016/j.jbusres.2021.05.009]
38. Li H, Manickam A, Samuel RDJ. RETRACTED ARTICLE: Automatic detection technology for sports players based on image recognition technology: the significance of big data technology in China's sports field. Ann Oper Res 2023 Jul;326(S1):97-97. [doi: 10.1007/s10479-021-04409-1]
39. Tsay CHH, Kofinas AK, Trivedi SK, Yang Y. Overcoming the novelty effect in online gamified learning systems: An empirical evaluation of student engagement and performance. J Comput Assist Learn 2020 Apr;36(2):128-146. [doi: 10.1111/jcal.12385]
40. Yang H, Li D. Understanding the dark side of gamification health management: A stress perspective. Inf Process Manag 2021 Sep;58(5):102649. [doi: 10.1016/j.ipm.2021.102649]
41. Jarvenpaa SL, Lang KR. Managing the Paradoxes of Mobile Technology. Inf Syst Manag 2005 Sep;22(4):7-23. [doi: 10.1201/1078.10580530/45520.22.4.20050901/90026.2]

Abbreviations

AI: artificial intelligence
MET: Metabolic equivalent of energy
PA: physical activity
SUS: System Usability Scale

Edited by R Baranyi; submitted 17.08.24; peer-reviewed by C Yuan, Y Yang; revised version received 08.01.25; accepted 17.02.25; published 24.03.25.

Please cite as:
Gao Y, Zhang J, He Z, Zhou Z
Feasibility and Usability of an Artificial Intelligence—Powered Gamification Intervention for Enhancing Physical Activity Among College Students: Quasi-Experimental Study
JMIR Serious Games 2025;13:e65498
URL: https://games.jmir.org/2025/1/e65498
doi:10.2196/65498

© Yanan Gao, Jinxi Zhang, Zhonghui He, Zhixiong Zhou. Originally published in JMIR Serious Games (https://games.jmir.org), 24.3.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

A Serious Game to Study Reduced Field of View in Keyhole Surgery: Development and Experimental Study

Phoebe Whitley[1], MSc; Connor Creasey[1], MSc; Matthew J Clarkson[2], PhD; Stephen Thompson[3], PhD

[1]Department of Medical Physics and Biomedical Engineering, Faculty of Engineering Sciences, University College London, London, United Kingdom
[2]UCL Hawkes Institute, Faculty of Engineering Sciences, University College London, London, United Kingdom
[3]Advanced Research Computing, University College London, London, United Kingdom

Corresponding Author:
Stephen Thompson, PhD
Advanced Research Computing
University College London
Gower Street
London, WC1E 6BT
United Kingdom
Phone: 44 2076792000
Email: s.thompson@ucl.ac.uk

Abstract

Background: During keyhole surgery, the surgeon is required to perform highly demanding tasks while only being able to see part of the patient's anatomy. This limited field of view is widely cited as a key limitation of the procedure, and many computational methods have been proposed to overcome it. However, the precise effects of a limited field of view on task performance remain unknown due to the lack of tools to study these effects effectively.

Objective: This paper describes our work on developing a serious game with 2 objectives: (1) to create an engaging game that communicates some of the challenges of keyhole surgery, and (2) to test the effect of a limited field of view on task performance. The development of a serious game that can be played by a wide range of participants will enable us to gather quantitative data on the effects of the reduced field of view on task performance. These data can inform the future development of technologies to help surgeons reduce the impact of a limited field of view on clinical outcomes for patients. The game is open source and may be adapted and used by other researchers to study related problems.

Methods: We implemented an open-source serious game in JavaScript, inspired by the surgical task of selectively cauterizing blood vessels during twin-to-twin transfusion surgery. During the game, the player is required to identify and cut the correct blood vessel under different fields of view and varying levels of vascular complexity. We conducted a quantitative analysis of task performance time under different conditions and a formative analysis of the game using participant questionnaires.

Results: We recruited 25 players to test the game and recorded their task performance time, accuracy, and qualitative metrics. Reducing the field of view resulted in participants taking significantly longer (P<.001) to perform otherwise identical tasks (mean 6.4 seconds, 95% CI 5.0-7.8 seconds vs mean 13.6 seconds, 95% CI 10.3-16.9 seconds). Participants found the game engaging and agreed that it enhanced their understanding of the limited field of view during keyhole surgery.

Conclusions: We recruited 25 players to test the game and recorded their task performance time, accuracy, and qualitative metrics. Reducing the field of view resulted in participants taking statistically significantly longer (16.4 vs 9.8 seconds; P=.05) to perform otherwise identical tasks. Participants found the game engaging and agreed that it enhanced their understanding of the limited field of view during keyhole surgery.

(JMIR Serious Games 2025;13:e56269) doi:10.2196/56269

KEYWORDS

keyhole surgery; laparoscopic surgery; serious games; image mosaicking; field of view; javascript; html; opensource

XSL•FO
RenderX

Introduction

Background

Keyhole surgery presents various advantages when compared with open surgery. The primary reason is that manipulation of abdominal tissue is reduced, resulting in less scarring, trauma, and hemorrhaging. This reduces the demand on health care services as patients require shorter hospital stays due to less postoperative pain [1-3]. Although keyhole surgery offers advantages, there are also limitations, such as a limited field of view, reduced depth perception, and increased procedure times.

Keyhole surgery is performed using endoscopes (and laparoscopes, a rigid endoscope used in abdominal procedures). Endoscopes consist of a long thin tube, with a camera and light source attached at the end. The surgeon is unable to see the anatomy directly but instead relies on video relayed from the endoscope camera [4]. This video presents a significantly reduced field of view in comparison to open surgery [5].

The monitor reduces depth perception of the operating scene as surgeons must map between the 2D image on the monitor and the 3D anatomical structure of the patient [6]. Visual misperceptions can occur from the loss of binocular vision due to a decrease in depth perception [7]. This can also be caused by surgeon fatigue as the laparoscopic setup is cerebrally intensive and increases the cognitive load of surgeons [8].

Modern endoscopes provide high spatial resolution, but at the expense of a limited field of view [9]. The surgeon's ability to view the surgical scene is limited by the narrow monocular field of view of the endoscopic camera [10]. In contrast to the panoramic view during open surgery, the endoscope only images small areas of the surgical scene at once [11]. The camera has a fixed field of view, requiring the surgeon to maneuver the laparoscope to the target region [12]. The limited field of view during endoscope surgery has been widely cited as a limitation, and this will be the focus of this research project.

Image Mosaicking

Image mosaicking is an established technique to construct a single image of the increased field of view by aligning various partially overlapped images of the same scene. Computational mosaicking can increase the field of view without compromising spatial resolution. Recently, this technique has been heavily researched, and its application is used in numerous industries, such as surveillance, satellite mapping, and agriculture. Mosaicking has also been used in endoscopic surgery to overcome the limited field of view and assists surgeons in manipulating the surgical scene and planning surgeries [13]. Daga et al [14] demonstrated the use of computational mosaicking for spatial orientation and anastomoses localization during endoscopic procedures. Computational mosaicking in endoscopic surgery remains challenging due to inhomogeneous lighting [15] and uncontrolled movement of the endoscope combined with geometric image distortion from the endoscope camera [16]. Because of these challenges, research to develop enhanced mosaicking algorithms is ongoing; however, there remains little understanding of the likely benefits of computational mosaicking in this field. A recent study has shown that experienced laparoscopic surgeons are proficient at "mentally mosaicking," which is the ability to effectively translate the 2D visual information into the 3D anatomical context [17]. Our research is inspired by the question of how to best deploy computational mosaicking taking into account the user's ability to do the same task mentally.

A key question that remains difficult to answer is "What precisely are the benefits of mosaicking or otherwise enlarging the field of view?." To put it in statistical terms, what is the expected effect size for a given change in the field of view? Estimating the effect size is essential for any power calculation required in a study on a proposed technology to enlarge the field of view. As the study becomes more realistic and onerous (eg, an in vivo study requiring human volunteers and expert surgeons), it becomes essential that a realistic required sample size can be calculated before gaining ethical approval. One way to estimate the effect size would be to perform a study measuring task performance versus field of view; however, such a study would require a large number of participants. Recruiting sufficient surgeons to perform such a study would be difficult and time-consuming.

User Studies With Nonexpert Users

Our work builds on the recent work of Yoo et al [18] who asked whether nonsurgical participants could stand in for surgical participants in user studies. They compared performance between participants with different levels of surgical training when interacting with a surgical augmented reality system (also see [19]). Comparing surgeons with different levels of experience with nonsurgeons, they found important differences in the performance of surgeons and nonsurgeons, but also similarities that can be used to inform system design, concluding that nonsurgical users could act as useful stand ins for surgical users, particularly in the early stages of device development. We wanted to see whether, by creating an abstract and fun representation of the mosaicking problem, we could lower the bar to recruitment, thus making a prerecruitment power study less important (ie, we can easily recruit many people, and there is no risk to the participants), so we can keep recruiting until we have enough data to show the statistical significance and calculate an effect size to inform future work. One way to achieve this may be through a serious game, which creates a simplified representation of the clinical problem, allowing us to recruit nonexpert users.

Serious Games

Serious games have become increasingly prevalent for educational purposes, partly due to advancements in technology [20]. Serious games fulfill an additional role beyond pure entertainment [21,22]. Research has shown that incorporating intrinsic motivation into games, such as challenges and curiosity, substantially increases user motivation [5]. Creating immersive game environments can generate a deeper understanding by allowing users to test their problem-solving and decision-making skills within a safe environment [23]. Serious games can be personalized and designed to support the acquisition of knowledge and skill development, showing the need for these games to evaluate learning progress through player feedback [24]. Providing an interactive learning environment with instant

XSL•FO
RenderX

visual feedback, such as a score, encourages more involvement and leads to a greater desire to complete the task at hand [25].

Serious games have previously been applied in the field of surgical training; for example, Underground is a serious game for the Nintendo Wii U platform, and the psychomotor skills required by users to complete the game objectives are closely related to the laparoscopic motor skills required by surgeons. Jalink et al [26] concluded that playing Underground increased laparoscopic skill development. A very important point when considering the use of games to represent a complex real-world procedure such as surgery is construct validity, that is, can it be shown that the skills used during the game correlate with performance during surgery. Construct validity for Underground was demonstrated by IJgosse et al [27], proving a link between in-game performance and surgical skills. Perhaps more interestingly, links between in-game performance and suturing skills have been shown for games with no apparent link to surgery [28].

Similarly, Ou et al [29] demonstrated that surgical trainees with previous gaming experience performed better in terms of laparoscopic simulation performance, compared with their nongaming counterparts. Surgical serious games must measure specific game metrics to quantify user performance [30]. The Kheiron Training System is a serious game designed to test basic psycho-motor skills required during laparoscopic surgery by utilizing real laparoscopic instruments. However, no studies have provided validation of this game as a training platform or obtained data to quantify the effect of the limited field of view in this game [23]. Although these studies show that serious games can be applied to skill development for surgical tasks, none of them attempt to answer questions on whether surgical technology development can be informed by users' task performance when playing serious games. In this paper, we demonstrate the use of a serious game to generate quantitative data to inform the ongoing development of computational mosaicking.

The remaining sections of this paper describe the development and testing of the serious game we are developing. The game is designed to enable gathering quantitative data on the effect of reduced field of view that will be applicable to keyhole surgery. These data will be useful for estimating effect sizes (and hence statistical power) for follow-up studies requiring more clinically representative participants and equipment. Alongside this, we also aimed to make the game engaging and fun to play, accessible to users of different abilities, and able to test the skills (hand-eye coordination and mental mosaicking) of different users.

Game Design and Implementation

The game has 2 aims: first, to study the effect of a reduced field of view on task performance, and second, to create a game for public engagement that communicates the challenges of keyhole surgery to a nontechnical audience and explores how image mosaicking may help address these challenges. These 2 aims are somewhat contradictory. For a strict study of a reduced field of view, a randomized-level structure coupled with a strictly defined training protocol would be ideal, to avoid comparison results being confounded with learning effects [31]. For a game aimed at public engagement, however, we want something that is easy to play from level 1 and engages the player with increased challenges at each level.

For this study, we decided to focus on the latter aim, so we use a set-level structure with increasing challenge at each level. This was to ensure the user was in a flow state by increasing the skill level required to successfully complete each game level [32]. To reduce the impact of learning effects, the levels used for comparison were placed at the end of the sequential-level structure. Table 1 summarizes the level structure, the objectives for each level, and the skills developed for each level.

Table 1. Level structure showing the game features and skills required in each level. The game gets more difficult with each level, while incrementally introducing 1 of 3 challenges (more vessels, more complex shapes, and limiting the field of view).

Level	Vessels	Vessels intertwined	Field of view	Skills tested
1	1	No	Full	Hand-eye coordination
2	2	No	Full	Hand-eye coordination and decision-making
3	1	No	Limited	Hand-eye coordination and image mosaicking
4	2	Intertwined	Full	Hand-eye coordination, visual perception, and decision-making
5	2	Intertwined	Limited	Hand-eye coordination, visual perception, decision-making, and image mosaicking
6	3	Intertwined	Full	Hand-eye coordination, visual perception, and decision-making
7	3	Intertwined	Limited	Hand-eye coordination, visual perception, decision-making, and image mosaicking

We identified 4 key skills that we wanted to address in the game. The first skill is hand-eye coordination, which is needed for all surgeries. Successful surgery requires that the surgeon is able to accurately cut in the intended location. In actual surgery, this is complicated by the need to use specialized tools. This is particularly difficult for keyhole surgery where the action of the tools is reversed. In a previous study [33] with nonexpert recruits, we observed that the mental load required to use laparoscopic tools can overwhelm any effect observed from changing the control variables. If the cutting mechanism is made too realistic, it is likely that we would not observe a change in performance with the field of view, as nonexpert users might find the game too difficult. Therefore, we decided that the hand-eye coordination skill would just require the use of a

XSL•FO
RenderX

mouse to position the cursor over the vessel and press the mouse button to commence a cut. This element of the game remains the same across all levels.

The second skill is decision-making in the form of curve tracing [34] (ie, given the choice of 2 vessels, which one should be cut?). The user is required to visually inspect each path and work out which one connects the 2 black dots. Our game was inspired by surgery to treat twin-to-twin transfusion, as this is an area where image mosaicking has been proposed to improve performance [14] in keyhole surgery. The surgeon selectively cauterizes placental blood vessels to separate the blood supply to each twin. This requires careful identification of each vessel and its path. We created a simplified representation of this to create a curve-tracing game. Our representation of this is abstract to enable nonexpert users to play, but elements such as multiple vessels and intertwining are introduced during the game. All levels apart from 1 and 3 have multiple vessels and the player must decide which is the correct vessel to cut.

The third skill is visual perception. In most keyhole procedures, differentiating one structure from another can be challenging, as human anatomy does not consist of regular shapes in high-contrast colors. Therefore, we designed the game with a low-color contrast between the vessels and the background. In levels 4-7, we introduced multiple intertwined vessels without color contrast to make it more difficult to distinguish between them.

The final skill is image mosaicking. During keyhole surgery, it is not possible to see the whole surgical scene, a situation referred to as the limited field of view. A skilled keyhole surgeon must be able to mentally reconstruct the whole anatomical scene from a series of partial views created as they move the endoscope around. We introduce this skill at level 3 using a spotlight effect, so the player can only see part of the scene at once and must move the spotlight around with the mouse to mentally reconstruct the scene and identify the correct vessel.

The 4 skills are combined in different ways as the game progresses. To study the potential effects of a limited field of view, we created 2 pairs of levels (4<>5 and 6<>7) that are identical except for the field of view, allowing for a comparison of results between these levels.

Game Implementation

Figure 1 (also see [35,36]) shows screenshots of 6 game levels, illustrating the main game mechanics. Figure 1A shows the game at its most basic level with a single vessel on the screen. The user has identified the vessel and drawn a black line across the vessel to cut it. The time (in seconds) it took them to do this is displayed at the top left along with the number of attempts. A reduced field of view is introduced at level 4 (Figure 1C), with levels getting progressively harder until level 7 (Figure 1F), which has 3 intertwined blood vessels with a reduced field of view. Finally, level 7 has 3 intertwined blood vessels combined with a limited field of view.

The game was implemented in HTML and JavaScript and can be run in most modern web browsers. It is hosted as a static web page on GitHub. Allowing users to run the game directly from their browsers provides instant feedback, enhancing user interaction and engagement by displaying results immediately [33]. Game elements were created using the Phaser (version 3.60.0) [37] game framework. Phaser offers a configurable, open-source development library that supports small build sizes and fast loading times [38]. Additionally, it provides a wide range of tutorials and community support to facilitate development.

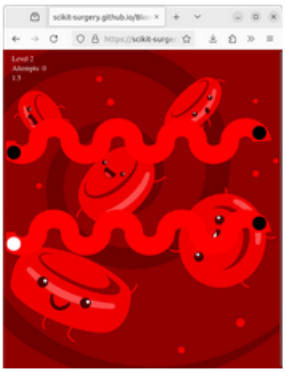
The game includes a timer to measure the time taken to complete each level but does not record results. Therefore, for the experiments, the results were recorded manually by the authors. For full technical details, the version of the game used in this paper, along with the data supporting the results, is archived on Zenodo [35].

XSL•FO
RenderX

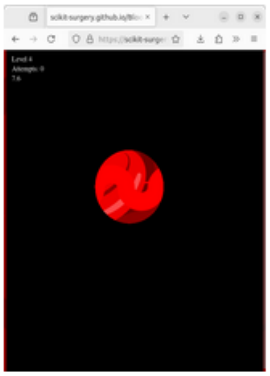
Figure 1. Examples of 6 game levels illustrating the incremental change in difficulty throughout the game. The full implementation of the game can be found at [35] or played directly at [36].



(a) Screenshot of Level 1 showing a single blood vessel across the screen.



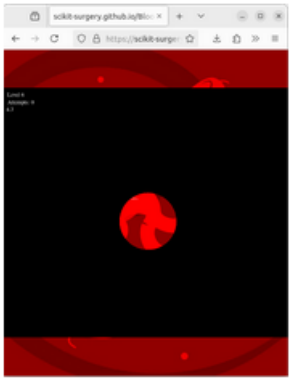
(b) Level 2 Introduces a second vessel, The player must cut the vessel with two black circles.



(c) Level 3 repeats the single vessel from level 1 and introduces a limited field of view.



(d) Level 4 adds intertwining of the vessels. The player is still required to cut the vessel with two black circles.



(e) Level 5 repeats level 4 but with a limited field of view. The player must now trace a complex path that they can only partially see.



(f) Level 6 introduces a third intertwined vessel. The user is still required to cut the vessel with two black circles.

Vessel Creation and Cutting Logic

Vessels were represented as polygons, colored red, and stretched to fit the window width. Users navigated a pointer around the scene using their mouse. On levels with a reduced field of view, the mouse also moved the viewport (implemented using a circular bitmap mask), ensuring that the mouse pointer remained at the center of the viewport. Pressing and holding the mouse button started drawing a black line across the scene, while releasing the button ended the line drawing and incremented the attempt counter by 1. Upon completing the line drawing, a check determined whether the drawn line completely intersected the target vessel. If successful, a level completion message was displayed, and the level timer stopped. To prevent users from drawing a line across the entire scene—inevitably intersecting both sides of the polygon—a maximum line length of 45 units was enforced.

For levels with multiple blood vessels, circles were added to the end of each path, and users were required to cut only the blood vessel labeled with 2 black circles (see Figure 1C for an illustration).

Figure 1D demonstrates how the limited field of view was incorporated. Users could only view the game scene within the mask by moving their mouse across the screen. They had to navigate around the screen to locate the blood vessel marked with 2 black circles and use their mouse to cut it. The timer, displayed at the top of the game interface, stopped only after both sides of the correct blood vessel were fully intersected. If the user failed to completely intersect both sides of the blood vessel, the accuracy counter was incremented, and the timer continued. Users then had to try again by drawing a new line to cut the blood vessel.

The complete implementation of the game used in this publication is archived online [35]. The archive also includes links to newer development versions of the game and the URL provided to participants for accessing the game. This participant

XSL•FO
RenderX

access URL leads to an index screen that contains links to the consent form, game instructions, and individual levels.

Methods

Study Design

We used the game to perform a single-arm user study with all users playing all levels of the game in the same order.

Participant Recruitment

As discussed in the "Introduction" section, a key aim of the game design was to ensure that clinical experience was not required to play. Therefore, we deliberately avoided recruiting surgeons at this stage, although they were not excluded. The only inclusion criteria were age (participants had to be between 18 and 65 years old) and residency in the United Kingdom. Participants were recruited through the departmental email list (Medical Physics and Biomedical Engineering) or via a direct approach among the first author's acquaintances.

Participant experiments were conducted either in person or via video calls and lasted approximately 15 minutes. Participants accessed the game through a URL [35].

Before playing the game, participants were asked, "What do you know about laparoscopic surgery?" and "Do you know the potential effects of a limited field of view in laparoscopic surgery?" to assess their prior knowledge of the domain. After completing a consent form, participants were provided with instructions on how to play the game.

Participants were instructed to play the game and complete each level in sequence, accessing each level through the game's home page.

Time to Complete Level

The in-game timer started automatically when participants clicked on a game level and stopped once they successfully cut the correct blood vessel. These data were recorded by the researcher for both in-person and remote experiments.

Accuracy

A counter variable recorded the number of attempts each participant needed to successfully complete each level, increasing with each mouse click used to draw a new line.

Participant Questionnaires

In addition to the pregame questionnaires, participants completed a postgame questionnaire, a NASA (National Aeronautics and Space Administration) Task Load Index [39] questionnaire, and a System Usability Scale [40] questionnaire after finishing all 7 levels. Full details of the questions can be found in the "Results" section.

Ethical Approval

This study was approved by University College London's Research Ethics Committee (reference number 24249_001). Informed consent was obtained from all individual participants included in the study.

Results

Participant Recruitment

We recruited 25 participants from a range of backgrounds, experiences, and ages for this study. The recruitment approach led to a high number of master's students with advanced knowledge of laparoscopic surgery and computer science, as well as students specializing in other sciences. Additionally, working professionals from various industries were recruited, some with little to no understanding of laparoscopic surgery. One participant had medical experience and reported knowledge of keyhole surgery. This diverse participant pool was selected to gather a broad range of responses and perspectives.

We did not record details of participants' prior gaming experience or their computer usage. However, as some participants were known to the authors, we can anecdotally state that those who used computers less tended to complete tasks more slowly and found the user interface harder to navigate. The participants with medical experience did not appear to perform differently from the main population.

Time to Complete Level

The mean and SD of the time to complete each level are shown on the left-hand side of Table 2, along with 95% CIs for the mean.

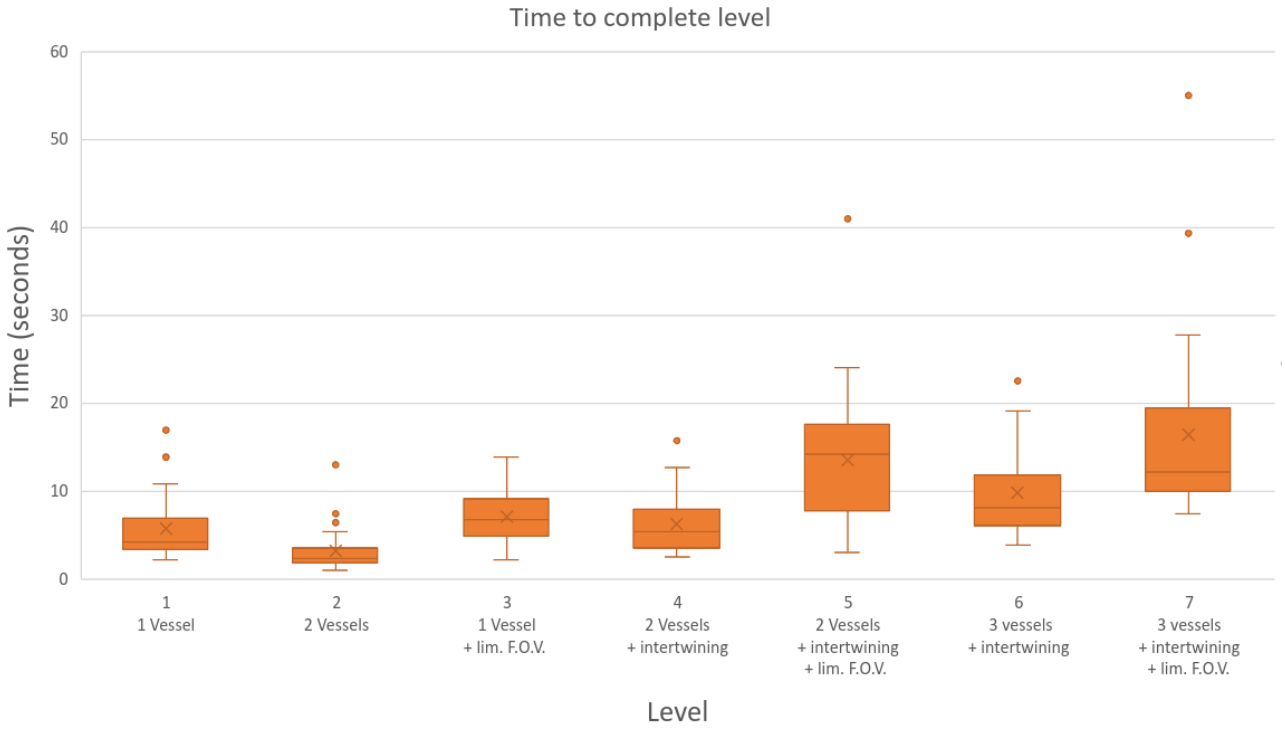
Table 2. Mean (SD) and 95% CI for time and number of attempts to complete each level.

Level	Participant time results (seconds)		Participant accuracy (attempts)	
	Mean (SD)	95% CI	Mean (SD)	95% CI
1	5.7 (3.6)	4.2-6.9	1.3 (2.0)	0.5-2.1
2	3.3 (2.6)	2.2-4.4	1.2 (0.7)	0.9-1.5
3	7.2 (2.8)	6.0-8.4	1.4 (1.1)	0.9-1.9
4	6.4 (3.3)	5.0-7.8	1.6 (1.1)	1.2-2.0
5	12.5 (5.7)	10.2-14.8	1.4 (0.7)	1.1-1.7
6	9.8 (4.8)	7.8-11.8	1.8 (0.9)	1.4-2.2
7	16.4 (11.0)	11.9-20.9	1.6 (1.2)	1.1-2.1

XSL•FO
RenderX

Figure 2 presents a boxplot of level completion times alongside a brief description of each level's features. The figure shows a general trend of increasing completion time from left to right, corresponding with an increasing level of complexity. On average, level 2 was completed the fastest (in 3.3 seconds). This level featured 2 blood vessels displayed on the game interface without a restricted field of view. By contrast, level 7—the most challenging level—took the longest time to complete, with an average time of 16.4 seconds. This level featured 3 intertwined blood vessels following a complex path and was constrained by a limited field of view.

Figure 2. Box plot displaying the average time to complete each level. F.O.V.: field of view.



While the general trend in Figure 2 shows increasing completion times as level complexity increases, there are 2 exceptions when the field of view is restricted. Level 4 (with 2 blood vessels) had a shorter completion time than level 3 (which featured a single blood vessel and a limited field of view). Similarly, level 6 (with 3 intertwined blood vessels) was completed faster than level 5 (which had 2 intertwined blood vessels and a limited field of view). This suggests that a restricted field of view has a greater impact on performance than on other factors examined in this study. The effect of limiting the field of view can be estimated by comparing results from levels that were identical except for this restriction (ie, comparing level 4 with level 5 and level 6 with level 7).

We compared the completion times for levels 4 and 5 using a Welch 2-sample *t* test, which yielded a *P* value of <.001 indicating a statistically significant impact of limiting the field of view. The Cohen *d* effect size was 1.19, suggesting a large effect. A similar comparison between levels 6 and 7 produced a *P* value of .009 and an effect size of 0.79, further supporting the significant impact of a restricted field of view.

Accuracy

The number of attempts is displayed on the right-hand side of Table 2. A Welch 2-sample *t* test confirmed no significant differences in the number of attempts across levels (the minimum *P* value was .23 between levels 5 and 6). The numerical results and analysis scripts referenced above are archived along with the game code in [35].

Participant Questionnaires

To assess participants' prior knowledge, all participants were asked, "What do you know about laparoscopic surgery?" and "Do you know the potential effects of a limited field of view in laparoscopic surgery?" before playing the game. All 25 participants recognized that laparoscopy is a type of surgical procedure; however, their level of understanding varied significantly depending on their occupation and personal experiences. Participants from the researcher's student cohort were highly knowledgeable about this topic and understood both the advantages and limitations of this minimally invasive procedure. By contrast, participants from a nonmedical background had limited awareness of the benefits of laparoscopic surgery and the potential impact of a restricted field of view.

The results of the postgame questionnaire are shown in Table 3. The questionnaire was completed by 22 participants. Three participants commented on the game's background color, suggesting that a greater contrast between the background and blood vessels would improve visibility. Four participants stated that they enjoyed the game timer, as it heightened their competitiveness under time pressure. Two participants mentioned feeling frustrated due to their limited experience with a Mac laptop and its built-in mouse, which hindered their ability to complete levels quickly.

XSL•FO
RenderX

Table 3. Participant questionnaire results: percentage and absolute number of participants who answered yes when answering the questionnaire after completing the game.

Question	Yes, n/N (%)
Did playing this game enhance your understanding of the limited field of view in laparoscopic surgery?	20/22 (91)
Do you think this game is clinically relevant?	19/22 (86)
Do you think the difficulty increased with each level?	20/22 (91)
Did you find the game engaging?	22/22 (100)
Did you find the game layout visually pleasing?	22/22 (100)

Participants also completed a NASA Task Load Index [39] questionnaire, rating each workload demand on a scale of 1-10. For the questions related to demand (1-3, 5-6) a score of 1 is described as “very low” and a score of 10 as “very high.” For question 4, a score of 1 is “perfect” and 10 is “failure.” Table 4 presents the average scores for the 6 workload demands. The results indicate that participants felt neutral about the mental demands required to complete the tasks in this serious game. Physical demand received the lowest rating, with an average score of 1.6, while performance workload was rated the highest. Effort workload scores varied significantly, with widely dispersed data resulting in a neutral average of 5.3. Participants rated their frustration levels relatively low, with an average score of 2.8.

Additionally, participants completed a System Usability Scale questionnaire [40], the results of which are shown in Table 5.

Table 4. NASAa Task Load Index results: the average score out of 10 for each demand.

NASA Task Load Index	Mean score
Mental: How mentally demanding was the task?	4.9
Physical: How physically demanding was the task?	1.6
Temporal: How hurried or rushed was the pace of the task?	5.6
Performance: How successful were you in accomplishing what you were asked to do?	7.7
Effort: How hard did you have to work to accomplish your level of performance?	5.3
Frustration: How insecure, discouraged, irritated, stressed, and annoyed were you?	2.8

[a]NASA: National Aeronautics and Space Administration.

Table 5. System Usability results: the average score out of 5 for each demand in the System Usability Scale.

System Usability Scale	Mean score
I think that I would like to use this system frequently	3.2
I found the system unnecessarily complex	1.3
I thought the system was easy to use	4.3
I think that I would need the support of a technical person to be able to use this system	1.5
I found that the various functions in this system were well integrated	3.8
I thought there was too much inconsistency in this system	1.7
I would imagine that most people would learn to use this system very quickly	4.6
I found the system very cumbersome to use	1.6
I felt very confident using the system	4.3
I needed to learn a lot of things before I could get going with this system	1.2

Discussion

Principal Findings

The qualitative results from participant questionnaires suggest that we have successfully created an engaging game that can facilitate discussions about the challenges of keyhole surgery with nontechnical audiences. The quantitative results, based on comparisons of level completion times, indicate that our game can provide valuable insights into the effects of a limited field of view on task performance.

Game Design

Our game design was a balance between creating an engaging experience by gradually increasing the difficulty of each level and enabling a paired comparison between different conditions. Our results indicate that we largely succeeded in both objectives. As shown in Table 3, all participants found the game engaging,

XSL•FO
RenderX

and we observed a statistically significant difference in performance when the field of view was reduced (see Figure 2).

Randomizing the level structure might have resulted in a more robust test of our hypotheses but would likely have come at the expense of participant engagement. Maintaining a continuous flow state throughout the game was crucial, as the flow has a positive impact on learning and is strongly linked to user attention and focus. Research suggests that when a user's attention is directed toward a limited set of stimuli, irrelevant perceptions and distractions are minimized [41]. Additionally, studies have demonstrated that sensory and cognitive curiosity increases when users find a game intrinsically interesting, thereby enhancing engagement [42,43]. Notably, all 22 participants who completed the questionnaire agreed that this serious game was engaging.

Three participants commented on the color contrast between the game background and blood vessels, suggesting that a greater difference in color would have helped them distinguish between the 2, especially in the harder levels. The low contrast between the background and blood vessels was an intentional design choice, as the illumination of the surgical scene is limited during laparoscopic surgery [15].

Additionally, bodily fluids, such as blood, can further obscure the surgeon's view of the anatomical context. The game design aimed to replicate these challenges by maintaining low contrast, requiring players to focus and use precision to identify and cut the correct blood vessels. Future work could explore fine-tuning the display for different applications of the game.

Learning the Anatomy

There is evidence of a learning effect between levels 1 and 2, as level 2 was completed faster on average and with fewer attempts than level 1 (see Table 2). Observations of participants attempting level 1 indicated that they were initially unaware of the maximum line length they could draw with their mouse. It is likely that the lower average number of attempts in level 2 resulted from participants becoming aware of this game mechanic.

Levels with a more complex vessel pattern (levels 4 and 6) required more attempts than those where the pattern remained the same or had no intertwining (levels 2, 3, 5, and 7); however, these differences were not statistically significant. The finding that levels with a reduced field of view took longer to complete but did not require more attempts suggests that the mechanisms for mental mosaicking and learning the anatomy may be distinct processes.

Accessibility

This game was designed to be accessible to users of all abilities. The participant pool was diverse, with varying levels of knowledge and experience in laparoscopic surgery and web-based games. Participants ranged in age from 18 to 57 years, with 2 reporting dyslexia and 1 reporting dyspraxia. All 25 participants successfully completed all levels, suggesting that the game was accessible to them. However, no participants had color blindness or vision deficiencies. Given the low contrast between the game background and blood vessels and the fact that the correct blood vessels were identifiable only by black circles, it cannot be concluded that the game is accessible to users with visual impairments. Results from the NASA Task Load Index indicated varied effort workload scores, suggesting that some participants had limited experience using a laptop. As a result, the game may be less accessible to players with no prior laptop or gaming experience.

Limitations and Future Work

At present, the game presents a very simplified representation of the surgical environment. This simplification limits the game's direct relevance to surgery. The simplification was a deliberate decision to keep the game accessible to the widest possible user base; however, future work may require a more clinically realistic environment. Repeating the experiment with a sample of surgeons, controlling for specialty and expertise, will help validate our methodology of recruiting nonexpert participants. Bearing in mind the results of Yoo et al [18], we still expect to see a correlation between the field of view and task completion time; however, we would not expect the results to be identical. Furthermore, the results would vary between levels of surgical experience and specialty. Surgeons would bring differing levels of prior knowledge that would alter game performance. Comparing the results between different user groups might yield useful information about the differences between trained surgeons and the general population.

The game's simplicity makes it impossible at present to fully understand the impact of learning effect on the results. The fact that the level pairs we used to compare between different field of view settings (4<>5 and 6<>7) were otherwise identical means that our results may underestimate the impact of reducing the field of view due to the participants learning from the preceding level. Future work could look at introducing more complex level progression to control for this.

To increase the clinical relevance of the game, the graphics and design could be changed to represent a more clinical environment. For example, it would be relatively easy to change the backdrop to a screenshot taken from a clinical procedure with vessels overlaid in more clinically realistic colors. Artifacts such as smoke and bleeding often seen in keyhole surgery could also be added, but would require significantly more programming work. It would then make interesting future work to compare the performance on this more realistic game between surgeons and nonsurgeons. We are also exploring ways to incorporate a negative scoring system to penalize mistakes, such as cutting the wrong blood vessel, and enable better analysis of how the field of view affects the errors made.

To investigate the impact of computational image mosaicking on task performance, additional levels will be required. At a basic level, users could use their mouse to "paint" on the scene, revealing the blood vessels underneath. This would create a larger field of view, simulating computational image mosaicking. As the game becomes more complicated, it is likely that the effect sizes will decrease, requiring a larger sample size to demonstrate statistical significance. A key advantage of our approach is its ability to support the recruitment of large numbers of participants.

XSL•FO
RenderX

Work is ongoing to improve scene management, making it easier for the user to move from one level to the next. Efforts are also underway to gather results using an automated database backend.

Conclusions

Our serious Blood Vessel Game was used to demonstrate a quantifiable effect of a limited field of view on task performance time, with the same task taking between 60% and 100% longer when the view was restricted. No effect on task accuracy was detected. Our results represent the first time this effect has been quantified in this way. The game also serves as an engaging educational tool for discussing the impact of a limited field of view on task performance, with 20 out of 22 (91%) participants agreeing that the game was educational and all (22/22, 100%) agreeing that it was engaging. The game is entirely open source, and we welcome contributions to enhance its usefulness.

Acknowledgments

This work is supported by the Wellcome/EPSRC Centre for Interventional and Surgical Sciences (WEISS) (grant 203145Z/16/Z). No artificial intelligence tools were used in the preparation of this work.

Data Availability

The game implementation and data used in this paper are freely available at [35].

Authors' Contributions

PW contributed to the conceptualization, formal analysis, investigation, software, visualization, and writing the original draft. CC contributed to writing reviews and editing. MJC contributed to conceptualization, funding acquisition, project administration, and supervision. ST contributed to conceptualization, investigation, project administration, software, supervision, and writing reviews and editing.

Conflicts of Interest

None declared.

References

1. Hayden P, Cowman S. Anaesthesia for laparoscopic surgery. Continuing Education in Anaesthesia Critical Care & Pain 2011 Oct;11(5):177-180. [doi: 10.1093/bjaceaccp/mkr027]
2. Kuoppala T, Tomás E, Heinonen P. Clinical outcome and complications of laparoscopic surgery compared with traditional surgery in women with endometrial cancer. Arch Gynecol Obstet 2004 Jul 1;270(1):25-30. [doi: 10.1007/s00404-003-0488-7] [Medline: 12728326]
3. Rege RV, Merriam LT, Joehl RJ. LAPAROSCOPIC SPLENECTOMY. Surgical Clinics of North America 1996 Jun;76(3):459-468. [doi: 10.1016/s0039-6109(05)70453-9]
4. Walczak DA, Pawełczak D, Piotrowski P, Trzeciak PW, Jędrzejczyk A, Pasieka Z. Video display during laparoscopy - where should it be placed? Wideochir Inne Tech Maloinwazyjne 2015 Apr;10(1):87-91 [FREE Full text] [doi: 10.5114/wiitm.2014.47434] [Medline: 25960798]
5. Kim J, Watras A, Liu H, Zeng Z, Greenberg J, Heise C, et al. Large-field-of-view visualization utilizing multiple miniaturized cameras for laparoscopic surgery. Micromachines (Basel) 2018 Aug 25;9(9):431. [doi: 10.3390/mi9090431] [Medline: 30424364]
6. Bogdanova R, Boulanger P, Zheng B. Depth perception of surgeons in minimally invasive surgery. Surg Innov 2016 Oct 09;23(5):515-524. [doi: 10.1177/1553350616639141] [Medline: 27009686]
7. Sinha R, Raje S, Rao G. Three-dimensional laparoscopy: Principles and practice. J Min Access Surg 2017;13(3):165. [doi: 10.4103/0972-9941.181761]
8. Xin H, Zelek JS, Carnahan H. Laparoscopic surgery, perceptual limitations and force: a review. In: First Canadian Student Conference on Biomedical Computing. 2006 Presented at: Queen’s University Kingston; September 28, 2006; Kingston, ON, Canada.
9. Qin Y, Hua H. Optical design and system engineering of a multiresolution foveated laparoscope. Appl Opt 2016 Apr 08;55(11):3058. [doi: 10.1364/ao.55.003058]
10. Tendick F, Jennings RW, Tharp G, Stark L. Sensing and Manipulation Problems in Endoscopic Surgery: Experiment, Analysis, and Observation. Presence: Teleoperators & Virtual Environments 1993 Jan;2(1):66-81. [doi: 10.1162/pres.1993.2.1.66]
11. Levy B, Mobasheri M. Principles of safe laparoscopic surgery. Surgery (Oxford) 2017 Apr;35(4):216-219. [doi: 10.1016/j.mpsur.2017.01.010]
12. Watras AJ, Kim J, Ke J, Liu H, Greenberg JA, Heise CP, et al. Large-Field-of-View Visualization with Small Blind Spots Utilizing Tilted Micro-Camera Array for Laparoscopic Surgery. Micromachines 2020 May 10;11(5):488. [doi: 10.3390/mi11050488]

XSL•FO
RenderX

13. Li L, Mazomenos E, Chandler JH, Obstein KL, Valdastri P, Stoyanov D, et al. Robust endoscopic image mosaicking via fusion of multimodal estimation. Med Image Anal 2023 Feb;84:102709 [FREE Full text] [doi: 10.1016/j.media.2022.102709] [Medline: 36549045]
14. Daga P, Chadebecq F, Shakir DI, Herrera L, Tella M, Dwyer G, et al. Real-time mosaicing of fetoscopic videos using sift. 2016 Mar 18 Presented at: SPIE Medical Imaging; March 18, 2016; San Diego, CA p. 533-539. [doi: 10.1117/12.2217172]
15. Pandey A, Pati UC. Image mosaicing: a deeper insight. Image and Vision Computing 2019 Sep;89:236-257. [doi: 10.1016/j.imavis.2019.07.002]
16. Bergen T, Wittenberg T. Stitching and surface reconstruction from endoscopic image sequences: a review of applications and methods. IEEE J Biomed Health Inform 2016 Jan;20(1):304-321. [doi: 10.1109/jbhi.2014.2384134]
17. Ten Dam E, Helder HM, van der Laan BFAM, Feijen RA, Korsten-Meijer AGW. The effect of three-dimensional visualisation on performance in endoscopic sinus surgery: a clinical training study using surgical navigation for movement analysis in a randomised crossover design. Clin Otolaryngol 2020 Mar 27;45(2):211-220 [FREE Full text] [doi: 10.1111/coa.13494] [Medline: 31846558]
18. Yoo S, Ramalhinho J, Dowrick T, Somasundaram M, Gurusamy K, Davidson B, et al. Can engineers represent surgeons in usability studies? Comparison of results from evaluating augmented reality guidance for laparoscopic surgery. Computers & Graphics 2024 Apr;119:103881. [doi: 10.1016/j.cag.2024.01.008]
19. Christoph D, Soojeong Y, Matthew M, Thompson S. Enhanced surgeons: understanding the design of augmented reality instructions for keyhole surgery. 2023 Presented at: IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW); March 25-29, 2023; Shanghai, China p. 123-127. [doi: 10.1109/vrw58643.2023.00031]
20. Khouna J, Ajana L, Rhazal A, El Mokri A. The use of educational software in teaching physics in the Moroccan context. Int J Emerg Technol Learn 2020 Sep 25;15(18):270. [doi: 10.3991/ijet.v15i18.15455]
21. Baker J. Designing immersive serious games. PhD thesis, University of Southampton. 2017 Jun 01. URL: https://eprints.soton.ac.uk/419472/1/JB_ThesisFinal.pdf [accessed 2025-02-03]
22. Gorbanev I, Agudelo-Londoño S, González RA, Cortes A, Pomares A, Delgadillo V, et al. A systematic review of serious games in medical education: quality of evidence and pedagogical strategy. Med Educ Online 2018 Dec 19;23(1):1438718 [FREE Full text] [doi: 10.1080/10872981.2018.1438718] [Medline: 29457760]
23. Olgers TJ, Bij de Weg AA, Ter Maaten JC. Serious games for improving technical skills in medicine: scoping review. JMIR Serious Games 2021 Jan 25;9(1):e24093 [FREE Full text] [doi: 10.2196/24093] [Medline: 33492234]
24. Bellotti F, Kapralos B, Lee K, Moreno-Ger P, Berta R. Assessment in and of serious games: an overview. Advances in Human-Computer Interaction 2013;2013:1-11. [doi: 10.1155/2013/136864]
25. Martins T, Carvalho V, Soares F, Moreira MF. Serious game as a tool to intellectual disabilities therapy: total challenge. New York, NY: IEEE; 2011 Presented at: IEEE 1st International Conference on Serious Games and Applications for Health (SeGAH); November 16-18, 2011; Braga, Portugal p. 1-7. [doi: 10.1109/segah.2011.6165444]
26. Jalink MB, Heineman E, Pierie JPEN, ten Cate Hoedemaker HO. The effect of a preoperative warm-up with a custom-made Nintendo video game on the performance of laparoscopic surgeons. Surg Endosc 2015 Aug;29(8):2284-2290. [doi: 10.1007/s00464-014-3943-6] [Medline: 25361658]
27. IJgosse W, van Goor H, Rosman C, Luursema J. Construct validity of a serious game for laparoscopic skills training: validation study. JMIR Serious Games 2020 May 07;8(2):e17222 [FREE Full text] [doi: 10.2196/17222] [Medline: 32379051]
28. Rosser JC, Lynch PJ, Cuddihy L, Gentile DA, Klonsky J, Merrell R. The impact of video games on training surgeons in the 21st century. Arch Surg 2007 Feb 01;142(2):181-6; discusssion 186. [doi: 10.1001/archsurg.142.2.181] [Medline: 17309970]
29. Ou Y, McGlone ER, Camm CF, Khan OA. Does playing video games improve laparoscopic skills? International Journal of Surgery 2013 Jun;11(5):365-369. [doi: 10.1016/j.ijsu.2013.02.020]
30. Graafland M, Schraagen JM, Schijven MP. Systematic review of serious games for medical education and surgical skills training. Br J Surg 2012 Oct;99(10):1322-1330. [doi: 10.1002/bjs.8819] [Medline: 22961509]
31. Blackburne T, Rodriguez A, Johnstone SJ. A serious game to increase healthy food consumption in overweight or obese adults: randomized controlled trial. JMIR Serious Games 2016 Jul 13;4(2):e10 [FREE Full text] [doi: 10.2196/games.5708] [Medline: 27417192]
32. Kiili K. Digital game-based learning: Towards an experiential gaming model. The Internet and Higher Education 2005 Jan;8(1):13-24. [doi: 10.1016/j.iheduc.2004.12.001]
33. Fairclough SH, Gilleade K, Ewing KC, Roberts J. Capturing user engagement via psychophysiology: measures and mechanisms for biocybernetic adaptation. IJAACS 2013;6(1):63. [doi: 10.1504/ijaacs.2013.050694]
34. Jolicoeur P, Ullman S, Mackay M. Curve tracing: a possible basic operation in the perception of spatial relations. Mem Cognit 1986 Mar;14(2):129-140. [doi: 10.3758/bf03198373] [Medline: 3724444]
35. Whitley P, Clarkson MJ, Thompson S. Blood vessel game. Zenodo. URL: https://zenodo.org/records/10489683 [accessed 2025-02-03]
36. Blood vessel game. GitHub. URL: https://scikit-surgery.github.io/Blood-Vessel-Game/ [accessed 2025-02-05]
37. Phaser Developers. Phaser Computer software. https://github.com/phaserjs/phaser. URL: https://github.com/phaserjs/phaser [accessed 2025-02-03]

XSL•FO
RenderX

38. Moreira RAF, Assunção WKG, Martinez J, Figueiredo E. Open-source software product line extraction processes: the ArgoUML-SPL and Phaser cases. Empir Software Eng 2022 Apr 08;27(4):1-35. [doi: 10.1007/s10664-021-10104-3]
39. Hart SG. Nasa-Task Load Index (NASA-TLX); 20 years later. Proceedings of the Human Factors and Ergonomics Society Annual Meeting 2006 Oct 01;50(9):904-908 [FREE Full text] [doi: 10.1177/154193120605000909]
40. Lewis JR. The System Usability Scale: past, present, and future. International Journal of Human–Computer Interaction 2018 Mar 30;34(7):577-590. [doi: 10.1080/10447318.2018.1455307]
41. Webster J, Trevino LK, Ryan L. The dimensionality and correlates of flow in human-computer interactions. Computers in Human Behavior 1993 Dec;9(4):411-426. [doi: 10.1016/0747-5632(93)90032-N]
42. Csikszentmihalyi M. Beyond Boredom and Anxiety. Hoboken, NJ: Jossey-Bass; 2000.
43. Li L, Mazomenos E, Chandler JH, Obstein KL, Valdastri P, Stoyanov D, et al. Robust endoscopic image mosaicking via fusion of multimodal estimation. Med Image Anal 2023 Feb;84:102709. [doi: 10.1016/j.media.2022.102709] [Medline: 36549045]

Abbreviations

NASA: National Aeronautics and Space Administration

Edited by R Baranyi; submitted 11.01.24; peer-reviewed by S Xiong, A Hassan, I Wani, M Aksoy, D Patel; comments to author 19.08.24; revised version received 30.09.24; accepted 17.01.25; published 11.02.25.

Please cite as:
Whitley P, Creasey C, Clarkson MJ, Thompson S
A Serious Game to Study Reduced Field of View in Keyhole Surgery: Development and Experimental Study
JMIR Serious Games 2025;13:e56269
URL: https://games.jmir.org/2025/1/e56269
doi:10.2196/56269
PMID:

©Phoebe Whitley, Connor Creasey, Matthew J Clarkson, Stephen Thompson. Originally published in JMIR Serious Games (https://games.jmir.org), 11.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Gamified mHealth System for Evaluating Upper Limb Motor Performance in Children: Cross-Sectional Feasibility Study

Md Raihan Mia[1], BSc; Sheikh Iqbal Ahamed[1], PhD; Samuel Nemanich[2], MSCI, PhD

[1]Department of Computer Science, Marquette University, Milwaukee, WI, United States

[2]Department of Occupational Therapy, Marquette University, Milwaukee, WI, United States

Corresponding Author:
Samuel Nemanich, MSCI, PhD
Department of Occupational Therapy
Marquette University
1700 West Wells St
Milwaukee, WI, 53233
United States
Phone: 1 4142883243
Email: sam.nemanich@marquette.edu

Abstract

Background: Approximately 17% of children in the United States have been diagnosed with a developmental or neurological disorder that affects upper limb (UL) movements needed for completing activities of daily living. Gold-standard laboratory assessments of the UL are objective and precise but may not be portable, while clinical assessments can be time-intensive. We developed MoEvGame, a mobile health (mHealth) gamification software system for the iPad, as a potential advanced technology to assess UL motor functions.

Objective: This feasibility study examines whether MoEvGame can assess children's whole-limb movement, fine motor skills, manual dexterity, and bimanual coordination. The specific aims were to (1) design and develop novel mHealth gamified software tools to examine theory-driven features of UL movement, (2) analyze spatiotemporal game data with new algorithms and statistical techniques to quantify movement performance as a parameter of speed, accuracy, and precision, and (3) validate assessment methods with healthy participants from schools.

Methods: Elementary school children (N=31, median 9.0, IQR 4.0-14.0 years old) participated by playing 5 games. The game tasks were focused on key features of skilled motor control: (1) whole limb reaching, (2) fine motor control and manual dexterity, and (3) bilateral coordination. Spatiotemporal game data were transferred and stored in a cloud-based data management server for further processing and analysis. We applied change point detection (ie, the pruned exact linear time method), signal processing techniques, and other algorithms to calculate movement speed and accuracy from spatiotemporal parameters. Different statistical methods (ie, Pearson correlation, mean, standard deviation, *P* value, 95% confidence interval) were used to compare speed-accuracy tradeoffs and evaluate the relationship between age and motor performance.

Results: A negative correlation was identified between speed and accuracy in the whole limb movement (r=–0.30 to –0.42). Significant relationships between age and upper limb performance were found: older participants exhibited lower errors with faster completion times compared to younger participants. Significant differences in bimanual coordination were found related to phase synchronization (in-phase congruent [mean 28.85, SD 18.97] vs antiphase congruent [mean 112.64, SD 25.82] and in-phase mirrored [mean 23.78, SD 16.07] vs antiphase mirrored [mean 121.39, SD 28.19]). Moreover, the average speed (revolutions per second) and travel distance (m) of the in-phase mode were significantly higher than those of the antiphase coordination.

Conclusions: Results of this feasibility study show that spatiotemporal data captured from the mHealth app can quantify motor performance. Moving beyond traditional assessments, MoEvGame incorporates gamification into ubiquitous and accessible technology as a fast, flexible, and objective tool for UL motor assessment.

(JMIR Serious Games 2025;13:e57802) doi:10.2196/57802

XSL•FO
RenderX

KEYWORDS

mobile health; mHealth; digital health; mobile apps; smartphones; iPad; gamification; serious games; digital interventions; digital technology; spatiotemporal; upper limb movement; motor performance; motor skills; pediatrics; toddler; children; youth

Introduction

Background

Upper limb (UL) movements such as reaching, grasping, and finger individuation are fundamental to a child's development and to becoming independent with activities of daily living. Furthermore, coordinated and purposeful UL movement enables participation in play and leisure activities. In the United States, more than 17% of children (1 in 6) aged 3-17 years were diagnosed with a developmental disability [1], many of which affect UL motor function and limit independence and quality of life. Importantly, motor impairments, if left unaddressed, will lead to chronic motor disabilities in adulthood that are linked to poor health outcomes [2-4].

To aid rehabilitation professionals such as physical and occupational therapists, timely and objective assessments of a child's motor abilities are pivotal to establishing treatment plans and monitoring the impact of ongoing neurorehabilitation interventions. Standardized play-based tests of gross and fine motor function are used clinically to assess motor outcomes [5-7]; however, these tests may have subjective scoring and require time to administer, some up to one hour, which challenges a child's motivation and attention. More objective laboratory tests use 3D motion capture technology for kinematic analysis of the UL movement. However, this laboratory equipment is expensive, and while newer portable solutions have been developed, it is usually restricted to a laboratory environment [8]. Furthermore, both laboratory and clinical assessments require in-person sessions wherein clinicians or researchers directly evaluate participants; such in-person visits can be a major limitation for families with travel, transportation, or time restrictions and thus may decrease research participation altogether [9]. Therefore, the current landscape of UL motor assessments shows gaps in meeting the needs of clinicians and researchers to be objective and the needs of families or participants to be flexible and accommodating [10,11].

Related Work

Mobile health (mHealth) platforms have the potential to address these gaps by integrating mobile technology with gamified software applications to test UL motor functions [12-15]. First, mHealth technologies are portable and would enable data collection in any setting with wireless or network connectivity. Second, they are accessible, with recent estimates indicating that a large majority of Americans, even those from underrepresented racial, ethnic, and socioeconomic groups, own and use mobile technology [16]. Lastly, they are customizable, primarily via flexibly programmed software packages that allow for many experimental tasks to be implemented seamlessly in a mobile app [17]. A recent survey on the feasibility and acceptability of mHealth technologies in clinical trials of movement disorders (n=54, by searching ClinicalTrials.gov) revealed the potential of these technologies to increase participation, reduce the cost and duration of procedures, overcome barriers to collect and validate data, and enable real-time monitoring of outcomes [18].

Specific to studying UL movement and motor control, two examples in the literature provide a framework for expanding experiments outside of the laboratory using mHealth approaches. First, the portable motor learning laboratory was developed to simulate lab equipment needed to study visuomotor adaptation and motor learning, showing that participants using a tablet reproduced similar learning curves and retention rates [19]. A second tablet-based (Android) software system designed to perform large-scale research outside of the laboratory illustrates various experimental protocols and tasks for quantifying and analyzing the potential mechanisms of motor behaviors and sensorimotor control [20]. The authors highlight the flexibility in their system to be fully customizable to the developer's needs and research interests. Although these and other studies have investigated game-based clinical diagnostics and motor skill assessment methods [21-30], none thus far have been designed specifically for pediatric populations and data collection inside and outside of a traditional clinical or laboratory environment.

Objective of This Study

For the current health problem of evaluating UL motor function, we developed a Gamified mHealth System (MoEvGame), a novel iPad-based tool consisting of games to assess and monitor UL motor function remotely without additional sensors or equipment. As an initial step in this research, we undertook a feasibility study with the following aims: (1) To design and develop the MoEvGame to capture UL movement, which consists of 5 games that examine theory-driven features of UL movement (whole-limb reaching, fine motor control, manual dexterity, and bilateral coordination) critical for performing functional tasks and daily activities. We introduced a software prototype for studying UL movement, consisting of an iOS app, a cloud-based data management and storage system, and a data analysis pipeline. Customizable features of the prototype allow for the design and implementation of game-based protocols considering different cohorts, age groups, and mechanisms of UL function. (2) To develop mathematical and analytical techniques to quantify spatiotemporal measures of whole-limb movement, fine motor precision, and bimanual coordination. As an additional novelty and highlight of MoEvGame, we included automated scripts (Multimedia Appendix 1) that are capable of performing real-time analysis of movement data collected with the iPad, thereby eliminating the need for time-consuming post-processing and analysis for movement quantification. (3) To collect preliminary data from typically developing children inside and outside a laboratory environment to show the prototype's initial feasibility and proof of concept.

This paper is structured as follows. In the Methods section, we describe the study participants, game prototype and description, data management, processing, and analysis methodologies. The Dependent Variables and Analytic Methods section briefly describes the analytical approach and algorithms to map

XSL•FO
RenderX

spatiotemporal movement variables to motor performance. In the Statistical Analysis section, we present the statistical methods to compare speed-accuracy tradeoffs using the participant's data. The Results section presents the motor performance data and correlation analysis. Then, the Discussion section describes the principal findings with limitations and opportunities of the MoEvGame. Finally, the paper concludes with planned future research directions and the translational potential of using mHealth in pediatric rehabilitation studies.

Methods

Participants

We collected data from 3 children in a laboratory setting and 28 children in a community setting (Discovery World Museum, Milwaukee, WI [31]) for initial feasibility testing of our prototype. The 3 children tested in the laboratory were run through a pilot and had to address any technical concerns before testing in a community setting. General guidance on how to complete each game was provided to each user. Anyone aged 5-14 years was eligible to participate. We used two Apple iPads (12.9" display, 7th generation) that were installed with MoEvGame. All participants were asked to play any of the games they preferred on the exhibited iPads. Research staff provided verbal instructions and cues if the participant did not understand or ignored the purpose of the game. Besides age, no other identifying information was collected from participants.

Ethical Considerations

The experimental protocol of the study involving human participants was carefully designed and implemented in strict compliance with ethical standards. The Maquette University Institutional Review Board (IRB) approved the overall protocol for this research project (IRB number 3675). In this feasibility study, we collected only anonymous hand movement data from the iPad application and did not collect any personal identifying information from participants to ensure privacy and confidentiality. Therefore, the data were completely anonymous and a waiver of informed consent was granted by the IRB. No compensation was provided to participants.

MoEvGame Prototype

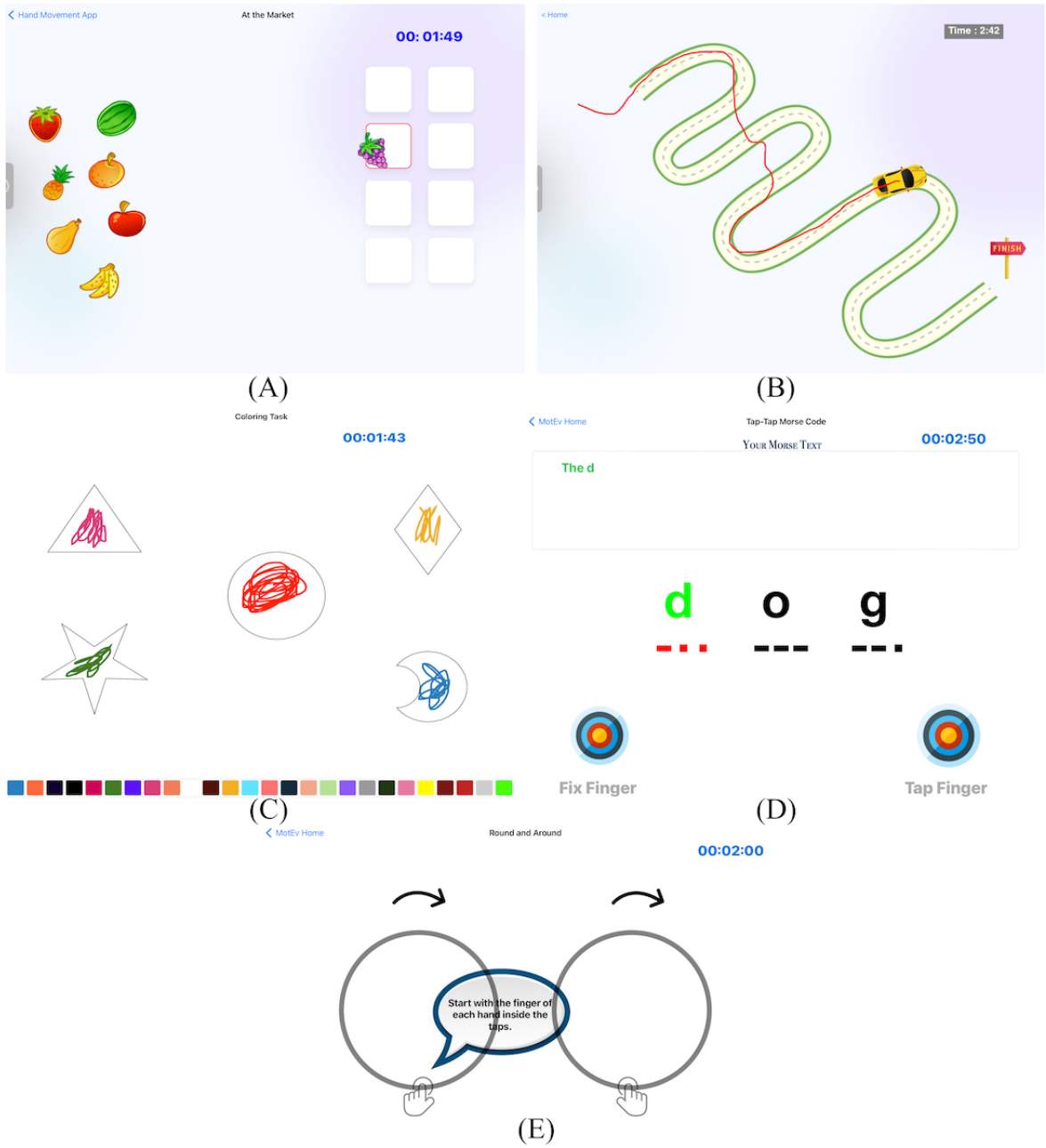
We developed MoEvGame targeting the Apple iPad platform and a cloud-based data management system. The system harnesses features of mobile devices and gaming technology to study real-world movement in different environments without additional equipment or sensors. The tasks were designed to aim at four key features of skilled motor control: (1) whole limb reaching, (2) fine motor and individual finger control, (3) manual dexterity, and (4) bilateral coordination. These features are essential for performing activities of daily living and for play and leisure participation [32]. Furthermore, these domains map onto discrete but overlapping neural substrates underlying motor control. Whole-limb reaching involves a sensorimotor network (posterior parietal cortex, premotor, and primary motor cortices) to identify objects in space and move the end-effector to the object [33]. Individual finger movement requires the integrity of the fast-conducting corticomotor projections to control individual muscles in the hand [34]. Bilateral coordination requires cooperation between the motor systems facilitated by the corpus callosum [35].

Game Descriptions

MoEvGame was developed as illustrated in Figure 1, considering a kinematic approach to studying UL motor performance. While playing, spatiotemporal data were recorded continuously to evaluate multiple variables related to speed, accuracy, precision, coordination, and skill learning. Of note, a speed-accuracy tradeoff framework was incorporated for the game design. Each game requires accurate movements to be completed within a time limit. Thus, participants were required to complete the games within the time limit to get feedback and "credit" for completing the game.

XSL•FO
RenderX

Figure 1. Representative user interface of movement games within MoEvGame. Together, the games test whole-limb movement (A and B), fine motor control and manual dexterity (C and D), and bilateral coordination (E).



At the Market

At the Market involves using a finger to drag and drop fruits into crates. The border of each crate will turn green if placed accurately and red if an error occurs (part of the fruit remains outside the crate, left panel of Figure 1A). A trial of this game block consists of 2 attempts using the right and left hands, respectively, taking 60 seconds for each attempt.

Race to the Finish

Race to the Finish involves using a finger to drag a race car along a curvilinear track (Figure 1B). An error occurs if the participant draws a direct path from the starting to the ending location to encourage adherence to the task. A trial consists of performing 3 races with each hand, where each race ends after 30 seconds.

Filling In

This is a coloring task with the Apple Pencil to color 5 shapes in 2 minutes. Those shapes reflect various curvilinear and rectilinear spaces (Figure 1C). The coloring task challenges fine motor control and spatial awareness. Similar coloring tasks are part of standardized developmental assessments [36].

Secret Message

The Secret Message involves tapping or holding a finger following a “dot” and “dash” signal of Morse Code to generate a hidden message while holding the contralateral finger in a fixed target. To test potential mirroring movements, the finger of the static hand must remain placed in a contralateral circular target (Figure 1D). A trial involves 180 seconds to perform tapping a secret message; after 90 seconds, the tapping finger and contralateral finger are required to switch.

XSL•FO
RenderX

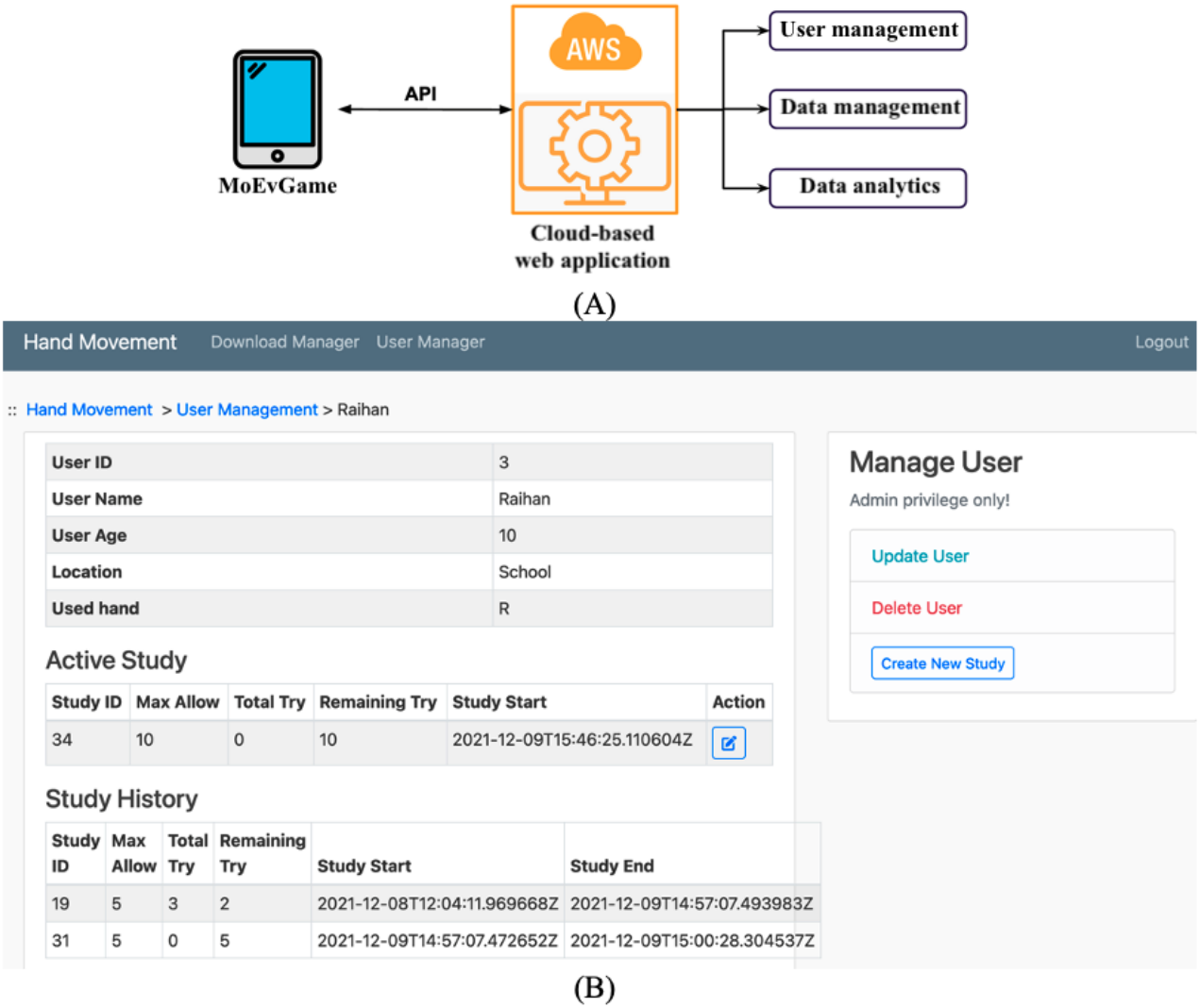
Round and Around

Round and Around involves drawing circles with both hands in 4 modes: in-phase congruent, antiphase congruent, in-phase mirrored, and antiphase mirrored. In-phase: starting finger positioning will be at the same location on the circle (eg, the top of the circle); antiphase: starting finger position will be 180 apart (eg, one finger at the top and a finger from the other hand at the bottom of the circle); congruent: movement direction is consistent (both fingers moving clockwise); and mirrored: movement direction is opposite (one finger moving clockwise and the other counterclockwise). Participants will trace around the circle for 30 seconds in each mode.

Data Management, Processing, and Analysis

A cloud-based data server has been designed and developed to manage the participants, game data storage, processing, and analysis (Figure 1). The architecture of the prototype is illustrated in Figure 2A, where the game data are transferred to a cloud server from the device via an application programming interface. The data and user management system prototype is implemented using the Python Django framework, JavaScript, and PostgreSQL database and deployed to an Amazon Web Service (AWS) EC2 cloud server. There are three core features of this management system as follows.

Figure 2. MoEvGame. (A) Prototype architecture. (B) Participant and game data management system in a web-based platform. API: application programming interface, AWS: Amazon Web Services.



User Management

Each participant's interaction with the app is controlled through a web-based dashboard. This is primarily used to link a participant to a study identified (eg, study ID) and to set how many attempts can be performed; this is necessary to limit practice and learning effects that might affect motor performance (Figure 2B). Moreover, the retention-based study design has been incorporated into the study management.

Data Management

The database has been designed by focusing on holistic data integration [37] that will enable data analysis from both individuals and specific groups. Data transmission depends on the data structure and storage representation. In the prototype, the raw data generated from games are mainly time series of spatiotemporal measures of the cursor position. Thus, we designed a database system with low latency to process a very high volume of JSON data [38]. Researchers can download user

XSL•FO
RenderX

data from the dashboard once the tasks are completed for analysis and further study.

Data Analysis

While the raw data are similar, each game has specific performance measures and outcome data parameters requiring varied analytical techniques. Based on the objective of the game and its dependent variables, we devised specific statistical analysis plans and implemented these as Python scripts. Data were downloaded from the server in CSV format and analyzed accordingly to generate the results. The primary goal of the analysis was to demonstrate the feasibility of the prototype's automated and real-time analysis capabilities. The following section will describe the analysis plan for each game and validate it with representative participants' data.

Dependent Variables and Analytic Methods

Whole-Limb Movement

Spatiotemporal variables such as movement path (x and y coordinates in 2D space) and time are used to quantify movement speed and accuracy. Dragging, dropping, and object tracking in the touchscreen were previously found to be effective methods to measure arm movement proficiency [39], where the completion time(s), accuracy, and deviation error from the shortest path trajectory (mm) were considered important dependent measures. In At the Market (Figure 3), suppose the movement path of an object (fruit) is (x_i,y_i); time is t_1, where i=1, 2, ..., N, with N the total number of samples; P_1 and P_2 are the coordinates of the starting position of the fruit and the target destination (box), respectively. The perpendicular deviation of each point (d_i) of the path from the P_1P_2 shortest path is calculated using equation (1). Then, the root mean square error (RMSE) of the whole movement path is calculated using equation (2).

Here, α, β, and δ are the shortest path (P_1P_2) coefficients.



(1)



(2)

The Race to the Finish task (Figure 4) tests visuomotor coordination of accurately tracing a curvilinear path. The first-order derivative (dy/dx) of the movement path has been used to identify the movement direction and segment to calculate the deviation error from the ideal path [40]. Let us consider a multivariate nonstationary first-order derivative signal $dy/dx = \{dy/dx_1, ..., dy/dx_T\}$ that takes value in $(-\leq dy/dx_i \leq)$ (and has T samples. The pruned exact linear time (PELT) method [41] identifies the ideal segmentation τ from a criterion $V(\tau, dy/dx)$ that is minimized for K, the known number of changes. In our case, equation (3) was minimized for K=12 (see Figure 4A) using the radial basis function ("rbf") of the PELT method. Figure 4B shows the detection of τ=13 segments using K breakpoints from a movement path (x, y) of a participants' trial.



(3)

where $f_{rbf}(.)$ is a Gaussian kernel-based cost function that measures the goodness of fit of the nonparametric subsignal  . The cost function f_{rbf} for a subsignal $z_{a...b}$ is shown in equation (4), where > 0 is the bandwidth parameter. The PELT minimization method identifies the segments of movement curvature from the criterion, V. These segments are used to calculate the error distance from the ideal racing track.



(4)

Each segment  . predicted by the PELT model has been used to calculate the perpendicular distance (d_i) between the movement point (x_i,y_i) and ideal curvature using equation (5).



(5)

where i=1, 2, ..., M, and M is the number of samples in the ideal curvature of  . segment boundary. Figure 4C shows the movement path of 6 trials (3 with the right hand and 3 with the left hand) and Figure 4D shows the respective RMSE and completion time of each trial. A speed-accuracy trade-off is demonstrated by the negative correlation between completion time and error; greater accuracy requires greater control and, thus, longer completion time. Furthermore, we can observe that the participant of this trial (12 years) performed better in the Right #2 and Left #2 blocks (Figure 4D), whereas the Right #1 block took the longest to complete with the smallest RMSE.

XSL•FO
RenderX

Figure 3. Analysis of At the Market game. (A) Movement trajectories for each item (a piece of fruit). (B) Root mean square error (red whisker) and completion time (bar) for each movement relative to the shortest path from item to target box.

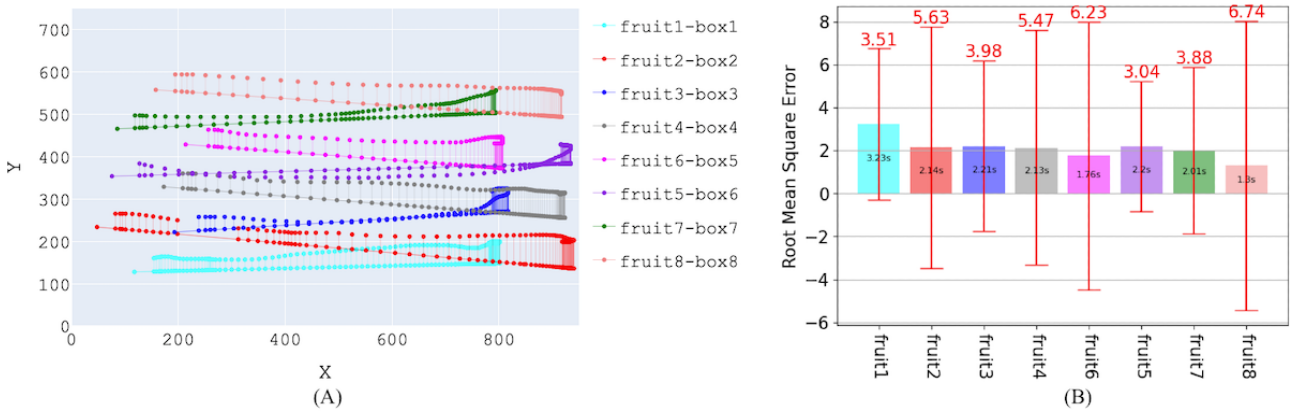
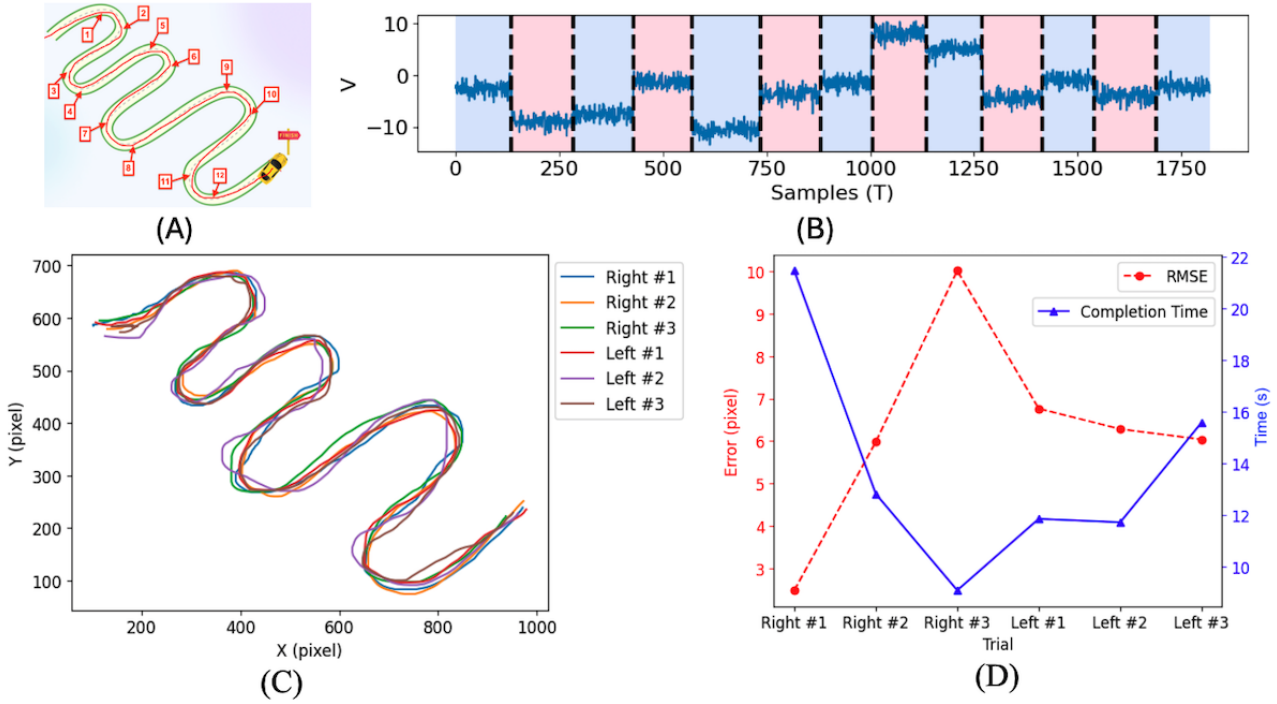


Figure 4. Analysis of Race to the Finish game. Changes in movement direction were determined using change point detection and the pruned exact linear time (PELT) method. (A) Break points in the curvature track were manually defined. (B) The PELT model detects the K=12 change points as noted by the change in criterion V. (C) Movement trajectories of 6 attempts. (D) Error (left y-axis) and completion time (right y-axis) are plotted for each attempt. Note the speed-accuracy trade-off for these 6 attempts.



Fine Motor Control and Manual Dexterity

Individual finger and fine motor control are needed for object manipulation and handwriting. Drawing or coloring objects (eg, square, circle, or other shapes) and handwriting tasks are common techniques for evaluating fine motor precision and manual dexterity in children with developmental disabilities [42,43]. In the Filling In task (Figure 5), the accuracy (% filled inside the shape; % filled in outside the shape) and completion time are considered dependent measures of fine motor control. To compute these, we reconstructed the colored shape from the game data as illustrated in Figure 5A. The pixels colored inside or outside the target shape were calculated using their geometric properties (eg, triangle, circle, star). Since the participant's movements during coloring likely contain duplicate pixel values, digital image processing techniques are applied to calculate the unique colored pixels from the reconstructed images. The percentage of coloring inside each shape is shown in Figure 5B, indicating the accuracy of fine motor control of an individual trial.

Speed and coordination of the distal UL correlate with corticospinal system integrity [44]. The Secret Message game (Figure 6) tests the motor performance of individual finger movements. Furthermore, because the contralateral (nontapping) finger is required to be engaged statically while the other finger is tapping, this game also tests response inhibition and mirroring movements. Tap and release speed, number of errors in tapping hand, and number of displacements (ie, nontapping finger movements inside or outside the target) are calculated as dependent measures. Errors are defined as inappropriate tap and release combinations corresponding to a "dot: or "dash" from Morse Code. For example, if a child pressed too long for a "dot," this would be counted as an error. Figures 6A and 6B show the

XSL•FO
RenderX

state of the tapping and nontapping fingers and the error rate with respect to time of a single trial in which the participant revealed 38 characters (right finger tapping 13 and left finger tapping 25 characters) of the secret message. In these figures, each data point represents a different state of the left and right fingers (eg, blue triangles represent the tapped state); data points within a state are offset in the figures for clarity. From these data, we observed that the tapping rate (blue triangle) increased, and the error rate (red x) decreased as the trial continued, perhaps indicating the participant understood the rules of the game. The amount of movement within the fixed finger target was less when the fingers were switched (gold plus symbols, states 2-3).

Figure 5. Analysis of Filling In game. (A) Sample coloring performance. (B) Bar plot of coloring accuracy, divided into the proportion colored inside and outside the target shape.

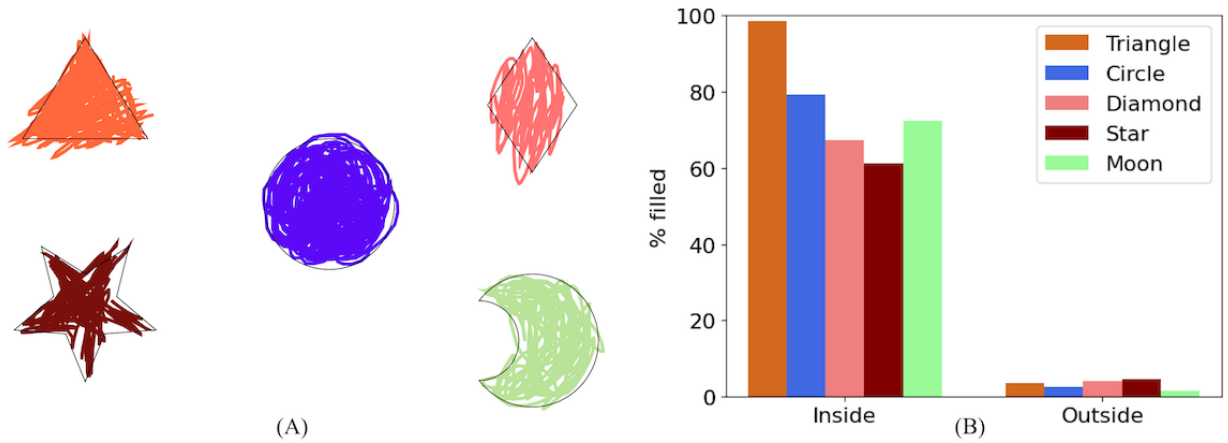
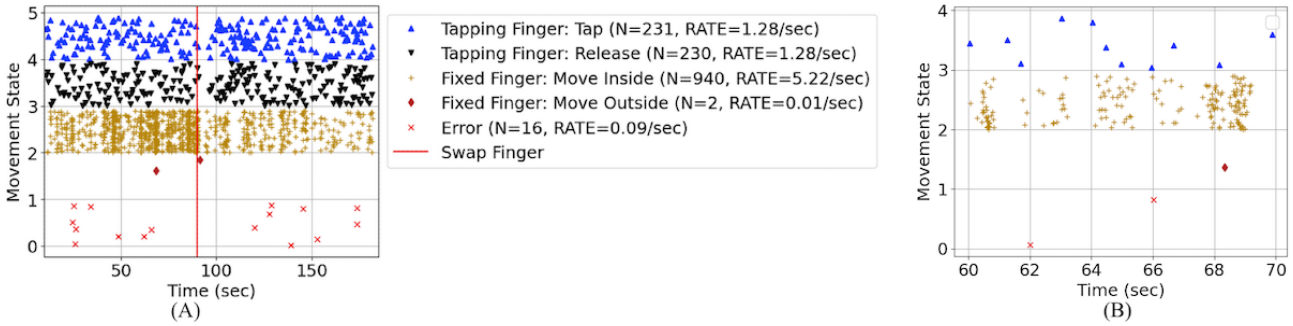


Figure 6. Analysis of Secret Message game. (A) Movement state of tapping finger and contralateral static (fixed) finger movement across time. (B) Magnified view of 10 seconds of tapping data. Note the user's non-tapping finger moved outside the target once (red diamond) and made two tapping errors (red x) with the tapping finger.



Bimanual Coordination

Bimanual coordination is important for completing many everyday activities. The Round and Around game is designed based on dynamical systems theory to challenge bimanual coordination with different levels of stability between the limbs [45,46]. We used 2 spatiotemporal measures to quantify bimanual coordination: (1) phase synchronization, which was defined as the instantaneous phase difference between the right hand and left hand during in-phase (ideal=0°) and antiphase (ideal=180°) conditions; and (2) speed synchronization, which was the difference in angular velocity of each hand. Let, $y_1=x(t_1)= \sin(\omega_1 t_1+\varphi_1)$ and $y_2 =x(t_2)= \sin(\omega_2 t_2+\varphi_2)$ are 2 nonlinear signals of the right finger and left finger, respectively. Here, ω, t, and φ represent angular velocity, time, and initial phase, respectively. The instantaneous phase has been calculated from the argument of the analytical signal based on Hilbert Transform [47], as defined in equation (6).

(6)

As y_1 and y_2 signals are sinusoidal, the Hilbert Transform method reconstructs the interpretable phases as  . and ., respectively. The phase difference between y_1 and y_2 has been estimated at a certain time, t, using equation (7), where $t = t_1 = t_2$.

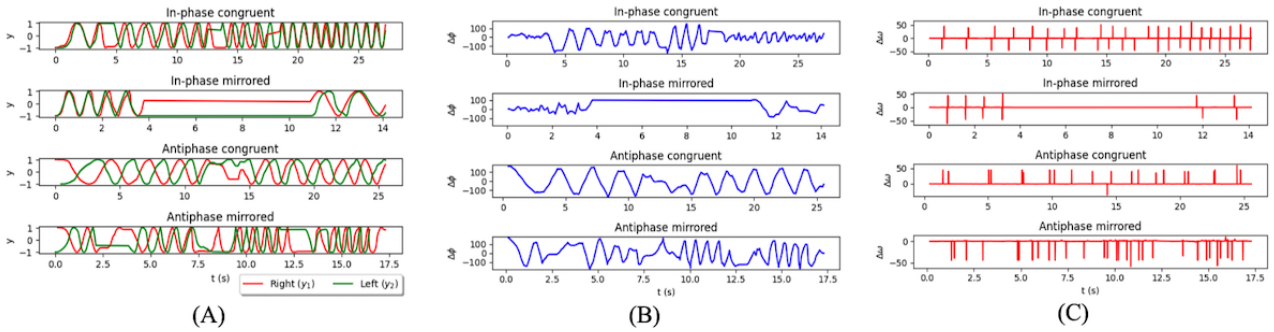


(7)

The velocity difference (Δω) between y_1 and y_2 has been calculated at a certain time, t, using $\Delta\omega= \omega_1-\omega_2$, where $t = t_1 = t_2$. Figure 7 displays the bimanual hand position (A), phase difference (B), and velocity difference (C) from a representative participant.

XSL•FO
RenderX

Figure 7. Analysis of Round and Around game. Each panel shows 4 plots corresponding to the 4 modes of coordination (in-phase congruent, in-phase mirrored, antiphase congruent, antiphase mirrored). (A) Raw angular position traces (y_1 and y_2) of each hand. (B) Phase difference (Δφ) between y_1 and y_2. (C) Velocity difference (Δω) between y_1 and y_2.



Statistical Analysis

As an initial proof-of-concept, we evaluated the feasibility of MoEvGame using 2 theory-based analytic approaches. First, we used a speed-accuracy tradeoff framework to compare motor performance [48]. In nonlearned tasks, users balance basic parameters of speed and accuracy to complete a task [49]. In the procedure, participants were not given specific instructions to use a "speed first" or "accuracy first" strategy. Therefore, different strategies that optimized movement speed and accuracy were likely employed. From this speed-accuracy framework, we predicted that participants moving faster would also show higher errors or error rates. In contrast, if both speed and accuracy were optimized, we concluded the participant performed a highly learned task. The speed-accuracy framework was used to analyze the data from At the Market, Race to the Finish, and Secret Message.

At the group level, analysis was performed based on developmental changes in motor abilities that would be expected as children mature. Thus, for a given assessment of motor development, one would expect to find age-related improvements in motor performance. A group analysis using the Pearson correlation coefficient was performed by combining data for each age and examining the relationship between changes in age and motor performance.

For bimanual coordination, a separate analysis was used based on the theoretical differences in difficulty for performing in-phase versus antiphase movements. Here, we examined the mean, standard deviation, and 95% confidence interval of the phase and angular velocity difference between the hands as measures of performance for each trial. We termed this "phase synchronization" or "speed synchronization," respectively, where values closer to 0 represent coherence between the two hands.

Results

Participants' Characteristics and Feasibility

A total of 56 trials were conducted by 31 participants. The participants' median age was 9.0 (IQR 5.0-14.0) years, where most were third- or fourth-grade students (n=18, 58.06%, 9≤age≤11). Some participants attempted more than one of the 5 games, whereas some just tried a single task. The summary of the trials and participation is detailed in Table 1. General observations during testing showed that children were willing and excited to try the games and were comfortable interacting with the iPad. A small percentage of children did not complete the entire game, and we excluded the incomplete trials from the analysis.

Table 1. Participant characteristics and engagement.

	Participants (N=31)
Characteristics	
Age (years), median (IQR)	9.0 (5.0-14.0)
Second or lower grade student, n (%)	9 (29.04)
Third- or fourth-grade student, n (%)	18 (58.06)
Fifth- or higher grade student, n (%)	4 (12.9)
Trials, n (%)	
At the Market	11 (19.6)
Race to the Finish	21 (37.6)
Filling In	2 (3.5)
Secret Message	13 (23.2)
Round and Around	9 (16.1)

XSL•FO
RenderX

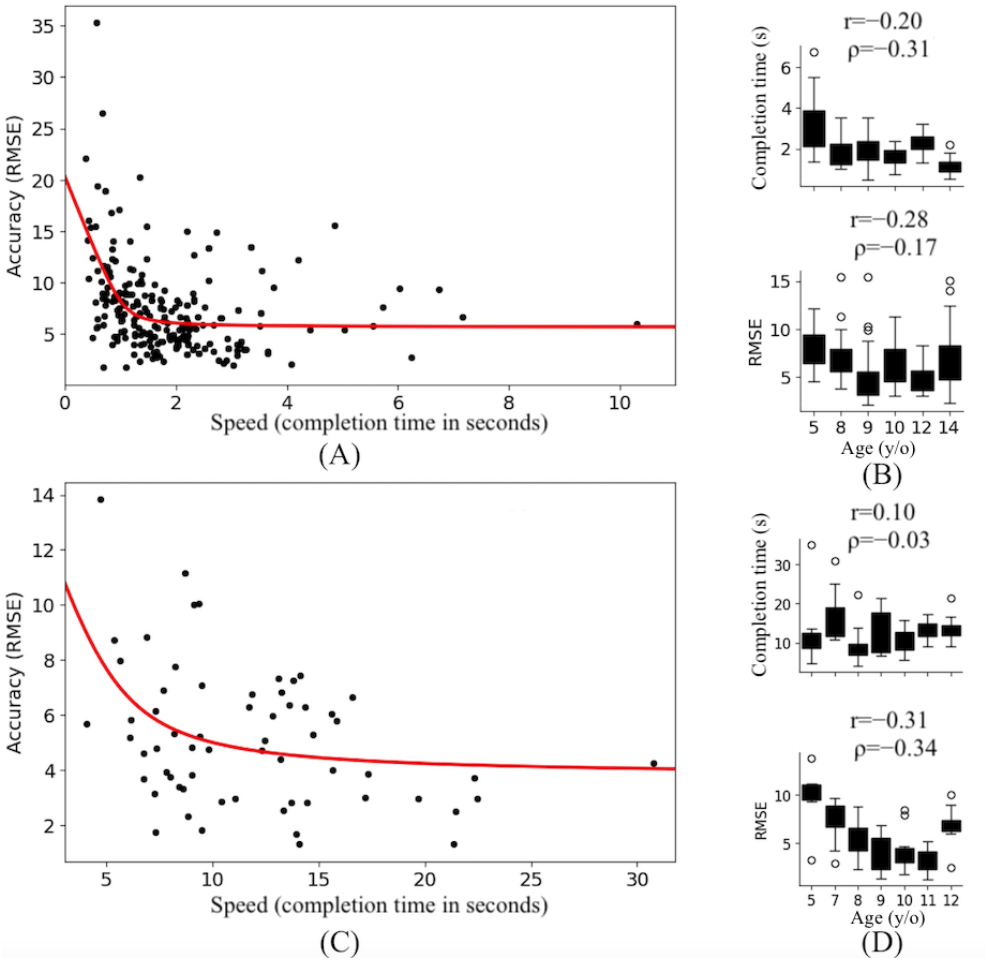
Motor Performance and Correlation Analysis

Figure 8 shows the whole-limb movement performance on two games, At the Market and Race to the Finish, and their association with age. These games had the most participants (N=23, 74.2%) across the widest age range. Among these data sets, 21.7% of participants played both games, whereas 60.8% and 17.5% of trials were Race to The Finish and At the Market trials, respectively. We found that there was a negative correlation between speed and accuracy in At the Market (r=–0.42, ρ=–0.29) and Race to the Finish (r=–0.30, ρ=–0.35) as illustrated in Figure 8A and 8C, respectively. The area under the curve of the quadric curve in these figures indicated that low RMSE (high accuracy) could be achieved with optimal speed. Participant's movement error (RMSE) decreased with increasing age (Figure 8B and 8D), indicating older participants had better movement accuracy.

Fine motor control and manual dexterity (finger movement) were evaluated from the speed-accuracy trade-off of Morse code finger tapping and error rates. The tapping rate was analyzed with respect to participants' age, as shown in Figure 9. While both older and younger participants exhibited higher tapping rates, error rates were higher in younger compared to older participants.

Finally, bilateral coordination results from the group data are shown in Table 2. The mean (μ) and standard deviation of phase synchronization for in-phase coordination were less than from antiphase coordination. Likewise, left- and right-hand speed was higher for in-phase than antiphase coordination. Moreover, the right hand traveled a shorter total distance during antiphase congruent (mean 3.31, SD 1.67 m; coefficient of variation [CV]=0.51) and mirrored (mean 3.21, SD 1.67 m; CV=0.52) compared to in-phase congruent (mean 5.26, SD 3.04 m; CV=0.58) and mirrored (mean 4.04, SD 3.39 m; CV=0.84) coordination. These results indicate participants performed better and completed more full revolutions per second in the in-phase coordination, consistent with prior observations that antiphase bilateral movements are more difficult to coordinate than in-phase [50,51].

Figure 8. Whole-limb movement assessment of group data. Plots (A, C) show movement speed versus accuracy, while plots (B, D) show the completion time and error (RMSE) as a function of the user's age when playing At the Market (A-B) and Race to the Finish (C-D). Note the general speed-accuracy trade-off across the group and the negative correlation of spatial effort with age. The red curve is an exponential best fit to the points. RMSE: root mean square error.



XSL•FO
RenderX

Figure 9. Tapping and error rate compared to user age for Secret Message. The finger tapping rate (speed) increased and the error rate (accuracy) decreased with increases in age.

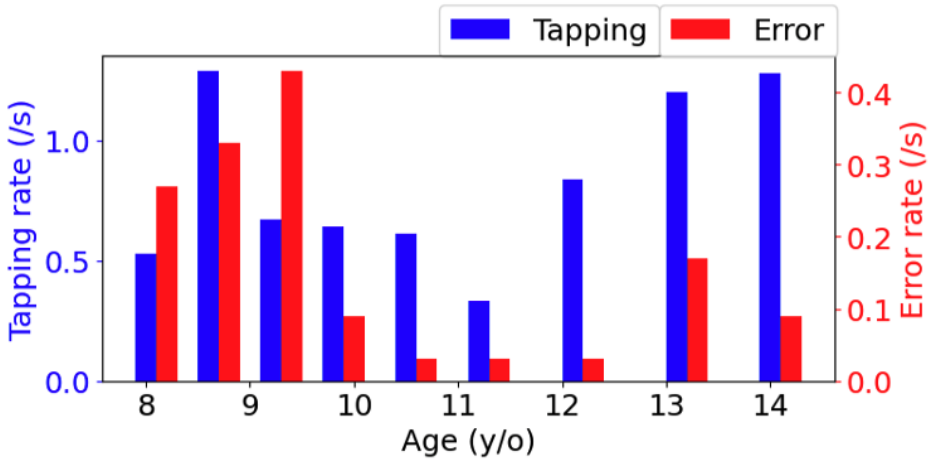


Table 2. Bimanual coordination outcomes from Round and Around game.

Mode	Phase synchronization (degree)		Speed synchronization (revolutions per second)					
	SD	Mean (95% CI)	Right hand		Left hand		Difference	
			SD	Mean (95% CI)	SD	Mean (95% CI)	SD	Mean (95% CI)
In-phase congruent	18.97	28.85 (28.49 to 29.21)	5.48	7.11 (7.02 to 7.21)	5.52	7.64 (7.54 to 7.74)	5.44	–0.52 (–0.62 to –0.43)
In-phase mirrored	16.07	23.78 (23.46 to 24.10)	4.97	7.60 (7.50 to 7.69)	4.97	7.24 (7.14 to 7.33)	5.02	0.36 (0.26 to 0.45)
Antiphase congruent	25.82	112.64 (111.98 to 113.29)	3.42	4.70 (4.63 to 4.77)	3.65	4.55 (4.48 to 4.63)	3.73	0.14 (0.06 to 0.22)
Antiphase mirrored	28.20	121.39 (120.53 to 122.24)	4.97	6.27 (6.15 to 6.39)	5.36	6.63 (6.50 to 6.77)	5.27	–0.36 (–0.49 to –0.23)

Discussion

Principal Results

This study aimed to demonstrate the feasibility of an mHealth app to quantify and evaluate UL movement. Pilot testing showed that MoEvGame can capture UL movement performance that is linked to aspects of motor control and development in children. Specifically, speed-accuracy tradeoffs were noted for At the Market, Race to the Finish, and Secret Message (Figures 8 and 9). These effects were observed because we did not ask participants to prioritize speed over accuracy, and thus, various strategies may have been used to complete the games. Notably, the speed-accuracy tradeoff is evidence for face and content validity, showing that the games test motor control processes and are suitable tasks for future studies of UL motor functions. In addition, age-related changes in UL performance were evident: in general, younger participants demonstrated higher spatial errors than older participants, consistent with the maturation and development of the motor system. These findings agree with prior results [52,53], where speed and accuracy were key factors in fine motor performance to complete a gamified toy. Other motor constructs, such as bimanual coordination, were examined. We observed differences in the left-hand and right-hand phases and speed synchronization between the four modes (in-phase or antiphase, congruent or mirrored) for the bimanual game that replicated prior studies in which in-phase coordination is more stable and therefore can be performed more efficiently than antiphase coordination [54]. These results are a promising first step toward validating this technology as an objective and standardized UL assessment tool.

Portable technology, such as mobile devices, advances modern technological research testing and protocol design. In a study by Mollà-Casanova et al [55], the authors implemented a series of tasks on an iPad for adults with stroke to test hand motor function and demonstrated strong correlations between variables measured with the iPad and outcomes from standardized clinical assessments. A host of other examples using tablets and mobile technology to study UL movement in children also exist in the literature [56]. However, some common limitations of these studies were object manipulation on a touchscreen, offline data analysis, real-world usability, and age-appropriate game design. The hand movement application described by Matic and Gomez-Marin and the portable motor learning laboratory described by Takiyama and Shinya [19] provide a strong foundation for using tablet devices to study motor control. Although these examples were not specifically designed for children, they could be adapted for pediatric populations. MoEvGame expands upon these ideas by adding gamification, automated analysis pipelines, and a data server connected to the application programming interface to facilitate data transfer and remote access.

XSL•FO
RenderX

Collectively, MoEvGame and other mHealth apps allow research to be conducted in contexts and environments. A common concern is that a participant's function observed in traditional laboratories or clinic settings may be different in their home or other environments. With mobile assessment technology, this means assessments might occur at home or when studying pediatric populations at school. Studies that collect multiple data points spread out across times and environments would demonstrate the benefits of mHealth technology and showcase robustness not commonly available with laboratory-based assessments. Overall, investigations that employ continuous monitoring of UL movements in varied environments will lead to more impactful and ecologically valid translational research.

Limitations and Opportunities

There are limitations and future variables to consider before further validating MoEvGame. First, the feasibility testing data were collected from a convenience sample of children attending a public museum exhibit. Testing in a larger and more representative sample of children would allow us to enhance the generalizability of our findings and begin to explore age-normative values. Second, since we were interested in learning how children interacted with MoEvGame, no standardized instructions were provided, resulting in varied strategies to achieve the goal of each game. For instance, while playing At the Market, some users moved the fruit directly to the box, while others moved it underneath and placed it into the crate from the opposite side. Considering we do not have other measures of motor development or function for comparison, overall this limits the interpretation of the MoEvGame data and how data from an individual child relates to their motor function. Considering these user behaviors is critical for further development and refinement of the MoEvGame. Finally, by using a 2D touchpad, we are limited to measuring the final endpoint of movement rather than quantifying UL movement kinematics in 3D, similar to what is obtained from laboratory motion capture systems. We plan to compare the MoEvGame movement variables to 3D kinematic variables in future validation studies. In spite of these limitations, we were able to achieve our primary objective of demonstrating feasibility and proof-of-concept. The goal of future work will be to systematically test the validity and reliability of MoEvGame in children with and without motor impairments with a large sample and wider age group. However, age-related changes can affect motor behavior [57], specifically older kids or adolescents with motor impairment who are capable of using a touchscreen device. Therefore, it may be necessary to capture an inclusive data set for further investigation.

Conclusions

In this feasibility study, we describe a gamified mHealth system, MoEvGame, as a tool to study UL movement in children with and without neuromotor impairments. Valid and objective outcomes are pivotal for clinical trials and provide clinicians with relevant information about a child's motor function and ability. Moving beyond traditional assessments that are time-consuming and require in-person administration, our approach for assessing UL movement skills incorporates gamification into ubiquitous and accessible technology. Research on the mechanisms of UL motor dysfunction is critical for advancing evidence-based assessment and treatment of children with motor impairments. The availability of fast, flexible, and objective testing tools using common mobile devices will be advantageous to researchers, clinicians, and families. Future work in participants with and without motor impairments across a wide age range in both laboratory and nonlaboratory settings will test the reliability and validity of MoEvGame.

Acknowledgments

We thank the members of the Pediatric Movement Neuroscience Laboratory and Ubicomp Lab for helping with data collection and input on the app and game design. Funding for the research reported in this publication was provided by the Eunice Kennedy Shriver National Institute of Child Health & Human Development of the National Institutes of Health under award numbers NIH/NICHD R21HD114094 and P2CHD101899. The content is solely the authors' responsibility and does not necessarily represent the official views of the National Institutes of Health.

Data Availability

The anonymous datasets generated and analyzed during this study are available in the GitHub repository [58].

Authors' Contributions

MRM was involved in data curation, formal analysis, methodology, software, validation, visualization, and writing (original draft and reviewing/editing); SIA was involved in conceptualization, funding acquisition, investigation, project administration, software, resources, supervision, and writing (original draft and reviewing/editing); and SN was involved in conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, software, resources, supervision, validation, and writing (original draft and reviewing/editing).

Conflicts of Interest

None declared.

Multimedia Appendix 1

XSL•FO
RenderX

Dataset and codebase.
[ZIP File (Zip Archive), 11259 KB - games_v13i1e57802_app1.zip]

References

1. Zablotsky B, Black LI, Maenner MJ, Schieve LA, Danielson ML, Bitsko RH, et al. Prevalence and trends of developmental disabilities among children in the United States: 2009-2017. Pediatrics 2019;144(4):e20190811 [FREE Full text] [doi: 10.1542/peds.2019-0811] [Medline: 31558576]
2. Manini T. Development of physical disability in older adults. Curr Aging Sci 2011;4(3):184-191 [FREE Full text] [doi: 10.2174/1874609811104030184] [Medline: 21529321]
3. Hill EL, Brown D. Mood impairments in adults previously diagnosed with developmental coordination disorder. J Ment Health 2013;22(4):334-340. [doi: 10.3109/09638237.2012.745187] [Medline: 23323694]
4. Hands B, Licari M, Piek J. A review of five tests to identify motor coordination difficulties in young adults. Res Dev Disabil 2015;41-42:40-51. [doi: 10.1016/j.ridd.2015.05.009] [Medline: 26057836]
5. Himmelmann K, Beckung E, Hagberg G, Uvebrant P. Gross and fine motor function and accompanying impairments in cerebral palsy. Dev Med Child Neurol 2006;48(6):417-423 [FREE Full text] [doi: 10.1017/S0012162206000922] [Medline: 16700930]
6. Alt Murphy M, Murphy S, Persson HC, Bergström UB, Sunnerhagen KS. Kinematic analysis using 3D motion capture of drinking task in people with and without upper-extremity impairments. J Vis Exp 2018(133):57228 [FREE Full text] [doi: 10.3791/57228] [Medline: 29658937]
7. Pfister A, West AM, Bronner S, Noah JA. Comparative abilities of Microsoft Kinect and Vicon 3D motion capture for gait analysis. J Med Eng Technol 2014;38(5):274-280. [doi: 10.3109/03091902.2014.909540] [Medline: 24878252]
8. Cagliari Tosin M, Machado JC, Balbinot A. sEMG-based upper limb movement classifier: current scenario and upcoming challenges. J Artif Intell Res 2022;75. [doi: 10.1613/jair.1.13999]
9. Tuttle KR. Impact of the COVID-19 pandemic on clinical research. Nat Rev Nephrol 2020;16(10):562-564 [FREE Full text] [doi: 10.1038/s41581-020-00336-9] [Medline: 32760016]
10. van Delden ALEQ, Peper CLE, Kwakkel G, Beek PJ. A systematic review of bilateral upper limb training devices for poststroke rehabilitation. Stroke Res Treat 2012;2012:972069 [FREE Full text] [doi: 10.1155/2012/972069] [Medline: 23251833]
11. McCall JV, Ludovice MC, Elliott C, Kamper DG. Hand function development of children with hemiplegic cerebral palsy: a scoping review. J Pediatr Rehabil Med 2022;15(1):211-228. [doi: 10.3233/PRM-200714] [Medline: 34864699]
12. Dienstmann R. Games for Motor Learning. Champaign, United States: Human Kinetics; 2008.
13. Coker C. Motor Learning and Control for Practitioners. New York, NY: Routledge; 2017.
14. Wardaszko M, Podgórski B. Mobile learning game effectiveness in cognitive learning by adults: a comparative study. Simul Gaming 2017;48(4):104687811770435. [doi: 10.1177/1046878117704350]
15. Jaramillo-Alcázar A, Salvador-Ullauri L, Luján-Mora S. A mobile serious games assessment tool for people with motor impairments. : ACM; 2017 Presented at: ICETC '17: Proceedings of the 9th International Conference on Education Technology and Computers; December 20-22, 2017; Barcelona, Spain p. 172-177. [doi: 10.1145/3175536.3175569]
16. Nemanich ST, Gillick BT, Sukal-Moulton T, Ahamed SI. Can technology improve participation from underserved children and families in rehabilitation research? WMJ 2022;121(3):174-176 [FREE Full text] [Medline: 36301641]
17. Agnew JMR, Hanratty CE, McVeigh JG, Nugent C, Kerr DP. An investigation into the use of mHealth in musculoskeletal physiotherapy: scoping review. JMIR Rehabil Assist Technol 2022;9(1):e33609 [FREE Full text] [doi: 10.2196/33609] [Medline: 35275089]
18. Artusi CA, Imbalzano G, Sturchio A, Pilotto A, Montanaro E, Padovani A, et al. Implementation of mobile health technologies in clinical trials of movement disorders: underutilized potential. Neurotherapeutics 2020;17(4):1736-1746 [FREE Full text] [doi: 10.1007/s13311-020-00901-x] [Medline: 32734442]
19. Takiyama K, Shinya M. Development of a portable motor learning laboratory (PoMLab). PLoS One 2016;11(6):e0157588 [FREE Full text] [doi: 10.1371/journal.pone.0157588] [Medline: 27348223]
20. Matic A, Gomez-Marin A. A customizable tablet app for hand movement research outside the lab. J Neurosci Methods 2019;328:108398 [FREE Full text] [doi: 10.1016/j.jneumeth.2019.108398] [Medline: 31412268]
21. de Menezes LDC, da Silva TD, Capelini CM, Tonks J, Watson S, de Moraes Í, et al. Can individuals with down syndrome improve their performance after practicing a game on a mobile phone? A new insight study. Learn Motiv 2020;72:101685. [doi: 10.1016/j.lmot.2020.101685]
22. Yang L, Grooten WJ, Forsman M. An iPhone application for upper arm posture and movement measurements. Appl Ergon 2017;65:492-500 [FREE Full text] [doi: 10.1016/j.apergo.2017.02.012] [Medline: 28274467]
23. Capelini CM, da Silva TD, Tonks J, Watson S, Alvarez MPB, de Menezes LDC, et al. Improvements in motor tasks through the use of smartphone technology for individuals with Duchenne muscular dystrophy. Neuropsychiatr Dis Treat 2017;13:2209-2217 [FREE Full text] [doi: 10.2147/NDT.S125466] [Medline: 28860778]

XSL•FO
RenderX

24. Coutinho F, Bosisio M, Brown E, Rishikof S, Skaf E, Zhang X, et al. Effectiveness of iPad apps on visual-motor skills among children with special needs between 4y0m-7y11m. Disabil Rehabil Assist Technol 2017;12(4):402-410. [doi: 10.1080/17483107.2016.1185648] [Medline: 27377131]
25. Giammarco E, Di Sano S, Aureli T, Cerratti P, Fanò-Illic G, Pietrangelo T. Psychological and physiological processes in figure-tracing abilities measured using a tablet computer: a study with 7 and 9 years old children. Front Psychol 2016;7:1528 [FREE Full text] [doi: 10.3389/fpsyg.2016.01528] [Medline: 27803678]
26. Dui LG, Lunardini F, Termine C, Matteucci M, Ferrante S. A tablet-based app to discriminate children at potential risk of handwriting alterations in a preliteracy stage. Annu Int Conf IEEE Eng Med Biol Soc 2020;2020:5856-5859. [doi: 10.1109/EMBC44109.2020.9176041] [Medline: 33019305]
27. Moreno Naya D, Vazquez-Araujo FJ, Castro PM, Vivas Costa J, Dapena A, González Doniz L. Utilization of a mobile application for motor skill evaluation in children. Appl Sci 2021;11(2):663. [doi: 10.3390/app11020663]
28. Junghans BM, Khuu SK. Populations norms for 'SLURP'-an iPad app for quantification of visuomotor coordination testing. Front Neurosci 2019;13:711 [FREE Full text] [doi: 10.3389/fnins.2019.00711] [Medline: 31354420]
29. Axford C, Joosten AV, Harris C. iPad applications that required a range of motor skills promoted motor coordination in children commencing primary school. Aust Occup Ther J 2018;65(2):146-155. [doi: 10.1111/1440-1630.12450] [Medline: 29355982]
30. Chua YW, Lu SC, Anzulewicz A, Sobota K, Tachtatzis C, Andonovic I, et al. Developmental differences in the prospective organisation of goal-directed movement between children with autism and typically developing children: a smart tablet serious game study. Dev Sci 2022;25(3):e13195 [FREE Full text] [doi: 10.1111/desc.13195] [Medline: 34800316]
31. Marquette brain games and demos. Discovery World. URL: https://discoveryworld.org/calendar-of-events/marquette-ot-demo-and-games/ [accessed 2024-02-26]
32. Irie K, Mukaiyama K, Yamashita R, Zeidan H, Bandara A, Nagai-Tanima M, et al. Investigating subtypes of motor skills and activities of daily living among young children with motor delay. Occup Ther Int 2023;2023:4031372 [FREE Full text] [doi: 10.1155/2023/4031372] [Medline: 37360553]
33. Chowdhury RH. The Contribution of Whole-Limb Kinematics to Proprioceptive Representations in the Central Nervous System. Evanston: Doctoral dissertation, Northwestern University; 2019.
34. Carneiro MIS, Russo C, Masson R, Rossi Sebastiano D, Baranello G, Turati C, et al. Motor learning in unilateral cerebral palsy and the influence of corticospinal tract reorganization. Eur J Paediatr Neurol 2020;27:49-59. [doi: 10.1016/j.ejpn.2020.04.013] [Medline: 32417186]
35. Ashori M, Norouzi G, Jalil-Abkenar SS. The effectiveness of motor therapy on motor skills and bilateral coordination of children with intellectual disability. Iran Rehabil J 2018;16(4):331-338. [doi: 10.32598/irj.16.4.331]
36. Cools W, Martelaer KD, Samaey C, Andries C. Movement skill assessment of typically developing preschool children: a review of seven movement skill assessment tools. J Sports Sci Med 2009;8(2):154-168 [FREE Full text] [Medline: 24149522]
37. Rahm E. The case for holistic data integration. 2016 Presented at: Advances in Databases and Information Systems. ADBIS 2016; 2016 August 28-31; Prague, Czech Republic. [doi: 10.1007/978-3-319-44039-2_2]
38. Bourhis P, Reutter J, Suárez F, Vrgoč D. JSON: data model, query languages and schema specification. 2017 Presented at: PODS '17: Proceedings of the 36th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems; May 14-19, 2017; Chicago, IL, USA p. 123-135. [doi: 10.1145/3034786.3056120]
39. Bertucco M, Sanger TD. Speed-accuracy testing on the Apple iPad provides a quantitative test of upper extremity motor performance in children with dystonia. J Child Neurol 2014;29(11):1460-1466 [FREE Full text] [doi: 10.1177/0883073813494265] [Medline: 23965400]
40. Truong C, Oudre L, Vayatis N. Selective review of offline change point detection methods. Signal Process 2020;167:107299. [doi: 10.1016/j.sigpro.2019.107299]
41. Killick R, Fearnhead P, Eckley IA. Optimal detection of changepoints with a linear computational cost. J Am Stat Assoc 2012;107(500):1590-1598. [doi: 10.1080/01621459.2012.737745]
42. Butler C, Pimenta R, Tommerdahl J, Fuchs CT, Caçola P. Using a handwriting app leads to improvement in manual dexterity in kindergarten children. Res Learn Technol 2019;27. [doi: 10.25304/rlt.v27.2135]
43. John S, Renumol VG. Design and development of an Android app (HanDex) to enhance hand dexterity in children with poor handwriting. IEEE Access 2022;10:48973-48993. [doi: 10.1109/access.2022.3172330]
44. Wolbrecht ET, Rowe JB, Chan V, Ingemanson ML, Cramer SC, Reinkensmeyer DJ. Finger strength, individuation, and their interaction: relationship to hand function and corticospinal tract injury after stroke. Clin Neurophysiol 2018;129(4):797-808 [FREE Full text] [doi: 10.1016/j.clinph.2018.01.057] [Medline: 29453171]
45. Hung YC, Charles J, Gordon AM. Bimanual coordination during a goal-directed task in children with hemiplegic cerebral palsy. Dev Med Child Neurol 2004;46(11):746-753 [FREE Full text] [doi: 10.1017/s0012162204001288] [Medline: 15540635]
46. Riquelme I, Arnould C, Hatem SM, Bleyenheuft Y. The two-arm coordination test: maturation of bimanual coordination in typically developing children and deficits in children with unilateral cerebral palsy. Dev Neurorehabil 2019;22(5):312-320. [doi: 10.1080/17518423.2018.1498552] [Medline: 30024779]

XSL•FO
RenderX

47. Matsuki A, Kori H, Kobayashi R. An extended Hilbert transform method for reconstructing the phase from an oscillatory signal. Sci Rep 2023;13(1):3535 [FREE Full text] [doi: 10.1038/s41598-023-30405-5] [Medline: 36864108]
48. Rival C, Olivier I, Ceyte H. Effects of temporal and/or spatial instructions on the speed-accuracy trade-off of pointing movements in children. Neurosci Lett 2003;336(1):65-69. [doi: 10.1016/s0304-3940(02)01246-6] [Medline: 12493603]
49. Plamondon R, Alimi AM. Speed/accuracy trade-offs in target-directed movements. Behav Brain Sci 1997;20(2):279-303; discussion 303. [doi: 10.1017/s0140525x97001441] [Medline: 10096999]
50. Robertson SD. Development of bimanual skill: the search for stable patterns of coordination. J Mot Behav 2001;33(2):114-126. [doi: 10.1080/00222890109603144] [Medline: 11404208]
51. Gerloff C, Andres FG. Bimanual coordination and interhemispheric interaction. Acta Psychol (Amst) 2002;110(2-3):161-186. [doi: 10.1016/s0001-6918(02)00032-x] [Medline: 12102104]
52. Rivera D, García A, Ortega JE, Alarcos B, van der Meulen K, Velasco JR, et al. Intraindividual variability measurement of fine manual motor skills in children using an electronic pegboard: cohort study. JMIR mHealth uHealth 2019;7(8):e12434 [FREE Full text] [doi: 10.2196/12434] [Medline: 31464193]
53. Brons A, de Schipper A, Mironcika S, Toussaint H, Schouten B, Bakkes S, et al. Assessing children's fine motor skills with sensor-augmented toys: machine learning approach. J Med Internet Res 2021;23(4):e24237 [FREE Full text] [doi: 10.2196/24237] [Medline: 33885371]
54. Wiesendanger M, Serrien D. The quest to understand bimanual coordination. Prog Brain Res 2004;143:491-505. [doi: 10.1016/S0079-6123(03)43046-X] [Medline: 14653191]
55. Mollà-Casanova S, Llorens R, Borrego A, Salinas-Martínez B, Serra-Añó P. Validity, reliability, and sensitivity to motor impairment severity of a multi-touch app designed to assess hand mobility, coordination, and function after stroke. J Neuroeng Rehabil 2021;18(1):70 [FREE Full text] [doi: 10.1186/s12984-021-00865-9] [Medline: 33892763]
56. Mia MR, Ahamed SI, Fial A, Nemanich S. A scoping review on mobile health technology for assessment and intervention of upper limb motor function in children with motor impairments. Games Health J 2024;13(3):135-148. [doi: 10.1089/g4h.2023.0224] [Medline: 38700552]
57. Findlater L, Zhang L. Input accessibility: a large dataset and summary analysis of age, motor ability and input performance. : ACM; 2020 Presented at: Proceedings of the 22nd International ACM SIGACCESS Conference on Computers and Accessibility; October 26-28, 2020; Virtual Event Greece p. 1-6. [doi: 10.1145/3373625.3417031]
58. MR M, SI A, S N. Data Analysis: Gamified mHealth System for Evaluating Upper Limb Motor Performance in Children. github.com. 2024. URL: https://github.com/RaihanSabique/MoEvGame-DataAnalysis [WebCite Cache ID https://github.com/RaihanSabique/MoEvGame-DataAnalysis]

Abbreviations

CV: coefficient of variation
IRB: Institutional Review Board
MoEvGame: Gamified mHealth System for Upper Limb Motor Evaluation
mHealth: mobile health
PELT: pruned exact linear time
rbf: radial basis function
RMSE: root mean square error
UL: upper limb

Edited by A Coristine, C Prahm; submitted 01.03.24; peer-reviewed by A Amini, A Cretual; comments to author 19.06.24; revised version received 30.07.24; accepted 16.10.24; published 28.02.25.

Please cite as:
Mia MR, Ahamed SI, Nemanich S
Gamified mHealth System for Evaluating Upper Limb Motor Performance in Children: Cross-Sectional Feasibility Study
JMIR Serious Games 2025;13:e57802
URL: https://games.jmir.org/2025/1/e57802
doi:10.2196/57802
PMID:40053722

©Md Raihan Mia, Sheikh Iqbal Ahamed, Samuel Nemanich. Originally published in JMIR Serious Games (https://games.jmir.org), 28.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium,

XSL•FO
RenderX

provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

A Novel Gamified Exercise Program Incorporating Stampede Training for Enhancing Functional Fitness, Physical Activity Levels, and Quality of Life in Community-Dwelling Older Adults: Randomized Parallel Exploratory Trial

Siao-Fen Huang*, BSc; Ching-Wen Fan*, MSc; Diana Khasna Nisrina*, MSc; Tsai-Wei Lin*, BSc; Wen Ching Huang, PhD

Department of Exercise and Health Science, National Taipei University of Nursing and Health Science, No. 365, Ming-te Rd., Peitou Dist., Taipei, Taiwan

*these authors contributed equally

Corresponding Author:
Wen Ching Huang, PhD
Department of Exercise and Health Science, National Taipei University of Nursing and Health Science, No. 365, Ming-te Rd., Peitou Dist., Taipei, Taiwan

Abstract

Background: With the growing prevalence of aging populations, improving the health and quality of life of older adults has become a critical concern globally. In this context, sports technology presents promising applications. The exergame-based training mat—an electronic exercise technology—integrates gamification with diverse training designs, offering a safe and engaging approach to promoting health and well-being in older adults.

Objective: This study aims to examine the impact of a novel exergame-based mat training program on community-dwelling older adults, particularly evaluating its effectiveness in enhancing physical activity levels, quality of life, and functional fitness. The primary end point was exploratory, focusing on the feasibility and effectiveness of the exergame-based training program on physical fitness, physical activity, and quality of life.

Methods: This randomized parallel-designed study enrolled 32 older adults aged 60-80 years from Taipei City. Participants were randomly assigned to either the experimental or control group. The experimental group underwent a 10-week exergame-based mat training program, consisting of 2 sessions per week (70 minutes per session), using gamified group-based exercise training. The control group maintained their usual daily activities. Pre- and postintervention assessments were conducted using the International Physical Activity Questionnaire (IPAQ), the World Health Organization Quality of Life Brief Version (WHOQOL-BREF), the Senior Fitness Test, and the AFAscan fitness assessment.

Results: The experimental group demonstrated significantly increased overall and high-intensity physical activity levels (P=.04; mean difference [MD] 439, 95% CI 28-914; d=0.72). Quality of life significantly enhanced across the physical (P=.01; r=0.53), psychological (P=.02; r=0.52), and social (P=.02; r=0.50) domains of the WHOQOL-BREF. Furthermore, functional fitness parameters, including upper limb muscular strength (P=.007; MD 5.33, 95% CI 1.59-9.07; d=1.06), lower limb muscular strength (P=.01; MD 4.73, 95% CI 1.15-8.32; d=0.98), core strength (P<.001; MD 13.1, 95% CI 7.90-18.2; d=1.89), lower limb flexibility (P=.008; MD 6.47, 95% CI 1.82-11.1; d=1.04), dynamic balance (P=.03; MD –0.72, 95% CI −1.36 to −0.07; d=0.84), static balance (P=.005; MD 15.1, 95% CI 5.01-25.3; d=1.12), agility (P=.001; MD 32.6, 95% CI 15.6-49.6; d=1.44), and cardiorespiratory endurance (P=.04; MD 9.73, 95% CI 0.37-19.1; d=0.78), showed significant enhancements with the exergame-based mat training program. There were no adverse events observed during the study.

Conclusions: In this exploratory trial, the exergame-based mat training program produced medium-to-large improvements (Cohen d ranging from 0.72 to 1.89) across physical activity, quality of life, and functional fitness domains. Although the precision of the CIs varied, the consistent direction of effects supports a meaningful impact of the intervention. These findings suggest that exergame-based mat training programs may serve as a practical community health promotion strategy, potentially preventing age-related frailty and enhancing independence and well-being among older adults.

Trial Registration: ClinicalTrials.gov NCT06898528; https://clinicaltrials.gov/study/NCT06898528

(JMIR Serious Games 2025;13:e73474) doi:10.2196/73474

KEYWORDS

exergame; functional fitness; sport technology; physical activity; stampede training

XSL•FO
RenderX

Introduction

As global populations age, the increasing proportions of older adults pose various risks and challenges for individuals and families, and for social, economic, and health care systems. The need for preventive health care strategies to reduce, delay, or prevent chronic diseases while maintaining independent living has become important [1]. The age-related loss of muscle mass, strength, and function is influenced by aging, disease, and other lifestyle factors. Physical inactivity, malnutrition, unhealthy habits, and chronic diseases further increase the risk of sarcopenia, leading to decreased survival rates, prolonged hospitalization, falls, fractures, metabolic disorders, and cognitive impairment, finally diminishing the quality of life [2]. In addition, aging is associated with sensory decline and cognitive deterioration, severely influencing independence, self-care abilities, and overall well-being among older adults. However, increased physical activity and structured exercise interventions can mitigate the adverse effects of aging, thereby improving the quality of life [3].

Regular exercise interventions enhance physical fitness in older adults, reduce fall risks, improve mobility, and prevent certain conditions, including cardiovascular diseases, cognitive impairment, osteoporosis, and muscle weakness, all of which enhance the quality of life [4]. Moreover, age has a significant impact on the quality of life among older adults, whereas engagement in physical activity contributes to enhancing their quality of life, thereby promoting healthier lifestyles and greater personal independence [5]. Compared with other types of exercise-based interventions, previous studies have indicated that resistance-based interventions are the most effective for enhancing physical activity and strength; however, age and sex may act as confounding factors influencing these outcomes [6]. Therefore, in the context of global aging, maintaining health-related physical fitness is essential for delaying muscle loss, preventing falls and injuries, and mitigating chronic diseases, all of which improve the quality of life. Exercise programs, consequently, play an indispensable role in promoting healthy aging.

With rapid technological advancements, active video games and exergames provide older adults with diverse training options beyond traditional fitness programs. Wii Fit (Nintendo) exercise games enhance balance functions in community-dwelling older adults while improving their self-perceived confidence, with high acceptance rates among older populations [7]. Based on the experiences and needs of older adults, exergames are perceived as enjoyable and engaging. In addition to promoting physical activity, they can help alleviate feelings of loneliness and are not constrained by specific usage conditions. The development of well-designed exergames may therefore provide valuable support for this population [8]. Compared with traditional exercise methods, Kinect-based (Microsoft Corporation) exergames have demonstrated superior efficacy in improving executive function, cognitive performance, and functional fitness [9]. In addition, exergames enhance motivation for physical activity, enabling older adults to experience exercise benefits engagingly and enjoyably [10]. Compared with conventional exercise training, exergames not only improve physical fitness but also provide cognitive and psychological benefits. Therefore, exergames are a promising strategy for health promotion and wellness among the aging population.

Recent meta-analyses have consistently reported that exergame interventions in older adults produce small-to-moderate improvements in balance, mobility, and engagement [11,12]. A more recent systematic review further highlighted that home-based exergames are generally well accepted and improve mobility-related outcomes such as the Timed Up and Go test [13]. These findings underscore the potential of gamified training to enhance functional fitness; however, most previous protocols have focused on platforms such as Wii or Kinect, often emphasizing balance or general mobility with limited attention to agility. In contrast, the present exploratory trial used a novel “stampede” agility-focused exergame delivered twice weekly via the mat exercise, integrating multidirectional stepping, reaction-based tasks, and functional movements. The specified exergames, such as Wii and Kinect, can be integrated with motion-sensing technology and have demonstrated applications beyond entertainment, extending to therapy, rehabilitation, and health promotion. However, as both Wii and Kinect have been discontinued, their long-term applicability is limited, underscoring the need for developing updated platforms tailored for diverse populations. However, assessing movement quality remains a critical concern in implementing exergames, because inappropriate compensatory movement patterns may impact both effectiveness and safety [14].

Therefore, this study seeks to introduce an exercise mat training program led by a certified coach and to investigate its effectiveness within a 10-week comprehensive training regimen designed within the exergame framework. This protocol design addresses agility and dynamic balance more directly than the majority of exergame interventions previously synthesized in meta-analyses, thereby contributing new evidence on the scope and feasibility of agility-oriented exergaming for older adults. We aimed to evaluate functional fitness in older adults using both the standard Senior Fitness Test and the sport technology–based AFAscan fitness assessment. Furthermore, we intended to explore overall physical activity levels and quality of life. We hypothesized that a well-designed exergame program incorporating novel exercise mat training may enhance not only physical fitness but also physical activity and quality of life.

Methods

Experimental Design

In this exploratory trial, we recruited older adults aged 60-80 years from the University Social Responsibility program at National Taipei University of Nursing and Health Sciences University (Taipei City), and used a randomized parallel design. A total of 32 participants were randomly assigned to either the experimental (n=16) or control group (n=16) through simple randomization using a random number table, with an allocation ratio of 1:1. The personnel (research assistant) responsible for enrolling participants and those (principal investigator) managing the random allocation sequence were distinct, ensuring that the allocation process remained blinded and

XSL•FO
RenderX

minimizing the potential for allocation bias. A pre- and postintervention design was implemented. The intervention consisted of a 70-minute exergame-based mat training session conducted twice weekly for 10 consecutive weeks. Each session included a 10-minute warm-up, approximately 45 minutes of primary training, and a 15-minute cooldown and stretching routine (Table 1). Participants completed the International Physical Activity Questionnaire (IPAQ) and the World Health Organization Quality of Life questionnaire (WHOQOL-BREF, Taiwanese Version) before and after the intervention. In addition, the Senior Functional Test and AFAscan fitness assessment (Scanleader, Scanleader Technologies Co. Ltd.) were conducted. Participants in the control group were instructed to maintain their usual daily activities without any additional structured exercise intervention during the study period. A no-intervention control condition was selected to provide a natural reference for functional fitness and quality-of-life changes without training. This approach allowed us to better isolate the effects of the exergame-based mat training and is consistent with previous exercise trials involving older adults.

Table . Training schedule, objectives, and equipment of the exergame-based mat program.

Week	Warm-up	Main course (training objectives)	Cooldown	Indicator light mode[a]	Equipment
1	Stretching exercise and built-in program[b]	Flexibility, agility, upper and lower limb strength, core muscles, cardio	Stretching	All light on, circle light on, 4 corners	Medicine ball
2	Stretching exercise and built-in program[b]	Agility, balance, lower limb strength	Stretching	All light on, circle light on, 4 corners	Medicine ball
3	Stretching exercise and built-in program[b]	Agility, cardio, upper and lower limb strength, flexibility	Stretching	All light on, circle light on, 4 corners	Medicine ball
4	Stretching exercise and built-in program[b]	Lower limb strength, cardio, flexibility	Stretching	All light on, circle light on, 4 corners, 2 lines, diamond	Elastic band, medicine ball
5	Stretching exercise and built-in program[b]	Upper and lower limb strength, balance, core muscles	Stretching	All light on, circle light on, 4 corners, 2 lines, diamond	Elastic band, medicine ball, kettlebell
6	Stretching exercise and built-in program[b]	Core muscles, upper and lower limb strength, agility	Stretching	All light on, circle light on, 4 corners, 2 lines, diamond	Elastic band, medicine ball, kettlebell
7	Stretching exercise and built-in program[b]	Core muscles, upper and lower limb strength, cardio	Stretching	All light on, circle light on, 4 corners, 2 lines, diamond	Elastic band, medicine ball, kettlebell
8	Stretching exercise and built-in program[b]	Upper and lower limb strength, core muscles, agility	Stretching	All light on, circle light on, 4 corners, 2 lines, diamond	Elastic band, medicine ball, kettlebell
9	Stretching exercise and built-in program[b]	Upper and lower limb strength, core muscles, balance	Stretching	All light on, circle light on, 4 corners, 2 lines, diamond	Elastic band, medicine ball, kettlebell, balance cushion
10	Stretching exercise and built-in program[b]	Upper and lower limb strength, lower limb explosive strength, balance, core muscles	Stretching	All light on, circle light on, 4 corners, 2 lines, diamond	Elastic band, medicine ball, kettlebell, balance cushion

[a]The lighting modes of the mat serve as visual cues, guiding participants to step on specific areas to turn off the lights as part of the exercise protocol.
[b]The built-in program comprises five regular exercise games (lights out game, jumping jacks, reactions, two-way jumping jacks, and whack-a-mole).

A randomized parallel-controlled clinical trial was conducted. A total of 32 eligible participants were randomly assigned to 2 groups: control group and experimental group. The participants completed questionnaires (the IPAQ and WHOQOL-BREF), along with physical fitness assessments (Senior Fitness Test and AFAscan Fitness Assessment), before and after the intervention. Finally, 30 participants completed all experimental procedures required for data analysis.

Exergame-Based Mat Training Program

The mat-based exercise system (Stampede, SHENNONA Co., Ltd) facilitates exercise training by using electronic light and sound effects to guide participants through various exercises, with intensity progressively increasing from moderate to high levels. This system facilitates customized movement programming tailored to specific training objectives, and its architectural diagram is presented in Figure 1A. The training program is divided into 2 phases: the first phase focuses on fundamental movements, whereas the second phase introduces

XSL•FO
RenderX

more advanced exercises. With progress, additional exercise equipment is incorporated, along with increased sets, duration, movement variations, and training intensity. Training sessions can be conducted in a simple progressive manner or a game-competitive format (Figure 1B). Throughout the training, participants wore a heart rate monitor (ALATHEC Obeat1) that recorded their exercise intensity and heart rate fluctuations (Figure 2B). Exercise intensity was indicated by lighting colors, corresponding to specific percentages of the maximum heart rate (HRmax) as follows: blue (0%-50% HRmax), cyan (50%-60% HRmax), green (60%-70% HRmax), yellow (70%-80% HRmax), orange (80%-90% HRmax), and red (>90% HRmax). However, the exercise intensity and frequency were adjusted based on participants' adaptation. This study encompassed 5 different lighting modes, namely all light on, circle on, 4 corners, 2 lines, and diamond (Figure 2A). Each mode was paired with different training protocols and movement patterns, supplemented by training equipment, such as medicine balls, resistance bands, kettlebells, and balance pads. This combination enhanced engagement while simultaneously promoting functional fitness, including muscular strength, muscular endurance, aerobic capacity, flexibility, agility, balance, and core strength. Detailed training protocols are provided in Multimedia Appendix 1. All participants achieved a 100% attendance rate, and adverse events were observed by qualified specialists. All training sessions were supervised and instructed by certified fitness instructors.

Figure 1. Stampede exercise mat system. (A) A central router facilitates data transmission among a cloud server, user interface devices (such as smartphones), and an analysis unit. The system highlights seamless data flow, emphasizing connectivity, data processing, and user interaction. (B) Stampede exercise mats use lighting and sound effects to guide participants through exergame implementation in the National Taipei University of Nursing and Health Sciences I203 laboratory. In addition, the system can be integrated with other exercise equipment to enhance fitness benefits while incorporating elements of enjoyment and gamification.

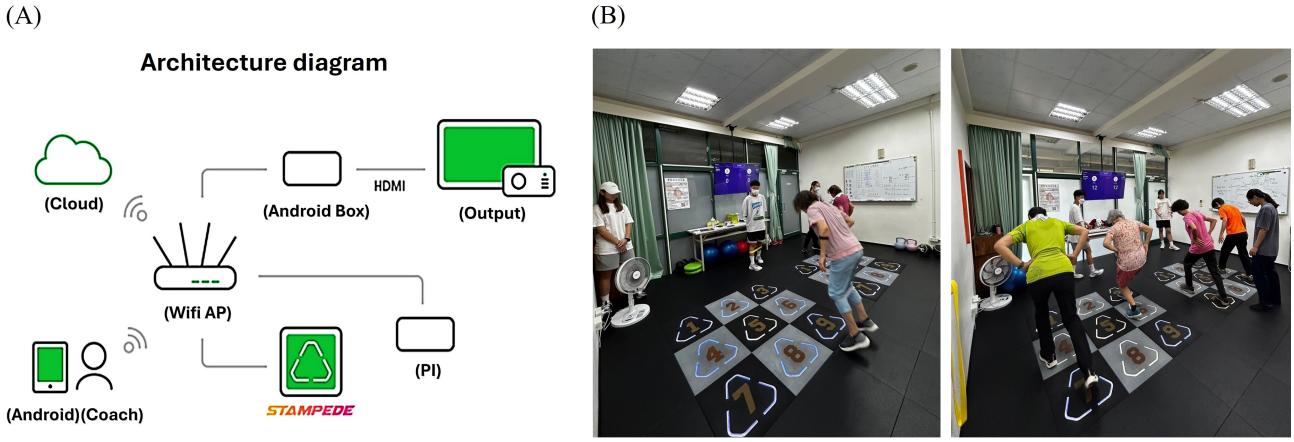
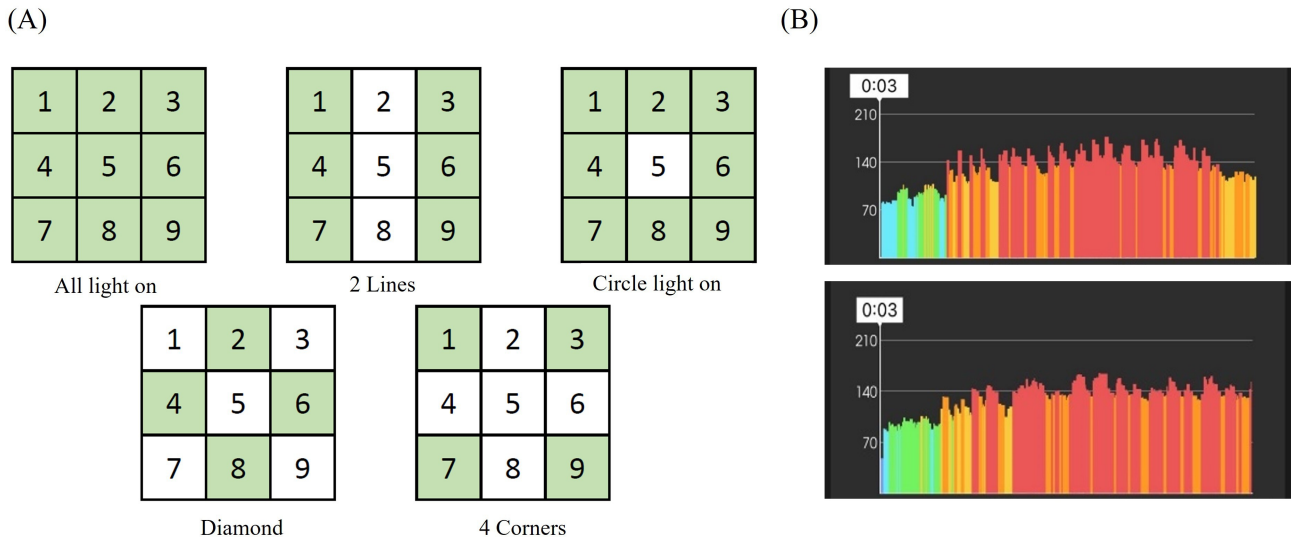


Figure 2. Exergame-based mat training modes and intensity monitoring. (A) This study used 5 lighting modes (all lights on, circle on, 4 corners, 2 lines, and diamond) for designated training protocols or in combination with other exercise equipment. (B) Heart rate was continuously monitored throughout the program.



Body Composition

InBody 270 (Biospace) was used as the body composition analyzer. It is a noninvasive bioelectrical impedance analysis device, which estimates body composition by transmitting a low-level electrical current through the body and measuring resistance. It provides accurate assessments of BMI, body fat percentage, obesity analysis, skeletal muscle mass, and visceral fat levels, serving as a comprehensive tool for evaluating body

XSL•FO
RenderX

composition. The detailed procedures have been described previously [15].

Senior Fitness Test

In older adults, various components of functional fitness were evaluated using standardized tests. Aerobic endurance was assessed through the 2-minute step test, whereas upper and lower limb muscular endurance were measured using the arm curl test and chair stand test, respectively. Lower limb flexibility was evaluated with the sit-and-reach test, and shoulder flexibility was assessed using the back-scratch test. In addition, static balance was measured using the one-leg standing test, and dynamic balance was evaluated using the 8-foot up-and-go test. Detailed procedures, devices, and manipulation have been described previously [16]. These senior functional tests are cost-effective and convenient for assessing functional fitness in older adults. They are suitable for independently living individuals aged 60-90 years and above, encompassing numerous fitness levels, ranging from frailty to high fitness [17]. All certified assessors, independent from the research team, conducted the Senior Fitness Assessments and AFAscan fitness assessment in a single-blind manner.

AFAscan Fitness Assessment

The AFAscan leader facility was used to assess various components of physical fitness, including lower body explosiveness, core strength, flexibility, agility, balance, and endurance. It used gyroscopes (positioned on the chest and both thighs) and a force platform to precisely and objectively measure all physical performance parameters.

Lower Body Explosiveness

Lower body explosiveness was evaluated using a countermovement jump with an arm swing. Participants jumped to the highest extent while attempting to land in the same position on the force mat from which they took off. During the airborne phase, they maintained full extension of their hip, knee, and ankle joints to prevent any additional flight time caused by leg bending. Flight time was recorded as the primary measurement reference.

Core Strength

Core strength was measured using the bent-knee sit-up exercise. Participants lay on their backs with their knees flexed and feet positioned 12 inches from their buttocks. To ensure stability, a partner could assist by holding their feet in place. Participants conducted curl-ups, raising their upper bodies to at least 35°, and completed the fullest possible repetitions within 60 seconds.

Flexibility

Flexibility was assessed using the standing toe touch exercise. Participants stood with their feet together, toes pointing forward on the force mat. They bent forward at the hips, attempting to touch their fingertips to that of their toes without bending their knees. They held the position for 3 seconds to allow for angle measurement, which served as the flexibility assessment outcome.

Agility

Agility was assessed using the quick feet test. Participants alternately raised their feet in rapid succession, taking the shortest possible steps. They tried completing the maximum number of repetitions within 15 seconds.

Balance

Balance was assessed using the single-leg stand test. Participants raised their preferred foot and placed it against the inner thigh of the standing leg. They maintained a steady center of gravity within the center of a designated circle. A higher score was achieved by remaining stable within the center for a longer duration, with a maximum assessment time of 30 seconds.

Endurance

Endurance was assessed using the high-knee test. Participants drove their right knee toward their chest and quickly returned it to the ground, continuously alternating with the opposite knee. Arm swinging was permitted to aid movement. Participants raised their knees to an indicated angle >50° and completed the maximum number of repetitions within 30 seconds.

Questionnaires

Physical activity levels were measured using the IPAQ, which evaluates walking, moderate-intensity, and vigorous-intensity activities over the past 7 days. The IPAQ has shown good test-retest reliability (Spearman ρ=0.76-0.94) and acceptable internal consistency (Cronbach α=0.80), with evidence of concurrent validity when compared against accelerometer data [18,19]. The Taiwanese version of the IPAQ was used, specifically the long-form Taiwan Physical Activity Survey Questionnaire, which consists of 26 items encompassing four domains: work, household activities, transportation, and leisure activities [20].

In addition, the Taiwanese version of the WHOQOL-BREF—a short version of the World Health Organization Quality of Life Assessment questionnaire—was used. This version consists of 28 items, encompassing 4 domains: physical health, psychological well-being, social relationships, and environmental factors. The WHOQOL-BREF has demonstrated acceptable internal consistency, with Cronbach α ranging from 0.66 to 0.84 across domains, and satisfactory construct validity when tested in different populations [21,22]. This shortened and adapted version serves as a valid alternative to the long-form WHOQOL questionnaire used in Taiwan [22].

Eligibility Criteria

The inclusion criteria were as follows: (1) older adults (aged 60 - 80 years) with no musculoskeletal injuries in the past month and capable of participating in training activities, (2) completion of the Physical Activity Readiness Questionnaire (PAR-Q) and eligibility for moderate-intensity exercise intervention, and (3) ability to adhere to the exercise frequency required by the study and independently commute to the training site. The exclusion criteria were as follows: (1) individuals with musculoskeletal disorders, joint diseases, unstable cardiovascular conditions, or cognitive impairments, as well as those with psychiatric disorders based on questionnaire evaluation, and (2) individuals whose PAR-Q responses indicate potential risks or concerns.

XSL•FO
RenderX

Statistical Analyses

Descriptive statistics and chi-square tests were used to analyze the anthropometric and sociodemographic characteristics of all participants. The normality of all dependent variables was evaluated using Kolmogorov-Smirnov tests to ensure the appropriate selection of parametric and nonparametric analyses. To examine significant between- and within-group differences, nonparametric methods (Wilcoxon signed-rank test and Mann-Whitney *U* test) and parametric methods (mixed 2-way analysis of variance, unpaired 2-tailed *t* tests, and paired *t* tests) were applied. Two participants withdrew before completing posttest assessments and thus provided only baseline data. Because no postintervention outcome data were available for these individuals, no imputation was applied, and analyses were performed using complete cases. Consequently, a total of 15 participants in the experimental group and 15 participants in the control group were included in the final statistical analysis. All statistical analyses were performed using SPSS software (version 22; IBM Corp), with statistical significance set at $P<.05$. Effect sizes (*d*) were calculated as Cohen *d* and interpreted according to conventional benchmarks: small (0.2), medium (0.5), and large (0.8) [23]. The 95% CIs were reported alongside all effect size estimates to enhance interpretability.

Ethical Considerations

This study was reviewed and approved by the Fu Jen Catholic University Institutional Review Board (IRB # C112161) per ethical guidelines for human research. Before participation, all participants were informed about the study procedures and precautions, and written informed consent was obtained. All acquired data will be analyzed and disseminated in deidentified format to ensure participant confidentiality. Study participants will receive individualized physical fitness assessment reports and monetary compensation consisting of US $32.77 (exchange rate: US $1=30.5 TWD) for transportation expenses. To ensure participant safety, all training sessions were supervised by qualified specialists, who continuously monitored for potential and unexpected adverse events. Any adverse event would have been documented immediately, managed with appropriate first aid, and referred for medical evaluation if necessary.

Results

Anthropometric Characteristics of Participants

The experimental procedure is illustrated in Figure 3. This study recruited 32 participants, with 16 allocated to each of the experimental and control groups. However, one participant from each group withdrew from the study because of personal reasons. Finally, 15 participants in both groups were analyzed for primary outcomes. The participants in the experimental group were paired and engaged in the training activities through alternating and cooperative interactions. Heart rate monitoring during the training indicated fluctuations characteristic of high-intensity interval training (Figure 2B). In addition, no significant differences were observed in sex, age, height, weight, systolic, diastolic pressure, and resting heart rate (t_{28}=0.065 - 1.121; all $P>.05$) before the intervention (Table 2).

XSL•FO
RenderX

Figure 3. Study flow diagram. PAR-Q: Physical Activity Readiness Questionnaire.

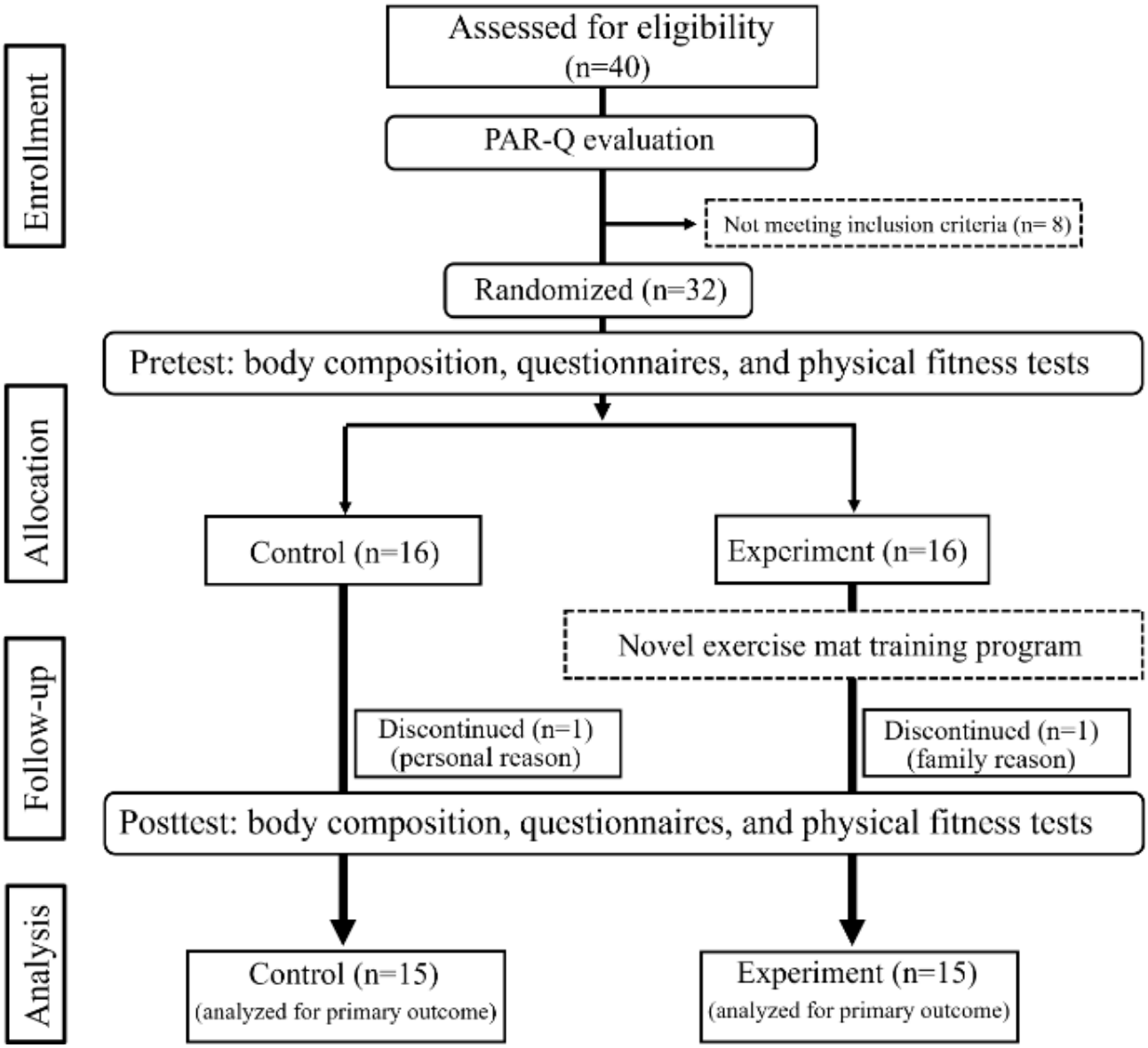


Table . Basal anthropometric characteristics.

Characteristics	Control (n=15)	Experiment (n=15)	Mean difference (95% CI)	Effect size (Cohen *d*)
Sex (male/female), n	2/13	2/13	—[a]	—
Age (years), mean (SD)	67.7 (4.4)	67.8 (3.9)	0.13 (–2.96 to 3.22)	0.03
Height (cm), mean (SD)	157.7 (3.5)	157.8 (5.7)	0.11 (–3.43 to 3.65)	0.02
Weight (kg), mean (SD)	50.2 (9.4)	55.4 (13.8)	5.26 (–3.63 to 14.2)	0.44
Systolic pressure (mm Hg), mean (SD)	121.6 (17)	127.8 (15)	6.13 (–6.17 to 18.4)	0.37
Diastolic pressure (mm Hg), mean (SD)	75.4 (10)	77.0 (9.2)	1.60 (–5.47 to 8.67)	0.17
Resting heart rate (bpm), mean (SD)	81.6 (11)	85.4 (12)	3.87 (–4.83 to 12.6)	0.33

[a]Not applicable.

XSL•FO
RenderX

Effects of the Exergame-Based Mat Program on Physical Activity

Pre- and postintervention results of the Physical Activity Scale indicated a significant main effect of time on total physical activity levels (Table 3). After the intervention, activity levels significantly increased in the experimental group, with significant within-group differences. Therefore, the mat exercise program effectively enhanced overall physical activity in older adults. For vigorous physical activity, the main effect of time ($F_{1,28}$=13.9; P=.001) and interaction effect ($F_{1,28}$=11.9; P=.002) were observed. The experimental group demonstrated a significant within-group increase in total metabolic equivalents (MET) (t_{14}=2.88; P=.01; mean difference [MD] 1266, 95% CI 325-2206; d=0.74) and vigorous activity (t_{14}=7.37; P<.001; MD 950, 95% CI 674-1227; d=1.90), indicating that the intervention effectively promoted physical activity. In addition, after the intervention, the experimental group demonstrated significantly higher vigorous activity levels than the control group (t_{28}=1.97; P=.04; MD 439, 95% CI 28-914; d=0.72), supporting the effectiveness of the exergame-based mat program. However, no significant main effects or interaction effects were observed for moderate-intensity activity, walking, or sedentary behavior ($F_{1,28}$=0.583 - 1.889; P>.05). Thus, the intervention exerted a relatively lower impact on moderate- and low-intensity physical activities.

Table . Effects of the exergame-based mat training program on physical activity levels in older adults: effect sizes and 95% CIs.

Physical activity intensity	Control (MET[a]-min/week[b]), mean (SD)		Experiment (MET-min/week[b]), mean (SD)		Mean difference (95% CI)	Effect size, Cohen *d*
	Pre	Post	Pre	Post		
Total MET	2866 (1540)	3328 (2045)	2052 (1650)	3318 (1440)[c]	–9.67 (–1332 to 1313)	0.01
Vigorous intensity	789 (692)[d]	828 (762)	317 (417)	1267 (406)[c,d]	439 (28 to 914)	0.72
Moderate intensity	752 (945)	1252 (1029)	560 (782)	757 (620)	–495 (–1131 to 141)	0.58
Walking	1153 (862)	1076 (715)	1112 (817)	1164 (732)	85.8 (–456 to 627)	0.12
Sitting (work)	297 (139)	211 (119)	206 (158)	230 (182)	19.3 (–96 to 134)	0.13
Sitting (holiday)	277 (134)	239 (111)	278 (157)	282 (144)	43.3 (−52 to 140)	0.34

[a]MET: metabolic equivalents.

[b]MET·min·wk ¹

[c]Indicates significant within-group differences.

[d]Indicates significant between-group differences.

Effects of Exergame-Based Mat Program on Quality of Life

The WHOQOL-BREF assessment evaluated overall health and health satisfaction, along with four domains: physical health, psychological health, social relationships, and environment (Table 4). Concerning physical health, the experimental group demonstrated significant differences both within (P=.02) and between (P=.01; Hodges-Lehmann median difference=1.2; 95% CI 0.6-2.3; r=0.53) groups. Specifically, significant within-group improvements were observed in pain and discomfort (P=.02), dependence on medical aids (P=.002), mobility (P=.008), and daily activities (P=.002). In addition, daily activities (P=.002) and work capacity (P=.006) showed a significant between-group difference in the experimental group, compared with the control group. Concerning psychological health, the experimental group exhibited a significantly higher overall score than the control group (P=.02; Hodges-Lehmann median difference=2.7, 95% CI 1.3-4; r=0.52). Significant between-group differences were observed in self-esteem (P=.02) and negative feelings (P=.001), with negative feelings showing significant differences within (P=.02).

XSL•FO
RenderX

Table . Effects of the exergame-based mat training program on quality of life in older adults with effect sizes and 95% CIs.

Dimensions	Control, mean (SD)		Experiment, mean (SD)		Hodges-Lehmann, median difference (95% CI)	Effect size (*r*)
	Pre	Post	Pre	Post		
Overall health	3.4 (0.6)	3.3 (0.5)	3.6 (0.5)	4 (0.5)[a]	1 (0 to 2)	0.64
Health satisfaction	3.3 (0.7)	3.0 (0.6)	3.4 (0.6)	3.7 (0.5)[a]	1 (0 to 2)	0.65
Domains						
Physical health[b]	13.3 (1.3)	14.3 (1.8)	13.5 (0.9)	16.1 (1.6)[a,c]	1.2 (0.6 to 2.3)	0.53
Pain and discomfort	3.3 (0.5)	3.5 (0.9)	3.0 (0.5)	4.1 (0.8)[c]	1 (0 to 2)	0.35
Dependence on medical substance and medical aids	3.3 (0.9)	3.3 (0.8)	3.3 (0.5)	4.1 (0.6)[c]	0 (–1 to 1)	0.17
Energy and fatigue	3.5 (0.5)	3.3 (0.6)	3.4 (0.5)	3.6 (0.5)	0 (1 to 1)	0.29
Mobility	3.5 (0.6)	3.9 (0.5)	3.5 (0.5)	4.2 (0.7)[c]	0 (–1 to 1)	0.28
Sleep and rest	3.2 (1.1)	3.1 (0.9)	3.5 (0.7)	3.3 (0.8)	0 (–1 to 1)	0.07
Activities and daily living	3.3 (0.6)	3.5 (0.6)	3.3 (0.5)	4.3 (0.5)[a,c]	1 (0 to 2)	0.56
Work capacity	3.9 (0.5)	3.5 (0.5)	4.1 (0.7)	4.1 (0.5)[a]	1 (0 to 2)	0.50
Psychological health[b]	13.8 (1.9)	13.4 (2.0)	14.3 (1.7)	15.3 (1.6)[a]	2.7 (1.3 to 4)	0.52
Positive feeling	3.5(0.6)	3.3 (0.6)	3.5 (0.6)	3.7 (0.7)	1 (0 to 2)	0.34
Spirituality, religion, and personal beliefs	3.7 (0.7)	3.7 (0.6)	3.7 (0.9)	3.8 (0.7)	0 (–1 to 1)	0.14
Thinking, learning, memory, and concentration	3.5 (0.7)	3.1 (0.7)	3.8 (0.7)	3.5 (0.5)	0 (–1 to 1)	0.30
Bodily image and appearance	3.7 (0.6)	3.5 (0.6)	3.9 (0.7)	4 (0.8)	1 (0 to 2)	0.31
Self-esteem	3.9 (0.6)	3.5 (0.6)	3.7 (0.6)	4.1 (0.5)[a]	1 (0 to 2)	0.42
Negative feelings	3.3 (0.6)	3.1 (0.7)	3.4 (0.9)	3.9 (0.5)[a,c]	1 (0 to 2)	0.64
Social relationship[b]	13.4 (2.1)	12.8 (1.9)	14.4 (2.6)	14.5 (1.5)[a]	2 (1 to 3)	0.50
Personal relationship	3.9 (0.6)	3.2 (0.6)[c]	3.9 (0.9)	3.7 (0.6)	1 (0 to 2)	0.38
Sexual activity	3.2 (1.1)	3.1 (0.8)	3.4 (0.7)	3.2 (1.1)	0 (–1 to 1)	0.23
Social support	3.7 (0.6)	3.3 (0.6)[c]	3.6 (0.7)	3.9 (0.5)[a]	1 (0 to 2)	0.56
Being respected	3.7 (0.6)	3.3 (0.5)[c]	3.8 (1.1)	3.7 (0.7)	1 (0 to 2)	0.31
Environment[b]	15.1 (1.7)	14.6 (1.8)	15.7 (1.7)	15.8 (1.4)	1.8 (0.9 to 2.7)	0.39
Freedom, physical safety, and security	3.7 (0.8)	3.6 (0.6)	4.1 (0.6)	4 (0.5)	0 (–1 to 1)	0.30

XSL•FO
RenderX

Dimensions	Control, mean (SD)		Experiment, mean (SD)		Hodges-Lehmann, median difference (95% CI)	Effect size (*r*)
	Pre	Post	Pre	Post		
Physical environment	3.6 (0.8)	3.4 (0.5)	3.6 (1.0)	3.4 (0.6)	0 (–1 to 1)	0.06
Financial resource	3.8 (0.7)	3.6 (0.7)	4.2 (0.7)	3.8 (0.9)	1 (0 to 2)	0.19
Opportunities for acquiring new information and skills	3.9 (0.6)	3.9 (0.6)	4.0 (0.5)	4.2 (0.6)	1 (0 to 2)	0.26
Participation and opportunities for recreation/leisure activities	3.8 (0.7)	3.7 (0.5)	4.2 (0.8)	3.8 (0.7)	1 (0 to 2)	0.15
Home environment	3.7 (1.1)	3.8 (0.6)	3.9 (0.7)	3.9 (0.5)	1 (0 to 2)	0.44
Health and social care: accessibility and quality	4.3 (0.7)	4.3 (0.5)	4.3 (0.5)	4.2 (0.5)	1 (0 to 2)	0.40
Transport	4.3 (0.6)	4.3 (0.5)	4.3 (0.5)	4.3 (0.6)	1 (0 to 2)	0.40
Eating	4.1 (0.6)	3.8 (0.7)	3.9 (0.6)	4.0 (0.8)	1 (0 to 2)	0.15

[a]Indicates significant between-group differences.

[b]Raw domain scores for the World Health Organization Quality of Life Brief Version are transformed to a 4 to 20 score according to guidelines (Total score of the participant in a specific domain×4) divided by (Total number of items in the domain).

[c]Indicates significant within-group differences.

Concerning social relationships, the experimental group demonstrated significantly higher overall scores than the control group (P=.02; Hodges-Lehmann median difference 2, 95% CI 1-3; r=0.50). Furthermore, the experimental group exhibited significantly higher scores in social support than the control group (P=.003). In contrast, regarding the environment, no significant within- or between-group differences were observed in overall scores or individual items, indicating that the intervention did not substantially affect this aspect of quality of life.

Effects of Exergame-Based Mat Program on Senior Fitness

Table 5 summarizes the results for systolic blood pressure, diastolic blood pressure, and heart rate. No significant interaction effects were observed between group and time ($F_{1,28}$=0.583 - 1.889; all P>.05). In addition, no significant main effect of group was observed ($F_{1,28}$=1.505 - 2.181; all P>.05), except for a significant main effect of time on diastolic blood pressure ($F_{1,28}$=4.350; P=.04). In the control group, systolic blood pressure and heart rate did not change, whereas diastolic blood pressure significantly reduced after the intervention. In the experimental group, systolic blood pressure, diastolic blood pressure, and heart rate showed no significant between- or within-group differences.

XSL•FO
RenderX

Table . Effects of the exergame-based mat training program on Senior Fitness Test outcomes in older adults with effect sizes and 95% CIs.

Variables	Control, mean (SD)			Experiment, mean (SD)			Mean difference (95% CI)	Effect size (Cohen *d*)
	Pre	Post	Gain score	Pre	Post	Gain score		
Systolic pressure (mm Hg)	121.6 (17)	116.6 (17)	–5.1 (15)	127.8 (15)	126.2 (16)	–1.6 (8.1)	9.60 (–2.83 to 22)	0.58
Diastolic pressure (mm Hg)	75.4 (10)	70.6 (8.4)[a]	–4.9 (6.7)	77 (9.2)	76.1 (7.6)	–1 (8.5)	5.47 (–0.54 to 11.5)	0.68
Resting heart rate (bpm)	81.6 (11)	78.1 (6.3)	–3.5 (11)	85.4 (12)	85.1 (13.5)	–0.4 (7.6)	7 (–0.89 to 14.9)	0.66
Arm curl test (times)	17.7 (5)	17.4 (5.3)	–0.3 (2.9)	17.7 (2.9)	22.7 (4.6)[a,b]	5.1 (3.9)[b]	5.33 (1.59 to 9.07)	1.06
Chair stand test (times)	23.7 (5.7)	22 (4.7)	–1.6 (4.1)	23.6 (4.5)	26.8 (4.8)[a,b]	3.2 (3.8)[b]	4.73 (1.15 to 8.32)	0.98
Chair sit-and-reach (cm)	7.8 (4.4)	5.8 (6.8)	–2 (4.1)	7.9 (6)	12.3 (5.5)[a,b]	4.4 (3.6)[b]	6.47 (1.82 to 11.1)	1.04
Back-scratch test (cm)	6.6 (4.2)	6.6 (5.1)	–0.1 (2.6)	5.8 (3.1)	6.4 (4.2)	0.5 (3.5)	–0.18 (–3.67 to 3.31)	0.04
One-leg standing test (s)	27.8 (5.5)	26.7 (7.2)	–1.1 (3.7)	26 (7.3)	28.7 (5.2)	2.7 (5.2)	2 (–2.67 to 6.67)	0.32
8-foot up-and-go test (s)	4.3 (0.6)	4.4 (0.6)	0.2 (0.3)	4.3 (0.5)	3.7 (1.1)[b]	–0.6 (1.1)[b]	–0.72 (–1.36 to –0.07)	0.84
2-minute step test (times)	109 (11)	109 (12)	– 0.3 (7.7)	107 (8.8)	119 (13)[a,b]	12 (10)[b]	9.73 (0.37 to 19.1)	0.78

[a]Indicates significant within-group differences.

[b]Indicates significant between-group differences.

For muscular endurance of both upper and lower limbs, a significant interaction effect was observed between time and group ($F_{1,28}$=18.1, *P*<.001; $F_{1,28}$=10.91, *P*=.003, respectively). Specifically, upper limb muscular endurance in the experimental group was significantly higher in the posttest compared with the pretest (t_{14}=4.92; *P*<.001; MD 5.07, 95% CI 2.80-7.27; *d*=1.27), and significantly higher compared with the control group (t_{28}=2.92; *P*=.007; MD 5.33, 95% CI 1.59-9.07; *d*=1.06). Moreover, the gain score for upper limb endurance was significantly higher in the experimental group than in the control group (t_{28}=4.26; *P*<.001; MD 5.4, 95% CI 2.80-7.99; *d*=1.55). Similarly, lower limb muscular endurance was significantly higher in the experimental group after the intervention compared with before the intervention (t_{14}=3.24; *P*=.006; MD 3.2, 95% CI 1.08-5.31; *d*=1.55) and the control group (postintervention values; t_{28}=2.70; *P*=.01; MD 4.73, 95% CI 1.15-8.32; *d*=0.98). The gain score for lower limb endurance was significantly higher in the experimental group than in the control group (t_{28}=3.31; *P*=.003; MD 4.80, 95% CI 1.83-7.77; *d*=1.21).

No significant interaction effect was observed between time and group for upper limb flexibility ($F_{1,28}$=0.284; *P*=.59). No significant between- or within-group differences were observed for upper limb flexibility after the intervention. However, a significant interaction effect was observed for lower limb flexibility ($F_{1,28}$=20.71; *P*<.001). However, lower limb flexibility in the experimental group was significantly higher after the intervention than before the intervention (t_{14}=4.74; *P*<.001; MD 4.41, 95% CI 2.41-6.40; *d*=1.04). Postintervention values were significantly higher than those in the control group (t_{28}=2.85, *P*=.008, MD 6.47, 95% CI 1.82-11.1; *d*=1.04). In addition, the gain score for lower limb flexibility was significantly higher in the experimental group than in the control group (t_{28}=4.55; *P*<.001; MD 6.45, 95% CI 3.54-9.36; *d*=1.66).

Regarding balance performance, no significant interaction effect was observed for static balance between time and group ($F_{1,28}$=1.308; *P*=.29), with no significant between- or within-group differences. Nonetheless, a significant interaction effect was observed for dynamic balance ($F_{1,28}$=6.646; *P*=.02). Despite no significant within-group difference in the experimental group before and after the intervention, the postintervention time was significantly lower than that in the control group (t_{28}=−2.29; *P*=.03; MD -0.72, 95% CI −1.36 to −.07; *d*=0.84). Furthermore, the gain score for dynamic balance was significantly lower in the experimental group than in the control group (t_{28}=−2.58; *P*=.02; MD −0.77, 95% CI −1.38 to −0.16; *d*=0.94).

For aerobic endurance, a significant interaction effect was observed between time and group ($F_{1,28}$=14.027; *P*=.001). In the experimental group, postintervention aerobic endurance was significantly higher than preintervention endurance (t_{14}=4.58; *P*<.001; MD 12, 95% CI 6.38-17.6; *d*=1.18). Moreover, postintervention values were significantly higher than those in the control group (t_{28}=2.13; *P*=.04; MD 9.73, 95% CI 0.37-19.1; *d*=0.78). In addition, the gain score for aerobic endurance was

XSL•FO
RenderX

significantly higher in the experimental group than in the control group (t_{28}=3.75; P=.001; MD 12.3, 95% CI 5.59-19.1; d=1.66).

Effects of Exergame-Based Mat Program on AFAscan Fitness Assessment

No significant interaction effect was observed for explosiveness, along with no significant between- or within-group differences ($F_{1,28}$=0.169; P=.68) (Table 6). For core strength, a significant interaction effect was observed between time and group ($F_{1,28}$=14.027; P=.001). In the experimental group, postintervention core strength was significantly higher than preintervention core strength (t_{14}=3.77; P=.002; MD 8.14, 95% CI 3.47-12.8; d=1). Moreover, postintervention values were significantly higher than those in the control group (t_{28}=5.18; P<.001; MD 13.1, 95% CI 7.90-18.2; d=1.89). In addition, the gain score for core strength was significantly higher in the experimental group than in the control group (t_{28}=2.49; P=.02; MD 6.69, 95% CI 1.15-12.2; d=1.04).

Table . Physical fitness outcomes evaluated by the AFAscan assessment before and after the exergame-based mat training program with effect sizes and 95% CIs.

Physical fitness variables	Control, mean (SD)			Experiment, mean (SD)			Mean difference (95% CI)	Effect size (Cohen *d*)
	Pre	Post	Gain score	Pre	Post	Gain score		
Vertical jump (cm)	15.8 (3)	16.7 (4.1)	0.9 (2.9)	16.9 (4.5)	18.2 (3.5)	1.3 (2.5)	1.53 (–1.32 to 4.39)	0.40
Sit-up (times)	11.9 (7)	13 (5.5)	1.5 (4.1)	18.1 (8.8)	26.3 (8.1)[a,b]	8.1 (8)[a]	13.1 (7.90-18.2)	1.89
Toe-touch test (degree)	125 (16)[a]	112 (18)[b]	–13.4 (13)	113 (12)	125 (17)[a,b]	12.7 (11.2)[a]	13.3 (0.34-26.3)	0.77
Agility (times)	57.5 (28)	61.1 (24)	3.7 (25)	59.5 (29)	93.7 (21)[a,b]	34.1 (28)[a]	32.6 (15.6-49.6)	1.44
Balance test (scores)	37.3 (12)	35.5 (15)	–1.8 (15)	39.6 (12)	50.6 (12)[a,b]	11 (15)[a]	15.1 (5.01-25.3)	1.12
Endurance test (times)	69 (17)	65 (12)	–3.6 (14)	76 (20)	92 (15)[a,b]	17 (15)[a]	27 (16.6-37.4)	1.95

[a]Indicates significant between-group differences.

[b]Indicates significant within-group differences.

Significant interaction effects were observed for flexibility, agility, balance, and endurance ($F_{1,28}$=5.27 - 34.078; all P<.05). Postintervention fitness characteristics in flexibility (t_{28}=2.10; P=.05; MD 13.3, 95% CI 0.34-26.3; d=0.77), agility (t_{28}=3.93; P=.001; MD 32.6, 95% CI 15.6-49.6; d=1.44), balance (t_{28}=3.06; P=.005; MD 15.1, 95% CI 5.01-25.3; d=1.12), and endurance (t_{28}=5.34; P<.001; MD 27, 95% CI 16.6-37.4; d=1.95) were significantly improved in the experimental group, compared with the control group. Moreover, significant within-group differences were observed in the experimental group, with significantly higher gain scores than in the control group.

Effects of Exergame-Based Mat Program on Body Composition

Regarding body composition (Table 7), no significant interaction effects were observed for body weight, BMI, skeletal muscle mass, body fat mass, or body fat percentage ($F_{1,28}$=0.179 - 0.347; all P>.05) between the control and experimental groups. In addition, no significant differences were observed in gain scores between groups. After the intervention, the experimental group exhibited an increasing trend in lower limb muscle mass, whereas the control group exhibited a minor decline in lower limb muscle strength. However, these changes were not statistically significant in terms of gain scores. Therefore, the intervention exerted a limited impact on the overall body composition, segmental muscles, and fat distribution.

XSL•FO
RenderX

Table . Effects of exergame-based mat training on body composition in older adults: effect sizes and 95% CIs.

Body composition variables	Control, mean (SD)			Experiment, mean (SD)			Mean difference (95% CI)	Effect size (Cohen *d*)
	Pre	Post	Gain score	Pre	Post	Gain score		
Body weight (kg)	50.2 (9.4)	50.1 (9.6)	–0.013 (1.3)	55.4 (13.8)	55.4 (13.7)	–0.013 (1.1)	5.62 (–3.59 to 14.1)	0.44
Body mass index	20.1 (3.6)	20.1 (3.7)	0 (0.5)	22.1 (4.1)	22.2 (4)	0.1 (0.4)	2.02 (–0.89 to 4.94)	0.52
Skeleton muscle (kg)	20.1 (2.8)	20.1 (3)	0.04 (0.5)	21.6 (4.7)	21.6 (4.7)	0.05 (0.7)	1.49 (–1.48 to 4.47)	0.37
Body fat mass (kg)	12.7 (6.2)	12.6 (6.2)	–0.12 (1.2)	15.4 (7.5)	15.5 (7.2)	0.14 (1.2)	2.93 (–2.11 to 7.97)	0.44
Body fat percent (%)	24.1 (8.6)	23.8 (8.6)	–0.31 (1.9)	26.7 (7.6)	26.9 (6.8)	0.11 (1.9)	3.10 (–2.73 to 8.93)	0.40
Segmental lean analysis (kg)								
Right arm	1.6 (0.4)	1.5 (0.4)	–0.098 (0.2)	1.9 (0.6)	1.8 (0.7)	–0.063 (0.1)	0.35 (–0.07 to 0.77)	0.62
Left arm	1.6 (0.4)	1.5 (0.4)	–0.117 (0.3)	1.8 (0.6)	1.8 (0.6)	–0.407 (0.1)	0.34 (–0.06 to 0.74)	0.62
Trunk	15.4 (2.7)	14.6 (3.8)	–0.740 (1.9)	17.1 (3.8)	16.9 (3.8)	–0.173 (0.7)	2.25 (–0.59 to 5.08)	0.59
Right leg	5.7 (0.8)	5.4 (1.3)	–0.277 (0.9)	6 (1.3)	6.1 (1.1)	0.072 (0.2)	0.67 (–0.28 to 1.65)	0.53
Left leg	5.7 (0.7)	5.4 (1.2)	–0.312 (0.9)	5.7 (1.8)	6 (1.3)	0.333 (1)	0.66 (–0.27 to 1.58)	0.53
Segmental fat analysis (kg)								
Right arm	0.9 (0.4)	0.8 (0.4)	–0.03 (0.1)	1 (0.6)	1.1 (0.6)	0.033 (0.1)	0.23 (–0.14 to 0.61)	0.47
Left arm	0.9 (0.4)	0.9 (0.4)	–0.04 (0.1)	1.1 (0.6)	1.1 (0.6)	0.013 (0.1)	0.23 (–0.14 to 0.61)	0.47
Trunk	5.7 (3.5)	5.5 (3.3)	–0.27 (0.8)	7.3 (4.2)	7.3 (4)	–0.033 (0.7)	1.83 (–0.92 to 4.59)	0.50
Right leg	2.1 (0.9)	2.1 (0.9)	–0.03 (0.3)	2.4 (1)	2.5 (1)	0.08 (0.2)	0.44 (–0.25 to 1.13)	0.48
Left leg	2.1 (0.9)	2.1 (0.9)	–0.03 (0.2)	2.4 (1)	2.5 (1)	0.09 (0.2)	0.43 (–0.26 to 1.11)	0.47

Pearson Correlation Between Functional Fitness in the Senior Fitness Test

The magnitude of correlations was interpreted according to Cohen conventions: small (0.10-0.29), moderate (0.30-0.49), and large (≥0.50). Positive values indicate direct associations, while negative values indicate inverse associations. Pearson correlation analysis of physical fitness assessments in older adults (Table 8) suggested a significantly moderate positive correlation between upper and lower limb muscular endurance (r=0.502; P<.05). In addition, a significantly low positive correlation between upper limb muscular endurance and lower limb flexibility was observed (r=0.377; P<.01). However, no significant correlations were observed between upper limb muscular endurance and either static, dynamic balance, or aerobic endurance. Lower limb muscular endurance demonstrated a significantly low to moderate positive correlation with both cardiorespiratory endurance (r=0.488; P<.01) and lower limb flexibility (r=0.386; P<.01), whereas a significantly moderate negative correlation with dynamic balance (r=−0.6; P<.01) was observed. In addition, lower limb flexibility exhibited a significantly moderate positive correlation with static balance (r=0.402; P<.01). Therefore, the intervention effectively enhanced multiple aspects of physical fitness, underscoring the interplay among fitness components.

XSL•FO
RenderX

Table . The Pearson correlation coefficients and 95% CIs for functional fitness as evaluated by the Senior Fitness Test.

Senior physical fitness	Arm curl	Chair stand	Sit-and-reach	One-leg standing	8 ft up-and-go test	2-minute step
Arm curl						
r (95% CI)	1	0.502 (0.17 to 0.73)	0.377 (0.02 to 0.65)	–0.063 (–0.39 to 0.33)	–0.063 (–0.39 to 0.33)	0.189 (–0.28 to 0.44)
P value	—[a]	.005	.04	.74	.74	.32
Chair stand						
r (95% CI)	0.502 (0.17 to 0.73)	1	0.386 (0.05 to 0.66)	0.072 (–0.33 to 0.39)	–0.600 (–0.79 to –0.31)	0.488 (0.16 to 0.72)
P value	.005	—	.04	.71	<.001	.006
Sit-and-reach						
r (95% CI)	0.377 (0.02 to 0.65)	0.386 (0.05 to 0.66)	1	0.402 (0.05 to 0.66)	–0.149 (–0.43 to 0.29)	0.242 (–0.23 to 0.47)
P value	.04	.04	—	.03	.43	.20
One-leg standing						
r (95% CI)	–0.063 (–0.39 to 0.33)	0.072 (–0.33 to 0.39)	0.402 (0.05 to 0.66)	1	–0.195 (–0.44 to .27)	0.009 (–0.35 to 0.36)
P value	.74	.71	.03	—	.30	.96
8 ft up-and-go test						
r (95% CI)	–0.063 (–0.39 to 0.33)	–0.600 (–0.79 to –0.31)	–0.149 (–0.43 to 0.29)	–0.195 (–0.44 to 0.27)	1	–0.207 (–0.44 to 0.27)
P value	.74	<.001	.43	.30	—	.27
2-minute step						
r (95% CI)	0.189 (–0.28 to 0.44)	0.488 (0.16 to 0.72)	0.242 (–0.23 to 0.47)	0.009 (–0.35 to 0.36)	–0.207 (–0.44 to 0.27)	1
P value	.32	.006	.20	.96	.27	—

[a]Not applicable.

Discussion

Principal Findings

Integrating exergames, motion-sensing technologies, and sports-based innovations is essential for the future development of physical fitness interventions. Well-structured training programs not only promote exercise participation and overall health across diverse populations but also enhance the regularity and acceptability of fitness behaviors. Accordingly, the present 10-week mat exercise program was designed based on an exergame strategy, incorporating a variety of exercise equipment to achieve multiple training objectives under coach supervision. The sessions used a 2-person exergame-based mat system that featured turn-taking, cooperative, and competitive game-based activities. In terms of functional fitness, significant improvements were observed in muscular endurance, lower limb flexibility, static and dynamic balance, core strength, agility, and aerobic endurance. Moreover, physical activity levels and quality of life—particularly aspects related to physical health, psychological well-being, and social relationships—were also enhanced as a result of this novel exergame-based mat training intervention.

Traditional community-based exercise programs, including aerobic, resistance, balance, and stretching training, have been shown to improve daily physical activity and aerobic capacity in older adults [24]. However, for those with mild disabilities, such interventions primarily enhance specific fitness components such as lower limb strength and flexibility without significantly affecting overall activity levels [25]. Alternative approaches, such as stationary cycling-based high-intensity interval training (HIIT), effectively increase moderate-to-vigorous activity and energy expenditure [26]. Similarly, motion-based exergames like the Nintendo Wii and Xbox Kinect increase energy expenditure and improve upper limb endurance, physical activity, and psychological well-being, particularly among sedentary or mobility-limited older adults [27,28]. These findings suggest that program design, intervention duration, and participant health status critically shape outcomes. Building on this evidence, our exergame-based mat training program integrated exergaming elements, progressive intensity, HIIT mode, and group-based sessions to provide a comprehensive physical activity experience, yielding significant gains in vigorous activity and functional fitness.

XSL•FO
RenderX

Regarding quality of life, a study [29] on Kinect-based exergames reported significant improvements in physical and psychological health, aligning with our findings. Both the exercise mat program and Kinect training enhanced mobility, reduced pain, and improved daily functioning. Psychologically, the exercise mat training boosted self-esteem and reduced negative emotions. Multicomponent exercise training has also been shown to enhance executive functioning in older adults, contributing to improved quality of life [30]. In addition, exergaming programs significantly improve self-efficacy in older adults, particularly by fostering confidence in maintaining long-term exercise participation [31]. This underscores the role of gamification and social interaction in overcoming psychological barriers and building confidence in physical activity. A systematic review further confirmed the effectiveness of exergames in promoting mental health and socialization among older adults, though identifying the most effective types of exergames remains an area for future research [32]. The exergame-based mat training in this study was designed to achieve multiple fitness goals by incorporating enjoyable, electronic elements to evoke positive emotions. Group-based sessions fostered social interaction and reduced loneliness, while the progressive intensity design introduced appropriate challenges to enhance self-efficacy. Together, these factors contributed to improved physical function and psychological adaptability in older adults.

Our findings also align with meta-analyses demonstrating that active video games improve balance, lower limb strength, and cardiorespiratory fitness in older adults, though effects on body composition and flexibility remain limited [33,34]. In the present study, the exergame-based mat training yielded similar results, with significant improvements in balance, flexibility, muscular strength, and aerobic capacity, despite no significant changes in body composition. Older adults without a history of falls exhibited superior performance, highlighting the role of static balance, lower limb strength, and flexibility in reducing fall risk in this population [35]. Previous studies consistently highlight the critical role of lower limb strength in older adults. Greater lower limb strength has been linked to enhanced static and dynamic balance [36,37], greater functional independence and self-care ability [38], as well as higher quality of life [39]. Together, these findings emphasize that maintaining lower limb strength is fundamental not only for mobility and fall prevention but also for preserving autonomy and overall well-being in later life. Our findings suggest that older adults who participated in the exergame-based mat training demonstrated functional improvements across functional fitness indicators, which may have implications for fall prevention, independence, and overall quality of life.

Exergame interventions provide multifaceted benefits for older adults, including improvements in strength, balance, flexibility, and cardiorespiratory endurance [40]. Although our mat-based exergame training did not reduce body fat, it effectively enhanced upper limb and core strength, aerobic endurance, agility, and balance, thereby reducing fall risk, consistent with prior mat-based studies [41]. Similar to Bacha et al [42], who reported superior cardiorespiratory fitness gains with Kinect-based exergames compared to conventional therapy, our findings further underscore the potential of exergame-based approaches for improving aerobic capacity and overall function. Given evidence linking physical fitness to fall frequency [43], the observed improvements in balance and muscular strength highlight clinically meaningful outcomes. Moreover, parallels with HIIT research indicate that the present exergame training program may facilitate favorable physiological adaptations, such as improved metabolic and cardiovascular health, while mitigating age-related decline [44-46].

This study has several limitations. Long-term quantitative tracking using accelerometry was not implemented, and participants' dietary caloric intake was neither controlled nor documented, limiting adjustments for physical activity and body composition effects. Upper limb flexibility requires further exploration through targeted exercises and equipment design. In addition, the benefits observed cannot be conclusively attributed to the program, and the lack of an attention-matched intervention for the control group may have introduced bias in self-reported outcomes, such as physical activity and quality of life. However, the inclusion of objective measures (eg, mat-based kinetics and standardized timed tests) provides supportive evidence for the observed effects. As an exploratory trial, the exergame-based training program demonstrated diverse benefits in physical fitness, physical activity, and quality of life. Future research should refine training intensity and modes, include appropriate comparator groups, and further investigate the program's physiological and therapeutic impacts to enhance its applicability for diverse older adult populations.

Conclusion

The implementation of exergame-based mat training positively influenced functional fitness in older adults. These gains are likely to translate into improved performance of daily physical activities, including enhanced walking capacity, prolonged standing tolerance, greater movement flexibility, and increased independence in activities of daily living. Beyond physical benefits, participation in exergame-based mat training positively influenced psychological well-being by boosting self-confidence and self-esteem while promoting greater social participation. In addition, the integration of group-based exergame and interactive experiences enhanced the willingness and acceptance of older adults to engage in training. These findings highlight the feasibility and benefits of the program in improving functional fitness and overall quality of life among older adults.

Acknowledgments

This work was financially supported by the National Science and Technology Council (NSTC 113-2423-H-227-001-MY3). We would like to state our appreciation to Uni-edit for editing and proofreading this manuscript and to Dr Hsuen-Ying Peng's University Social Responsibility program for providing participant resources.

XSL•FO
RenderX

The authors acknowledge the use of ChatGPT (OpenAI) as an artificial intelligence–assisted tool during the manuscript preparation process. The tool was employed exclusively for text editing to ensure clarity, precision, and adherence to scientific rigor. After using this service, the manuscript underwent additional professional editing by a certified English editing service to further ensure fluency, grammatical accuracy, and appropriateness of academic expression.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

WCH was responsible for the original draft, conceptualization, methodology, supervision, validation, visualization, writing (reviewing & editing), and funding acquisition for this study. SFH, CWF, DKN, and TWL contributed to the investigation, methodology, project administration, and visualization on this study. All authors have read and approved the submitted version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Training protocol.
[PDF File, 64 KB - games_v13i1e73474_app1.pdf]

Checklist 1
CONSORT checklist.
[PDF File, 121 KB - games_v13i1e73474_app2.pdf]

References

1. Lunenfeld B, Stratton P. The clinical consequences of an ageing world and preventive strategies. Best Pract Res Clin Obstet Gynaecol 2013 Oct;27(5):643-659. [doi: 10.1016/j.bpobgyn.2013.02.005] [Medline: 23541823]
2. Battista F, Duregon F, Vecchiato M, Ermolao A, Neunhaeuserer D. Sedentary lifestyle and physical inactivity: a mutual interplay with early and overt frailty. Nutr Metab Cardiovasc Dis 2025 Jun;35(6):103971. [doi: 10.1016/j.numecd.2025.103971] [Medline: 40180827]
3. Schwartz BD, Liu H, MacDonald EE, Mekari S, O'Brien MW. Impact of physical activity and exercise training on health-related quality of life in older adults: an umbrella review. Geroscience 2025 Jun;47(3):2879-2893. [doi: 10.1007/s11357-024-01493-6] [Medline: 39754009]
4. Ayenigbara IO. The contributions of physical activity and fitness for the optimal health and wellness of the elderly people. J Gerontol Geriatr 2020;68(1):40-46. [doi: 10.36150/2499-6564-351]
5. Kiełtyka-Słowik A, Michalik-Marcinkowska U, Zawadzka B. The association between physical activity and quality of life among people aged 60-89 living in own homes and nursing homes. BMC Geriatr 2024 Mar 23;24(1):280. [doi: 10.1186/s12877-024-04898-2] [Medline: 38521918]
6. Sánchez JLC, Gallardo-Gómez D, Alfonso-Rosa RM, Cruz BDP, Ramos-Munell J, del Pozo-Cruz J. Effectiveness of different types of exercise based-interventions in sarcopenia: a systematic review and meta-analysis. Geriatr Nurs (Lond) 2025 May;63:635-642. [doi: 10.1016/j.gerinurse.2025.04.019]
7. Williams MA, Soiza RL, Jenkinson AM, Stewart A. EXercising with Computers in Later Life (EXCELL) - pilot and feasibility study of the acceptability of the Nintendo® WiiFit in community-dwelling fallers. BMC Res Notes 2010 Sep 13;3:238. [doi: 10.1186/1756-0500-3-238] [Medline: 20831836]
8. Chen X, Jiang D, Ning H, et al. The experience of and needs for exergames in older adults with mild cognitive impairment: qualitative interview study. JMIR Serious Games 2025 Jun 17;13:e53631. [doi: 10.2196/53631] [Medline: 40526915]
9. Ramnath U, Rauch L, Lambert EV, Kolbe-Alexander T. Efficacy of interactive video gaming in older adults with memory complaints: a cluster-randomized exercise intervention. PLoS ONE 2021;16(5):e0252016. [doi: 10.1371/journal.pone.0252016] [Medline: 34032799]
10. Rytterström P, Strömberg A, Jaarsma T, Klompstra L. Exergaming to increase physical activity in older adults: feasibility and practical implications. Curr Heart Fail Rep 2024 Aug;21(4):439-459. [doi: 10.1007/s11897-024-00675-9] [Medline: 39023808]
11. Pacheco TBF, de Medeiros CSP, de Oliveira VHB, Vieira ER, de Cavalcanti FAC. Effectiveness of exergames for improving mobility and balance in older adults: a systematic review and meta-analysis. Syst Rev 2020 Jul 18;9(1):163. [doi: 10.1186/s13643-020-01421-7] [Medline: 32682439]

XSL•FO
RenderX

12. Buyle M, Jung Y, Pavlou M, Gonzalez SC, Bamiou DE. The role of motivation factors in exergame interventions for fall prevention in older adults: a systematic review and meta-analysis. Front Neurol 2022;13:903673. [doi: 10.3389/fneur.2022.903673] [Medline: 35989930]
13. Medeiros CSPD, Farias LBA, Santana MCDL, Pacheco TBF, Dantas RR, Cavalcanti FADC. A systematic review of exergame usability as home-based balance training tool for older adults usability of exergames as home-based balance training. PLoS One 2024;19(8):e0306816. [doi: 10.1371/journal.pone.0306816] [Medline: 39172915]
14. Dill S, Müller PN, Caserman P, Göbel S, Hoog Antink C, Tregel T. Sensing in exergames for efficacy and motion quality: scoping review of recent publications. JMIR Serious Games 2024 Nov 5;12:e52153. [doi: 10.2196/52153] [Medline: 39499916]
15. Larsen MN, Krustrup P, Araújo Póvoas SC, Castagna C. Accuracy and reliability of the InBody 270 multi-frequency body composition analyser in 10-12-year-old children. PLoS One 2021;16(3):e0247362. [doi: 10.1371/journal.pone.0247362] [Medline: 33770106]
16. Rikli RE, Jones CJ. Functional fitness normative scores for community residing older adults ages 60-94. J Aging Phys Act 1999;7(2):162-181. [doi: 10.1123/japa.7.2.162]
17. Rikli RE, Jones CJ. Development and validation of criterion-referenced clinically relevant fitness standards for maintaining physical independence in later years. Gerontologist 2013 Apr;53(2):255-267. [doi: 10.1093/geront/gns071] [Medline: 22613940]
18. Craig CL, Marshall AL, Sjöström M, et al. International physical activity questionnaire: 12-country reliability and validity. Med Sci Sports Exerc 2003 Aug;35(8):1381-1395. [doi: 10.1249/01.MSS.0000078924.61453.FB] [Medline: 12900694]
19. Deng HB, Macfarlane DJ, Thomas GN, et al. Reliability and validity of the IPAQ-Chinese: the Guangzhou Biobank Cohort study. Med Sci Sports Exerc 2008 Feb;40(2):303-307. [doi: 10.1249/mss.0b013e31815b0db5] [Medline: 18202571]
20. Tsai CH, Liao Y, Chang SH. Cross-sectional association of physical activity levels with risks of sarcopenia among older Taiwanese adults. BMC Geriatr 2024 Jun 27;24(1):560. [doi: 10.1186/s12877-024-05087-x] [Medline: 38937702]
21. Skevington SM, Lotfy M, O'Connell KA, WHOQOL Group. The World Health Organization's WHOQOL-BREF quality of life assessment: psychometric properties and results of the international field trial. A report from the WHOQOL group. Qual Life Res 2004 Mar;13(2):299-310. [doi: 10.1023/B:QURE.0000018486.91360.00] [Medline: 15085902]
22. Yao G, Chung CW, Yu CF, Wang JD. Development and verification of validity and reliability of the WHOQOL-BREF Taiwan version. J Formos Med Assoc 2002 May;101(5):342-351. [Medline: 12101852]
23. Cohen J. Statistical Power Analysis for the Behavioral Sciences, 2nd edition: Lawrence Erlbaum Associates; 1988.
24. Hernandes NA, Probst VS, Da Silva RA Jr, Januário RSB, Pitta F, Teixeira DC. Physical activity in daily life in physically independent elderly participating in community-based exercise program. Braz J Phys Ther 2013;17(1):57-63. [doi: 10.1590/s1413-35552012005000055] [Medline: 23117651]
25. Taguchi N, Higaki Y, Inoue S, Kimura H, Tanaka K. Effects of a 12-month multicomponent exercise program on physical performance, daily physical activity, and quality of life in very elderly people with minor disabilities: an intervention study. J Epidemiol 2010;20(1):21-29. [doi: 10.2188/jea.je20081033] [Medline: 19897943]
26. Bruseghini P, Tam E, Calabria E, Milanese C, Capelli C, Galvani C. High intensity interval training does not have compensatory effects on physical activity levels in older adults. Int J Environ Res Public Health 2020 Feb 8;17(3):1083. [doi: 10.3390/ijerph17031083] [Medline: 32046311]
27. Taylor LM, Maddison R, Pfaeffli LA, Rawstorn JC, Gant N, Kerse NM. Activity and energy expenditure in older people playing active video games. Arch Phys Med Rehabil 2012 Dec;93(12):2281-2286. [doi: 10.1016/j.apmr.2012.03.034] [Medline: 22522217]
28. Keogh JWL, Power N, Wooller L, Lucas P, Whatman C. Physical and psychosocial function in residential aged-care elders: effect of Nintendo Wii Sports games. J Aging Phys Act 2014 Apr;22(2):235-244. [doi: 10.1123/japa.2012-0272] [Medline: 23752164]
29. Rica RL, Shimojo GL, Gomes MC, et al. Effects of a Kinect-based physical training program on body composition, functional fitness and depression in institutionalized older adults. Geriatr Gerontol Int 2020 Mar;20(3):195-200. [doi: 10.1111/ggi.13857] [Medline: 31923924]
30. Kaushal N, Desjardins-Crépeau L, Langlois F, Bherer L. The effects of multi-component exercise training on cognitive functioning and health-related quality of life in older adults. Int J Behav Med 2018 Dec;25(6):617-625. [doi: 10.1007/s12529-018-9733-0] [Medline: 29926316]
31. Sowle AJ, Francis SL, Margrett JA, Shelley MC, Franke WD. A community-based exergaming physical activity program improves readiness-to-change and self-efficacy among rural-residing older adults. J Aging Phys Act 2017 Jul;25(3):432-437. [doi: 10.1123/japa.2015-0278] [Medline: 27992237]
32. Lima A, Moreira MT, Ferreira MS, et al. Efficacy of the use of exergames in promoting the mental health of the elderly: protocol of a systematic review. J Ageing Longev 2023;3(3):191-202. [doi: 10.3390/jal3030015]
33. Taylor LM, Kerse N, Frakking T, Maddison R. Active video games for improving physical performance measures in older people: a meta-analysis. J Geriatr Phys Ther 2018;41(2):108-123. [doi: 10.1519/JPT.0000000000000078] [Medline: 26974212]

XSL•FO
RenderX

34. Deng N, Soh KG, Abdullah BB, Tan H, Huang D. Active video games for improving health-related physical fitness in older adults: a systematic review and meta-analysis. Front Public Health 2024;12:1345244. [doi: 10.3389/fpubh.2024.1345244] [Medline: 38694976]
35. Gehlsen GM, Whaley MH. Falls in the elderly: part II, balance, strength, and flexibility. Arch Phys Med Rehabil 1990 Sep;71(10):739-741. [Medline: 2403279]
36. Astari KLW, Jaya IPP, Paramurthi IAP, et al. The relationship of leg muscle strength on dynamic balance among elderly. Phys Ther J Indones 2021;2(2):67-72. [doi: 10.51559/ptji.v2i2.27]
37. Yeh PC, Syu DK, Ho CC, Lee TS. Associations of lower-limb muscle strength performance with static and dynamic balance control among older adults in Taiwan. Front Public Health 2024;12:1226239. [doi: 10.3389/fpubh.2024.1226239] [Medline: 38414890]
38. Batista FS, Gomes GADO, D'Elboux MJ, et al. Relationship between lower-limb muscle strength and functional independence among elderly people according to frailty criteria: a cross-sectional study. Sao Paulo Med J 2014;132(5):282-289. [doi: 10.1590/1516-3180.2014.1325669]
39. da Costa Pereira JP, Freire YA, da Silva AMB, et al. Associations of upper- and lower-limb muscle strength, mass, and quality with health-related quality of life in community-dwelling older adults. Geriatr Gerontol Int 2024 Jul;24(7):683-692. [doi: 10.1111/ggi.14912] [Medline: 38840315]
40. Hernandez-Martinez J, Cid-Calfucura I, Herrera-Valenzuela T, et al. Xbox Kinect Sports vs. Nintendo Switch Sports and their effects on body composition and physical performance in older females: a randomized controlled trial. J Clin Med 2024 Aug 23;13(17):4987. [doi: 10.3390/jcm13174987] [Medline: 39274198]
41. Peng HT, Tien CW, Lin PS, Peng HY, Song CY. Novel mat exergaming to improve the physical performance, cognitive function, and dual-task walking and decrease the fall risk of community-dwelling older adults. Front Psychol 2020;11:1620. [doi: 10.3389/fpsyg.2020.01620] [Medline: 32793044]
42. Bacha JMR, Gomes GCV, de Freitas TB, et al. Effects of kinect adventures games versus conventional physical therapy on postural control in elderly people: a randomized controlled trial. Games Health J 2018 Feb;7(1):24-36. [doi: 10.1089/g4h.2017.0065] [Medline: 29239677]
43. Lin WS, Hsu NW, Lee MJ, Lin YY, Tsai CC, Pan PJ. Correlation analysis of physical fitness and its impact on falls in 2130 community- dwelling older adults: a retrospective cross-sectional study. BMC Geriatr 2022 May 23;22(1):447. [doi: 10.1186/s12877-022-03138-9] [Medline: 35606737]
44. Wu ZJ, Wang ZY, Gao HE, Zhou XF, Li FH. Impact of high-intensity interval training on cardiorespiratory fitness, body composition, physical fitness, and metabolic parameters in older adults: a meta-analysis of randomized controlled trials. Exp Gerontol 2021 Jul 15;150:111345. [doi: 10.1016/j.exger.2021.111345] [Medline: 33836261]
45. Carpes L, Costa R, Schaarschmidt B, Reichert T, Ferrari R. High-intensity interval training reduces blood pressure in older adults: a systematic review and meta-analysis. Exp Gerontol 2022 Feb;158:111657. [doi: 10.1016/j.exger.2021.111657] [Medline: 34921916]
46. Bouaziz W, Malgoyre A, Schmitt E, Lang PO, Vogel T, Kanagaratnam L. Effect of high-intensity interval training and continuous endurance training on peak oxygen uptake among seniors aged 65 or older: a meta-analysis of randomized controlled trials. Int J Clin Pract 2020 Jun;74(6):e13490. [doi: 10.1111/ijcp.13490] [Medline: 32083390]

Abbreviations

HIIT: high-intensity interval training
HRmax: maximum heart rate
IPAQ: International Physical Activity Questionnaire
MD: mean difference
MET: metabolic equivalents
NTUNHS: National Taipei University of Nursing and Health Sciences
PAR-Q: Physical Activity Readiness Questionnaire
USR: University Social Responsibility
WHOQOL-BREF: World Health Organization Quality of Life Brief Version

XSL•FO
RenderX

Edited by S Brini; submitted 05.03.25; peer-reviewed by MN Shalaby, PE Roos; revised version received 25.09.25; accepted 25.09.25; published 03.11.25.

Please cite as:
Huang SF, Fan CW, Nisrina DK, Lin TW, Huang WC
A Novel Gamified Exercise Program Incorporating Stampede Training for Enhancing Functional Fitness, Physical Activity Levels, and Quality of Life in Community-Dwelling Older Adults: Randomized Parallel Exploratory Trial
JMIR Serious Games 2025;13:e73474
URL: https://games.jmir.org/2025/1/e73474
doi:10.2196/73474

© Siao-Fen Huang, Ching-Wen Fan, Diana Khasna Nisrina, Tsai-Wei Lin, Wen Ching Huang. Originally published in JMIR Serious Games (https://games.jmir.org), 3.11.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

FITLIGHT Training and Its Influence on Visual-Motor Reactions and Dribbling Speed in Female Basketball Players: Prospective Evaluation Study

Ahmed K Hassan[1,2], PhD

[1]Department of Physical Education, College of Education, King Faisal University, P.O. Box 400, Al-Ahsa, Saudi Arabia

[2]Department of Team Sports and Racket Games, Faculty of Physical Education, Minia University, Minia, Egypt

Corresponding Author:
Ahmed K Hassan, PhD
Department of Physical Education, College of Education, King Faisal University, P.O. Box 400, Al-Ahsa, Saudi Arabia

Related Article:

This is a corrected version. See correction statement: https://games.jmir.org/2025/1/e82614

Abstract

Background: Basketball exists as a team-based sport played on a court involving intense physical demands because players need continuous movement between offensive and defensive zones.

Objective: The aim of the current research was to investigate the impact of a training program implementing FITLIGHT (FITLIGHT Corp) technology on female basketball players' visual-motor interactions and dribbling speed.

Methods: The study included 28 female basketball players. Participants were divided into 2 groups, experimental group (EG; n=14, mean age 18.29, SD 0.99 y; mean height 174.50, SD 2.10 cm; mean weight 75.36, SD 2.10 kg; mean training 5.64, SD 0.50 y) and control group (CG; n=14, mean age 18.50, SD 0.65 y; mean height 175.64, SD 1.55 cm; mean weight 75.57, SD 1.79 kg; mean training 5.43, SD 0.51 y), in a random manner. Pre- and post-measurements were used, and FITLIGHT training was conducted for 10 weeks with 4 sessions per week. The FITLIGHT training program targeted the elements of visual-motor interactions and dribbling speed for the EG. The CG experienced the same training regimen, but did not participate in the FITLIGHT training. The tests that were adopted and administered were the reaction time in the right hand and left hand, the reactive time (*t* test), and the reactive time with dribbling (*t* test D) tests. Statistical analysis included the calculation of descriptive statistics for minimum and maximum values and mean and SD, which were used for paired *t* tests for within-group comparison alongside independent *t* tests for between-group differences. Effect sizes (Cohen *d*) measured small effects at less than 0.2, medium effects at 0.2 - 0.5, and large effects at greater than 0.8 at $P<.05$.

Results: All the EG improvements were statistically significant ($P<.001$), with reaction time in the right hand improving by 0.04 ms more than that in CG (0.016 ms; effect sizes $d=0.87$ for the EG and $d=0.79$ for the CG). The reaction time in the left hand of the EG was 0.05 ms faster ($d=0.97$), compared with a difference of 0.019 ms in the CG ($d=0.71$). Participants in the EG had 1.90 seconds more reactive time ($d=0.98$) than those in the CG, who had 0.85 seconds ($d=0.97$). For dribbling in the EG, reactive time improved by 4.08 seconds ($d=0.98$), much better than the improvement seen in the CG of 1.71 seconds ($d=0.93$) when compared with using effect sizes. The analysis of the data is finished, showing that the EG had much larger effects than the CG. Study participants achieved large effect sizes during reaction time in the right hand ($d=0.51$), reaction time in the left hand ($d=0.71$), reactive time ($d=0.84$), and reactive time with dribbling ($d=0.88$).

Conclusions: A pairwise comparison between the EG and the CG also revealed a statistically significant difference in the EG, which proved that the EG outperformed the CG. This study provides evidence for the enhanced visual-motor interactions and dribbling speeds of female basketball players who received training that included the use of the FITLIGHT technique.

(JMIR Serious Games 2025;13:e70519) doi:10.2196/70519

KEYWORDS

reaction time; reactive time; FITLIGHT technology; visual-motor interaction; sports performance

XSL•FO
RenderX

Introduction

Basketball is a team sport that is played on a court. The activity is extremely strenuous, particularly due to the constant movement between offense and defense, and entails highly energetic movements such as jumping, running, and shuffling [1]. Considering the grueling nature of the exercises involved in basketball, it is appropriate for coaches to provide the essential measures to monitor the physical pressures that players experience [2]. Montella et al [3] confirmed that, as a team sport, basketball is quantitative in nature due to the elements of performance, strength, and endurance, as the focus is on the statistical and operational dimensions of quality. Basketball may be characterized as a somewhat intermittent team sport in which different players' movements and activities are performed repeatedly in the form of short bouts of high intensity followed by periods of low to moderate intensity [1,4]. Modern basketball can be described as a game that involves both rhythm and skills [5], as shown in the global level of honed skills observed across different basketball tournaments, revisiting current trends in training young talents [6].

FITLIGHT (FITLIGHT Corp) training is a modern trend in the training of sportspeople. It involves the use of wireless light units connected to a tablet to promote thinking processes and motor skills. This technology has been incorporated into several sports due to the improvements in the athletes and in their mental functions. After conducting an analysis of judo athletes, the researcher demonstrated that a 5-week cognitive-motor training program using FITLIGHT enhanced their executive functions, their cognition, and their motor skills [7]. The FITLIGHT training system also demonstrated acceptable intertest reactivity for the assessment of reaction times. This is a crucial parameter for the evaluation of cognitive states following traumatic brain injury [8]. A study of basketball players showed that small side games using FITLIGHT training enhanced abilities, basic skills, and performances in the experimental group (EG) [9]. FITLIGHT technology helped to provide feedback during training, thus improving the athletes' strategies and their ability to correct their own techniques in the same way that a personal trainer would do [10]. FITLIGHT is a training tool that can be used to improve basketball players' physical skills and visual development [11]. FITLIGHT includes an internal contact sensor that can be used to turn each light unit off or on. These lights can be switched on individually or several at a time, with the option of altering the patterns and durations of the stimulation (FITLIGHT Corp). In addition to their design that enables different light sequences, it is possible to mount the FITLIGHT to walls, columns, floors, and other training equipment [12].

The nature of performance in basketball requires the integration of motor and hand-eye coordination (visual training) with the aim of enhancing interactions at the level of visualization skills, developing skills, and increasing the level of focused attention. This link is authentic and mimics the dynamics of an actual player during a game. The faster a player becomes aware of the ball, their teammates, or the space on the field, the better and faster they are at making decisions and executing them [13]. Visualization via modern techniques and instruments has a positive effect on the enhancement of visual neural networks and some cerebral functions. Visual training programs also improve athletic performances and decrease weaknesses by sharpening hand-eye coordination, decreasing reaction times, expanding peripheral vision, and enhancing the playing field [14]. The researcher noted that basketball coaches hardly considered the aspect of hand speed, which made it easier for the opposing team to counter and score points or to score in the basket because basketball requires extensive hand-eye coordination with regard to receiving, passing, dribbling, and shooting the ball at the basket. It also requires rapid visual reactions in switching between the attacking and the defending lines, as previous studies have indicated [1,9,11-22]. Thus, the researcher conducted a survey study of the tests and then confirmed the poor performances of players in those tests; hence, there was the need to undertake a study to determine the impact of FITLIGHT exercises on enhancing the visual-motor coordination and dribbling speeds of female basketball players to assist coaches in understanding the correct approaches and strategies for enhancing their performance levels.

This study aimed to look at female basketball players' visual-motor coordination after a 10-week FITLIGHT-based training program, by using the Lafayette Instrument 63035A Response Panel, and in addition, to evaluate in skillful dribbling speed with the *t* test protocol. The methodology focused on a 10-week intervention with 4 weekly sessions, designed to optimize skill acquisition through structured, repetitive practice and real-time feedback mechanisms [10]. This approach diverges from previous studies that used shorter, less frequent training protocols. To quantify outcomes, visual-motor coordination was assessed using standardized reaction time and hand-eye coordination tests, while dribbling speed improvements were measured via timed obstacle courses and motion analysis software [7,9,11,21,22]. These metrics align with kinematic analysis, a validated tool for evaluating movement patterns and comparing performance across interventions [6]. The study's design builds on recent advancements in digital training technologies. By extending the training duration to 10 weeks, this research creates conditions to assess long-term skill development, contrasting with previous work that often focused on short-term interventions. In addition, the exclusive focus on female athletes addresses a critical gap in sports science literature, where gender-specific adaptations to training protocols remain underexplored.

We hypothesized that females in the EG would show significantly improved visual-motor coordination and dribbling speed compared with those in the control group (CG), as measured at the beginning and end of the intervention. In addition, we hypothesized that the prolonged practice period would make FITLIGHT more efficient, so the differences in dribbling speed between the EG and CG after measurement would favor the EG, because of better visual-motor coordination. This structure highlights how research in this area helps improve the training of women athletes, provides practical guidance for coaches and practitioners, and advances the science surrounding skill retention and performance when using technologies.

XSL•FO
RenderX

Methods

Study Design

The research design implemented a FITLIGHT system training protocol, which ran for 10 weeks to enhance visual-motor coordination ability and dribbling speed for Al-Way Club's female basketball team players from Minya Governorate. Furthermore, 2 research groups formed the EG and CG. The EG participated in FITLIGHT-based training, whereas the CG executed the traditional program. G*Power software from the University of Düsseldorf enabled statistical power analysis to determine the needed sample size [23], which became 28 amateur participants after excluding 6 people from the initial group of 34 participants because of certain exclusion criteria. The researchers strictly chose female participants who had more than 4 years of sports experience and commitment to assessment and training activities. The first round of evaluation, known as pretests, was administered to participants on August 4 and August 5, 2024. The research relied on dependable assessment instruments, including the visual response board and the FITLIGHT technology kit for data collection. The training program lasted from August 6 to October 13, 2024. The experimental participants underwent FITLIGHT training throughout 10 weeks in 4 scheduled sessions, yet the CG underwent conventional training without access to FITLIGHT devices. The supervisors monitored the gradual and timed progress of both the intensity level and duration while ensuring safety, along with correct performance by participants throughout the protocol. Both posttests took place from October 15 to October 16, 2024, according to the same assessment process as the initial tests. The study participants achieved satisfactory statistical significance due to their 28-person sample, which yielded an effect power of 0.81 through a power analysis.

Participants

An experimental approach was employed with 28 female players. Participants were divided into 2 groups, experimental and control, randomly: the EG (n=14, mean age 18.29, SD 0.99 y; mean height 174.50, SD 2.10 cm; mean weight 75.36, SD 2.10 kg; mean training 5.64, SD 0.50 y) and the CG (n=14, mean age 18.50, SD 0.65 y; mean height 175.64, SD 1.55 cm; mean weight 75.57, SD 1.79 kg; mean training 5.43, SD 0.51 y). The research inclusion criteria were established specifically to generate precise findings about the study. All participants needed to fulfill distinct qualifications, which included 6 female and active involvement in sports as well as healthy status, a minimum of 4 years of athletic experience, and a set age range. The study participants needed to exhibit continuous training participation until the end of the research period and successfully finish each assessment at the start and conclusion of the study. The study included fundamental exclusion criteria for eliminating factors that might produce confounding results during the research period. Athletes using medications or drugs were excluded because their substances could affect both their physiology and performance abilities. All participants who exhibited inconsistent training attendance or failed to perform baseline tests were automatically removed from the research. The established inclusion and exclusion criteria ensured homogeneity of the research sample as well as accurate data collection, which therefore strengthened the scientific rigor of this research.

Ethical Considerations

The study did not need preinitiation registration. A clinical trial registry registration before initiation did not occur because this study falls outside the registered scope for clinical trials that mainly encompass drug or medical intervention studies. The study adhered to ethical guidelines and scientific standards for all procedures, from participant recruitment to randomization of participants and intervention delivery and data analysis, to maintain reliable results. All athletes received comprehensive knowledge about both risks and benefits that they would encounter during study participation before they joined. Each participant received an institutionally approved informed consent form while voluntarily signing it for written consent that established voluntary and uncoerced participation. The participants were not compensated or paid any monetary amounts since they volunteered to take part. The Research Ethics Committee at King Faisal University, Saudi Arabia, awarded ethical approval for the study protocol (approval KFU-REC-2024-MAR-ETHICS2088) (Multimedia Appendix 1). This study followed the international standards from the Declaration of Helsinki for protecting the rights, privacy, and welfare of human participants. The researchers took steps to ensure that participant data would stay confidential and serve only the research needs. Individuals who enrolled in the research received full disclosure about withdrawal rights as well as protection for their confidentiality as part of the procedure to safeguard their anonymity.

Tools and Devices

The study incorporated various pieces of equipment for height and weight measurements, a restameter, a basketball, and the Device Response Panel 63035A combined reaction time (visual-motor interaction) obtained from the Lafayette Instrument Company (Figure 1) [9]. Furthermore, the modified *t* test was used to assess the speed of skillful dribbling (Figure 2) [11,24], and the FITLIGHT system (Figure 3), a stopwatch, and stands for the FITLIGHT disks and ropes were also used. The validity and reliability of the devices were further established via comparison with other instruments, such as the Visual Response Board 63035A (Lafayette Instrument Company) from another model exploding. To eliminate all doubts, these measurements were applied to additional participants in the current exploratory study. The Lafayette Instrument 63035A Reaction Time (visual-motor interaction) response board is illustrated in Figure 1; the instrument consists of 4 stimulus lamps, a Sonalert, and 5 response keys. The specifications are a line voltage of 105/125 V, AC/50/60 Hz power supply of 12 V DC from a wall-mounted transformer timer, a relay contact rating of 0.5 A/30 V DC and 0.5 A/50 V AC with an inductive load to be protected with arc suppression. The lamp used in the stimulus was 47/6.3 V Chicago Miniature [11]. Table 1 provides a summary of the training program (For further details, refer to Multimedia Appendices 2 and 3).

XSL•FO
RenderX

Figure 1. Visual reaction time apparatus 63035 (Multimedia Appendix 2).

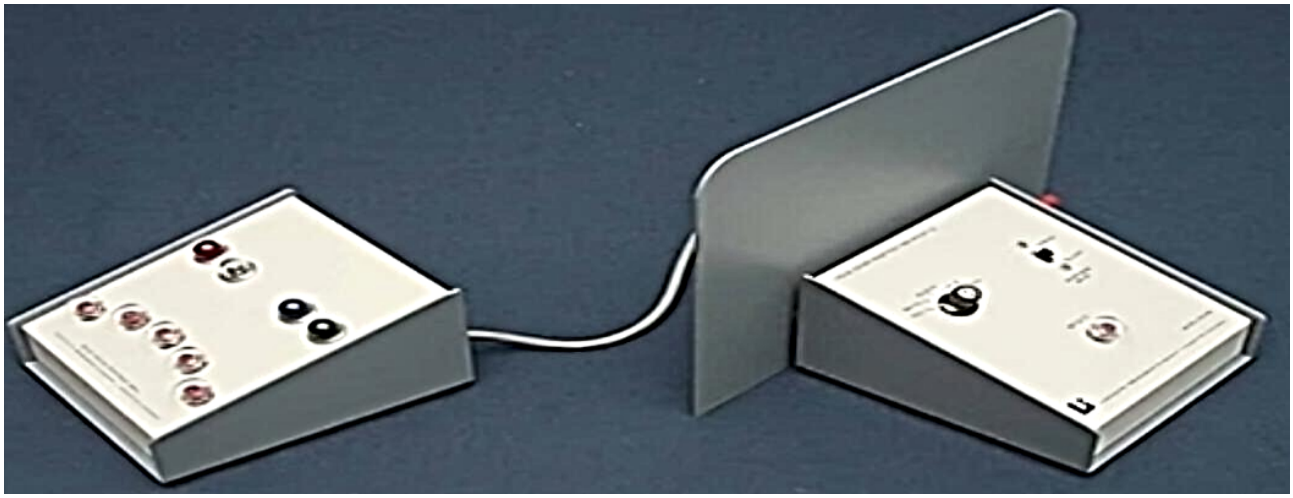


Figure 2. Modified *t* test with dribbling.

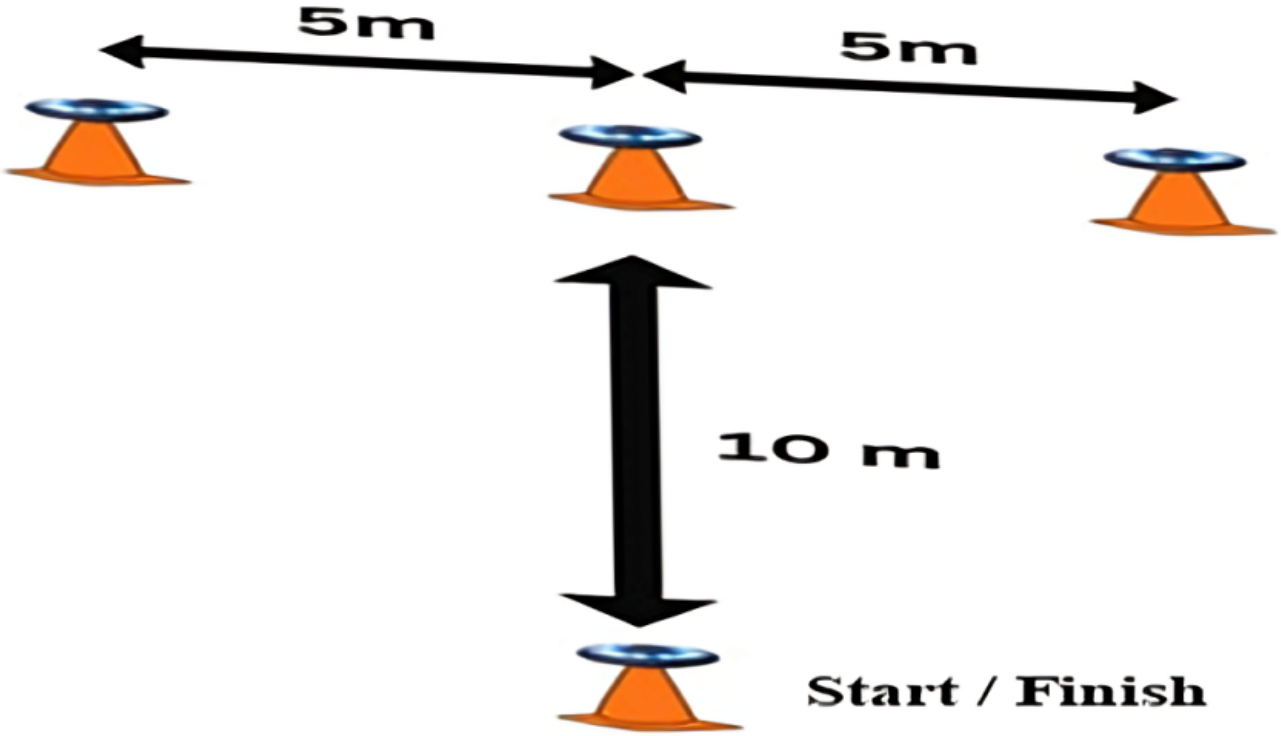


Figure 3. FITLIGHT technology.



XSL•FO
RenderX

Table . Distribution of 10-week FITLIGHT exercise program for basketball players (total program time=4000 min).

Stage	General preparation			Special preparation				Precompetition preparation		
Week	1	2	3	4	5	6	7	8	9	10
Load[a]										
Maximum					☒		☒	☒		
High	☒	☒		☒						☒
Medium			☒			☒			☒	
Week time (min)	320	320	240	320	400	240	400	400	240	320

[a]The graphics represent training loads for maximum, high, and medium loads.

Experimental Procedures

The researcher developed an exercise program using FITLIGHT technology that lasted for 10 weeks; there were 40 sessions in total, with 4 training sessions per week. The average duration of each session ranged from 60 to 100 minutes, of which the FITLIGHT exercises lasted for 25 - 40 minutes. The level of exertion was checked and adjusted to guarantee an appropriate level of load that was varied and augmented in each phase of the program.

The EG followed these steps in each training unit:

Every session began with a set of generic warm-up exercises to prepare separate groups of muscles in the body. Different, specific warm-up exercises were then conducted to increase preparedness for the planned activities.

During the session, the researcher explained the nature of the exercises that would be performed using the FITLIGHT technique in relation to the objectives of the program.

Members of the EG completed the FITLIGHT exercises in the pre- and post–warm-up phases, in addition to the inter- and intramain session exercises to enhance visual-motor coordination and speed when dribbling. During the training, the researcher was always present for the EG and provided extra help, corrected mistakes, ensured that the participants were performing the exercises correctly, and ensured their safety. Finally, every session concluded with a cool-down stretching session to assist in recovery and subsequent flexibility.

In this program, the exercises were progressive in nature to ensure that the increase in the intensity and duration of the exercises was based on the improvements in the participants' performances and their adaptation levels during the 10-week program (Table 1).

Statistical Analysis

IBM-SPSS Statistics 26 served as the tool to conduct statistical analysis for achieving reliable results. The researchers performed descriptive statistical analysis to find minimum and maximum, mean, SD, variance, skewness, and coefficient of variation (CV) for the EG and CG. The CV measurement tool calculates group homogeneity by dividing the SD by mean values to determine percentages that show homogeneity, <30% indicating homogeneity and >30%. For analysis between and across groups, the inferential statistics involved paired *t* tests for measuring intragroup differences and independent *t* tests for evaluating performance distinctions between the EG and CG. The research calculated effect sizes through Cohen *d* and interpreted 3 effect magnitude levels, including small (<0.2), medium (0.2 - 0.5), and large (>0.8). All statistical tests used a *P* value lower than .05 as the threshold to maintain consistency and reduce the concerns about multiple comparisons. Statistical computations of 95% CIs and mean differences, *t* values, and *P* values provided a deep analysis of the obtained results. The analysis followed standard conventions through these statistical methods, and the researchers applied these methods consistently during the entire analysis to ensure clear results and measurement reliability.

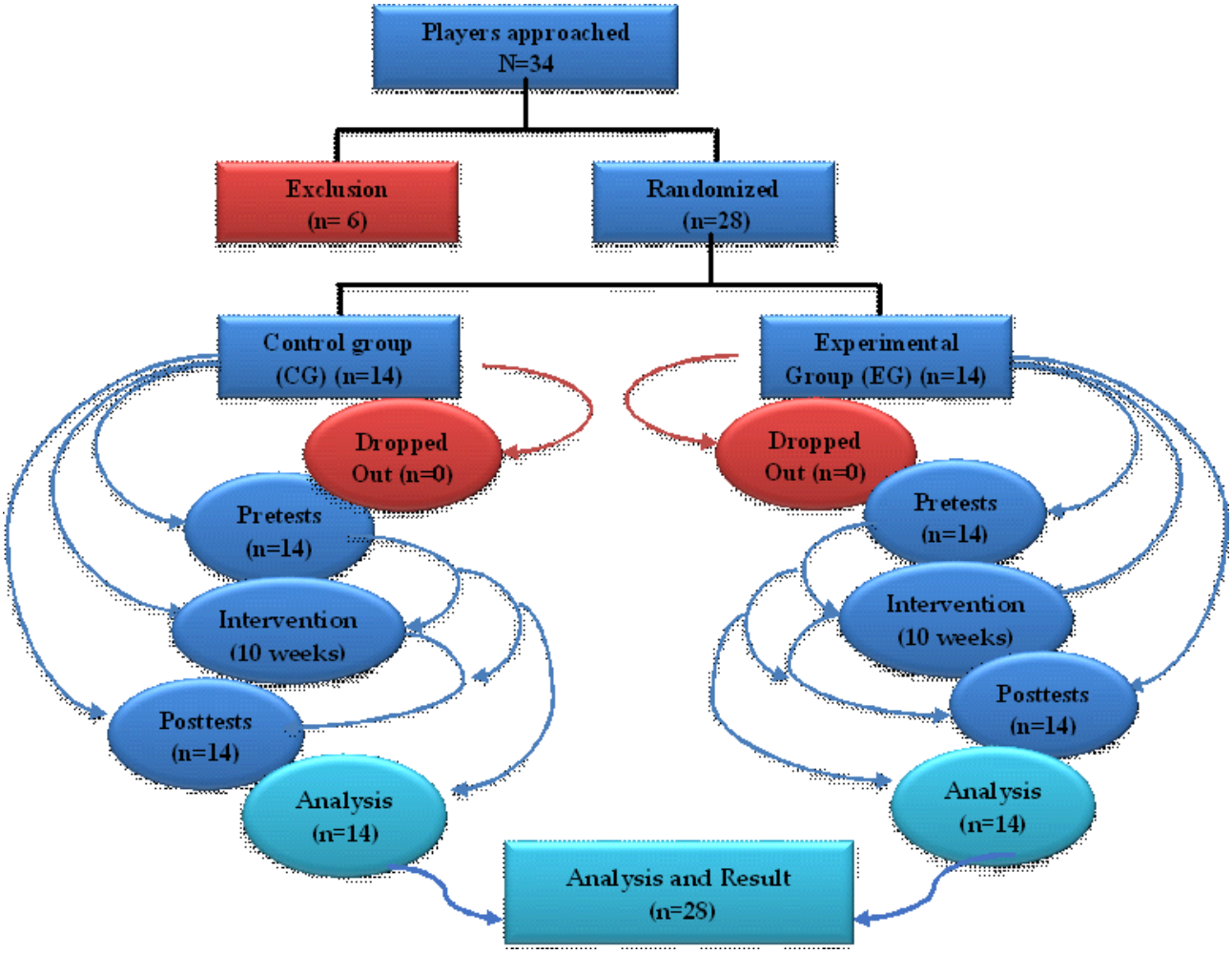
Results

Descriptive Statistics

The flowchart of this study is illustrated in Figure 4. Analyzing the data in Table 2, an improvement in the homogeneity of performance within the EG compared with the CG during the tests was observed. Compared with the premeasurements, it is observable that the CV of the EG ranged from 0.63% to 5.41%. This reduction is indicative of increased reliability of the results that are generated, while the CG CV ranged from 0.39% to 2.26%, which can also be considered a good sign as it indicates good results consistency.

XSL•FO
RenderX

Figure 4. Flowchart in this study.



XSL•FO
RenderX

Table . Descriptive statistics for reaction tests and reactive tests.

Group and outcome measures	Test	Range, minimum-maximum	Mean (SD)	Variance	Skewness, statistic (SE)	CV[a]
EG[b]						
Age	Pre	17-20	18.29 (0.99)	0.99	−0.12 (0.60)	—[c]
Height	Pre	170-177	174.50 (2.10)	4.42	−1.42 (0.60)	—
Weight	Pre	70-77	75.36 (2.10)	4.40	−1.60 (0.60)	—
Training	Pre	5-6	5.64 (0.50)	0.25	−0.67 (0.60)	—
Reaction time in the right hand (ms)	Pre	0.401-0.420	0.41 (0.01)	0.00	0.65 (0.60)	2.44%
Reaction time in the right hand (ms)	Post	0.347-0.398	0.37 (0.02)	0.00	0.23 (0.60)	5.41%
Reaction time in the left hand (ms)	Pre	0.433-0.453	0.44 (0.01)	0.00	−0.18 (0.60)	2.27%
Reaction time in the left hand (ms)	Post	0.387-0.408	0.40 (0.01)	0.00	0.65 (0.60)	2.5%
Reactive time (sec)	Pre	12.57-12.88	12.78 (0.08)	0.01	−1.15 (0.60)	0.63%
Reactive time (sec)	Post	10.34-11.25	10.88 (0.32)	0.11	−0.35 (0.60)	2.94%
Reactive time with dribbling (s)	Pre	21.68-22.88	22.03 (0.43)	0.19	1.44 (0.60)	1.95%
Reactive time with dribbling (s)	Post	17.18-18.89	17.94 (0.45)	0.20	0.28 (0.60)	2.51%
CG[d]						
Age	Pre	18-20	18.50 (0.65)	0.42	0.98 (0.60)	—
Height	Pre	173-178	175.64 (1.55)	2.40	−0.17 (0.60)	—
Weight	Pre	72-77	75.57 (1.79)	3.19	−1.13 (0.60)	—
Training	Pre	5-6	5.43 (0.51)	0.26	0.32 (0.60)	—
Reaction time in the right hand (ms)	Pre	0.401-0.420	0.410 (0.006)	0.00	−0.34 (0.60)	1.46%
Reaction time in the right hand (ms)	Post	0.382-0.401	0.394 (0.004)	0.00	−1.15 (0.60)	1.02%
Reaction time in the left hand (ms)	Pre	0.433-0.453	0.444 (0.007)	0.00	−0.18 (0.60)	1.58%
Reaction time in the left hand (ms)	Post	0.401-0.445	0.425 (0.013)	0.00	−0.29 (0.60)	3.06%
Reactive time (s)	Pre	12.68-12.88	12.83 (0.05)	0.00	−1.76 (0.60)	0.39%
Reactive time (s)	Post	11.73-12.21	11.98 (0.14)	0.02	−0.02 (0.60)	1.17%
Reactive time with dribbling (s)	Pre	21.76-22.88	22.05 (0.43)	0.18	1.48 (0.60)	1.95%
Reactive time with dribbling (s)	Post	19.84-21.45	20.34 (0.46)	0.21	0.91 (0.60)	2.26%

[a]CV: coefficient of variation.

[b]EG: experimental group.

[c]Not applicable.

[d]CG: control group.

XSL•FO
RenderX

Paired Sample *t* Test

Table 3 shows the results of the assessment. The statistical analysis of the obtained scores indicates significant superiority of the EG over the CG for all the assessments carried out. Significant improvements were observed in the EG, as the differences in performance pre- and postmeasurements were higher (P<.001). The Cohen *d* of all the tests was significantly higher for the EG, ranging between 0.87 and 0.98. This means that the group conducted an experiment and made a significant level of achievements, as shown by the size of the effect. Specifically, the CG had a comparatively lower magnitude of change in the outcome variables. The effect size for the CG varied from 0.71 to 0.97, which is much lower than for the EG. The steady increase in the difference of improvement between the EG and the CG adds credence to the training program conducted for the EG.

Table . Statistical analysis of paired sample *t* test for reaction time and reactive time.

Outcome measures and group	Test	Mean difference (SD)	SE	95% CI	*t* test (df)	*P* value	Cohen *d*[a]
Reaction time in the right hand (ms)							
EG[b]	Pre-post	0.04 (0.02)	0.00	0.03-0.05	9.22 (13)	<.001	0.87
CG[c]	Pre-post	0.016 (0.009)	0.002	0.01-0.021	7.10 (13)	<.001	0.79
Reaction time in the left hand (ms)							
EG	Pre-post	0.05 (0.01)	0.00	0.04-0.05	19.07 (13)	<.001	0.97
CG	Pre-post	0.019 (0.012)	0.003	0.012-0.03	5.71 (13)	<.001	0.71
Reactive time (s)							
EG	Pre-post	1.90 (0.31)	0.08	1.72-2.08	22.70 (13)	<.001	0.98
CG	Pre-post	0.85 (0.15)	0.04	0.76-0.94	20.52 (13)	<.001	0.97
Reactive time with dribbling (s)							
EG	Pre-post	4.08 (0.67)	0.18	3.70-4.47	22.85 (13)	<.001	0.98
CG	Pre-post	1.71 (0.49)	0.13	1.43-1.99	13.04 (13)	<.001	0.93

[a]d: effect size.

[b]EG: experimental group.

[c]CG: control group.

Independent *t* Test

In the analysis of the data presented in Table 4, while comparing the performance of the 2 groups, the premeasurements suggest that both the EG as well as the CG offered similar performance in all the tests conducted. This similarity at the start of the study means there are no differences in the visual-motor interaction and dribbling speed between the 2 groups that would give one an unreasonable edge over the other. On the other hand, the post measures showed significant differences in the performance of the EG in comparison with the CG. This improvement is evidenced by considerably better performance in all the assessed tests. Cohen *d* is ranging between 0.51 and 0.88. The probability values post measurements work as pointers toward a fair and highly significant improvement that is not arbitrary and that is attributed to the revelation of a structured training program intended for the improvement of visual-motor coordination and dribbling speed among the female basketball players.

XSL•FO
RenderX

Table . Statistical analysis of independent *t* test for reaction time and reactive time for experimental group and control group.

Outcome measures and test	Levene test		*t* test		Mean difference (SE)	95% CI	Cohen *d*
	F test (df)	*P* value	*t* test (df)	*P* value			
Reaction time in the right hand (ms)							
Pre	0.14 (26)	.71	1.18 (26)	.25	0.00 (0.00)	−0.01 to 0.00	0.05
Post	26.49 (26)	<.001	5.19 (26)	<.001	−0.02 (0.00)	−0.03 to −0.01	0.51
Reaction time in the left hand (ms)							
Pre	0.00 (26)	>.99	0.00 (26)	>.99	0.00 (0.00)	−0.01 to 0.01	0.00
Post	11.41 (26)	<.001	7.91 (26)	<.001	−0.03 (0.00)	−0.04 to −0.02	0.71
Reactive time (s)							
Pre	2.25 (26)	.15	1.89 (26)	.07	−0.05 (0.03)	−0.10 to 0.00	0.12
Post	13.32 (26)	<.001	11.66 (26)	<.001	−1.10 (0.09)	−1.29 to −0.91	0.84
Reactive time with dribbling (s)							
Pre	0.01 (26)	.93	0.15 (26)	.88	−0.02 (0.16)	−0.36 to 0.31	0.00
Post	0.002 (26)	.97	13.88 (26)	<.001	−2.39 (0.17)	−2.75 to −2.04	0.88

Discussion

Principal Findings

The purpose of this study was to evaluate the effectiveness of a basketball training program involving the use of FITLIGHT technology on female basketball players' visual-motor coordination and dribbling speeds. A comparative analysis of the results of each of the tests indicated the positive impact of the experimental training program on the aspects of visual-motor coordination and dribbling. When evaluating the results obtained before and after the implementation of the program for the EG and CG, statistically meaningful changes throughout the assessments in this research were evident. The EG's achievements in all the assessments were not only better but were statistically greater than the CG achievements. In addition, higher Cohen *d* values for the EG supported and highlighted not only the basic statistical evidence of the treatment's effect but also quantified the impact of the FITLIGHT training program on the improvements in the visual-motor coordination levels and dribbling speeds of the female basketball players.

The study has demonstrated that FITLIGHT exercises enhance visual-motor coordination along with dribbling speeds, primarily because they engage light stimuli that improve motor skills and reaction times. The specific brain centers that process visual information and control motor responses become activated because of this enhancement. FITLIGHT technology demonstrates the successful development of reactive coordination and movement combination capacity in basketball players since EGs showed more notable gains than CGs [24]. The application of light-based training methods in soccer enhances visual and visuomotor capabilities through improved performance in dynamic visual acuity and sensory reaction time, according to the findings of Rodrigues et al [25]. Small-sided games that incorporate FITLIGHT training have proven effective for improving basketball performance at both harmonic abilities and basic skills to benefit a wide range of sports training [9]. FITLIGHT, during reactive agility training, produced noteworthy improvements in both visual reaction times and dribbling skills because of the system's success in athlete performance strengthening [11].

The visual cortex processing of visual information is enhanced for athletes following their engagement with FITLIGHT exercises because these workouts activate the visual cortex. The available research demonstrates that reaction times, together with cognitive functionality and motor skills, improve after such training becomes part of athletes' preparation programs across various sports like basketball, soccer, and tennis. A visual stimuli program led young soccer players to achieve an 11.8% enhancement of their reaction times alongside important cognitive improvement, according to research [26]. Tennis players who belonged to the 10-year-old category showed advancements in both reaction times and motor performance after using FITLIGHT technology for their visual training [15]. The research shows that athletic performance benefits from FITLIGHT exercises that activate the visual cortex [27].

Neuroplasticity enables significant neural efficiency enhancement through repetitive FITLIGHT training because it allows the brain to restructure and reinforce neural connections as a result of learning and exposure. Research shows that such educational programs enhance cognitive-motor abilities by improving athletes' reactive coordination, combined with better reaction times and visual-motor responses. The reactive abilities and dribbling skills of basketball players substantially improved after FITLIGHT training based on effect sizes that showed a large impact [11,24,28]. Judo athletes achieved better executive functions and motor performance results after completing FITLIGHT cognitive-motor training because the technology helps brains adapt sensory-motor integrative mechanisms [7,27].

The cerebellum functions to refine motor control for precise movements, especially when athletes need quick responses to visual cues when dribbling basketballs. Scientific evidence shows that the cerebellum has exceptional precision when encoding dynamic motor frequencies to produce uniform motor kinematics across all people needed for accurate and timed

XSL•FO
RenderX

activities [29-32]. When athletes take part in FITLIGHT sessions, their improved responses allow the cerebellum, together with the cerebral cortex, to adapt action precision and timing that result in better athletic performance [31]. The cerebellum plays additional tasks across motor control by affecting cognitive and emotional processes, leading to improved athlete performance during pressure situations [33]. The cerebellum plays a key role in complex motor tasks, together with its role in sports performance [34].

While the FITLIGHT system is not age-specific, it is used as a tool for determining response times, attention, concentration, physical activity, and overall fitness. The training program itself was developed to enhance reactive ability, hand-eye coordination, and the ability to perform movements in a direct and highly intensive manner using the FITLIGHT technologies. The application of these technologies in the framework of the experimental program was time-sensitive, which enabled the further evolution of the training methodology depending on the outcomes that the athletes achieved. As a result, the players improved their reaction to stimuli, and their control of movement patterns in basketball-specific training situations had higher accuracy and coordination. The results of our study are confirmed by previous studies [9,11,12,19,35-38]. The physical activity and the quality of a sport, together with the developmental progress of a player, are related to the improvements in response times that were also identified in this study. Reaction time plays an important role in different game conditions in which the athlete should make the correct decision in the shortest time to guarantee the successful execution of an action. Hence, improvements in response times are correlated positively with the quality and development of physical activity and sports.

Several papers have compared the effectiveness of basketball players' training and changes in training programs using FITLIGHT technology [24,39]. When incorporating visual stimuli using different techniques and parallel types of stimulation via touch and movement, the result was a positive increase in the athletic abilities that are demanded in sports disciplines [40,41]. Badau et al [42] integrated 3 months of sports training using FITLIGHT technology into open-skill sports for adolescents, such as basketball, handball, and volleyball. This led to the athletes having significantly faster cognitive reaction rates, which are a measure of cognitive flexibility [42]. The study by Silvestri et al [27] found that the use of FITLIGHT technology in training enhanced the executive functions without reducing the enjoyment aspect of training sessions. A separate study, which involved 154 male and female volleyball and basketball players [43], also revealed that using FITLIGHT technology helped to enhance the athletes' interactions.

Thus, this work substantially enriches the existing knowledge about the application of FITLIGHT technology to optimize visual-motor coordination and the dribbling speeds of female basketball players. Coordination and rapid reaction capabilities are crucial components in the training of basketball athletes; these abilities can be improved using FITLIGHT in a way that is compatible with the challenges that basketball players encounter. Designing a specific program based on the use of information technologies that concern visual-motor coordination and response rates can significantly enhance athletic abilities. According to FITLIGHT, this allows athletes to enhance their coordination and responsiveness in a setting that mimics real-time basketball training. Furthermore, it enhances motivation for training and ensures the achievement of mastery in the technical aspects of basketball. With practice in the visual-motor skills and the reactions that are needed in training being customized, athletes can complete complicated actions and then make sound, rapid decisions in sports. Therefore, our results support the feasibility of applying sophisticated training technologies to boost or elevate sports performance and overall basketball training.

Strength

The study benefits from its usage of FITLIGHT technology to introduce a modern method for enhancing both visual-motor coordination and dribbling speed abilities. The research design used controlled experiments to study the FITLIGHT intervention through comparisons between the EG and CG. Analysis methods containing complete statistical methods provided reliable findings across the 10-week training cycle. Near-instant data collection tools measured reaction and reactive times, which showed substantial improvements in the EG results. The research results indicate that structured training approaches effectively boost athlete competency alongside performance reliability for coaches to consider. The established training method minimizes performance alterations while helping female basketball players improve their skills.

Limitation

Sample Size

The research included just 28 participants, whereby 14 participants received FITLIGHT training in the EG while 14 participants were allocated as the CG. The study's restricted participant population reduces the ability of the results to extend to different female athletes in society. A bigger participant number, exceeding the current 28, would help increase the quality of study findings and deliver a more detailed view of FITLIGHT training effects.

Duration of the Training Program

The 10-week duration of training seemed inadequate to determine how FITLIGHT affects visual-motor coordination and ball control speed exclusively. The training program results could show different outcomes with prolonged training duration at each session.

Nature of the Sporting Context

The authors applied their research exclusively to basketball because the study investigated this sport specifically. The findings cannot extend to sports that require dissimilar visual-motor abilities and different skill requirements.

Recommendations and Future Studies

The investigation shows that training techniques incorporating FITLIGHT systems should be integrated into programs for female basketball athletes because this technology helps develop their visual-motor coordination while speeding up ball handling.

XSL•FO
RenderX

FITLIGHT allows players to receive personalized training plans from coaches who use this system to achieve balanced management of team and individual sessions. An extension of visual response time and motor coordination training will boost total player performance by strengthening the connection between the brain and muscles while minimizing athletic performance inconsistencies. Training and developing institutions should create instructions for FITLIGHT use and selection of advanced fitness technologies, which will enable coaches to enhance their performance results. The study will gain increased validity if future research analyzes more female basketball players through the same methodology while including additional comparative evaluation of male and female performance. The assessment should analyze how anthropometric markers (height and weight) affect performance results following a FITLIGHT implementation, along with evaluating extended training procedures (over 10 wk) to determine their effect on athletic skill advancement. The long-term effects of this technology on athletic performance and female athlete mental health need evaluation, while researchers should compare it with established visual-motor coordination improvement techniques. An investigation of how FITLIGHT training affects mental resilience and mental focus development during competitions needs further evaluation. Research involving this technology can establish a complete assessment of its performance-enhancing impact on athletic prowess and mental well-being.

Conclusions

The study investigated the power of FITLIGHT technology-based basketball training methods as they enhanced visual-motor coordination capabilities and ball control speed for female basketball athletes. The 10-week duration of the program generated statistical differences between those participating in the EG compared with the CG participants. The EG enhancements exceeded those of the CG because research participants achieved CV improvements between 0.63% and 5.41% during the intervention period, but the CG CV stayed in the 0.39% to 2.26% range. Data analysis through paired sample *t* tests demonstrated that the EG surpassed the CG in all assessments, while obtaining large effect sizes between 0.87 and 0.98 for the EG compared with 0.71 to 0.97 for the control CG. Independent *t* test analysis confirmed no preintervention differences, yet the intervention led to substantial and statistically important improvements for both groups, and Cohen *d* values spanned between 0.51 and 0.88. The training program specifically improves visual response time, together with visual-motor coordination capabilities. The obtained data support implementing systematic training programs anchored in modern methods, including FITLIGHT, for minimizing performance variability and enhancing skill consistency.

Acknowledgments

The author expresses deep gratitude to the female basketball players who joined the study and generously provided feedback, which led to the improvement of the intervention. This study was funded by the Deanship of Scientific Research at King Faisal University, Saudi Arabia (GRANT KFU252207).

Data Availability

The study-produced data cannot be made available to the public due to privacy laws although they can be accessed by reaching out to the primary author.

Authors' Contributions

The author led the research process starting from its development to funding, reviewing, data collection and review tasks. The author finalized his evaluation of this research work.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Ethical clearance.
[PDF File, 375 KB - games_v13i1e70519_app1.pdf]

Multimedia Appendix 2
Materials used for the visual-motor interaction test.
[DOCX File, 72 KB - games_v13i1e70519_app2.docx]

Multimedia Appendix 3
Training program for exercises using the Fitlight system.
[DOCX File, 45 KB - games_v13i1e70519_app3.docx]

XSL•FO
RenderX

References

1. Stojanović E, Stojiljković N, Scanlan AT, Dalbo VJ, Berkelmans DM, Milanović Z. The activity demands and physiological responses encountered during basketball match-play: a systematic review. Sports Med 2018 Jan;48(1):111-135. [doi: 10.1007/s40279-017-0794-z] [Medline: 29039018]
2. Shalom A, Gottlieb R, Alcaraz PE, Calleja-Gonzalez J. A narrative review of the dominant physiological energy systems in basketball and the importance of specificity and uniqueness in measuring basketball players. Appl Sci (Basel) 2023;13(23):12849. [doi: 10.3390/app132312849]
3. Montella M, Ceciliani A, Morsanuto S, Federici A, Altavilla G. Development of motor skills applied to basketball in the developmental age. J Hum Sport Exerc 2019;14. [doi: 10.14198/jhse.2019.14.Proc4.46]
4. Ben Abdelkrim N, El Fazaa S, El Ati J. Time-motion analysis and physiological data of elite under-19-year-old basketball players during competition. Br J Sports Med 2007 Feb;41(2):69-75. [doi: 10.1136/bjsm.2006.032318] [Medline: 17138630]
5. Sun W, Chee C, Kok L, Lim F, Samsudin S. Evaluation of differences in the performance strategies of top and bottom basketball teams utilizing rank-sum ratio comprehensive. Front Sports Act Living 2022;4:1052909. [doi: 10.3389/fspor.2022.1052909] [Medline: 36506722]
6. Pastrana-Brincones JL, Troyano-Gallegos B, Morillo-Baro JP, et al. Mixed methods in tactical analysis through polar coordinates and function estimation: the transition play in ACB basketball. Front Sports Act Living 2021;3:739308. [doi: 10.3389/fspor.2021.739308] [Medline: 34632379]
7. Campanella M, Cardinali L, Ferrari D, et al. Effects of Fitlight training on cognitive-motor performance in élite judo athletes. Heliyon 2024 Apr 15;10(7):e28712. [doi: 10.1016/j.heliyon.2024.e28712] [Medline: 38689957]
8. Myers LR, Toonstra JL, Cripps AE. The test–retest reliability and minimal detectable change of the FitLight trainer. International Journal of Athletic Therapy and Training 2023;28(2):84-88. [doi: 10.1123/ijatt.2022-0012]
9. Hassan AK, Alibrahim MS, Sayed Ahmed YAR. The effect of small-sided games using the FIT LIGHT training system on some harmonic abilities and some basic skills of basketball players. Front Sports Act Living 2023;5:1080526. [doi: 10.3389/fspor.2023.1080526] [Medline: 36761781]
10. Kotte H, Daiber F, Kravcik M, Duong-Trung N. FitSight: tracking and feedback engine for personalized fitness training. Presented at: Proceedings of the 32nd ACM Conference on User Modeling, Adaptation and Personalization (UMAP '24); Jul 1-4, 2024; Cagliari Italy p. 223-231. [doi: 10.1145/3627043.3659547]
11. Hassan AK, Alhumaid MM, Hamad BE. The effect of using reactive agility exercises with the FITLIGHT training system on the speed of visual reaction time and dribbling skill of basketball players. Sports (Basel) 2022 Nov 14;10(11):176. [doi: 10.3390/sports10110176] [Medline: 36422946]
12. Perroni F, Mol E, Walker A, et al. Reaction time to visual stimulus in firefighters and healthy trained subjects: a preliminary comparative study. Open Sports Sci J 2018 Nov 27;11(1):69-77. [doi: 10.2174/1875399X01811010069]
13. Oudejans RRD. Effects of visual control training on the shooting performance of elite female basketball players. Int J Sports Sci Coach 2012 Sep;7(3):469-480. [doi: 10.1260/1747-9541.7.3.469]
14. Clark J, Betz B, Borders L, Kuehn-Himmler A, Hasselfeld K, Divine J. Vision training and reaction training for improving performance and reducing injury risk in athletes. J Sports Perform Vis 2020;2(1):e8-e16. [doi: 10.22374/jspv.v2i1.4]
15. Forni F, Farinini E, Leardi R, Rinaldo A. Effects of visual training on motor performance in young tennis players using FitLight trainer. J Sports Med Phys Fitness 2022;62(4):585-592. [doi: 10.23736/S0022-4707.21.12145-0]
16. Horníková H, Jeleň M, Zemková E. Determinants of reactive agility in tests with different demands on sensory and motor components in handball players. Appl Sci (Basel) 2021;11(14):6531. [doi: 10.3390/app11146531]
17. Jin P, Li X, Ma B, Guo H, Zhang Z, Mao L. Dynamic visual attention characteristics and their relationship to match performance in skilled basketball players. PeerJ 2020;8:e9803. [doi: 10.7717/peerj.9803] [Medline: 32879809]
18. Candra O. The contribution of eye-hand coordination to basketball lay up shoot skills. Presented at: 1st Progress in Social Science, Humanities and Education Research Symposium (PSSHERS 2019); 2019; Padang, Indonesia p. 864-869. [doi: 10.2991/assehr.k.200824.192]
19. Lynall RC, Blackburn JT, Guskiewicz KM, Marshall SW, Plummer P, Mihalik JP. Reaction time and joint kinematics during functional movement in recently concussed individuals. Arch Phys Med Rehabil 2018 May;99(5):880-886. [doi: 10.1016/j.apmr.2017.12.011] [Medline: 29337022]
20. Candra O, Asmawi M, Tangkudung J. The effect leg muscle explosive power, flexibility, hand eye coordination and confidence of skill lay up shoot basketball. Jipes - J Indones Phys Educ Sport 2017;3(2):162-192. [doi: 10.21009/JIPES.032.13]
21. Mangine GT, Hoffman JR, Wells AJ, et al. Visual tracking speed is related to basketball-specific measures of performance in NBA players. J Strength Cond Res 2014 Sep;28(9):2406-2414. [doi: 10.1519/JSC.0000000000000550] [Medline: 24875429]
22. Zurek M, Comi S, Cicchelli A, Roi GS. Simple and complex reaction time at visual stimulation, before and after a rehabilitation after knee surgery in football players. 2015 Presented at: International Conference on Sports Rehabilitation and Traumatology URL: https://cris.unibo.it/handle/11585/467770 [accessed 2024-10-08]
23. Kang H. Sample size determination and power analysis using the G*Power software. J Educ Eval Health Prof 2021;18:17. [doi: 10.3352/jeehp.2021.18.17] [Medline: 34325496]

XSL•FO
RenderX

24. Steff N, Badau D, Badau A. Study on the impact of implementing an exercise program using Fitlight technology for the development of upper limb coordinative abilities in basketball players. Sensors (Basel) 2024 May 28;24(11):3482. [doi: 10.3390/s24113482] [Medline: 38894272]
25. Rodrigues P, Woodburn J, Bond AJ, Stockman A, Vera J. Light-based manipulation of visual processing speed during soccer-specific training has a positive impact on visual and visuomotor abilities in professional soccer players. Ophthalmic Physiol Opt 2025 Mar;45(2):504-513. [doi: 10.1111/opo.13423] [Medline: 39564808]
26. Theofilou G, Ladakis I, Mavroidi C, et al. The effects of a visual stimuli training program on reaction time, cognitive function, and fitness in young soccer players. Sensors (Basel) 2022 Sep 3;22(17):6680. [doi: 10.3390/s22176680] [Medline: 36081136]
27. Silvestri F, Campanella M, Bertollo M, et al. Acute effects of Fitlight training on cognitive-motor processes in young basketball players. Int J Environ Res Public Health 2023 Jan 1;20(1):817. [doi: 10.3390/ijerph20010817] [Medline: 36613139]
28. Kim HJ, Lee JW, Choi G, Huh J, Han DH. Differences in brain activity and body movements between virtual reality and offline exercise: randomized crossover trial. JMIR Serious Games 2023 Jan 5;11:e40421. [doi: 10.2196/40421] [Medline: 36602842]
29. Widhalm K, Maul L, Durstberger S, Putz P, Klupper C, Werner F. Real-time digitized visual feedback in exercise therapy for lower extremity functional deficits: qualitative study of usability factors during prototype testing. JMIR Serious Games 2024 Dec 10;12:e51771. [doi: 10.2196/51771] [Medline: 39658006]
30. Pan MK, Liu CW, Chen SY, et al. The cerebellum computes frequency dynamics for motions with numerical precision and cross-individual uniformity. Research Square. Preprint posted online on 2024. [doi: 10.21203/rs.3.rs-4615547/v1]
31. Kim LH, Heck DH, Sillitoe RV. Cerebellar functions beyond movement and learning. Annu Rev Neurosci 2024 Aug;47(1):145-166. [doi: 10.1146/annurev-neuro-100423-104943] [Medline: 38663092]
32. Herren S, Seebacher B, Mildner S, et al. Exergame (ExerG)-based physical-cognitive training for rehabilitation in adults with motor and balance impairments: usability study. JMIR Serious Games 2025 Feb 14;13:e66515. [doi: 10.2196/66515] [Medline: 39951650]
33. Rudolph S, Badura A, Lutzu S, et al. Cognitive-affective functions of the cerebellum. J Neurosci 2023 Nov 8;43(45):7554-7564. [doi: 10.1523/JNEUROSCI.1451-23.2023] [Medline: 37940582]
34. Caccianiga G, Mooney RA, Celnik PA, Cantarero GL, Brown JD. Anodal cerebellar t-DCS impacts skill learning and transfer on a robotic surgery training task. Sci Rep 2023;13(1):1394. [doi: 10.1038/s41598-023-47404-1]
35. Hadžović MM, Đorđević SN, Jorgić BM, et al. Innovative protocols for determining the non-reactive agility of female basketball players based on familiarization and validity tests. Appl Sci (Basel) 2023;13(10):6023. [doi: 10.3390/app13106023]
36. Badau D, Badau A. Optimizing reaction time in relation to manual and foot laterality in children using the Fitlight technological systems. Sensors (Basel) 2022 Nov 14;22(22):8785. [doi: 10.3390/s22228785] [Medline: 36433379]
37. Katanić B, Ilić P, Stojmenović A, Vitasović M. The application of Fitlight trainer system in sports. Fizička kultura 2020;74(2):115-126. [doi: 10.5937/fizkul74-27189]
38. Shelly Z, Stewart E, Fonville T, et al. Helmet prototype response time assessment using NCAA division 1 collegiate football athletes. Int J Kinesiol Sports Sci 2019;7(4):53. [doi: 10.7575/aiac.ijkss.v.7n.4p.53]
39. Hurwitz J, Rich D, Agnew C, Harry J. Workload management strategies to optimize athlete performance in collegiate men's and women's basketball. Int'l Journal of Strength & Conditioning 2022;2(1). [doi: 10.47206/ijsc.v2i1.136] [Medline: 32241487]
40. Martin-Niedecken AL, Bucher V, Adcock M, de Bruin ED, Schättin A. Impact of an exergame intervention on cognitive-motor functions and training experience in young team sports athletes: a non-randomized controlled trial. Front Sports Act Living 2023;5:1170783. [doi: 10.3389/fspor.2023.1170783] [Medline: 38077282]
41. Rogozhnikov MA, Baturin AE, YuV Y, Kuritsyna AE. FITLIGHT' training system benefits for neuromuscular control training in basketball. Theory Pract Phys Cult 2020;9:10-12 [FREE Full text]
42. Badau D, Badau A, Ene-Voiculescu C, et al. The impact of implementing an exergame program on the level of reaction time optimization in handball, volleyball, and basketball players. Int J Environ Res Public Health 2022 May 5;19(9):5598. [doi: 10.3390/ijerph19095598] [Medline: 35564993]
43. Willberg C, Kohler A, Zentgraf K. Construct validity and applicability of a team-sport-specific change of direction test. J Hum Kinet 2022 Dec;85:115-126. [doi: 10.2478/hukin-2022-0115] [Medline: 36643841]

Abbreviations

CG: control group
CV: coefficient of variation
EG: experimental group

XSL•FO
RenderX

Edited by A Coristine; submitted 23.12.24; peer-reviewed by A Badau, Y Tabrizi; revised version received 04.06.25; accepted 05.06.25; published 04.07.25.

Please cite as:
Hassan AK
FITLIGHT Training and Its Influence on Visual-Motor Reactions and Dribbling Speed in Female Basketball Players: Prospective Evaluation Study
JMIR Serious Games 2025;13:e70519
URL: https://games.jmir.org/2025/1/e70519
doi:10.2196/70519

© Ahmed K Hassan. Originally published in JMIR Serious Games (https://games.jmir.org), 4.7.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Insights on the User Experience and Feasibility of an Electromyography-Driven Exergame Combined With Blood Flow Restriction for Strength Training in Hospitalized Older Adults: Mixed Methods Randomized Controlled Feasibility Study

Ruben Debeuf[1,2,3], MSc; Reinhard Claeys[1,2,3], MSc; Margo Berlanger[1], BSc; Myrthe Bunt[1], BSc; Aziz Debain[4], PhD; Daan De Vlieger[1,2,3,5], MSc; Matthias Eggermont[1], MSc; Mahyar Firouzi[1,2,3,6], PhD; Stefania Guida[1,7], MSc; Katarína Kostková[8], PhD; Siddhartha Lieten[4], PhD; Lubos Omelina[8], PhD; Silvia Zaccardi[1,8], MSc; Bart Jansen[8], PhD; Eva Swinnen[1,2,3], PhD; David Beckwée[1,3,9], PhD

[1]Rehabilitation Research Group, Department of Physiotherapy, Human Physiology and Anatomy, Faculty of Physical Education and Physiotherapy, Vrije Universiteit Brussel, Laarbeeklaan 121, Brussels, Belgium

[2]Center for Neurosciences (C4N), Vrije Universiteit Brussel, Brussels, Belgium

[3]Brussels Human Robotic Research Center (BruBotics), Vrije Universiteit Brussel, Brussels, Belgium

[4]Geriatrics Department, Universitair Ziekenhuis Brussel, Brussels, Belgium

[5]Department of Rehabilitation Sciences, Ghent University, Ghent, Belgium

[6]Brain, Body and Cognition, Department of Psychology and Educational Sciences, Vrije Universiteit Brussel, Brussels, Belgium

[7]Unit of Clinical Epidemiology, Istituto di Ricovero e Cura a Carattere Scientifico Ortopedico Galeazzi, Milan, Italy

[8]Department of Electronics and Informatics, Vrije Universiteit Brussel, Brussels, Belgium

[9]Research Group Movement Antwerp (MOVANT), Department of Rehabilitation Sciences and Physiotherapy, University of Antwerp, Antwerp, Belgium

Corresponding Author:
Eva Swinnen, PhD
Rehabilitation Research Group, Department of Physiotherapy, Human Physiology and Anatomy, Faculty of Physical Education and Physiotherapy, Vrije Universiteit Brussel, Laarbeeklaan 121, Brussels, Belgium

Abstract

Background: Hospitalized older adults often spend prolonged periods of time bedridden, leading to decreased muscle strength and function. To tackle this, rehabilitation aims to keep patients active and train affected muscles. Exergames have proven to be effective in the rehabilitation of different patient populations and offer a motivating solution to combat inactivity associated with hospitalization. Furthermore, blood flow restriction (BFR) is effective in therapy for weakened patients, so combining BFR and exergames might be promising.

Objective: As part of an iterative process of user-centered development, this mixed method study investigates the acceptability and feasibility of the Ghostly game as a stand-alone added therapy or combined with BFR in strength training of hospitalized older adults.

Methods: A mixed methods study was conducted on 15 hospitalized older adults. Participants were randomized into 3 groups and received daily interventions from the moment they were included in the geriatric ward, until discharge from the hospital. The Ghostly group received daily conventional therapy with the Ghostly game as added therapy, the Ghostly + BFR group received daily conventional therapy with Ghostly in combination with BFR as added therapy and last, the control group received daily conventional therapy with dose-matched isometric exercises as added therapy. The primary outcome, user experience, was assessed before discharge from the hospital using the Usefulness, Satisfaction, and Ease of Use questionnaire and through expert observations. Clinical outcomes such as muscle strength, muscle architecture, and segmental body composition were assessed at baseline and before discharge from the hospital to test the feasibility of the research protocol in preparation for future randomized controlled trials.

Results: A total of 15 hospitalized older adults (11 female participants, 73.33%) were included in this study with an average age of 84.53 (range: 78 - 94) years. Participants received an average of 3.47 (range: 3 - 5) intervention sessions after transferring to the geriatric ward of the hospital. Results on user experience revealed high scores on all subcategories of the Usefulness, Satisfaction, and Ease of Use questionnaire (usefulness: 78.93%, ease of use: 82.99%, ease of learning: 85.36%, and satisfaction: 87.55%). Furthermore, expert observations identified issues with color contrast, reaction time speed, and the need to tailor the

XSL•FO
RenderX

game to accommodate the diverse requirements of different patient populations. All outcomes and procedures were found feasible for a future randomized controlled trial.

Conclusions: This mixed methods study combines the innovative aspects of an electromyography-driven exergame with strength training principles of BFR and reveals the acceptability and feasibility of the Ghostly game as a stand-alone added therapy modality for strength training in hospitalized older adults and in combination with BFR. Future improvements of the exergame could focus on addressing expert-identified issues, including optimizing color contrast, adjusting reaction time speeds, and tailoring the game to meet the needs of different patient populations.

Trial Registration: ClinicalTrials.gov NCT05258500; https://clinicaltrials.gov/study/NCT05258500

(JMIR Serious Games 2025;13:e69400) doi:10.2196/69400

KEYWORDS

exergames; older adults; blood flow restriction; feasibility; user experience

Introduction

Hospitalization in older adults poses a high risk for functional decline due to acute inflammatory conditions, prolonged periods of bed rest, and reduced activity [1,2]. They are especially prone to inactivity and sedentary behavior during nontherapy time, as time spent in scheduled rehabilitation is often limited [3,4]. To engage older adults in rehabilitation, especially during nontherapy time, it is important to increase patient motivation to train. Evidence already showed that even short periods of disuse (ie, 10 d) induce a decline in rates of muscle protein synthesis and a rise in muscle protein breakdown, which, in turn, results in a rapid rate of muscle atrophy over the first few days of disuse [5]. Furthermore, in an older population, we see a reduced capacity to regain lost muscle tissue during recovery from disuse [5]. Moreover, this loss of functional independence, often attributed to muscle mass loss, is known as hospital-associated disability and affects 30% of hospitalized adults aged 65 years and older [6,7]. Recognizing this reduced functional capacity is detrimental to implementing evidence-based interventions to preserve and potentially improve outcomes in hospitalized older adults. Strategies to mitigate these effects of disuse include targeted physical activity [8]. However, older adults, especially those who are bedridden, are often unable to perform high-load, dynamic functional exercises, further limiting recovery potential. Nonetheless, low-load interventions, such as isometric strength exercises in combination with blood flow restriction (BFR), might provide a feasible solution for bedridden patients to commence rehabilitation in an early stage of hospitalization [9-11]. BFR entails the partial occlusion of blood flow towards the exercising limb, which results in a partial inflow of arterial blood and complete restriction of venous outflow [12,13]. This novel strength training modality has proven to have beneficial effects on muscle strength compared with conventional low-load strength training in older adults [14]. Evidence on the safety of BFR in adults aged 65 years and older shows that low-load BFR training increases heart rate and blood pressure (systolic and diastolic) to levels similar to those observed during conventional high-load training, indicating no greater risk [15]. However, to minimize the risk of adverse events and ensure safe training, contraindications such as poor circulatory health, open fractures, and infections should be carefully considered [12]. Despite the evidence on the safety and effectiveness of BFR, motivation to exercise remains an important limiting factor.

A novel therapy modality that has proven effects on motivation is exergames for rehabilitation. Exergames combine exercise with the motivational aspect of interactive video games [16]. In older adults, it was found that exergames have a positive effect on physical function [17]. Additionally, Rytterström et al [18] reviewed the feasibility and practical implications of exergames in this population and reported that older adults found exergaming physically and cognitively demanding, yet enjoyable. Focusing on physical function in older adults, the systematic review and meta-analysis of Pacheco et al [19] showed evidence that exergames can improve balance and mobility and motivate patients to keep performing the exercise. However, the combination of BFR and rehabilitation technology, such as exergames, is not addressed in the existing literature.

To provide a solution to the earlier-mentioned issues of inactivity and low motivation to train in nontherapy time and the need for low-load strength training in bedridden patients, we developed a prototype version of an electromyography (EMG)-driven exergame, the Ghostly game. The Ghostly game is a novel isometric strength training tool for rehabilitation, cocreated based on the principles of user-centered design. The goal of user-centered design is to incorporate the knowledge from different experts in the field, as well as insights from the end users to maximize adoption, user experience, and clinical relevance [20]. The development process of the Ghostly game commenced in 2016 and is marked by progressive iterations and advancements from its inception through various stages of development, testing, and refinement. The main goal for the Ghostly game throughout this entire process remains to incorporate evidence-based guidelines on strength training and rehabilitation in the game. As mentioned earlier, isometric strength training can prove relevant in the early stages of hospitalization [10,11], so guidelines on isometric strength training were incorporated in the level design to ensure effective state-of-the-art strength training. To further increase the potential effectiveness, it was decided to combine the Ghostly game with the evidence-based principles of BFR training to ensure positive effects during low-load isometric strength training.

Various game-based EMG biofeedback systems for rehabilitation have already been investigated, but evidence on clinical effectiveness remains limited [21,22]. A study by

XSL•FO
RenderX

Garcia-Hernandez et al [23] describes an EMG-controlled exergame to enhance grip strength. The exergame used in this study requires users to contract their muscles at 100% of their maximal voluntary contraction (MVC) to complete the different goals set in the game. A total of 16 healthy participants completed 2 weeks of strength training using the exergame, which resulted in a significant increase in grip strength. Additionally, Garcia-Hernandez et al [23] reported significantly higher intrinsic motivation in participants who trained with the exergame compared with participants performing conventional therapy. Most importantly, the results demonstrated that the exergames group adhered significantly better to the work-rest time proposed in the study, and the MVC was significantly higher in the exergames group compared with the conventional therapy group [23]. Enciso et al [24] focused on exergaming for people with lower limb impairments who use a wheelchair. The exergame "Workout on Wheels" is aimed at encouraging and facilitating exercise at home. Three games were incorporated within the app to strengthen the muscles of the upper limb using surface EMG sensors to register muscle contractions. Enciso et al [24] tested the feasibility of this exergame in 4 participants with varying levels of spinal cord injury and found that the participants perceived the game as highly useful and easy to use. Additionally, the results from the open-ended questionnaire revealed that the exergames helped participants exercise animatedly and engagingly and motivated them to work out. No clinical outcomes were detected for this study [24]. Both of these studies highlight the potential of an EMG-driven exergame but also indicate the need for more research into its effectiveness in rehabilitation.

This study investigates an EMG-driven exergame combined with evidence-based principles of BFR training for strength training in a hospital environment. As an early part of a larger iterative process of user-centered development of rehabilitation technology [20], the goal was to gather relevant insights on the needs and barriers of end users to maximize user experience and effective adoption in rehabilitation. The study focused on the development of the Ghostly game, as well as the innovative aspects of Ghostly combined with BFR. Additionally, the feasibility of the research protocol in preparation for a larger randomized controlled trial (RCT) was assessed. The final aim of this study is to evaluate the data gathered to determine whether an updated version of the Ghostly game is needed.

Methods

Study Design

This mixed method study used an RCT design with 3 study arms and focused on gaining insights into the acceptability and feasibility of the Ghostly game as an added therapy modality on top of the conventional therapy in strength training of hospitalized older adults. Qualitative and quantitative data were collected during and after the intervention period in which participants received their daily conventional therapy, with an added therapy modality based on group allocation. Patients were randomized into 3 groups using sealed envelopes. The first experimental group received the conventional therapy supervised by the physiotherapist of the geriatric ward, and the Ghostly game as added therapy supervised by the researchers. Similarly, the second experimental group received conventional therapy and the Ghostly game combined with BFR as an added therapy. Last, the control group received their conventional therapy and the added therapy of isometric exercises dose-matched to the Ghostly game in terms of repetitions, sets, and interset rest. All exercises were guided by the researchers and solely focused on both quadriceps muscles. The researchers instructed the participant to keep the lower limb fixed on the bed throughout the entire exercise session to ensure activation of the quadriceps muscle without external fixation of the lower limb. In our approach, we prioritized eliciting strong and consistent contractions of the quadriceps femoris. However, we did not implement any specific containment or isolation procedures to prevent the cocontraction of other lower limb muscles. This decision reflects our intent to maintain a more functionally relevant setup, as such cocontractions naturally occur during daily activities. The quadriceps were selected as the primary focus due to their critical role in functional movements such as standing, walking, and stair climbing, as well as their relatively large muscle mass compared with other lower limb muscles. Participants received daily interventions from the moment they entered the geriatric ward and were deemed fit for training with the Ghostly game by the health care professionals of the ward until discharge from the hospital. Data were collected at baseline and before discharge from the hospital.

Ethical Considerations

The study was conducted according to the guidelines of the Declaration of Helsinki [25], approved by the ethics committee of the University Hospital of Brussels (Eudamed: CIV-22-02-038967), and the details of the protocol were registered a priori (registration number: NCT05258500). All participants read and signed the informed consent form before inclusion in the study. All study data were anonymized to remove identifiable information of the participants and ensure privacy and confidentiality.

Participants

Patients were assessed for study eligibility by the physiotherapist of the geriatric ward on the day they were transferred to the geriatric ward and were included in this study if they met all of the following criteria: hospitalized, 65 years or older, and unable to perform 14 repetitions on the 30 seconds sit to stand test [26]. Patients were excluded from this study if they were unable to understand the basic instructions of the Ghostly game or had other severe conditions that prohibited them from performing lower limb strength training in combination with BFR. Relevant demographic information was collected from the medical files of the participants before discharge from the hospital.

Since the study was a feasibility study, no a priori sample size calculations were performed, and a sample of 15 participants was chosen based on practical considerations and the number of participants needed to reasonably evaluate feasibility goals and gather meaningful data on user experience [27].

XSL•FO
RenderX

Intervention

Ghostly

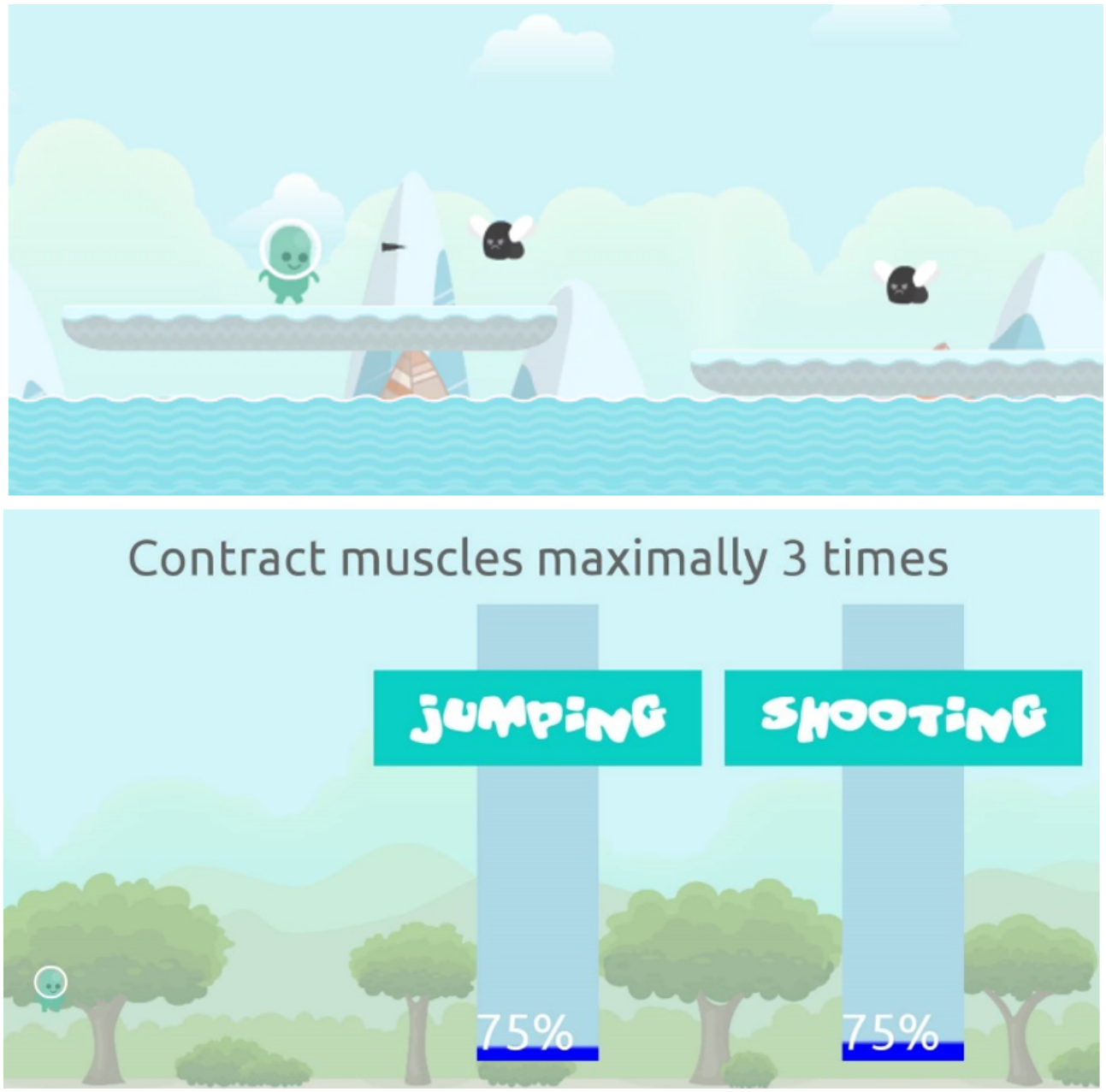
The Ghostly game (Figure 1) is an EMG-driven exergame in which the main character is controlled through muscle contractions measured by 2 surface EMG sensors (Figure 2) placed on the skin over the targeted muscles using double-sided adhesive tape. The Ghostly game was developed as an individually tailored strength training modality for bedridden patients that adheres to evidence-based guidelines on strength training. To complete levels in the Ghostly game, the participant needs to lead the main character to the end of the level by shooting enemies (ie, contraction of the muscle of one limb) and jumping over obstacles (ie, contraction of the muscle of the other limb). The duration of the contraction is based on the evidence-based guidelines that have been implemented in the in-game animations, which last on average one second. All levels were designed to require at least 12 contractions with each limb to reach the end. This way, as participants were asked to complete 3 levels, they performed 3 sets of 12 repetitions [28].

Since training at one's personal maximal capacity is a priority, the Ghostly game incorporates a baseline assessment level (Figure 1) in which the MVC of the user is measured while the participant remains in a supine position in bed with the legs extended. Two surface EMG sensors were placed over the belly of the target muscles of the participant by the researcher. First, the assessment level, consisting of 3 maximal contractions, with each contraction lasting 3 seconds and with 90 seconds of rest between each contraction, needs to be completed [29]. After the first maximal contraction, an MVC value is set to match the intensity of this first contraction. If subsequent contractions exceed the MVC value, the value is raised accordingly. Based on this MVC value, the intensity at which the user must contract the muscles to control the main character of the game is set to match an intensity of 75% of MVC [30,31] as this has been proven to increase muscle hypertrophy [30] and muscle strength [31] in isometric strength training. After completion of the assessment level, a training level was selected, and the participants started playing the game.

As the target population for this feasibility study consists of bedridden patients who benefit from lower limb isometric strength training in preparation for more functional gait training, the sensors were placed over the belly of the quadriceps muscles of both legs (Figure 2) following the SENIAM guidelines [32].

XSL•FO
RenderX

Figure 1. The Ghostly game and the assessment level (showing the intensity sliders) to set the training intensity of the game based on maximal voluntary contraction.



XSL•FO
RenderX

Figure 2. The surface electromyography sensors used to control the main character, and the placement of the surface electromyography sensors over the belly of the quadriceps muscle.





Blood Flow Restriction Therapy

The second therapy arm of this study involves the addition of BFR therapy to the Ghostly game. Cuffs (width: 10 cm, length: 58 cm) were placed over the proximal part of both lower limbs to partially occlude the blood flow towards the quadriceps muscles. To achieve the desired level of occlusion to the target muscle, a percentage of the arterial occlusion pressure (AOP) was used, measured by the SmartCuffs pro system (Smart Tools). AOP is defined as the amount of pressure required to completely cease arterial blood flow toward the limb [33]. To determine AOP, the cuffs are filled with air until a pressure is reached at which no arterial blood pressure is registered by the SmartCuffs pro system. This pressure is then set as the AOP, and all air within the cuffs is automatically released. Afterward, the cuffs are reinflated until a certain percentage of the

XSL•FO
RenderX

personalized AOP is reached. As pressures ranging from 40% to 80% of AOP have been proven to be effective in strength training [33,34], and with the goal of minimizing discomfort for the participants, an occlusion pressure equal to 50% of AOP was used in this feasibility study. A comprehensive overview of the complete intervention is given in Multimedia Appendix 1.

Outcome Measures

Primary Outcome: User Experience (USE)

The primary outcome of this study was the user experience of hospitalized older adults after playing the Ghostly game, which was assessed before discharge from the hospital using the Dutch and French versions of Usefulness, Satisfaction, and Ease of Use (USE) questionnaire [35]. This 30-item questionnaire is separated into 4 subcategories: usefulness, ease of use, ease of learning, and satisfaction. Each item is scored on a 7-point Likert scale, with a score of 1 indicating "strongly disagree" and a score of 7 indicating "strongly agree." A high score on the USE questionnaire indicates good usability of the device [36]. Results are reported as a mean percentage of the total score of the USE questionnaire, as well as a mean percentage of the scores of each subcategory. The USE questionnaire shows high reliability and validity, in line with other usability questionnaires [36]. Both the Dutch and French versions of this questionnaire have been used in research previously [37-39]. On top of this, open questions were added to the questionnaire that focused on relevant changes that should be made to the Ghostly game.

Secondary Outcome: End User Feedback and Expert Observations

To ensure all data on user experience and important barriers were gathered, a structure similar to think-aloud protocols was used in which expert observations and patient feedback were documented through field notes [40,41]. A think-aloud protocol aims to gather information on how the participants interact with the intervention by asking them to think aloud as a means of capturing the problem-solving process and refining the intervention to be more effective [41]. In this study, the researchers and experts observed and noted the thoughts and experiences of the participants while they interacted with the Ghostly game.

Patient feedback and real-time feedback while engaging with the Ghostly game, either as a stand-alone added therapy or combined with BFR, were collected throughout all intervention sessions. Comments and recommendations were documented using field notes and subsequently compared across participants.

Expert observations by different health care professionals (ie, physiotherapists and physicians), engineers, and researchers were conducted at various time points throughout the study to identify key barriers and elements within the game that require a change to maximize user experience. The experts followed multiple participants across different sessions to gather data on both difficulties during the initial learning phase, as well as difficulties in later sessions when participants had become familiar with the game. Experts focused on game design, difficulties while playing, relevance for rehabilitation, etc. Furthermore, the experts prioritized gathering input on the feasibility of applying and explaining the Ghostly game and BFR therapy in relation to future unsupervised training with the game.

Feasibility of the Intervention and Clinical Outcomes

The feasibility of the clinical study protocol was assessed as a preliminary step in preparation for a fully powered RCT. For this, the following clinical outcomes were measured: muscle strength of the quadriceps muscle, muscle architecture of the rectus femoris muscle, and body composition of both lower limbs at baseline (T0) and before discharge from the hospital (T1). Muscle strength was measured using the MicroFET 2 handheld dynamometer (Hoggan Health) and the 30-second sit-to-stand test [26]. Muscle architecture (cross-sectional area [CSA], muscle thickness, and subcutaneous fat) was measured using the Viamo sv7 ultrasound device (Canon Medical Systems, Japan). And last, body composition was measured using the Quadscan 4000 bio-impedance analysis device (Bodystat, Douglas). Because of the aim of the Ghostly and BFR interventions (targeting muscle mass loss), we were specifically interested in appendicular lean mass (ALM). To calculate ALM, the following formula was used:

$$ALM = 0.827 + 0.19 * (length^2/impedance\ 50\ Hz) + 2.101 * sex + 0.079 * weight$$

where sex equals 0 for female participants and 1 for male participants [42].

The feasibility criteria for the success of the study protocol were for all participants to be able to complete the intervention. Full adherence to the study protocol is reached when participants complete daily interventions from baseline until discharge from the hospital. Furthermore, 80% or more of participants should complete all primary and secondary measurements. Last, if adverse events occurred, they were noted. The premises of the threshold were set to ensure sufficient therapy adherence and compliance in preparation for the future larger RCT.

Statistical Analyses

The primary outcome measures were analyzed using Microsoft Excel (Microsoft Office 2024). Mean scores, SDs, and medians were calculated for each subcategory of the questionnaire. All clinical outcomes were statistically analyzed using SPSS statistics software (version 28; IBM Corp). Descriptive statistics were used to determine the baseline values of the participants. Mean difference and 95% CIs were calculated. Last, box plots of the most relevant clinical outcomes were created to visualize the distribution of the data.

Results

Participant Characteristics

In total, 15 participants (11 female participants, 73.33%), 5 in each study arm, were included in this feasibility study with an average age of 84.53 (range: 78 - 94) years (Figure 3). Participants received an average of 3.47 (range: 3 - 5) intervention sessions after transferring to the geriatric ward of the hospital. Table 1 gives an overview of the baseline characteristics of the participants.

XSL•FO
RenderX

Figure 3. CONSORT diagram for the 3 intervention groups. BFR: blood flow restriction.

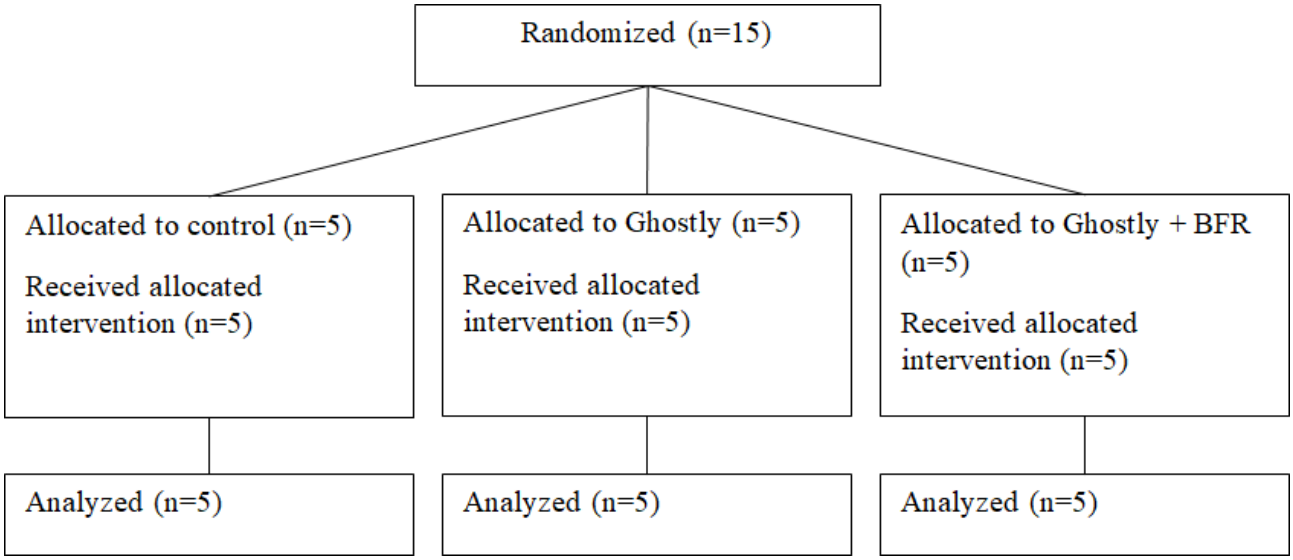


Table . Baseline characteristics of all participants.

Baseline characteristics of all participants (n=15)	Values
Age (years), mean (SD)	84.53 (4.53)
Sex, n (%)	
Female	11 (73.33)
Male	4 (26.67)
Number of intervention sessions, mean (SD)	3.47 (0.72)
30-second sit-to-stand (repetitions), mean (SD)	3.33 (2.55)
Quadriceps muscle strength (Newton), mean (SD)	75.75 (20.92)
Rectus femoris muscle thickness (cm), mean (SD)	0.87 (0.27)
Rectus femoris muscle cross-sectional area (cm²), mean (SD)	2.89 (1.26)
Lower limb subcutaneous fat (cm), mean (SD)	0.85 (0.52)
Lower limb lean mass (kg), mean (SD)	24.10 (5.20)

Primary Outcome: User Experience

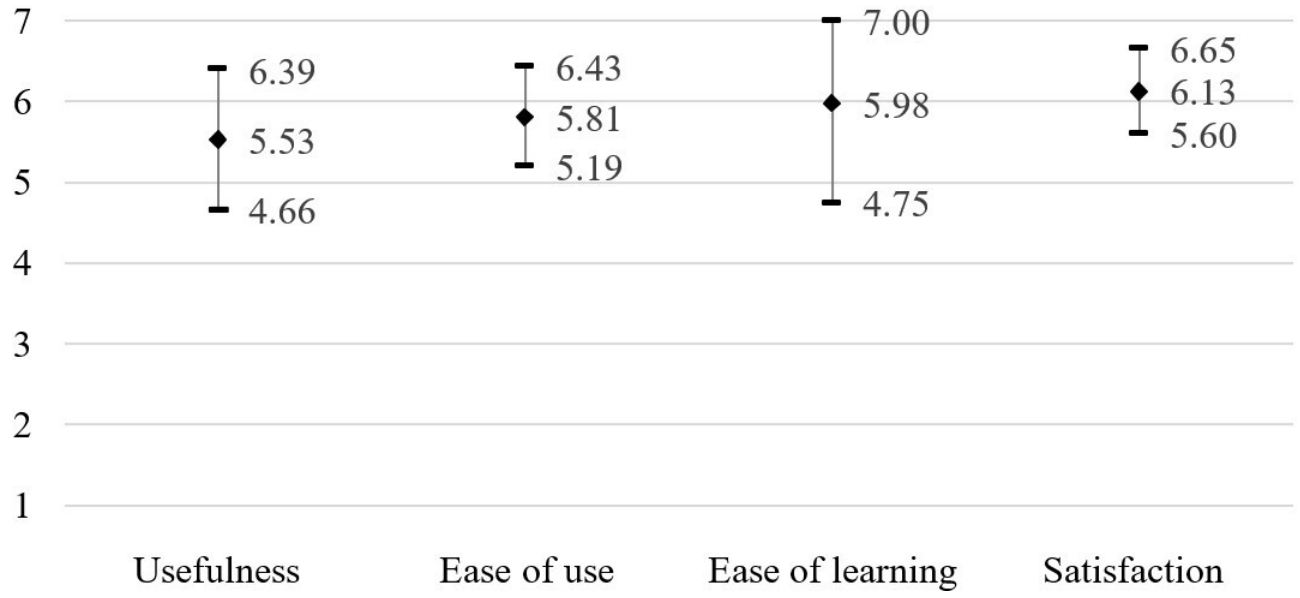
After completion of the intervention, all 10 participants who trained with the Ghostly game (study arms 1 and 2) completed the USE questionnaire. The results (Table 2 and Figure 4) show an overall high mean score on the USE questionnaire (83.29%). Dividing the results into the subcategories shows a mean score of 78.93% for usefulness, 82.99% for ease of use, 85.36% for ease of learning, and 87.55% for satisfaction, indicating a high mean score across all subcategories. When separating the results of the Ghostly group and the Ghostly + BFR group (Multimedia Appendix 2), an overall high mean score on the USE questionnaire can be found for both groups (Ghostly: 86.38%; Ghostly + BFR: 80.19%), implying no influence of BFR on user experience.

Table . Results of the Usefulness, Satisfaction, and Ease of Use questionnaire (n=10).

	Usefulness (n=56)	Ease of use (n=77)	Ease of learning (n=28)	Satisfaction (n=49)	Total (n=210)
Mean (SD)	44.20 (6.94)	63.90 (6.80)	23.90 (4.91)	42.90 (3.67)	174.90 (17.33)
Mean percentage (SD)	78.93 (12.39)	82.99 (8.84)	85.36 (17.53)	87.55 (7.50)	83.29 (8.25)
Median (IQR)	42.5 (13)	63 (9.8)	26 (8)	43.5 (5.8)	170 (28.8)

XSL•FO
RenderX

Figure 4. Mean scores for each subcategory of the Usefulness, Satisfaction, and Ease of Use questionnaire. Total mean scores for each subcategory were transformed to fit a 7-point Likert scale. The dot indicates the mean score for each subcategory, and the whiskers represent the mean score plus and minus 1 SD. A score of 7 indicates participants highly agree with the statement, and a score of 1 indicates that participants highly disagree with the statement.



Secondary Outcome: End User Feedback and Expert Observations

End User Feedback

First, at the start of the intervention period, many participants expressed concerns about their lack of knowledge and experience with video games and technology. They reported feeling unsure about their ability to complete or understand the intervention. However, by the end of the intervention period, this feeling of uncertainty disappeared as all participants reported that they felt like they were able to play and understand the game easily.

Second, difficulty perceiving the main character was often mentioned. Due to the color scheme used in the Ghostly game, participants experienced difficulties in discriminating the main character from the background. This often made it difficult to train efficiently and complete levels as they were unable to contract the muscle at the required time.

Last, participants who received Ghostly combined with BFR reported no discomfort during gameplay or postintervention, although participants mentioned feeling the cuff pressure during the intervention.

Expert Observations

In total, 5 experts observed the participants and collected their findings through field notes. First, experts noticed that older adults had difficulties with the reaction speed element integrated into the Ghostly game. As mentioned before, the participants are required to contract their muscles at the correct time for the main character to shoot or jump over obstacles. So, when users were unable to contract their muscles in time to avoid the obstacle, they were unable to complete the levels. In line with this, it was observed that when participants were unable to complete a level, the Ghostly game would restart the level. This led to a higher number of repetitions than initially prescribed and created a highly repetitive experience as participants continuously played the same level. Since participants were advised to complete 3 levels, after having to restart a level multiple times, it was sometimes unclear to participants how many levels still had to be completed before finishing the prescribed intervention session.

Second, with future unsupervised training in mind, the experts observed that participants understood the Ghostly game relatively quickly and understood the application of the sensors.

Next, experts noted that the difficulties experienced by the older adults in this study were specific to this population. They also observed variability within the group, with some participants struggling with fatigue, while others required additional time and support to learn and understand the Ghostly game. When expanding the Ghostly game beyond the population of older adults, the experts mentioned that a one-size-fits-all approach to game design will not fulfill the needs of all end users. Other patient populations might not experience the same difficulties and need different changes to enhance the usability of the game.

Last, the stability of game control occasionally posed challenges during the intervention. Specifically, 2 patients experienced difficulties with accurate registration of EMG signals by the Ghostly game during one of the intervention sessions, causing complications in completing certain levels. Nonetheless, despite these technical difficulties, all participants successfully completed the required intervention levels.

Feasibility of the Intervention and Clinical Outcomes

Feasibility of the Intervention

After completion of data collection, all participants (15 out of 15 participants) completed all daily intervention sessions. Moreover, 13 out of 15 participants (86.67%) completed all measurements at baseline and before discharge from the hospital. Two participants were unable to complete the measurements at discharge due to feeling unwell at the time of measurement.

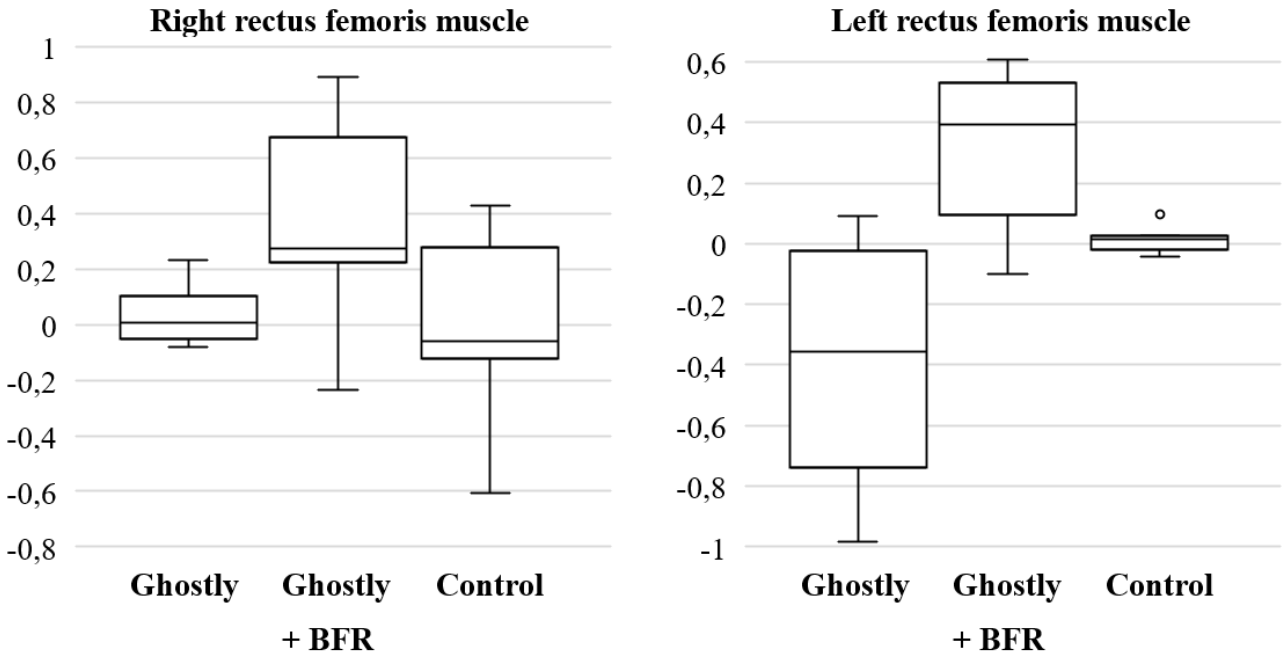
XSL•FO
RenderX

This feeling was unrelated to the intervention. Last, no adverse events were recorded during or after the intervention. Based on these findings, it can be concluded that the study protocol is feasible.

Clinical Outcomes

Clinical outcomes were assessed mainly focusing on the feasibility of the research protocol for a future fully powered RCT. Boxplots were used to visualize the distribution of the data on rectus femoris CSA (Figure 5) and quadriceps muscle strength (Figure 6) of both legs. After the intervention, the Ghostly + BFR group showed the largest mean increase in CSA of 0.34 cm (95% CI 0.08 to 0.59) for the rectus femoris muscle compared with other 2 groups. For muscle strength, a mean difference of 12.18 N (95% CI 2.43 to 21.92) was found for the quadriceps muscle in the Ghostly + BFR group. This was similar to the mean difference in muscle strength of both other groups, with 16.71 N (95% CI –12.09 to 45.51) for the quadriceps muscle in the Ghostly group, and 19.29 N (95% CI 11.70 to 26.88) for the quadriceps muscle in the control group. Results on rectus femoris muscle thickness revealed the largest increase in the Ghostly + BFR group with a mean difference of 0.01 cm (95% CI –0.12 to 0.13). Lower limb subcutaneous fat showed the largest mean difference in the control group, with a decrease of 0.02 cm (95% CI –0.07 to 0.10). Last, the largest mean difference in lower limb lean mass was found in the Ghostly + BFR group (mean difference: 1.83 kg, 95% CI –0.12 to 3.77). A full overview of the mean differences of all outcomes can be found in Table 3.

Figure 5. Boxplots of the mean difference in muscle cross-sectional area of the rectus femoris muscle for each group. The y-axis shows the mean difference in rectus femoris muscle cross-sectional area in cm. The x-axis shows the 3 intervention groups. BFR: blood flow restriction.



XSL•FO
RenderX

Figure 6. Boxplots of the mean difference in muscle strength of the quadriceps muscle for each group. The y-axis shows the mean change in quadriceps muscle strength in Newton. The x-axis shows the 3 intervention groups. BFR: blood flow restriction.

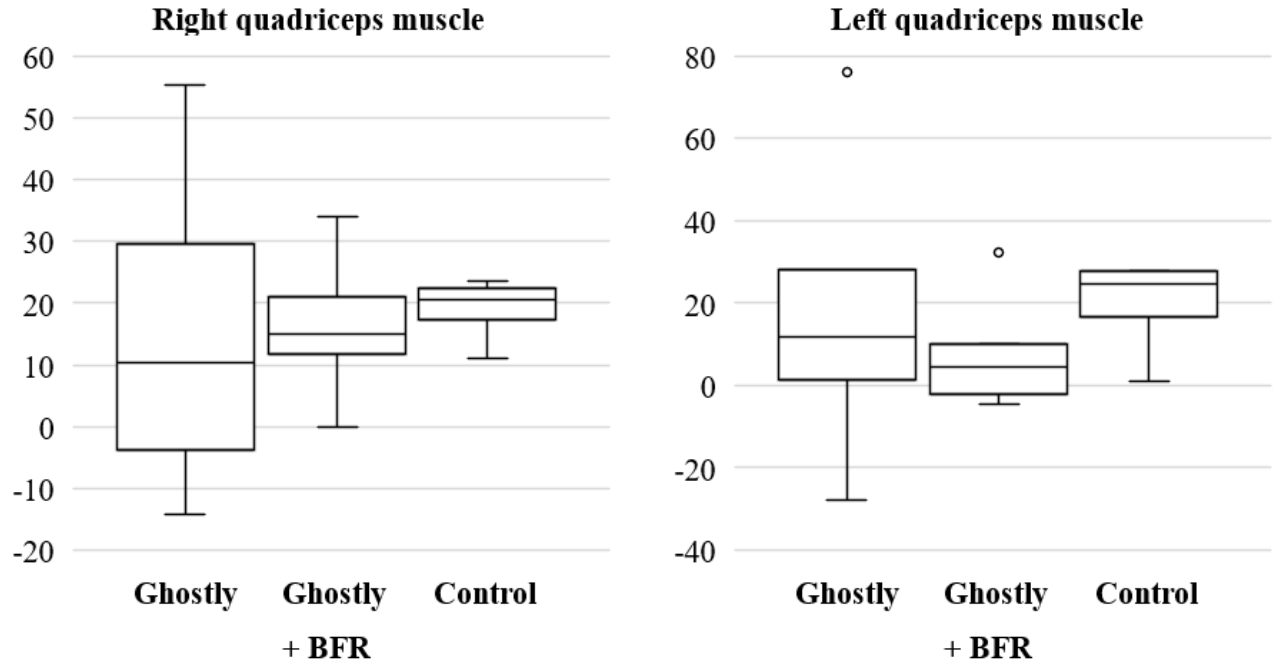


Table . Clinical outcome measures of participants separated into the 3 intervention groups. Mean difference and 95% CI of muscle characteristics for each group

	Ghostly (n=4)		Ghostly + BFR[a] (n=5)		Control (n=4)	
	Δ[b]	95% CI	Δ	95% CI	Δ	95% CI
30-second sit-to-stand (repetitions)	2.00	–0.25 to 4.25	1.00	–0.96 to 2.96	1.00	–0.30 to 2.30
Quadriceps muscle strength (Newton)	16.71	–12.09 to 45.51	12.18	2.43 to 21.92	19.29	11.70 to 26.88
Rectus femoris muscle thickness (cm)	–0.05	–0.18 to 0.08	0.01	–0.12 to 0.13	–0.03	–0.09 to 0.04
Rectus femoris muscle cross-sectional area (cm²)	–0.18	–0.53 to 0.17	0.34	0.08 to 0.59	–0.00	–0.20 to 0.19
Lower limb subcutaneous fat (cm)	0.01	–0.07 to 0.09	–0.03	–0.09 to 0.03	0.02	–0.07 to 0.10
Lower limb lean mass (kg)	0.84	–1.46 to 3.13	1.83	–0.12 to 3.77	0.94	–1.65 to 3.53

[a]BFR: blood flow restriction.

[b]Δ: mean difference from T0 to T1.

Discussion

Principal Findings

This mixed method feasibility study investigated the usability and feasibility of an EMG-driven exergame combined with BFR in relation to strength training in hospitalized older adults. Currently, no other studies have combined an EMG-driven exergame with BFR, highlighting the innovative aspect and relevance of our results in the field of exergames. This study is an early part of a larger iterative process of user-centered technology development. The main focus of this study was to gather information on the user experience of older adults playing the Ghostly game as a stand-alone added therapy or in combination with BFR, and how the game can be updated to further adhere to the needs and recommendations of the end users while simultaneously adhering to evidence-based guidelines for strength training. Second, clinical outcome measures were assessed to test the feasibility of the study protocol in preparation for a future larger RCT aimed at investigating the clinical effectiveness of the Ghostly game.

The results of the user experience questionnaire indicate overall high satisfaction with the Ghostly game, as well as good scores

XSL•FO
RenderX

on usability and ease of learning. When only focusing on the user experience results from the Ghostly + BFR group, high scores remained, implying that BFR does not adversely influence user experience. These findings strengthen the evidence for the Ghostly game combined with BFR as a viable rehabilitation approach in this target population. In this literature, only one study by Mousavi et al [43] investigated the effects of BFR in combination with a boxing exergame for the Xbox 360 Kinect, on enjoyment and energy expenditure. However, this study reported decreased, although not significant, perceived enjoyment and no difference in energy expenditure in 14 healthy participants after a single session of the boxing exergame combined with BFR compared with a single session with the exergame alone [43]. This hiatus in literature further highlights the relevance of this study and the need for more research into BFR combined with exergames. A potential explanation for the limited implementation of exergames and BFR could be the concern of increased hemodynamic stress and disturbances in local blood flow [44], however, correct implementation of BFR has been shown to present no greater risk than conventional exercise [45].

Comparison to Prior Work

Further expanding on the user experience while training with exergames, our results are in line with literature showing that exergames have a beneficial effect on intrinsic motivation and thus might increase therapy adherence [46-48]. As aforementioned, Garcia-Hernandez et al [23] reported significantly higher intrinsic motivation in patients who trained with an EMG biofeedback exergame compared with patients performing conventional therapy. The exergames group showed higher mean scores for interest and enjoyment, as well as perceived competence [23]. Similarly, Fitzgerald et al [49] reported high scores on interest and enjoyment in healthy adults who performed 12 sessions of postural stability training with a wobble board-based exergame. These findings highlight the potential of exergames to motivate patients to remain active during and after rehabilitation. As revealed by Ong et al [50], exergames can prove to be an efficient tool for rehabilitation, but there is a need for end user perspectives in the development of evidence-based exergames to ensure sufficient implementation. A user-centered design approach is required to support the development of rehabilitation technologies that maximize both user acceptance and functional outcomes [51].

Data from end user feedback and expert observations highlighted a few key elements that should be adjusted to further enhance the user experience of the end users such as a change of color contrast, different level difficulties, etc. Expanding on these findings, the noticeable decline in color contrast perception experienced by older adults due to perceived yellowing of the human lens warrants the need for a change in color contrast in the Ghostly game [52]. Different color combinations produce varying degrees of difficulty in perceiving contrast. For example, the contrast between green and blue is often more difficult to perceive in older adults [53,54]. Moreover, this decline in contrast perception strongly influences reaction time as older adults need more time to detect, discriminate, and recognize objects in the visual field [54,55]. Noticeably, the main color scheme of the Ghostly game is blue and green, which potentially explains some of the difficulties participants experienced while playing the Ghostly game. Research recommends color combinations with medium to long wavelengths, such as yellow, orange, and red, as these are easier to perceive, or to implement a large contrast in lightness when using blue or green to ensure sufficient contrast perception [54].

Furthermore, as people grow older, there is a generalized slowing in information processing, with a more specific increase in the amount of time necessary to make a correct judgment about visual stimuli, also known as visual processing speed [56]. Older adults often experience difficulties in tasks that involve visual clutter, secondary task demands, and time-sensitive responses [56], all of which are incorporated into the Ghostly game. With this in mind, it is important to add changes to the Ghostly game that take into account this decline in information processing speed by creating levels that limit the need for quick processing of visual information. Nonetheless, the cognitive-motor dual-task aspect of the Ghostly game should not be fully removed, as dual-task training has positive effects on cognitive and physical function [57]. A review by Joubert and Chainay [58] reported a positive effect of cognitive-motor dual-task training on cognitive functioning. Similarly, a study by Uematsu et al [59] on the effect of a cognitive-motor dual-task standing balance exergame reported a beneficial effect of exergame balance training on one-leg standing time, functional reach distance, and hip and knee extensor strength in 25 healthy older adults.

Focusing on the secondary outcomes of this feasibility study, descriptive statistics revealed very limited effects on clinical outcomes. Nonetheless, these outcomes are relevant in hospitalized older adults. Within this population, muscle weakness and muscle mass loss due to inactivity can lead to a functional decline and loss of independence, which is not directly related to the condition for which they were admitted, also defined as hospital-associated disability [7]. Due to the short intervention period, no changes in clinical parameters were expected. Literature has shown that early changes in muscle strength and mass can occur after 2 to 4 weeks of training, but the adaptations after resistance training are generally evident after 8 to 12 weeks of training [60]. The early adaptations can be explained by neuromuscular and connective tissue adaptations, and edema-induced swelling of the muscle [60,61]. Nonetheless, as described in the results, all interventions and outcomes were feasible in relation to a future larger RCT testing the effectiveness of the Ghostly game.

Future Directions

With future development and research in mind, an updated version of the Ghostly game combined with BFR might prove to be a relevant tool for rehabilitation to promote the Aging in Place framework proposed within the Healthy Aging model of the World Health Organization [62,63]. Aging in Place is defined as the ability of older adults to remain in their community during later life to preserve a sense of connection, security, and familiarity, as well as their sense of identity and autonomy [62,63]. As stated earlier, exergames have proven to enhance motivation and willingness to exercise in older adults, even when unsupervised [46-48]. Thus, the Ghostly game can

XSL•FO
RenderX

provide opportunities for older adults living independently or in community homes and residential settings who are unable to train in a high-load, more dynamic way to remain active in a motivating and challenging manner. Moreover, implementing BFR into the Ghostly game in a comprehensive manner ensures evidence-based training at an intensity that is beneficial for rehabilitation.

Strengths and Limitations

To date, no other studies combining an EMG-driven exergame with the strength training principles of BFR training in hospitalized patients have been found in the literature, highlighting the relevance of our results in the field of exergame research. This combination of evidence-based game design and the already-established effectiveness of BFR training in older adults maximizes the potential for rehabilitation in hospitalized older adults. Second, another strength of this study is the mixed methods nature of the study design, which provides a comprehensive understanding of both the user experience of the participants while using the Ghostly game, as well as the clinical effects of the Ghostly game for rehabilitation. Third, the implementation of a wide variety of target groups (ie, patients and health care professionals) in the investigation of the user experience ensures an all-encompassing vision of the needs and recommendations, without being limited to the expertise of one specific demographic.

Nonetheless, this study has a few limitations that warrant caution when interpreting the results. First, it is important to note that the intervention period of this feasibility study was short, ranging from 3 to 5 sessions. However, as the main goal of this study is to gather information on user experience, 3 to 5 sessions were sufficient to gather reliable user experience data from all the participants. Second, the study sample involved 15 participants, with 10 participants training with the Ghostly game. This small sample may restrict the trial's ability to detect meaningful effect sizes, but it is sufficient to gather user experience data. Last, as mentioned in the results, game control was not always stable while completing the intervention due to inaccurate registration of EMG signals. It seems plausible to hypothesize that the EMG signal issues may partly result from variations in participants' subcutaneous fat thickness. Subcutaneous fat is known to reduce the EMG amplitude above active muscles [64]. Kuiken et al [64] reported a 30% - 80% reduction in EMG amplitude with fat layers of 0.3 - 0.9 cm. In our study, the average baseline subcutaneous fat was 0.85 cm (SD 0.52 cm), with 4 participants exceeding 0.9 cm, of which only one experienced issues. However, by implementing an assessment level before the Ghostly game is played, inaccurate EMG signals are filtered, and potential signal issues could be identified before the start of the intervention. When signal issues occurred, the supervising researcher addressed them by disconnecting and reconnecting the EMG sensors and restarting the assessment procedure until stable EMG signals were observed. Additionally, observations regarding EMG signal stability were documented to inform future development of the Ghostly game.

Conclusions

As part of a larger iterative development process, this mixed method randomized feasibility study investigated the user experience and feasibility of an evidence-based exergame, the Ghostly game, as a stand-alone added therapy or in combination with BFR for rehabilitation of bedridden hospitalized older adults. The results revealed a high overall level of satisfaction and usability of the Ghostly game. Furthermore, analysis of the USE questionnaire demonstrated good scores when only focusing on the Ghostly + BFR group. Our findings also highlighted important changes required to maximize user experience based on expert opinions and participants' recommendations. Since this study is part of a larger user-centered design approach, the results further emphasize the necessity of incorporating all end users in the development of rehabilitation technology with the aim of maximizing the functional capabilities of the users, as well as maximizing user experience. Additionally, this study shows the potential of exergames combined with BFR to maximize clinical outcome effects based on state-of-the-art evidence on strength training. In regard to feasibility, it was found that all outcomes and procedures were feasible for a larger RCT testing the effectiveness of the Ghostly game in combination with BFR.

Acknowledgments

The authors would like to thank the health care professionals of the geriatric ward of the University Hospital Brussels for their contribution to the recruitment of the participants. Furthermore, the authors would like to thank Claudia Zaldívar, Veronika Schatz, and Abbi Needham for their support in collecting data. MF (grant 11G9622N), DDV (grant 11PKX24N), and SZ (grant FWOSB139) are research fellows funded by the Research Foundation Flanders (Fonds Wetenschappelijk Onderzoek; FWO). This work was supported by the Willy Gepts Fonds of the Vrije Universiteit Brussel.

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

RD, KK, LO, BJ, ES, and DB contributed to the conception and design of the work. RD, MB, and MB collected the data. RD, AD, and SL contributed to the recruitment of participants. RD, RC, ME, ES, and DB contributed to the analysis of the results. RD, RC, DDV, ME, MF, SG, SZ, ES, and DB contributed to the interpretation of the results. RD, SG, BJ, ES, and DB drafted

XSL•FO
RenderX

the manuscript. All authors critically reviewed the manuscript for important intellectual content. All authors have read and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Template for intervention description and replication checklist for the 3 intervention arms.
[DOCX File, 17 KB - games_v13i1e69400_app1.docx]

Multimedia Appendix 2
Results of the Usefulness, Satisfaction, and Ease of Use questionnaire for each intervention group.
[DOCX File, 15 KB - games_v13i1e69400_app2.docx]

Checklist 1
The CONSORT-eHealth checklist (V 1.6.1).
[PDF File, 1090 KB - games_v13i1e69400_app3.pdf]

References

1. Deer RR, Dickinson JM, Fisher SR, Ju H, Volpi E. Identifying effective and feasible interventions to accelerate functional recovery from hospitalization in older adults: a randomized controlled pilot trial. Contemp Clin Trials 2016 Jul;49:6-14. [doi: 10.1016/j.cct.2016.05.001] [Medline: 27178766]
2. Alley DE, Koster A, Mackey D, et al. Hospitalization and change in body composition and strength in a population-based cohort of older persons. J Am Geriatr Soc 2010 Nov;58(11):2085-2091. [doi: 10.1111/j.1532-5415.2010.03144.x] [Medline: 21054288]
3. Peel NM, Paul SK, Cameron ID, Crotty M, Kurrle SE, Gray LC. Promoting activity in geriatric rehabilitation: a randomized controlled trial of accelerometry. PLoS One 2016;11(8):e0160906. [doi: 10.1371/journal.pone.0160906] [Medline: 27564857]
4. van Grootel J, Bor P, Netjes JA, Veenhof C, Valkenet K. Improving physical activity in hospitalized patients: the preliminary effectiveness of a goal-directed movement intervention. Clin Rehabil 2023 Nov;37(11):1501-1509. [doi: 10.1177/02692155231189607] [Medline: 37487188]
5. Wall BT, Dirks ML, van Loon LJC. Skeletal muscle atrophy during short-term disuse: implications for age-related sarcopenia. Ageing Res Rev 2013 Sep;12(4):898-906. [doi: 10.1016/j.arr.2013.07.003] [Medline: 23948422]
6. Saitoh M, Takahashi Y, Okamura D, et al. Prognostic impact of hospital-acquired disability in elderly patients with heart failure. ESC Heart Fail 2021 Jun;8(3):1767-1774. [doi: 10.1002/ehf2.13356] [Medline: 33838022]
7. Giacomino K, Hilfiker R, Beckwée D, Taeymans J, Sattelmayer KM. Assessment tools and incidence of hospital-associated disability in older adults: a rapid systematic review. PeerJ 2023;11:e16036. [doi: 10.7717/peerj.16036] [Medline: 37872951]
8. Dent E, Morley JE, Cruz-Jentoft AJ, et al. International clinical practice guidelines for sarcopenia (ICFSR): screening, diagnosis and management. J Nutr Health Aging 2018;22(10):1148-1161. [doi: 10.1007/s12603-018-1139-9] [Medline: 30498820]
9. Nicholson VP, McKean MR, Burkett BJ. Low-load high-repetition resistance training improves strength and gait speed in middle-aged and older adults. J Sci Med Sport 2015 Sep;18(5):596-600. [doi: 10.1016/j.jsams.2014.07.018] [Medline: 25154701]
10. Sommers J, Van Den Boorn M, Engelbert RHH, Nollet F, Van Der Schaaf M, Horn J. Feasibility of muscle activity assessment with surface electromyography during bed cycling exercise in intensive care unit patients. Muscle Nerve 2018 Nov;58(5):688-693. [doi: 10.1002/mus.26330] [Medline: 30160070]
11. Burtin C, Clerckx B, Robbeets C, et al. Early exercise in critically ill patients enhances short-term functional recovery. Crit Care Med 2009 Sep;37(9):2499-2505. [doi: 10.1097/CCM.0b013e3181a38937] [Medline: 19623052]
12. Martin PM, Bart RM, Ashley RL, Velasco T, Wise SR. An overview of blood flow restriction physiology and clinical considerations. Curr Sports Med Rep 2022 Apr 1;21(4):123-128. [doi: 10.1249/JSR.0000000000000948] [Medline: 35394953]
13. Wedig IJ, Durocher JJ, McDaniel J, Elmer SJ. Blood flow restriction as a potential therapy to restore physical function following COVID-19 infection. Front Physiol 2023;14:1235172. [doi: 10.3389/fphys.2023.1235172] [Medline: 37546539]
14. Centner C, Wiegel P, Gollhofer A, König D. Effects of blood flow restriction training on muscular strength and hypertrophy in older individuals: a systematic review and meta-analysis. Sports Med 2019 Jan;49(1):95-108. [doi: 10.1007/s40279-018-0994-1] [Medline: 30306467]
15. Zhang T, Tian G, Wang X. Effects of low-load blood flow restriction training on hemodynamic responses and vascular function in older adults: a meta-analysis. IJERPH 2022;19(11):6750. [doi: 10.3390/ijerph19116750]

XSL•FO
RenderX

16. Nawaz A, Skjæret N, Helbostad JL, Vereijken B, Boulton E, Svanaes D. Usability and acceptability of balance exergames in older adults: a scoping review. Health Informatics J 2016 Dec;22(4):911-931. [doi: 10.1177/1460458215598638] [Medline: 26303810]
17. Ismail NA, Hashim HA, Ahmad Yusof H. Physical activity and exergames among older adults: a scoping review. Games Health J 2022 Feb;11(1):1-17. [doi: 10.1089/g4h.2021.0104] [Medline: 34851732]
18. Rytterström P, Strömberg A, Jaarsma T, Klompstra L. Exergaming to increase physical activity in older adults: feasibility and practical implications. Curr Heart Fail Rep 2024 Aug;21(4):439-459. [doi: 10.1007/s11897-024-00675-9] [Medline: 39023808]
19. Pacheco TBF, de Medeiros CSP, de Oliveira VHB, Vieira ER, de Cavalcanti FAC. Effectiveness of exergames for improving mobility and balance in older adults: a systematic review and meta-analysis. Syst Rev 2020 Jul 18;9(1):163. [doi: 10.1186/s13643-020-01421-7] [Medline: 32682439]
20. International Organization for Standardization. Ergonomics of Human-System Interaction, in ISO 9241-210: Human Centred Design for Interactive Systems: European Committee for Standardization; 2019.
21. MacIntosh A, Vignais N, Vigneron V, et al. The design and evaluation of electromyography and inertial biofeedback in hand motor therapy gaming. Assist Technol 2022 Mar 4;34(2):213-221. [doi: 10.1080/10400435.2020.1744770] [Medline: 32207635]
22. Troka P, Toczko H, Przvstup P, Kaczmarek M. A biofeedback system that uses the game to study electrical muscle activity. Presented at: 2018 11th International Conference on Human System Interaction (HSI); Jul 4-6, 2018; Gdansk, Poland. [doi: 10.1109/HSI.2018.8431367]
23. Garcia-Hernandez N, Garza-Martinez K, Parra-Vega V. Electromyography biofeedback exergames to enhance grip strength and motivation. Games Health J 2018 Feb;7(1):75-82. [doi: 10.1089/g4h.2017.0054] [Medline: 29227162]
24. Enciso J, Variya D, Sunthonlap J, et al. Electromyography-driven exergaming in wheelchairs on a mobile platform: bench and pilot testing of the WOW-mobile fitness system. JMIR Rehabil Assist Technol 2021 Jan 19;8(1):e16054. [doi: 10.2196/16054] [Medline: 33464221]
25. World Medical Association. Declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA 2013;310(20):2191-2194. [doi: 10.1001/jama.2013.281053]
26. Rikli RE, Jones CJ. Development and validation of criterion-referenced clinically relevant fitness standards for maintaining physical independence in later years. Gerontologist 2013 Apr;53(2):255-267. [doi: 10.1093/geront/gns071] [Medline: 22613940]
27. Teresi JA, Yu X, Stewart AL, Hays RD. Guidelines for designing and evaluating feasibility pilot studies. Med Care 2022 Jan 1;60(1):95-103. [doi: 10.1097/MLR.0000000000001664] [Medline: 34812790]
28. Nelson ME, Rejeski WJ, Blair SN, et al. Physical activity and public health in older adults: recommendation from the American College of Sports Medicine and the American Heart Association. Med Sci Sports Exerc 2007 Aug;39(8):1435-1445. [doi: 10.1249/mss.0b013e3180616aa2] [Medline: 17762378]
29. Nuzzo JL, Barry BK, Jones MD, Gandevia SC, Taylor JL. Effects of four weeks of strength training on the corticomotoneuronal pathway. Med Sci Sports Exerc 2017 Nov;49(11):2286-2296. [doi: 10.1249/MSS.0000000000001367] [Medline: 28692630]
30. Lum D, Barbosa TM. Brief review: effects of isometric strength training on strength and dynamic performance. Int J Sports Med 2019 May;40(6):363-375. [doi: 10.1055/a-0863-4539] [Medline: 30943568]
31. Carr JC, Ye X, Stock MS, Bemben MG, DeFreitas JM. The time course of cross-education during short-term isometric strength training. Eur J Appl Physiol 2019 Jun;119(6):1395-1407. [doi: 10.1007/s00421-019-04130-9] [Medline: 30949806]
32. Hermens HJ, Freriks B, Disselhorst-Klug C, Rau G. Development of recommendations for SEMG sensors and sensor placement procedures. J Electromyogr Kinesiol 2000 Oct;10(5):361-374. [doi: 10.1016/s1050-6411(00)00027-4] [Medline: 11018445]
33. Patterson SD, Hughes L, Warmington S, et al. Blood flow restriction exercise: considerations of methodology, application, and safety. Front Physiol 2019;10:533. [doi: 10.3389/fphys.2019.00533] [Medline: 31156448]
34. Libardi CA, Chacon-Mikahil MPT, Cavaglieri CR, et al. Effect of concurrent training with blood flow restriction in the elderly. Int J Sports Med 2015 May;36(5):395-399. [doi: 10.1055/s-0034-1390496] [Medline: 25700103]
35. Lund AM. Measuring usability with the USE questionnaire. Usability Interface 2001;8(2):3-6 [FREE Full text]
36. Gao M, Kortum P, Oswald F. Psychometric evaluation of the USE (Usefulness, Satisfaction, and Ease of Use) questionnaire for reliability and validity. SAGE J 2018 Sep;62(1):1414-1418. [doi: 10.1177/1541931218621322]
37. Plouvier P, Marcilly R, Robin G, et al. Evaluation of satisfaction with a secure, connected mobile app for women in assisted reproductive technology programs: prospective observational study. JMIR Hum Factors 2025 Feb 24;12:e63570. [doi: 10.2196/63570] [Medline: 39993725]
38. Vanmulken DAMM, Spooren AIF, Bongers HMH, Seelen HAM. Robot-assisted task-oriented upper extremity skill training in cervical spinal cord injury: a feasibility study. Spinal Cord 2015 Jul;53(7):547-551. [doi: 10.1038/sc.2014.250] [Medline: 25644390]

XSL•FO
RenderX

39. Swinnen E, Lefeber N, Willaert W, et al. Motivation, expectations, and usability of a driven gait orthosis in stroke patients and their therapists. Top Stroke Rehabil 2017 May;24(4):299-308. [doi: 10.1080/10749357.2016.1266750] [Medline: 27996894]
40. Fossey E, Harvey C, McDermott F, Davidson L. Understanding and evaluating qualitative research. Aust N Z J Psychiatry 2002 Dec;36(6):717-732. [doi: 10.1046/j.1440-1614.2002.01100.x] [Medline: 12406114]
41. Wolcott MD, Lobczowski NG. Using cognitive interviews and think-aloud protocols to understand thought processes. Curr Pharm Teach Learn 2021 Feb;13(2):181-188. [doi: 10.1016/j.cptl.2020.09.005] [Medline: 33454077]
42. Vermeiren S, Beckwée D, Vella-Azzopardi R, et al. Evaluation of appendicular lean mass using bio impedance in persons aged 80+: a new equation based on the BUTTERFLY-study. Clin Nutr 2019 Aug;38(4):1756-1764. [doi: 10.1016/j.clnu.2018.07.029] [Medline: 30119982]
43. Mousavi Z, Karimi Z, Wong A, Cheraghloo N, Bagheri H, Bagheri R. Comparison of enjoyment and energy expenditure of exergame with and without blood flow restriction in men and women. Sci Rep 2024 Apr 19;14(1):9030. [doi: 10.1038/s41598-024-59379-8] [Medline: 38641659]
44. Waclawovsky G, Lehnen AM. Hemodynamics of aerobic and resistance blood flow restriction exercise in young and older adults. Eur J Appl Physiol 2016 Apr;116(4):859-860. [doi: 10.1007/s00421-015-3318-2] [Medline: 26705246]
45. Loenneke JP, Wilson JM, Wilson GJ, Pujol TJ, Bemben MG. Potential safety issues with blood flow restriction training. Scand J Med Sci Sports 2011 Aug;21(4):510-518. [doi: 10.1111/j.1600-0838.2010.01290.x] [Medline: 21410544]
46. Joronen K, Aikasalo A, Suvitie A. Nonphysical effects of exergames on child and adolescent well-being: a comprehensive systematic review. Scand J Caring Sci 2017 Sep;31(3):449-461. [doi: 10.1111/scs.12393] [Medline: 27862145]
47. Marques LM, Uchida PM, Barbosa SP. The impact of exergames on emotional experience: a systematic review. Front Public Health 2023;11:1209520. [doi: 10.3389/fpubh.2023.1209520] [Medline: 37744509]
48. Alfieri FM, da Silva Dias C, de Oliveira NC, Battistella LR. Gamification in musculoskeletal rehabilitation. Curr Rev Musculoskelet Med 2022 Dec;15(6):629-636. [doi: 10.1007/s12178-022-09797-w] [Medline: 36301514]
49. Fitzgerald D, Trakarnratanakul N, Smyth B, Caulfield B. Effects of a wobble board-based therapeutic exergaming system for balance training on dynamic postural stability and intrinsic motivation levels. J Orthop Sports Phys Ther 2010 Jan;40(1):11-19. [doi: 10.2519/jospt.2010.3121] [Medline: 20044704]
50. Ong DSM, Weibin MZ, Vallabhajosyula R. Serious games as rehabilitation tools in neurological conditions: a comprehensive review. Technol Health Care 2021;29(1):15-31. [doi: 10.3233/THC-202333] [Medline: 32804107]
51. Vaezipour A, Whelan BM, Wall K, Theodoros D. Acceptance of rehabilitation technology in adults with moderate to severe traumatic brain injury, their caregivers, and healthcare professionals: a systematic review. J Head Trauma Rehabil 2019;34(4):E67-E82. [doi: 10.1097/HTR.0000000000000462] [Medline: 30608310]
52. Artigas JM, Felipe A, Navea A, Fandiño A, Artigas C. Spectral transmission of the human crystalline lens in adult and elderly persons: color and total transmission of visible light. Invest Ophthalmol Vis Sci 2012 Jun 26;53(7):4076-4084. [doi: 10.1167/iovs.12-9471] [Medline: 22491402]
53. Dodd C, Athauda R, Adam M. Designing user interfaces for the elderly: a systematic literature review. Presented at: 28th ACIS 2017; Dec 5-6, 2017; Hobart, Australia URL: https://aisel.aisnet.org/acis2017/ [accessed 2025-05-20]
54. Delcampo‐Carda A, Torres‐Barchino A, Serra‐Lluch J. Chromatic interior environments for the elderly: a literature review. Col Res Appl 2019 Jun;44(3):381-395. [doi: 10.1002/col.22358]
55. Suzuki TA, Qiang Y, Sakuragawa S, Tamura H, Okajima K. Age-related changes of reaction time and p300 for low-contrast color stimuli: effects of yellowing of the aging human lens. J Physiol Anthropol 2006 Mar;25(2):179-187. [doi: 10.2114/jpa2.25.179] [Medline: 16679715]
56. Owsley C. Visual processing speed. Vision Res 2013 Sep 20;90:52-56. [doi: 10.1016/j.visres.2012.11.014] [Medline: 23231958]
57. Gallou-Guyot M, Mandigout S, Bherer L, Perrochon A. Effects of exergames and cognitive-motor dual-task training on cognitive, physical and dual-task functions in cognitively healthy older adults: an overview. Ageing Res Rev 2020 Nov;63:101135. [doi: 10.1016/j.arr.2020.101135] [Medline: 32768594]
58. Joubert C, Chainay H. Aging brain: the effect of combined cognitive and physical training on cognition as compared to cognitive and physical training alone - a systematic review. Clin Interv Aging 2018;13:1267-1301. [doi: 10.2147/CIA.S165399] [Medline: 30057444]
59. Uematsu A, Tsuchiya K, Fukushima H, Hortobágyi T. Effects of motor-cognitive dual-task standing balance exergaming training on healthy older adults' standing balance and walking performance. Games Health J 2023 Aug;12(4):302-309. [doi: 10.1089/g4h.2022.0203] [Medline: 36944151]
60. Hughes DC, Ellefsen S, Baar K. Adaptations to endurance and strength training. Cold Spring Harb Perspect Med 2018 Jun 1;8(6):a029769. [doi: 10.1101/cshperspect.a029769] [Medline: 28490537]
61. Damas F, Phillips SM, Lixandrão ME, et al. Early resistance training-induced increases in muscle cross-sectional area are concomitant with edema-induced muscle swelling. Eur J Appl Physiol 2016 Jan;116(1):49-56. [doi: 10.1007/s00421-015-3243-4] [Medline: 26280652]
62. Ratnayake M, LPCMH, ATR, et al. Aging in place:: are we prepared? Dela J Public Health 2022 Aug;8(3):28-31. [doi: 10.32481/djph.2022.08.007] [Medline: 36177171]

XSL•FO
RenderX

63. World Report on Ageing and Health: World Health Organization; 2015.
64. Kuiken TA, Lowery MM, Stoykov NS. The effect of subcutaneous fat on myoelectric signal amplitude and cross-talk. Prosthet Orthot Int 2003 Apr;27(1):48-54. [doi: 10.3109/03093640309167976] [Medline: 12812327]

Abbreviations

ALM: appendicular lean mass
AOP: arterial occlusion pressure
BFR: blood flow restriction
CSA: cross-sectional area
EMG: electromyography
MVC: maximal voluntary contraction
RCT: randomized controlled trial
USE: Usefulness, Satisfaction, and Ease of Use

Edited by A Coristine; submitted 29.11.24; peer-reviewed by F Noveletto, L Carcreff; revised version received 04.04.25; accepted 06.04.25; published 10.06.25.

Please cite as:
Debeuf R, Claeys R, Berlanger M, Bunt M, Debain A, De Vlieger D, Eggermont M, Firouzi M, Guida S, Kostková K, Lieten S, Omelina L, Zaccardi S, Jansen B, Swinnen E, Beckwée D
Insights on the User Experience and Feasibility of an Electromyography-Driven Exergame Combined With Blood Flow Restriction for Strength Training in Hospitalized Older Adults: Mixed Methods Randomized Controlled Feasibility Study
JMIR Serious Games 2025;13:e69400
URL: https://games.jmir.org/2025/1/e69400
doi:10.2196/69400

©Ruben Debeuf, Reinhard Claeys, Margo Berlanger, Myrthe Bunt, Aziz Debain, Daan De Vlieger, Matthias Eggermont, Mahyar Firouzi, Stefania Guida, Katarína Kostková, Siddhartha Lieten, Lubos Omelina, Silvia Zaccardi, Bart Jansen, Eva Swinnen, David Beckwée. Originally published in JMIR Serious Games (https://games.jmir.org), 10.6.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Effects of a Session of Exergames and Traditional Games on Inhibitory Control in Children With Autism Spectrum Disorder: Randomized Controlled Crossover Trial

Juliana Macedo Miranda, MSc; Rodrigo Alberto Vieira Browne, PhD; Weslley Quirino Alves da Silva, PhD; João Paulo Rodrigues dos Santos, BPEd; Carmen Silvia Grubert Campbell, PhD; Isabela Almeida Ramos, PhD

Graduate Program in Physical Education, Catholic University of Brasilia, QS 07 Lote 01 EPCT Águas Claras, Brasilia, Brazil

Corresponding Author:
Juliana Macedo Miranda, MSc
Graduate Program in Physical Education, Catholic University of Brasilia, QS 07 Lote 01 EPCT Águas Claras, Brasilia, Brazil

Abstract

Background: Autism spectrum disorder (ASD) is characterized by deficits in executive functions, such as inhibitory control, which affect behavior and social adaptation. Although physical activity–based interventions, such as exergames, have shown potential to improve these functions, their comparative effects with active traditional games remain underexplored, particularly regarding inhibitory control in children with ASD.

Objective: We aim to analyze the effects of a session of exergames and active traditional games on inhibitory control in children with ASD.

Methods: This randomized controlled crossover trial included 9 male children with ASD (mean age 8.6, SD 1.4 y). Participants completed three 20-minute experimental sessions in random order, with a minimum interval of 48 hours: (1) active traditional games, (2) exergames using Just Dance 2022, and (3) a control session with manual painting activities. Inhibitory control was assessed 5 minutes postsession using a modified flanker task in the E-Prime (version 3.0; Psychological Software Tools Inc) program, recording reaction time (RT) and accuracy in congruent and incongruent phases. Repeated measures ANOVA was used to compare RT and accuracy between experimental and control conditions. Data are presented as means and 95% CIs.

Results: There was a statistically significant effect of condition on RT in the incongruent phase (P=.02). RT in the exergame session (849 ms, 95% CI 642 to 1057) was lower compared to the traditional games (938 ms, 95% CI 684 to 1191; P=.02) and control (969 ms, 95% CI 742, 1196 to P=.01) sessions. No significant differences were observed in RT during the congruent phase or in accuracy across either phase.

Conclusions: A 20-minute session of exergame improved inhibitory control performance in children with ASD compared to active traditional games and painting activities.

Trial Registration: Brazilian Registry of Clinical Trials (ReBEC) RBR-5r9xzbq, Universal Trial Number U1111-1302-3490; https://ensaiosclinicos.gov.br/rg/RBR-5r9xzbq

(JMIR Serious Games 2025;13:e65562) doi:10.2196/65562

KEYWORDS

children; pediatric; autism; ASD, autistic; behavior; exergame; physical education; exercise; physical activity; cognition; anthropometric; Flanker test; inhibitory control; randomized control trial; crossover

Introduction

Autism spectrum disorder (ASD) is characterized by significant challenges in reciprocal social communication and interactions, as well as restricted and repetitive behavioral patterns [1,2]. These challenges are closely associated with pronounced impairments in executive functions, particularly inhibitory control, which is linked to deficits in social communication [3] and repetitive behaviors [4]. Moreover, impairments in inhibitory control and working memory in children with ASD are associated with low school readiness [5]. Inhibitory control plays a crucial role in academic performance by regulating attention, behavior, and thought [6] and managing emotions to overcome internal impulses or external stimuli, thereby enabling adaptation to challenging contexts [7]. This highlights the importance of investigating inhibitory processes in ASD and the factors that may influence their outcomes.

Studies have demonstrated the positive effects of physical exercise on inhibitory control in both neurotypical children [8] and children with ASD [9]. For instance, a study involving children with ASD conducted three 20-minute exercise sessions—circuit training, treadmill walking, and a sedentary

XSL•FO
RenderX

control—and observed significant improvements in cerebral oxygenation and inhibitory control during the exercise sessions [9]. These acute benefits of exercise are attributed to the release of catecholamines and neurotrophic factors [10], as well as enhanced cerebral oxygenation regulation, which improves cognitive processing [9,11]. However, many children fail to meet recommended levels of physical activity [12,13], and children with ASD are particularly less likely to engage in school-based movement activities than their neurotypical peers [14]. In light of the World Health Organization's campaign emphasizing that "every move counts," as long as it is enjoyable, accessible, safe, and valued [15,16], exploring alternative strategies to promote physical activity and assess their potential effects on inhibitory control in children with ASD is essential.

Playful activities, such as traditional games, can effectively encourage physical activity and enhance inhibitory control during classroom time [17]. However, it remains unclear whether exercise sessions based on play produce similar effects on inhibitory control in children with ASD. In Brazil, physical education classes, which are integral to the school curriculum, play a crucial role in child development by fostering exploration of emotions, reasoning, imagination, and creativity through play and games [18,19]. Yet, children with ASD often struggle to participate in symbolic play due to delays in behavioral skills such as communication and socialization, as well as learning barriers such as rigidity and repetitive behaviors [20]. These challenges may result in social isolation and limit opportunities to benefit from traditional Physical Education alongside peers [21,22]. This underscores the need to adapt school physical education sessions to accommodate the behavioral needs of children with ASD and explore alternative practices that provide comparable benefits for executive functioning, particularly inhibitory control.

Exergaming, which integrates physical exercise with digital games, has emerged as a promising strategy to engage individuals in activities combining physical and cognitive stimulation, aligning with educational and therapeutic goals [23,24]. These interventions address the unique challenges faced by children with ASD, offering an enjoyable and accessible medium to foster motor and executive skills [25]. Platforms such as Dance Dance Revolution have garnered attention for merging physical activity with cognitive engagement in an interactive format. Studies indicate that such platforms enhance cognitive performance, including executive functions, while also providing physiological and motivational benefits [26]. Features such as dynamic adaptation to movement and immediate feedback make exergames particularly relevant in educational and therapeutic contexts. This study specifically used Just Dance 2022 due to its ability to adapt dynamically to users' movements, providing immediate visual and auditory feedback, which can enhance engagement and motivation. The game's structure, which includes a wide range of songs with varying levels of difficulty, allows participants to select options that align with their abilities and preferences, promoting inclusion and reducing potential barriers to participation. These features make Just Dance 2022 an ideal platform for addressing the unique behavioral and cognitive challenges faced by children with ASD.

To address the need for further robust evidence on exergame interventions in school settings, this study aimed to analyze the acute effects of a session of exergames, specifically Just Dance 2022, and active traditional games on inhibitory control performance of school-aged children with ASD. By exploring these 2 distinct approaches, this study seeks to provide contributions to alternative strategies for promoting executive functioning and physical activity among children with ASD, particularly in environments such as schools, where inclusive and engaging practices are essential.

Methods

Study Design

This randomized controlled crossover trial was registered in the ReBEC (Brazilian Registry of Clinical Trials; protocol RBR-5r9xzbq). This study was conducted in accordance with the CONSORT (Consolidated Standards of Reporting Trials) guidelines for randomized clinical trials [27]. Given the within-subjects crossover design, measures were implemented to address potential day-to-day variations and carryover effects in performance. The 48-hour washout period between sessions was selected based on the available literature, including studies involving children with ASD and typically developing children, which demonstrated the adequacy of this interval to minimize residual effects [8,9]. Additionally, randomization of the session order was employed to counterbalance potential biases introduced by session sequencing. All sessions were conducted at the same time of day for each participant, in a controlled environment, and by the same research team to ensure consistency.

Ethical Considerations

Ethical approval was obtained from the Ethics Committee on Human Research at the Catholic University of Brasília (protocol 5448930/2022). Prior to participation, informed consent was obtained from parents or legal guardians, who received a detailed explanation of the study's objectives, procedures, potential risks, and benefits. In addition, assent was obtained from the child participants, ensuring their voluntary participation. To ensure participant privacy and confidentiality, all collected data were fully anonymized and stored securely. No financial or material compensation was provided, as the study adhered to Brazilian ethical guidelines, which prohibit compensation for research participation. The study was conducted in accordance with the Helsinki Declaration and Resolution 466/2012 of the Brazilian National Health Council.

Randomization

Participants were randomly assigned to the experimental and control sessions to minimize bias and enhance the reliability of the results. Randomization was conducted by a blinded researcher who was not involved in data collection, ensuring impartiality. The allocation list was generated electronically using the GraphPad website [28], guaranteeing a rigorous random selection process. The exercise protocol was pretested in a pilot study and was subsequently implemented with a 48-hour interval between sessions. The protocol included (1) an exergame session, (2) a physical education session with

XSL•FO
RenderX

active traditional games, and (3) a control session with manual painting activities.

Participants

Nine male Brazilian schoolchildren were recruited from Escola Classe 10, a public elementary school located in the administrative region of Taguatinga, Federal District, Brazil, to participate in this study. The inclusion criteria were: children aged 6 to 11 years, enrolled in school, diagnosed with ASD according to the *DSM-5* (*Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition*) [29], and without a history of physical, pulmonary, or cardiac malformations. The exclusion criteria included nonverbal ASD, intellectual disability (IQ below 70), and voluntary withdrawal from this study. Additionally, parents or guardians provided the child's medical history.

Procedures

The evaluations and interventions were conducted at the premises of Escola Classe 10. Data collection took place between August and October of 2022. The first meeting involved anthropometric measurements and the administration of the Körperkoordinationstest für Kinder (KTK) battery [30]. In the second meeting, Raven Colored Progressive Matrices Test was applied [31,32]. The third meeting was dedicated to familiarizing participants with the flanker task. The final 3 meetings involved the exergame, traditional games, and control sessions, each followed by the application of the flanker task. A cognitive test was administered within 5 minutes after each session. The sessions were carried out during the school hours of each child, in the respective shift (morning or afternoon) in which they were enrolled.

Initial Assessments

Anthropometry

Weight and height were measured using an electronic scale (Tec-Silver, Techline) and a fixed wall stadiometer (Sanny ES2040, Sanny), respectively, following the procedures outlined by Heyward et al [33]. The BMI was calculated by dividing weight (in kilograms) by the square of height (in meters, kg/m^2). The calculations were performed according to the guidelines provided by the Centers for Disease Control and Prevention [34].

Body Coordination Test

The motor skills of the participating children were assessed using the body coordination test for children (KTK). The KTK is a standardized instrument designed to evaluate various aspects of body coordination. The test battery includes 4 tasks—balancing, jump on one foot, lateral jump, and lateral transposition—each targeting specific components of body coordination, such as balance, rhythm, strength, laterality, speed, and agility. These tasks collectively aim to assess the overall body coordination and mastery of movement [30]. The final score of the KTK is obtained by summing the points from the 4 tasks, which are compared to normative data for the age group. Based on performance, participants were categorized into 2 groups: normal coordination for those within the expected range of motor skills, and nonnormal coordination for those whose results indicated potential delays or difficulties in motor development.

Raven Colored Progressive Matrices Test

The Raven Colored Progressive Matrices test, a nonverbal intelligence assessment, was employed to evaluate cognitive abilities. The test consists of 36 items divided into 3 series (A, Ab, and B), each containing 12 items. The total score is the sum of the correct answers across all 3 series, with series B being more difficult than Ab, and Ab being more difficult than A. Each item presents a figure with a missing part, and the child must select the correct option from 6 alternatives to complete the figure. One point was awarded for each correct answer, and 0 for incorrect answers, yielding a total score range from 0 to 36 points. The first section of the test involves completing missing parts of continuous patterns, focusing on identical or sequential patterns. Series Ab and B involve tasks related to analogies, permutations, pattern alterations, and logical relationships, with no time limit for completion [31,32]. The raw score was calculated by summing the correct answers. This score was then converted into a percentile based on the child's age and total points. The conversion of the raw score into percentile and standard scores enabled the interpretation of the child's intelligence level.

Experimental and Control Sessions

Three types of sessions were conducted: the exergame session with Just Dance 2022, the active traditional games session, and the sedentary control session. The first 2 sessions involved physical activity, while the control session was sedentary. All sessions were conducted with the participants' classmates, totaling 17 students per class, which included 15 typically developing children and 2 children with ASD. However, only the children with ASD were evaluated. Each session was administered by 2 researchers with assistance from the classroom teacher and the monitor assigned to children with ASD, totaling 4 facilitators.

The Just Dance 2022 exergame session was conducted in an air-conditioned classroom at the school. The Xbox One console, Kinect sensor, and Samsung TV (model UN32J4290) were used for the intervention. The Kinect sensor and TV were positioned 0.8 m from the ground, and the player stood at least 1.8 m from the television. The session involved 5 songs, totaling 20 minutes and 50 seconds of activity, with 15-second breaks between each song, bringing the total session time to 22 minutes and 5 seconds. Songs were selected based on their difficulty level (from easy to difficult; for more details, see Table S1 in Multimedia Appendix 1). Just Dance 2022 is a digital dance game that recognizes body movements through Kinect (Xbox) and presents various dance movements with different levels of difficulty, resembling moderate to high-intensity interval exercise. Players are scored based on the accuracy of their movements; the closer the player's movements align with the ones shown on the screen, the higher the score.

The active traditional games session was held on the school's multisport court and included 4 games: statue of the body parts, save yourself with a hug, hula hoop dance, and even-odd tag (for more details, see Table S2 in Multimedia Appendix 1).

XSL•FO
RenderX

These games required various perceptual-motor skills, such as coordination, balance, and laterality, and emphasize cooperation and respect for diversity. The activities were taught according to the Movement Curriculum of the Federal District for Elementary School Initial and Final Years [35]. The session lasted 20 minutes, and a JBL Flip 4 Bluetooth speaker, along with 15 hula hoops, was used to support the games.

The sedentary control session was applied in an air-conditioned classroom. In this session, the children engaged in a quiet activity of coloring superhero pictures for 20 minutes. No physical activity was involved, and no heart rate measurements were taken during this session.

Heart Rate Monitoring

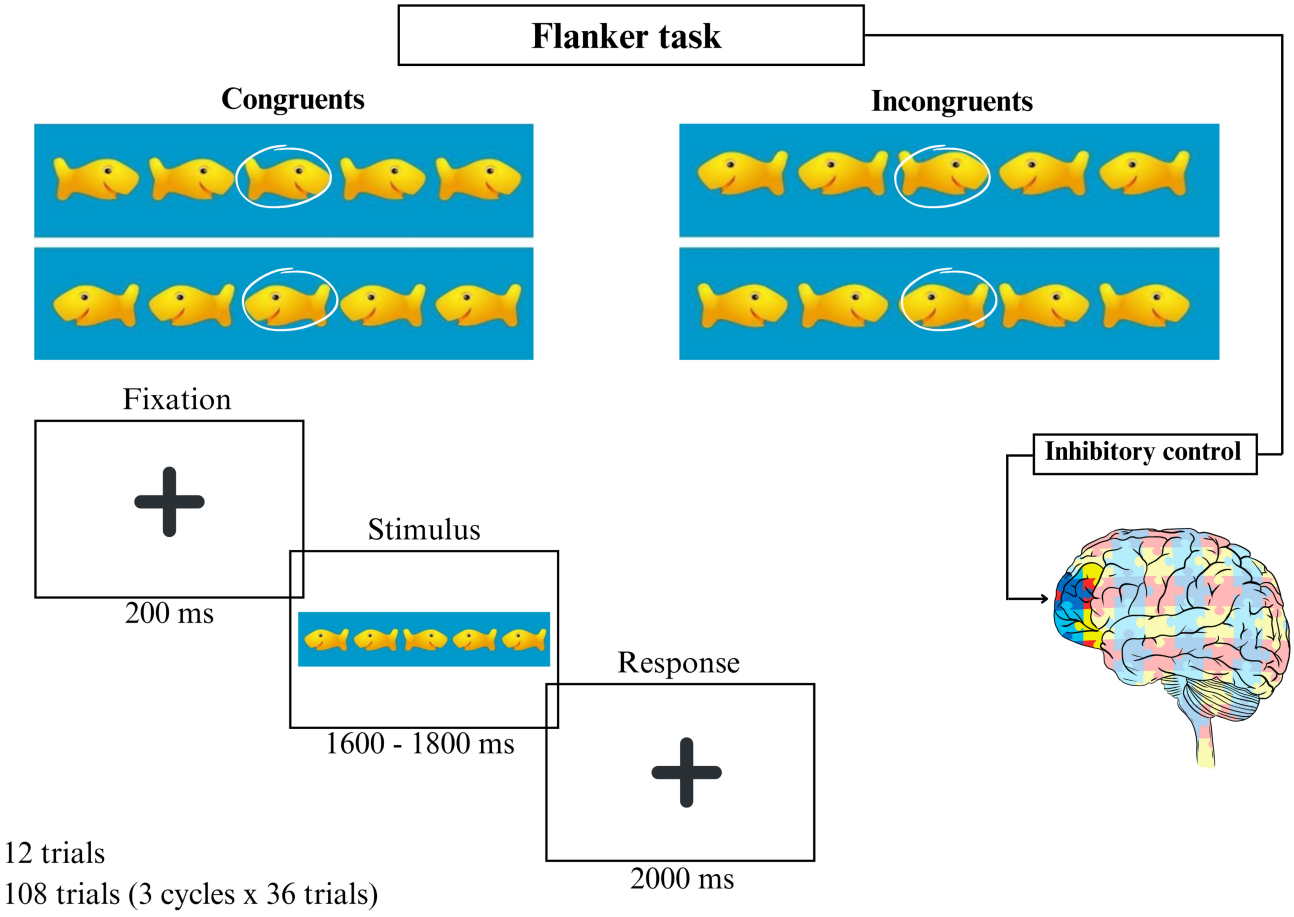
Heart rate was measured during the experimental sessions using a Galaxy Watch4 BT 40mm smartwatch. Due to the reluctance of several children with ASD to wear a chest strap, the original plan to use a Polar FT1 heart rate monitor was adjusted. Instead, heart rate was recorded at 5-minute intervals during the sessions (at 5, 10, 15, and 20 min of exercise). The mean heart rate was calculated from all recorded readings, and the percentage of maximum heart rate (HRmax) was assessed during the sessions. Activity intensity was determined based on HRmax, calculated using the Tanaka formula: 208 - 0.7 × age [36]. Although the Galaxy Watch4 smartwatch has not been specifically validated for children with ASD, studies have demonstrated its validity for monitoring heart rate during physical activity in young adult populations [35,37].

Inhibitory Control Task

A modified version of the Eriksen flanker task [38,39] was used to assess attentional inhibition. In this task, the stimuli consisted of visual representations of fish, with a target fish presented at the center, facing either to the right or left. The child's attention was directed toward the target fish, surrounded by 4 flanker fish on either side. The flanker fish, which needed to be ignored, could either face the same direction as the target (congruent condition) or the opposite direction (incongruent condition). Children were instructed to indicate the direction of the target fish by pressing the corresponding keys (← for left-facing target fish and → for right-facing target fish). The task consisted of 2 types of trials: congruent, where the direction of the flanking fish matched the target fish, and incongruent, where the direction of the flanking fish was opposite to the target fish. The experiment took place in a room with ambient light and air conditioning, and children were positioned in front of a computer programmed with the E-Prime (version 3.0; Psychological Software Tools Inc) software to present the stimuli. Before the test, the examiner explained the procedure, and a familiarization session was conducted with 12 trials (3 trials of each type: congruent and incongruent for both right- and left-facing target fish). Following the familiarization session, the experimental task was administered, consisting of 108 test trials presented in 3 blocks (36 trials per block). The trials were randomly assigned to include an equal number of congruent and incongruent trials, with the target fish presented randomly on the left or right. Each test began with the presentation of a central fixation point, followed by the display of 5 fish (2.5 cm in size) on a blue background for 200 ms. The stimuli remained on the screen until a response was made or for up to 2000 ms. The intervals between stimuli were randomly varied at 1600, 1700, or 1800 ms. The design followed previous studies using the flanker task to analyze inhibitory control in children with ASD [40-42]. For each trial, children were instructed to respond as quickly and accurately as possible by pressing a key to indicate the direction of the central target fish (left or right; Figure 1). The analyzed data included reaction time (RT) in milliseconds and the percentage of correct responses (accuracy) for both congruent and incongruent phases.

XSL•FO
RenderX

Figure 1. Illustration of the modified flanker task. Top panel: Possible combinations of congruent conditions (fish facing the same direction) or incongruent conditions (central fish facing the opposite direction). Bottom panel: Sequence of events and duration of each stimulus.



Data Analysis

The descriptive statistics for sample characteristics were presented as mean (SD; for continuous variables) and as absolute and relative frequencies (for categorical variables). The normality of the data distribution was assessed using the Shapiro-Wilk test. Repeated measures ANOVA was conducted to compare RT and accuracy between the experimental and control sessions. Following a significant ANOVA, post hoc analyses using the least significant difference test were performed. The assumption of homogeneity of variance was assessed with Levene test. To improve the reliability of the results and correct for deviations from normality, bootstrapping procedures with 1000 resamples were applied, and 95% CIs for the differences between means were obtained. Descriptive ANOVA results were reported as means and 95% bootstrapping CIs. Effect size was calculated using partial eta squared (η^2), and Cohen *d* was computed for pairwise comparisons in the post hoc analysis. Paired 2-tailed *t* tests, also conducted with bootstrapping procedures with 1000 resamples, were used to compare heart rate between experimental sessions. A $P<.05$ was considered statistically significant. All statistical analyses were performed using SPSS for Windows (version 27; IBM Corp).

Results

Sample Characterization

The sample characterization is summarized in Table 1. The mean age was 8.6 (SD 1.4) years, and the BMI indicated a wide range of nutritional statuses, including thinness (n=2, 22.2%), normal weight (n=3, 33.3%), and overweight or obesity (n=4, 44.4%). Motor coordination, assessed by the KTK, was below the normative range in 33.3% (n=3) of the participants. Cognitive assessment using the Raven test revealed a mean score within the expected range for this age group. Comorbidities were prevalent, with 44.4% (n=4) of the children diagnosed with attention-deficit/hyperactivity disorder, 11.1% (n=1) presenting with microcephaly, and 11.1% (n=1) with central auditory processing disorder. Psychological support was the most common (n=7, 77.8%), followed by speech therapy (n=5, 55.6%) and occupational therapy (n=3, 33.3%). Other forms of support included education assistance (n=3, 33.3%), psychomotricity (n=2, 22.2%), and, less frequently, music therapy (n=1, 11.1%) and ecotherapy (n=1, 11.1%).

XSL•FO
RenderX

Table . Characterization of children with autism spectrum disorder (N=9). The data are presented as mean (SD; continuous variables) and absolute and relative frequencies (categorical variables).

	Values
Age (years), mean (SD)	8.6 (1.4)
Weight (kg), mean (SD)	35.7 (14.4)
Height (cm), mean (SD)	1.38 (0.12)
BMI (kg/m^2), mean (SD)	18 (4.9)
Thinness, n (%)	2 (22.2)
Normal weight, n (%)	3 (33.3)
Overweight or obese, n (%)	4 (44.4)
KTK[a] score, mean (SD)	95.8 (16.5)
Normal coordination, n (%)	6 (66.7)
Nonnormal coordination, n (%)	3 (33.3)
Raven[b] score, mean (SD)	111.7 (16.6)
Comorbidities, n (%)	
Microcephaly	1 (11.1)
ADHD[c]	4 (44.4)
CAPD[d]	1 (11.1)
Therapeutic approaches, n (%)	
Psychologist	7 (77.8)
Speech therapist	5 (55.6)
Occupational therapist	3 (33.3)
Education support	3 (33.3)
Psychomotricity	2 (22.2)
Nutritionist	1 (11.1)
Music therapy	1 (11.1)
Ecotherapy	1 (11.1)

[a]KTK: K rperkoordinationstest für Kinder (body coordination test for children).

[b]Raven Colored Progressive Matrices test.

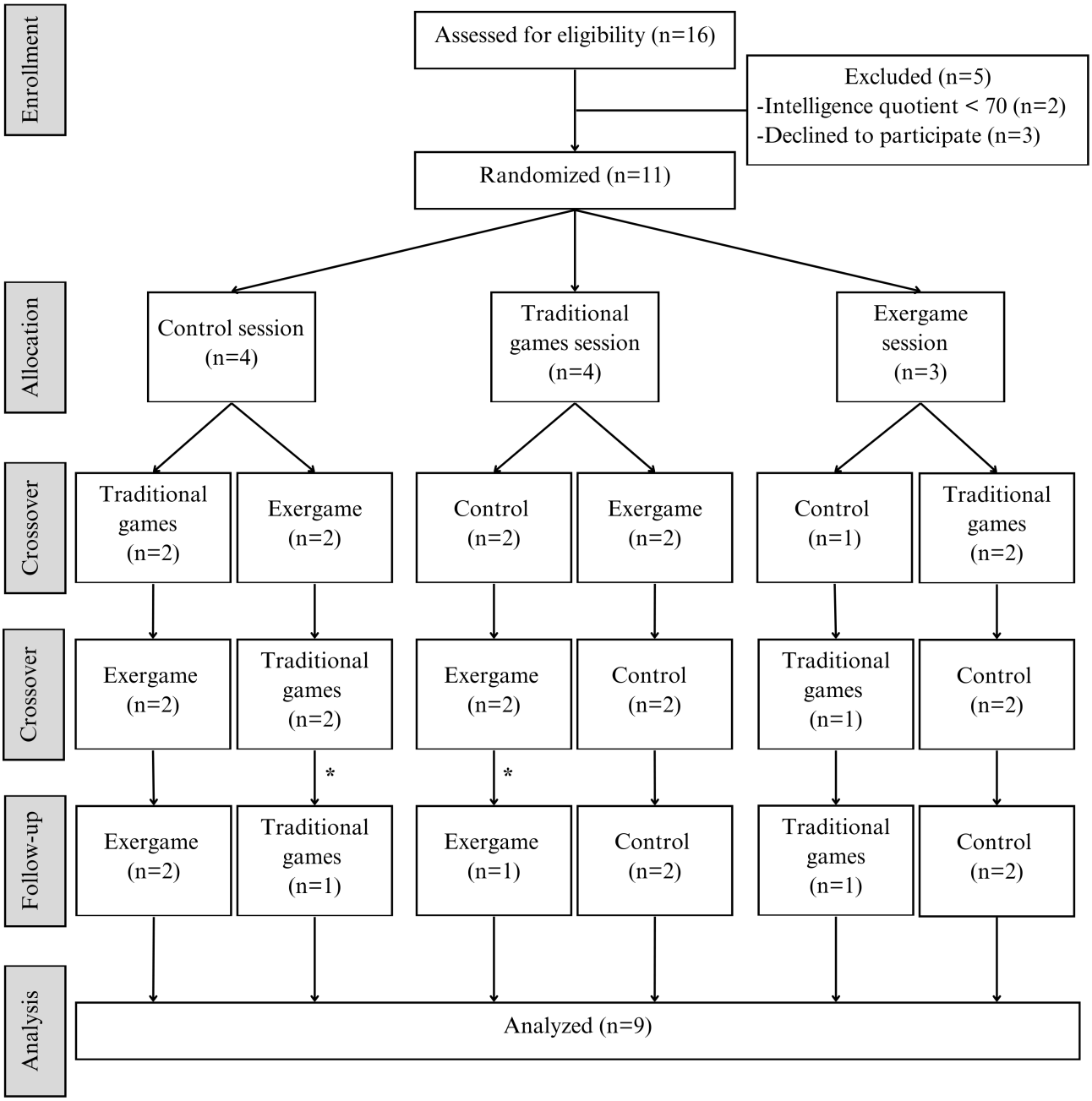
[c]ADHD: attention-deficit/hyperactivity disorder.

[d]CAPD: central auditory processing disorder.

Figure 2 depicts this study's flowchart, including all 3 sessions—exergame, traditional games, and control—and their respective order of implementation. Each participant completed all 3 conditions in a randomized sequence, with a 48-hour washout period between sessions to mitigate carryover effects and ensure data reliability.

XSL•FO
RenderX

Figure 2. Study flowchart. *Loss of follow-up.



Heart Rate Monitoring

Heart rate during the active traditional games session (153.8 bpm; 95% CI 139.4 to 168) showed a trend toward higher values compared to the exergame session (129 bpm, 95% CI 121.8 to 135.6), with *P*=.052. Similarly, the active traditional games session reached higher values of HRmax (76.1%, 95% CI 69.1 to 83.3) compared to the exergame session (63.9%, 95% CI 60.3 to 67.2), with *P*=.053. Both conditions reached moderate intensity.

Inhibitory Control Task

Table 2 and Figure 3 presents the results of RT and accuracy for the congruent and incongruent phases between the experimental and control sessions. A statistically significant effect of the condition was found for incongruent RT (*P*=.02; η^2=0.38). Post hoc analysis revealed that the exergame session resulted in a shorter RT compared to the control session (849, SD 270 vs 969, SD 295 ms; mean difference: −120 ms, 95% CI −202 to −38; *P*=.01; Cohen *d*=−1.1, 95% CI −2.0 to −0.3), as well as compared to the traditional games session (849, SD 270 vs 938, SD 330 ms; mean difference: −89 ms, 95% CI −159 to −19; *P*=.02; Cohen *d*=−1.0, 95% CI −1.8 to −0.2). No statistically significant effects were found for congruent RT, nor for congruent and incongruent accuracy (*P*>.05).

XSL•FO
RenderX

Table . Effects of a session of exergames and traditional games on inhibitory control performance in children with autism spectrum disorder (N=9). The data are presented as mean, 95% bootstrapping CI, and partial eta squared (η^2).

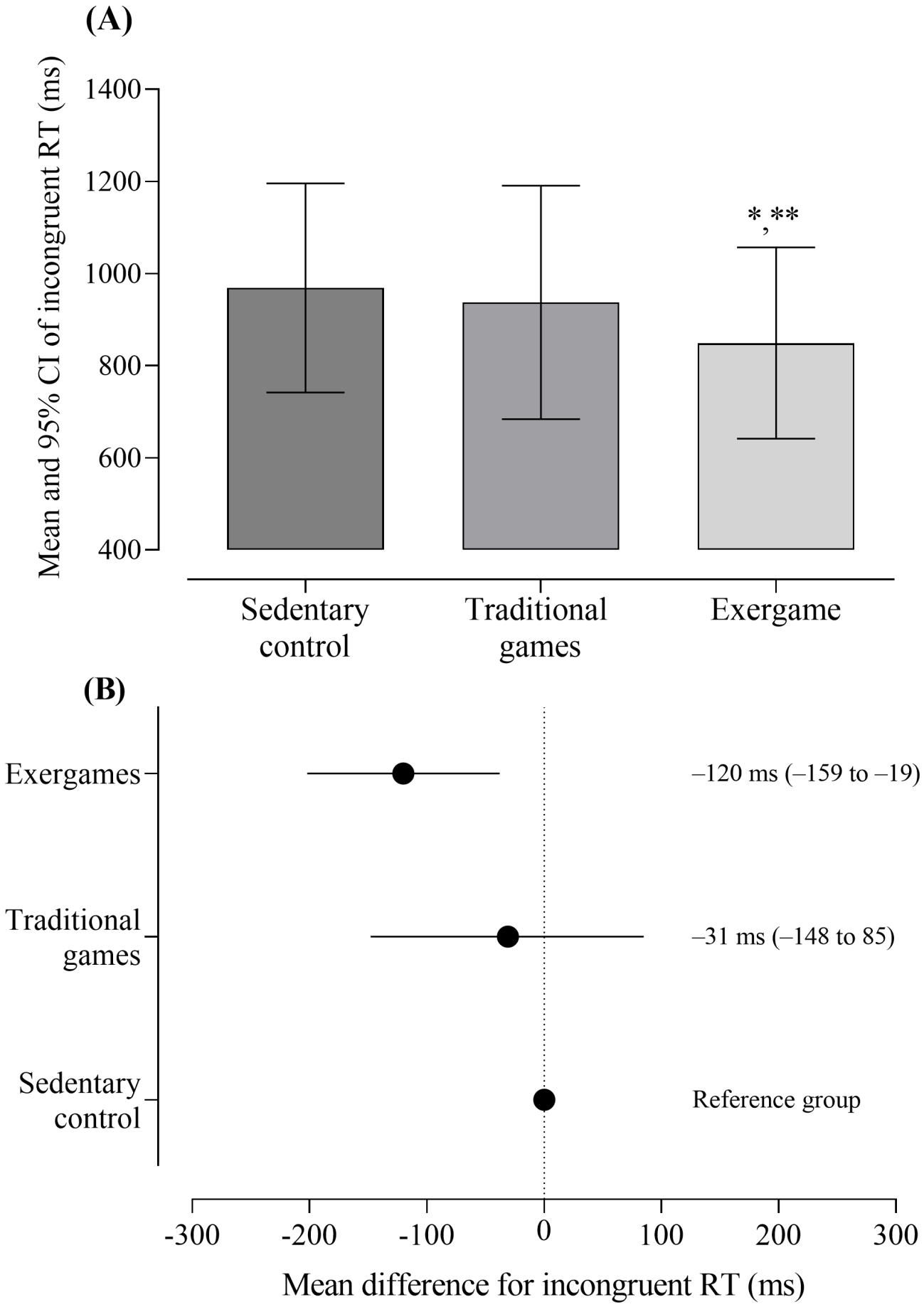
	Control session, mean (95% CI)	Traditional games session, mean (95% CI)	Exergame session, mean (95% CI)	*P* value	η^2
Congruent					
Reaction time (ms)	879 (673 to 1085)	829 (623 to 1035)	775 (571 to 978)	.08	0.27
Accuracy (%)	90.9 (77.7 to 104.2)	87.9 (73.1 to 102.6)	89.1 (77.2 to 101)	.30	0.14
Incongruent					
Reaction time (ms)	969 (742 to 1196)	938 (684 to 1191)	849 (642 to 1057)[a]	.02	0.38
Accuracy (%)	83.8 (72.7 to 95)	72.8 (54.9 to 90.8)	69.1 (43.5 to 94.7)	.06	0.30

[a]*P*=.01 compared to the control session. *P*=.02 compared to the traditional games session.

XSL•FO
RenderX

Figure 3. Effects of a session of exergames and traditional games on inhibitory control performance, specifically on incongruent RT, in children with autism spectrum disorder (N=9). (A) Mean values and CIs for incongruent RT in each experimental and control session. (B) Mean difference of the exergames and traditional games sessions compared to the reference session (sedentary control). The data are presented as mean, mean difference, and 95% CI. *P=.01 compared to the control session. **P=.02 compared to the traditional games session. RT: reaction time.

XSL•FO
RenderX



XSL•FO
RenderX

Discussion

Principal Findings

This study was designed to analyze the effects of a single session of exergames and active traditional games on the inhibitory control of school-aged children with ASD. The results highlight that the exergame session with Just Dance 2022 led to significant improvements in children's RTs, particularly in the incongruent condition, compared to both the sedentary control session and active traditional games. Notably, the exergame session proved more effective than the active traditional games session in enhancing inhibitory control, even with a lower exercise intensity. Importantly, the active traditional games session did not differ from the sedentary control session in terms of its effects on inhibitory control. These findings suggest that exergames could be a valuable resource in educational and clinical settings, with considerable potential for improving inhibitory control in children with ASD.

Exergaming has been recognized as an alternative practice that offers substantial benefits for children with ASD, including improvements in physical fitness, executive function, and self-perception, as well as promoting greater participation in moderate to vigorous physical activities [43]. Evidence suggests that even a single 20-minute session of exergames, such as those using platforms such as Dance Dance Revolution and Cybercycling, can lead to significant improvements in executive function in young individuals with ASD [44]. Our study supports these findings, demonstrating improvements in inhibitory control following an exergame session. These results underscore the potential of exergames as a viable alternative to active traditional games, particularly in school settings. Incorporating exergames into regular physical activity programs in schools can create a stimulating and engaging environment tailored to the specific needs of children with ASD, fostering their active participation and sustained engagement.

Previous studies suggest that moderate-intensity exercise can enhance inhibitory control in children, with exercise intensity playing a crucial role in this improvement. However, our study found that exergame sessions, despite being performed at a lower intensity (64% of HRmax), resulted in superior inhibitory control compared to traditional physical education sessions (76% of HRmax). This finding indicates that the benefits of exergames may go beyond physical intensity, potentially due to the cognitive demands involved. Exergames require children to engage in both physical and virtual environments, process multiple stimuli, and make quick decisions, thereby providing intense cognitive training. This type of training may activate brain regions associated with executive control more effectively than activities that primarily emphasize physical effort [45]. Consequently, future research should further explore the role of cognitive load during physical activities and its impact on inhibitory control and academic performance.

Integrating playful activities, such as games, into exercise sessions has been identified as a promising strategy for enhancing inhibitory control and, consequently, academic performance in neurotypical children [17]. While our study did not find statistically significant effects of active traditional games on inhibitory control in children with ASD, this approach should not be dismissed as it may still offer meaningful benefits. These games create opportunities for children with ASD to engage in social interactions, improve motor skills, and experience cooperative and symbolic play—areas where they often encounter difficulties due to challenges in communication, socialization, and behavioral flexibility [20]. These activities can help reduce social isolation and foster a sense of inclusion, providing an environment conducive to skill development alongside peers [21,22]. This underscores the importance of designing physical education classes that accommodate the specific needs of children with ASD, incorporating diverse practices that support their overall development.

This study has limitations that warrant consideration in interpreting the findings and guiding future research. The small sample size, consisting of 9 Brazilian children with ASD, limits the generalizability of the results to broader populations. Nearly half of the participants had comorbid attention-deficit/hyperactivity disorder, which may have influenced the findings due to the interaction of traits such as impulsivity with inhibitory control measures. Additionally, the lack of detailed information regarding the severity of ASD symptoms and participants' medication use constrains the interpretation of results. The testing protocol, despite including a familiarization session to mitigate novelty effects, could not fully address practice-related biases or the risk of inattentiveness during repeated neurocognitive assessments. Furthermore, the potential for day-to-day variations in performance, particularly among children with ASD, remains a concern, even with randomization and the crossover design. The 48-hour washout period, selected based on existing literature, may not have completely eliminated carryover effects, especially given the variability in cognitive responses among this population. Another limitation relates to the personalization of the intervention. Although the choice of Just Dance 2022 was justified by its adaptability to participants' movements and immediate feedback, fostering engagement, the lack of individual customization in song selection and difficulty progression may have impacted outcomes. Finally, while a significant improvement in RT was observed following the exergame session, this was accompanied by a nonsignificant reduction in accuracy compared to the control session, suggesting a potential trade-off between speed and accuracy.

The findings of this study hold significant implications for educational and clinical settings, as they demonstrate that exergaming-based exercise activities positively impact inhibitory control in children with ASD. This is particularly important because it highlights the potential of incorporating these activities into school environments, offering opportunities for both cognitive and physical development in an inclusive and adaptive setting. Integrating exergames into the school curriculum can support the academic and social growth of these children while encouraging more active and meaningful participation in physical activities, which are vital for their overall well-being. Future research should address these limitations by incorporating larger and more diverse samples, stratifying participants based on comorbidities, collecting detailed clinical data, exploring longer washout periods, and

XSL•FO
RenderX

implementing individualized interventions to ensure more robust and reliable findings. Additionally, further studies should investigate possible trade-offs between speed and accuracy in inhibitory control tasks to enhance our understanding of the observed effects.

Conclusions

In conclusion, even a single session of the Just Dance 2022 exergame showed a significant and promising benefit for improving inhibitory control in school-aged children with ASD. The inclusion of exergames in the school routine holds great potential, as physical activity has been shown to enhance executive functions. Despite this, the use of exergames remains uncommon in many physical education programs. These games, however, can serve as an innovative pedagogical tool to enhance executive functions and help children with ASD meet recommended levels of healthy physical activity. Moreover, the use of exergames does not necessarily require a video game console; they can be implemented by projecting internet-available videos in the classroom, making them more accessible and feasible for widespread use.

Acknowledgments

We express our gratitude to Dr. Clay Brites, PhD (in memoriam), and Dr. Raiane Maiara dos Santos Pereira, PhD, for their invaluable contributions through the critical review of this paper. We also thank researchers Valéria dos Santos Pedrosa, MSc, and Rafael dos Santos Cruz, MSc, for their essential support in data collection. Our appreciation extends to the administration of Escola Classe 10 in Taguatinga, Federal District (Brazil), for graciously providing the school space for this research. Lastly, we acknowledge the Fundação de Apoio à Pesquisa do Distrito Federal (FAP/DF, Research Support Foundation of the Federal District) for funding this publication (grant 00193-00001758/2023-11), as well as the Catholic University of Brasília (UCB) for providing a partial master's scholarship to the first author.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Protocols of the experimental sessions.
[DOCX File, 18 KB - games_v13i1e65562_app1.docx]

Checklist 1
CONSORT (Consolidated Standards of Reporting Trials) checklist.
[PDF File, 65 KB - games_v13i1e65562_app2.pdf]

References

1. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, Text Revision: American Psychiatric Association Publishing; 2022. [doi: 10.1176/appi.books.9780890425787]
2. First MB, Clarke DE, Yousif L, Eng AM, Gogtay N, Appelbaum PS. DSM-5-TR: rationale, process, and overview of changes. Psychiatr Serv 2023 Aug 1;74(8):869-875. [doi: 10.1176/appi.ps.20220334] [Medline: 36510761]
3. Parrella NF, Hill AT, Dipnall LM, Loke YJ, Enticott PG, Ford TC. Inhibitory dysfunction and social processing difficulties in autism: a comprehensive narrative review. J Psychiatr Res 2024 Jan;169:113-125. [doi: 10.1016/j.jpsychires.2023.11.014] [Medline: 38016393]
4. Faja S, Darling LN. Variation in restricted and repetitive behaviors and interests relates to inhibitory control and shifting in children with autism spectrum disorder. Autism 2019 Jul;23(5):1262-1272. [doi: 10.1177/1362361318804192] [Medline: 30394786]
5. Pellicano E, Kenny L, Brede J, Klaric E, Lichwa H, McMillin R. Executive function predicts school readiness in autistic and typical preschool children. Cogn Dev 2017 Jul;43:1-13. [doi: 10.1016/j.cogdev.2017.02.003]
6. Whedon M, Perry NB, Bell MA. Relations between frontal EEG maturation and inhibitory control in preschool in the prediction of children's early academic skills. Brain Cogn 2020 Dec;146:105636. [doi: 10.1016/j.bandc.2020.105636] [Medline: 33197766]
7. Marques S, Correia-de-Sá T, Guardiano M, Sampaio-Maia B, Ferreira-Gomes J. Emotion dysregulation and depressive symptoms mediate the association between inhibitory control difficulties and aggressive behaviour in children with ADHD. Front Psychiatry 2024;15:1329401. [doi: 10.3389/fpsyt.2024.1329401] [Medline: 38690203]
8. Browne RAV, Costa EC, Sales MM, Fonteles AI, de Moraes J, de França Barros J. Acute effect of vigorous aerobic exercise on the inhibitory control in adolescents. Revista Paulista de Pediatria (English Edition) 2016 Jun;34(2):154-161. [doi: 10.1016/j.rppede.2016.01.005]

XSL•FO
RenderX

9. Bremer E, Graham JD, Heisz JJ, Cairney J. Effect of acute exercise on prefrontal oxygenation and inhibitory control among male children with autism spectrum disorder: an exploratory study. Front Behav Neurosci 2020;14:84. [doi: 10.3389/fnbeh.2020.00084] [Medline: 32655381]
10. Stillman CM, Erickson KI. Physical activity as a model for health neuroscience. Ann N Y Acad Sci 2018 Sep;1428(1):103-111. [doi: 10.1111/nyas.13669] [Medline: 29732566]
11. da Silva WQA, Fontes EB, Forti RM, et al. Affect during incremental exercise: the role of inhibitory cognition, autonomic cardiac function, and cerebral oxygenation. PLoS ONE 2017;12(11):e0186926. [doi: 10.1371/journal.pone.0186926] [Medline: 29091915]
12. Aubert S, Barnes JD, Abdeta C, et al. Global Matrix 3.0 physical activity report card grades for children and youth: results and analysis from 49 countries. J Phys Act Health 2018 Nov 1;15(S2):S251-S273. [doi: 10.1123/jpah.2018-0472] [Medline: 30475137]
13. Martins C, Mota J, Goulart N, et al. Compliance with the 24-h movement behaviors guidelines among urban and rural Brazilian preschoolers. Am J Hum Biol 2024 Jun;36(6):e24045. [doi: 10.1002/ajhb.24045] [Medline: 38284305]
14. Oliveira KSC, Fontes DE, Longo E, Leite HR, Camargos ACR. Motor skills are associated with participation of children with autism spectrum disorder. J Autism Dev Disord 2023 Apr;53(4):1403-1412. [doi: 10.1007/s10803-021-05318-9] [Medline: 34669103]
15. Varela AR, Hallal PC. Does every move really count towards better health? Lancet Glob Health 2024 Aug;12(8):e1215-e1216. [doi: 10.1016/S2214-109X(24)00173-6]
16. Every move counts towards better health – says WHO. World Health Organization. URL: https://www.who.int/news/item/25-11-2020-every-move-counts-towards-better-health-says-who [accessed 2025-02-06]
17. Koepp AE, Gershoff ET, Castelli DM, Bryan AE. Preschoolers' executive functions following indoor and outdoor free play. Trends Neurosci Educ 2022 Sep;28:100182. [doi: 10.1016/j.tine.2022.100182] [Medline: 35999013]
18. Jylänki P, Mbay T, Hakkarainen A, Sääkslahti A, Aunio P. The effects of motor skill and physical activity interventions on preschoolers' cognitive and academic skills: a systematic review. Prev Med 2022 Feb;155:106948. [doi: 10.1016/j.ypmed.2021.106948] [Medline: 34974071]
19. Silva A, Ferraz R, Forte P, Teixeira JE, Branquinho L, Marinho DA. Multivariate training programs during physical education classes in school context: theoretical considerations and future perspectives. Sports (Basel) 2022 Jun 3;10(6):89. [doi: 10.3390/sports10060089] [Medline: 35736829]
20. González-Sala F, Gómez-Marí I, Tárraga-Mínguez R, Vicente-Carvajal A, Pastor-Cerezuela G. Symbolic play among children with autism spectrum disorder: a scoping review. Children (Basel) 2021 Sep 12;8(9):801. [doi: 10.3390/children8090801] [Medline: 34572233]
21. Hedgecock JB, Dannemiller LA, Shui AM, Rapport MJ, Katz T. Associations of gross motor delay, behavior, and quality of life in young children with autism spectrum disorder. Phys Ther 2018 Apr 1;98(4):251-259. [doi: 10.1093/ptj/pzy006] [Medline: 29325143]
22. Hickingbotham MR, Wong CJ, Bowling AB. Barriers and facilitators to physical education, sport, and physical activity program participation among children and adolescents with psychiatric disorders: a systematic review. Transl Behav Med 2021 Sep 15;11(9):1739-1750. [doi: 10.1093/tbm/ibab085] [Medline: 34166515]
23. Eng CM, Flynn RM, Thiessen ED, Fisher AV. A literature review on the effects of exergames on executive function in youth. Technol Mind Behav 2023;4(3). [doi: 10.1037/tmb0000118] [Medline: 38162378]
24. Warsinsky S, Schmidt-Kraepelin M, Rank S, Thiebes S, Sunyaev A. Conceptual ambiguity surrounding gamification and serious games in health care: literature review and development of game-based intervention reporting guidelines (GAMING). J Med Internet Res 2021 Sep 10;23(9):e30390. [doi: 10.2196/30390] [Medline: 34505840]
25. Chaidi I, Pergantis P, Drigas A, Karagiannidis C. Gaming platforms for people with ASD. J Intell 2024 Nov 27;12(12):122. [doi: 10.3390/jintelligence12120122] [Medline: 39728090]
26. López-Serrano S, Ruiz-Ariza A, Suarez-Manzano S, de la Torre Cruz MJ, University of Jaen. "Dance dance" education. A real "revolution" for the classroom? [Article in Spanish]. MLSER 2017;1(1):7-18. [doi: 10.29314/mlser.v1i1.22]
27. Dwan K, Li T, Altman DG, Elbourne D. CONSORT 2010 statement: extension to randomised crossover trials. BMJ 2019 Jul 31;366:l4378. [doi: 10.1136/bmj.l4378] [Medline: 31366597]
28. Random number calculators. GraphPad. URL: https://www.graphpad.com/quickcalcs/randomn1/ [accessed 2025-02-06]
29. American Psychiatric Association. Manual Diagnóstico e Estatístico de Transtornos Mentais: DSM-5 [Book in Spanish], 5th edition: Artmed; 2014.
30. Kiphard EJ, Schilling VF. Korper-Koordinations-Test für Kinder KTK: Manual Von Fridhelm Schilling [Book in German]: Beltz Test; 1974.
31. Raven JC, Court JH. Raven's Progressive Matrices and Raven's Coloured Matrices: HK Lewis; 1986.
32. Sbicigo JB, Bosa CA, Bandeira DR, Teixeira M, Salles JD. Desempenho no teste de raven: diferenças entre crianças-adolescentes com transtorno do espectro autista e com dificuldades de leitura [Article in Portugese]. AvalPsicol 2019;18(2):192-200. [doi: 10.15689/ap.2019.1802.16917.10]
33. Heyward VH. In: Dornelles M DS, Ugrinowitsch C, editors. Avaliação Física e Prescrição de Exercício: Técnicas Avançadas [Book in Portugese], 6th edition: Artmed; 2013.

XSL•FO
RenderX

34. Data and statistics on autism spectrum disorder. US Centers for Disease Control and Prevention. URL: https://www.cdc.gov/autism/data-research/index.html [accessed 2025-02-06]
35. Lima CS, Bertocco FC, de Oliveira JIV, de Souza TMF, da Silva EP, Von Zuben FJ. Assessment of Samsung Galaxy Watch4 PPG-based heart rate during light-to-vigorous physical activities. IEEE Sens Lett 2024;8(7):1-4. [doi: 10.1109/LSENS.2024.3408089]
36. Spitznagel AS, Dos Santos DS, Branco KR, Schiwe D, De Oliveira V, Heinzmann-Filho JP. Equações utilizadas para a predição de frequência cardíaca máxima na população pediátrica não atlética: uma revisão sistemática [Article in Portugese]. Sci Med 2019;29(4):e34286. [doi: 10.15448/1980-6108.2019.4.34286]
37. Nissen M, Slim S, Jäger K, et al. Heart rate measurement accuracy of Fitbit Charge 4 and Samsung Galaxy Watch Active2: device evaluation study. JMIR Form Res 2022 Mar 1;6(3):e33635. [doi: 10.2196/33635] [Medline: 35230250]
38. Eriksen BA, Eriksen CW. Effects of noise letters upon the identification of a target letter in a nonsearch task. Percept Psychophys 1974 Jan;16(1):143-149. [doi: 10.3758/BF03203267]
39. Eriksen CW. The flankers task and response competition: a useful tool for investigating a variety of cognitive problems. Vis cogn 1995 Jun;2(2-3):101-118. [doi: 10.1080/13506289508401726]
40. Christ SE, Holt DD, White DA, Green L. Inhibitory control in children with autism spectrum disorder. J Autism Dev Disord 2007 Jul;37(6):1155-1165. [doi: 10.1007/s10803-006-0259-y] [Medline: 17066307]
41. Cremone-Caira A, Vaidyanathan A, Hyatt D, Gilbert R, Clarkson T, Faja S. Test-retest reliability of the N2 event-related potential in school-aged children with autism spectrum disorder (ASD). Clin Neurophysiol 2020 Feb;131(2):406-413. [doi: 10.1016/j.clinph.2019.09.024] [Medline: 31877490]
42. Geurts HM, Luman M, van Meel CS. What's in a game: the effect of social motivation on interference control in boys with ADHD and autism spectrum disorders. J Child Psychol Psychiatry 2008 Aug;49(8):848-857. [doi: 10.1111/j.1469-7610.2008.01916.x] [Medline: 18564068]
43. Fang Q, Aiken CA, Fang C, Pan Z. Effects of exergaming on physical and cognitive functions in individuals with autism spectrum disorder: a systematic review. Games Health J 2019 Apr;8(2):74-84. [doi: 10.1089/g4h.2018.0032] [Medline: 30332294]
44. Anderson-Hanley C, Tureck K, Schneiderman RL. Autism and exergaming: effects on repetitive behaviors and cognition. Psychol Res Behav Manag 2011;4:129-137. [doi: 10.2147/PRBM.S24016] [Medline: 22114543]
45. Chen J, Zhou X, Wu X, Gao Z, Ye S. Effects of exergaming on executive functions of children: a systematic review and meta-analysis from 2010 to 2023. Arch Public Health 2023 Oct 13;81(1):182. [doi: 10.1186/s13690-023-01195-z] [Medline: 37833741]

Abbreviations

ASD: autism spectrum disorder
CONSORT: Consolidated Standards of Reporting Trials
***DSM-5*:** *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition*
HRmax: maximum heart rate
KTK: K rperkoordinationstest für Kinder
ReBEC: Brazilian Registry of Clinical Trials
RT: reaction time

Edited by R Baranyi; submitted 20.08.24; peer-reviewed by N Tottori, WC Chin; revised version received 24.01.25; accepted 30.01.25; published 05.03.25.

Please cite as:
Miranda JM, Browne RAV, da Silva WQA, Rodrigues dos Santos JP, Campbell CSG, Ramos IA
Effects of a Session of Exergames and Traditional Games on Inhibitory Control in Children With Autism Spectrum Disorder: Randomized Controlled Crossover Trial
JMIR Serious Games 2025;13:e65562
URL: https://games.jmir.org/2025/1/e65562
doi:10.2196/65562

© Juliana Macedo Miranda, Rodrigo Alberto Vieira Browne, Weslley Quirino Alves da Silva, João Paulo Rodrigues dos Santos, Carmen Silvia Grubert Campbell, Isabela Almeida Ramos. Originally published in JMIR Serious Games (https://games.jmir.org), 5.3.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium,

XSL•FO
RenderX

provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Effect of a Gamified Family-Based Exercise Intervention on Adherence to 24-Hour Movement Behavior Recommendations in Preschool Children: Single-Center Pragmatic Trial

Gaizka Legarra-Gorgoñon[1*], MSc; Yesenia García-Alonso[1*], PhD; Robinson Ramírez-Vélez[1,2*], PhD; Loreto Alonso-Martínez[1*], MSc; Mikel Izquierdo[1,2*], Prof Dr; Alicia M Alonso-Martínez[1*], PhD

[1]Navarrabiomed, Hospital Universitario de Navarra, Universidad Pública de Navarra, Instituto de Investigación Sanitaria de Navarra, Avda Barañain s/n, Pamplona, Spain

[2]Biomedical Research Center (CIBER) of Frailty and Healthy Aging, Instituto de Salud Carlos III, Madrid, Spain

[*]all authors contributed equally

Corresponding Author:
Mikel Izquierdo, Prof Dr
Navarrabiomed, Hospital Universitario de Navarra, Universidad Pública de Navarra, Instituto de Investigación Sanitaria de Navarra, Avda Barañain s/n, Pamplona, Spain

Abstract

Background: Adherence to 24-hour movement behavior recommendations, including physical activity (PA), sedentary time, and sleep, is essential for the healthy development of preschool children. Gamified family-based interventions have shown the potential to improve adherence to these guidelines, but evidence of their effectiveness among children is limited.

Objective: This study aimed to evaluate the effectiveness of a gamified family-based exercise intervention in promoting adherence to 24-hour movement behavior recommendations among preschool-aged children.

Methods: This 12-week study is a single-center, pragmatic randomized controlled trial that included 80 preschool children (56% boys) and their families, who were randomly assigned to either the gamification group (n=40) or the control group (n=40). The "3, 2, 1 Move on Study" incorporates family-oriented physical activities and gamification techniques to increase PA domains, reduce sedentary behavior, and improve sleep patterns. The primary outcome was to increase moderate to vigorous PA (MVPA) by 5 minutes/day, as measured by accelerometer at follow-up. Accelerometer-determined daily time spent (PA domains, sedentary behavior, and sleep), physical fitness (cardiorespiratory, speed-agility, muscular, physical fitness z-score), basic motor competencies (self-movement and object movement), and executive function (memory, cognitive flexibility, and inhibitory control) were also included as secondary outcomes.

Results: The 71 participants included in the per-protocol analyses (32 girls, 45%; 39 boys, 55%) had a mean (SD) age of 5.0 (0.5) years. Change in MVPA per day after the intervention (12 weeks) increased in both groups by +25.3 (SD 24.6) minutes/day in the gamification group and +10.0 (SD 31.4) minutes/day in the routine care group, but no significant between-group differences were observed (8.62, 95% CI –5.72 to 22.95 minutes/day, ηp^2=.025; *P*=.23). The analysis of secondary outcomes showed significant between-group mean differences in the change in physical behaviors derived from the accelerometers from baseline to follow-up of 26.44 (95% CI 8.93 to 43.94) minutes/day in favor of light PA (ηp^2=.138; *P*=.01) and 30.88 (95% CI 4.36 to 57.41) minutes/day in favor of total PA, which corresponds to a large effect size (ηp^2=.087; *P*=.02). Likewise, the gamification group substantially increased their score in standing long jump and physical fitness z-score from baseline (*P*<.05).

Conclusions: In the "3, 2, 1 Move on Study," a gamified intervention showed a modest but relevant increase in MVPA and other domains of 24-hour movement behavior among preschool-aged children. Therefore, gamified family-based interventions may provide a viable alternative to improve adherence to 24-hour movement behavior recommendations.

Trial Registration: ClinicalTrials.gov NCT05741879; https://clinicaltrials.gov/study/NCT05741879?tab=history

(JMIR Serious Games 2025;13:e60185) doi:10.2196/60185

KEYWORDS

children; gamification; exercise; physical fitness; domains of physical activity; game; fitness; child; family-based; exercise program; randomized controlled trial; strength

XSL•FO
RenderX

Introduction

Physical activity (PA) is essential for early childhood development [1] and positively impacts skeletal growth [2], metabolism [3], and cardiovascular health [4]. At the psychological level, PA has been shown to enhance self-esteem [5], reduce stress [6], and improve social behaviors. In 2019, the World Health Organization (WHO) issued guidelines for children under 5 years of age. These underlined the importance of the early habits of PA adoption for health promotion and prevention of obesity throughout life. For children aged 3 - 4 years, the WHO recommends at least 180 minutes of daily PA, including 60 minutes of moderate to vigorous intensity, to support physical, cognitive, and motor development [7]. Additionally, children should not remain sedentary for more than an hour at a time, screen time should be limited to less than 1 hour per day, and they should get 10 - 13 hours of sleep with regular sleep and wake times [7]. Preschool-aged children have a unique need to establish healthy activity patterns, as they rely on structured, supervised activities. Effective interventions for this age group should offer engaging and developmentally appropriate activities that encourage movement and active play.

Adherence to these PA recommendations is associated with improved mental health outcomes, including reduced symptoms of anxiety and depression [8,9] and the promotion of healthier lifestyles [10]. However, an upward trend in childhood sedentarism has been observed in recent decades [11]. A systematic review by Tucker et al [12] reported that only 54% of children aged 2 - 6 years met the PA guidelines of the National Association of Sport Officials. Research further indicates a decline in PA from early childhood to adolescence [13,14], which contributes to the increased rates of overweight and obesity in this population, as well as a decline in cardiorespiratory fitness (CRF) [15,16].

Physical inactivity and sedentary behavior (SB) have contributed to declining physical fitness (PF) levels among children [17]. Given the established health benefits of engaging in regular moderate-to-high-intensity PA, this level of PA is recommended to support improvements in PF components, including CRF, musculoskeletal fitness (muscle strength and endurance), and flexibility [18]. By emphasizing these intensity levels, one can maximize their potential fitness gains to help with overall health and functional capacity [19].

PA is also important for the development of motor skills in young children because it provides the practice and experience needed to improve coordination, balance, agility, and other essential motor competencies. Indeed, previous cross-sectional studies have noted a positive relationship between PA and PF in preschoolers, with higher PA levels also linked to superior motor skills performance [20]. Motor skill competency is considered an integral part of PF in children since it enables them to successfully engage in various activities—skills that form the foundation for lifelong active habits. The development of these motor skills tends to contribute to overall PF by enhancing cardiovascular health, muscular strength, and endurance; increasing a child's confidence and enjoyment of physical activities; and encouraging the development of positive attitudes toward regular PA. In this study, we considered motor skill competency as an integral part of PF, emphasizing the importance of early interventions targeting PA levels and motor development to promote holistic PF outcomes in preschool-aged children [21].

Gamification techniques, including badges, leaderboards, streaks, team-based competition, and story elements, have been proven to improve many health behaviors beyond weight loss [22]. Specifically, mobile health interventions using "exergames" have successfully embedded gamified elements to enhance enjoyment and increase engagement in PA [23-25]. Evidence on younger populations has shown that gamified interventions have the potential to enhance PA levels, motor skills, social interaction, and emotional well-being [26]. For example, virtual reality combined with gamification in PA programs has been reported to enhance motivation, thereby increasing the extent of engagement in exercise [27]. In adolescents, gamification particularly enhances motivation and adherence to PA via social or competitive elements appealing to this phase of development [28]. These results emphasize the need for age-appropriate gamified interventions that allow each age group to engage optimally and achieve even better improvements in PA levels and health outcomes. To date, few attempts to gamify weight loss have prioritized increasing PA, and they have shown only modest success [22]. Thus, gamification strategies can bolster health habits and promote appropriate PA to inspire behavioral changes through enhanced motivation [29]. In this context, Sailer et al [30] posit that gamification can potentially improve motivational challenges while being supported by robust implementation models.

Therefore, integrating motivational tactics is likely to positively affect PA behaviors and the related motivational outcomes. To better understand under which circumstances gamification improves adherence to 24-hour movement behavior recommendations, for whom the effects would be most beneficial, and which aspects of physical health would be best targeted, we designed a nature-based intervention, the "3, 2, 1 Move on Study" trial, consisting of gamified training for 15-37 minutes per week for 12 weeks. We hypothesized that the "3, 2, 1 Move on Study" would lead to improved PA levels, reduced SB, and enhanced sleep patterns.

Methods

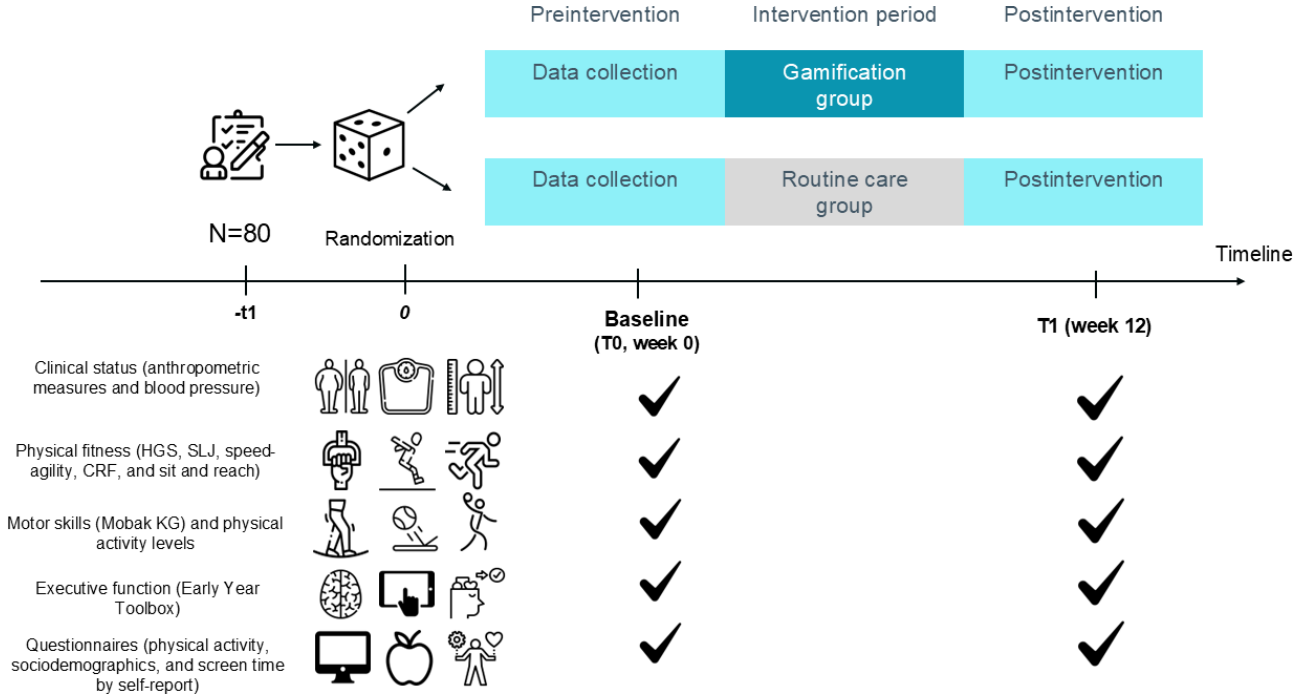
Study Design and Participants

The "3, 2, 1 Move on Study" is a single-center, pragmatic randomized controlled trial (registered on ClinicialTrials.gov; NCT05741879) of 12 weeks of gamified training from November 2022 to February 2023, delivered through Iturrama Primary Care Center (Pamplona, Spain).

This trial followed the CONSORT (Consolidated Standards of Reporting Trials) reporting guideline, including reporting of protocol deviations [31]. The overall study procedure is illustrated in Figure 1.

XSL•FO
RenderX

Figure 1. Study design of the "3, 2, 1 Move on Study." Participating children were randomly assigned to either the 12 weeks of routine care group (n=40) or the gamification group (n=40). CRF: cardiorespiratory fitness; HGS: handgrip strength; SLJ: standing long jump.



A brief description of the methodology was peer reviewed and published [31]. All participants were from the Iturrama Primary Care Center (Pamplona, Spain). The parents or legal guardians of the children provided written informed consent to participate in the trial. Children must not have a diagnosis of any musculoskeletal, cardiovascular, pulmonary, or orthopedic problems; disabilities precluding the participant from being physically active; or any other physical condition that precludes the participant from being physically active. Children who did not comply with established procedures and those who did not understand Spanish were excluded from the study. Throughout this protocol, the trial staff was responsible for weekly monitoring of various aspects, including meetings discussing topics related to managing the platform and performing exercises. Additionally, dedicated email addresses and phone numbers were provided to address any questions or concerns.

Ethical Considerations

The study was conducted in accordance with the Declaration of Helsinki and approved by the Clinical Research Ethics Committee of Navarra (PI_2021/111). Participation of children was voluntary, and their parents provided written informed consent. No compensation was provided to participants.

Randomization

The participants were randomly assigned to 1 of the 2 groups before the commencement of the study. The stratified randomization (1:1 balance between the number of children in the gamification and control groups) was computerized by an independent, blinded statistician not involved in the study (Research Randomizer V.4). Allocation assignment was concealed to the researchers at the analysis stage.

Intervention and Control

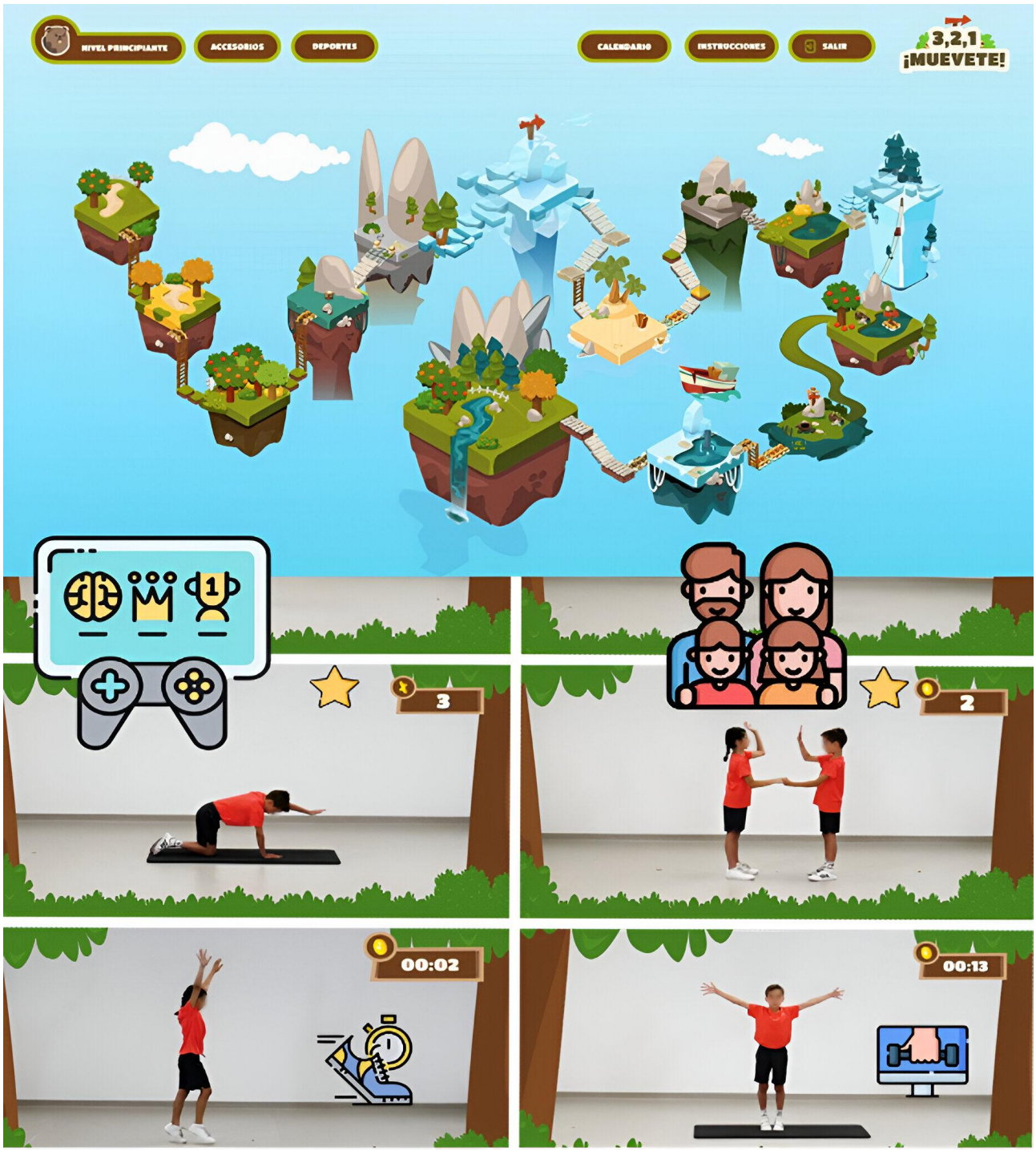
A family-based, gamified intervention was developed by experts in education and sport sciences to create a structured PA program specifically designed for preschool children. This approach emphasizes the use of playfulness, which is important to engage young children, while integrating game mechanics such as point systems, rewards, and progress tracking. These elements were strategically implemented to enhance motivation and promote adherence to the PA program. The intervention activities incorporated customizable features, including adjustable difficulty levels and personalized avatars, to sustain participants' interest and engagement. Family involvement was emphasized as a core component, recognizing that young children often require adult support and encouragement to fully engage in the program. The intervention design was based on the premise that gamified family-centered activities can establish a supportive setting, thereby making PA both enjoyable and sustainable for children and their parents. Participants in this group were invited to a 12-week exercise program on an online gamified platform, with the idea of encouraging children to engage in PA with their families either at home or in an outdoor space. Each child had a user account to log in and complete exercises twice per week, with one session available from Monday to Wednesday and a repeat session available from Thursday to Sunday. The program comprised 3 phases: a warm-up/activation phase, a training phase (featuring body weight strength exercises, CRF exercises, and color-coded exercises), and a cooldown phase. The number of exercises and repetitions varied from week to week. The warm-up phase consisted of 3 minutes of continuous light-intensity training, which lasted 10 - 26 minutes with 6 - 12 repetitions, 20 - 40 seconds of work, and 10 - 20 seconds of recovery. The cooldown phase involved 3 minutes of static and dynamic exercises. Exercises were changed weekly, gradually increasing

XSL•FO
RenderX

in length and intensity. The parents or guardians of the children supervised the program.

Figure 2 illustrates the main screen of the platform viewed by participants upon completing each session. At the beginning of the game, the children were prompted to select and customize their avatars by entering their name, choosing the avatar type, and customizing its appearance based on their preferences. The personalization feature was designed to instill a sense of ownership and, consequently, enhance participant engagement. As participants progressed through the program, they unlocked new "worlds" and earned additional rewards. The platform consisted of 14 worlds, with each milestone providing opportunities to access new accessories and further personalize avatars (after completing worlds 3, 6, 9, and 12). This reward system was designed to maintain interest and provide continuous progress reinforcement, which is crucial for sustained motivation among younger users.

Figure 2. Framework screen of the gamified online platform.



Children advanced through a structured level system, progressing from the "novice level" to the "expert level." This structure ensured they were consistently challenged in line with their developmental skills while fostering a systematic sense of progress. The platform also featured a calendar that guided participants on the appropriate session or world to engage with,

XSL•FO
RenderX

based on the program's week. This scheduling tool facilitated participants in organizing their training progression, establishing clear and attainable goals, and monitoring their improvement over time. Overall, the gamified platform integrated PA with an engaging and dynamic experience, encouraging children to remain active while enjoying the process.

To increase motivation and adherence, the research team was responsible for ensuring participants' compliance with the study requirements by supervising their participation in 2 weekly sessions. Each week, access to the exercise platform was monitored by evaluating the total usage time per session and verifying the completion of the prescribed training program. Participants who did not adhere to these guidelines were contacted via telephone to identify the reasons for noncompliance and to provide encouragement and support to continue their participation.

The control group continued with usual routine care according to the international guidelines set by the WHO [7]. Activities included regular physical education classes at school and opportunities to play with friends at parks or in backyards. The parents or guardians in the control group were not prohibited from noting the children's PA levels on their own.

Outcome Measures

The primary outcome was a change in the child's 24-hour movement behavior recommendations between baseline and 12 weeks (increasing PA levels, reducing SB, and enhancing sleep patterns) as measured by the GENEActiv Original triaxial accelerometer, with a frequency of 87.5 Hz, over 7 consecutive days on the nondominant wrist to record PA and sleep information. Specifically, the "3, 2, 1 Move on Study" will be able to detect a difference in the mean daily MVPA between the gamification group and routine care group of ±5 minutes at follow-up. Activity patterns at baseline and during the intervention were collected through GENEActiv PC software (version 3.3) and processed and analyzed using the R package GGIR as described elsewhere [32]. Briefly, thresholds for children aged 4 - 6 years were as follows: <56.3 (mg) for sedentary time, ±56.3 mg light PA (LPA), 191.6 mg moderate PA (MPA), and ≥695.8 mg vigorous PA (VPA). Sleep patterns were calculated using the algorithm developed by van Hees et al [33].

Secondary outcomes were changes in health-related PF as measured by the PREFIT battery [34]. Participants completed each test consecutively, except for the 20-meter shuttle run, where multiple children ran simultaneously in small groups. An analog handgrip dynamometer was used to measure upper body muscular strength (grip strength Dynamometer T.K.K. 5001 Grip). The handgrip strength (HGS) test was performed twice for each hand, where the children gradually squeezed the dynamometer for 2 - 3 seconds. The highest value for each hand was recorded, and the average of these 2 results was calculated to analyze the upper-body muscular strength measurements. The standing long jump (SLJ) test was used to assess lower body strength. The children performed a maximal horizontal jump from a standing position, landing on both feet, while maintaining an upright position. The test was conducted 3 times, and the best result, in centimeters, was recorded. Motor competence was assessed using the 4 × 10-meter shuttle run test of speed of movement, agility, and coordination. Lower scores indicated better performance (in seconds) and these were recorded for analysis. CRF was assessed using the 20-meter shuttle run (laps and estimated maximum oxygen consumption). An incremental audio signal was used, starting at 6.5 km/h and increasing by 0.5 km/h per minute. The test ended when the child failed to reach one of the lines on 2 consecutive occasions with the audio signal or when the child stopped owing to exhaustion [35]. Finally, a previously validated overall PF score was calculated [36]. The individual score of each PF component was transformed into sex-specific standardized values (z-scores). An overall PF z-score was calculated as the mean of the z-score values for HGS, SLJ, the 4 × 10-meter shuttle run test (for analytic purposes, values were multiplied by −1, so higher scores indicate better motor competence), and the CRF test. Higher z-score values in PF indicate better fitness performance.

The MOBAK KG test battery was used to evaluate basic motor competencies (BMC) in preschool children aged 4 - 6 years [37]. This assessment comprises 8 test items to measure 2 key domains: self-movement (SM: balancing, rolling, jumping, running) and object movement (OM: throwing, catching, bouncing, dribbling). Each domain includes 4 items, with a maximum score of 8 points per domain, resulting in a total possible score of 16 points (highest BMC). Children were not provided with practice attempts before the assessment. For the "throwing" and "catching" tasks, children had 6 attempts, with scoring based on success rates: 0 points for 0 - 2 successful attempts, 1 point for 3 - 4 attempts, and 2 points for 5 - 6 attempts. In the "bouncing," "dribbling," "balancing," "rolling," "jumping," and "running" tasks, participants had 2 attempts each, scored dichotomously: 0 points for no success, 1 point for one success, and 2 points for two successful attempts.

Executive function was evaluated using a digital testing platform on iPads, administered by trained research assistants. The assessment incorporated tasks from "Early Tools" (Years Toolbox YET-2017), including "Mr. Ant" and "Not This" to assess memory, "Card Sorting" to measure cognitive flexibility, and "Go/No-Go" to evaluate inhibitory control [38]. These tasks are designed to provide a comprehensive profile of executive function in children.

Anthropometric measurements were taken to determine the weight (kg), height (cm), waist circumference (WC), and BMI of the children. These assessments were conducted following the protocol established by the Centers for Disease Control and Prevention's National Health and Nutrition Examination Survey (CDC-NHANES), with trained evaluators performing the measurements [39]. Height was measured using a stadiometer with an accuracy of 1 mm (SECA 213 model from Seca Ltd) in the Frankfurt position. Weight was measured using a Tanita DC-430 body composition analyzer with an accuracy of 100 g (Tanita DC430 MA model, Tanita Corporation). BMI was calculated by dividing weight (kg) by the square of height (m). WC was measured twice to the nearest 1 mm with a nonelastic tape applied horizontally midway between the lowest rib margin and the iliac crest, at the end of gentle expiration with the children in a standing position. Triceps thickness (mm) was measured twice on the right side of the body to the nearest 0.2

XSL•FO
RenderX

mm with a skinfold calliper (Holtain, range 0 - 40 mm, Holtain Ltd) halfway between the acromion and the olecranon process at the back of the arm. A validated automatic oscillometric device (Omron 705-IT) was used to assess blood pressure. Last, a questionnaire including self-reported variables was completed by the participants' parent(s) or guardian(s) via Google Forms. This questionnaire will include information about the child as well as details about the parent(s) or guardian(s), such as maternal and paternal education, socioeconomic status, or clinical child data records.

Statistical Analysis

Descriptive sample characteristics were presented as mean (SD) or frequency (percentage). All variables were checked for normality using both graphical (normal probability plots) and statistical (Kolmogorov–Smirnov test) procedures. We conducted per-protocol analyses using a repeated measures ANOVA to test for significant differences across the time points. The interaction effect between group and time was assessed using repeated measures analysis of covariance (ANCOVA), with the baseline as the covariate. To describe the differences in the related treatments, the effect size between-group differences were calculated using the partial eta squared (η_p^2), which was interpreted considering the η_p^2 values of 0.01, 0.06, and 0.14, which correspond to small, moderate, and large effect sizes, respectively [40]. Statistical analyses were conducted using IBM SPSS statistical software (version 26.0; IBM Corp). The level of statistical significance was set at $P<.05$.

Protocol Deviations

There were several protocol deviations from the trial registration. First, we planned to exclusively recruit from a list of 120 participants enrolled in a cross-sectional study[31]. We began the recruitment process using this list, but responses were substantially lower than anticipated. Second, the 24-week follow-up could not be conducted due to many parents reporting challenges in coordinating the follow-up appointments because of their work and personal commitments. Other common reasons that parents declined to participate included (1) not wanting to add strain to the curriculum because of a lack of time at the end of the course, (2) difficulty obtaining parental consent forms, and (3) not wanting to add to their workload. Third, some families had difficulties maintaining consistent communication. Although parental involvement was crucial for achieving a high level of compliance with the intervention, parents and children in this study may represent a subgroup with a particularly high motivation to reduce screen media use, as families volunteered to participate, which could influence the generalizability of the results at 24-week follow-up. Finally, the response rate to the questionnaires was unexpectedly low, leading to incomplete data collection and posing a challenge to acquiring reliable information within the intended time frame.

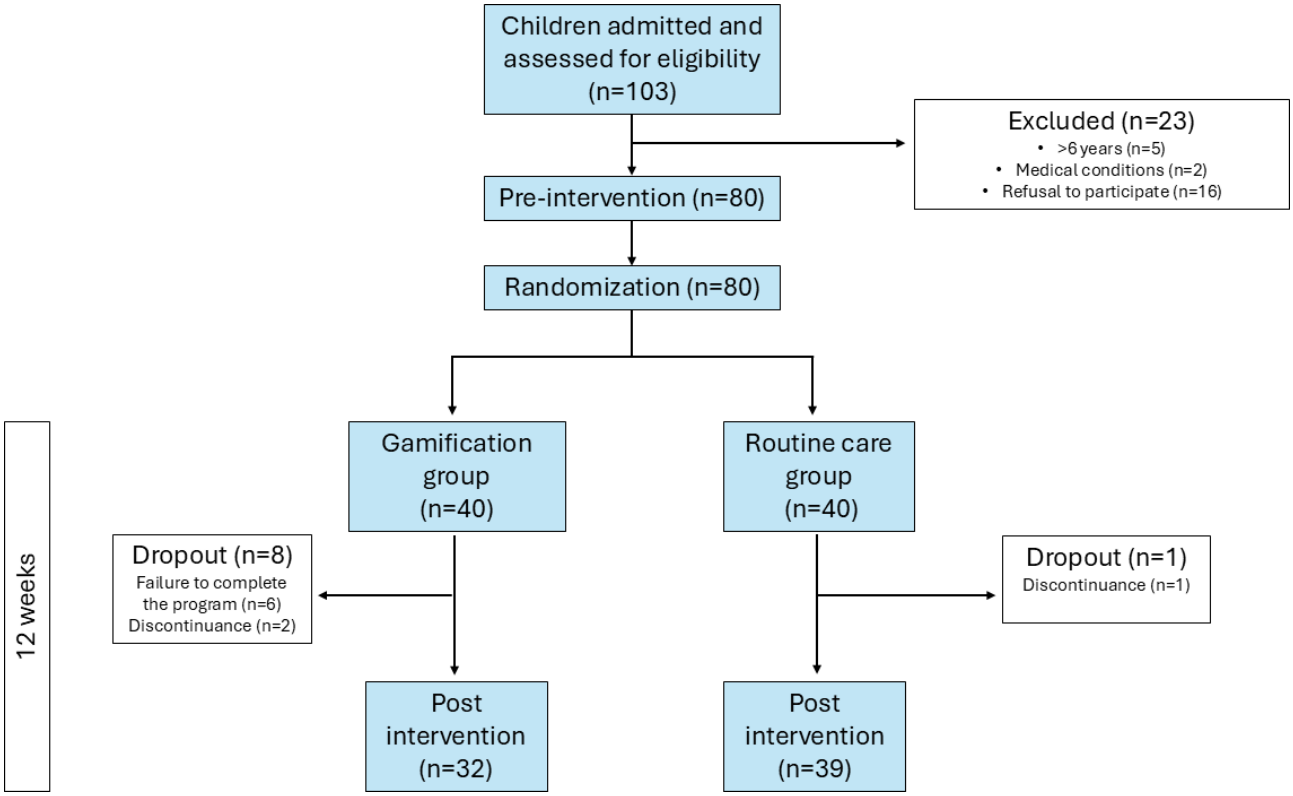
Results

Participants

The flow of participants through the trial is described in Figure 3, showing that 40 children were randomly assigned to the routine care group and 40 were assigned to the gamification group. After 12 weeks of intervention, 32 and 39 children in the gamification and routine care groups, respectively, were assessed. A total of 32 participants completed the gamification sessions. Among them, 28 achieved an adherence rate of over 70% to the exercise sessions, while the remaining 4 demonstrated adherence above 50%. A total of 8 dropouts were recorded in the gamification group because they failed to meet the minimum number of sessions required. One dropout occurred in the routine care group due to a change in the primary care center during the intervention.

XSL•FO
RenderX

Figure 3. CONSORT flow diagram. CONSORT: Consolidated Standards of Reporting Trials.



The baseline characteristics of the participants are presented in Table 1. Among the children, most were boys (n=44, 55%) and aged from 4.1 to 5.7 years, with an average age of 5.0 (SD 0.5) years. Baseline characteristics were generally similar across the 2 groups.

Table . Participant characteristics at baseline. Data are presented as mean (SD) or frequency (percentage).

Characteristics	Gamification group (n=40)	Routine care group (n=40)	Full sample (n=80)
Sex (boys/girls), n	18/22	26/14	44/36
Age (years), mean (SD)	5.1 (0.5)	4.9 (0.5)	5.0 (0.5)
Height (cm), mean (SD)	110.3 (5.3)	109.3 (5.1)	109.8 (5.2)
Weight (kg), mean (SD)	20.0 (3.6)	19.4 (4.0)	19.7 (3.8)
BMI (kg/m^2), mean (SD)	16.4 (1.8)	16.1 (2.4)	16.2 (2.1)
Waist circumference (cm), mean (SD)	53.8 (4.6)	54.0 (2.4)	53.9 (5.6)
Fat mass (%), mean (SD)	21.5 (4.3)	20.8 (4.7)	21.1 (4.5)
Lean mass (%), mean (SD)	73.8 (4.1)	74.5 (4.4)	74.2 (4.2)
Triceps skinfold thickness (mm), mean (SD)	12.9 (3.8)	13.8 (4.1)	13.3 (4.0)
Systolic blood pressure (mm Hg), mean (SD)	97.9 (8.8)	99.5 (6.2)	98.7 (7.6)
Diastolic blood pressure (mm Hg), mean (SD)	60.8 (7.0)	63.8 (7.4)	62.3 (7.3)
Maternal university education, %	66	82	74
Paternal university education, %	40	43	42
Monthly family income ≥€3000 (≥US $3146.43), %	50	70	60

XSL•FO
RenderX

Outcomes

Table 2 lists the effects of the groups on primary and secondary outcomes. In per-protocol analyses, the primary outcome (mean minutes change in MVPA per day, after the 12-week intervention) increased in both groups, by +25.3 (SD 24.6) minutes/day in the gamification group, and +10.0 (SD 31.4) minutes/day in the routine care group, but no significant between-group differences were observed (8.62, 95% CI –5.72 to 22.95 minutes/day; ηp^2=.025; *P*=.23). The analysis of secondary outcomes showed significant between-group mean differences in the change in physical behaviors derived from the accelerometers from baseline to follow-up of 26.44 (95% CI, 8.93 to 43.94) minutes/day in favor of LPA (ηp^2=.138; *P*=.004) and 30.88 (95% CI 4.36 to 57.41) minutes/day in favor of TPA, which corresponds to a large effect size (ηp^2=.087; *P*=.02). Likewise, the gamification group substantially increased their score on the SLJ and overall PF (z-score) from baseline, with a large effect size between groups for the SLJ of +12.26 (95% CI 6.23 to 18.28) cm (ηp^2=.205; *P*<.001) and overall PF by 0.15 (95% CI 0.01 to 0.29) z-score (ηp^2=.076; *P*=.04). No significant differences between groups were observed in the change in other 24-hour movement behavior composition (MPA, VPA, SB, sleep times), PF domains (HGS, 4 × 10-meter test), or BMC outcomes. Similarly, per-protocol analyses showed no effects on executive function scores (Table 2).

XSL•FO
RenderX

Table . Per-protocol analysis of primary and secondary outcomes at baseline and changes after 12 weeks.

Outcome	Groups				Within-group difference		Between-group difference	
	Baseline, mean (SD)		Follow-up, mean (SD)		Mean differences (SD)		Mean differences (95% CI)	
	Gamification (n=40)	Routine care (n=40)	Gamification (n=32)	Routine care (n=39)	Gamification	Routine care	Gamification online minus routine care	ηp^2 (*P* value)
Primary outcome								
MVPA[a] (min/day)	66.8 (25.1)	79.7 (32.4)	89.9 (33.9)	90.4 (30.6)	25.3 (24.6)	10.0 (31.4)	8.62 (–5.72 to 22.95)	.025 (.23)
Secondary outcomes								
LPA[b] (min/day)	242.7 (40.6)	249.1 (36.7)	269 (34.5)	253.6 (41.3)	33.4 (33.3)	0.3 (39.9)	26.44 (8.93 to 43.94)	.138 (.01)
MPA[c] (min/day)	57.8 (20.1)	68.5 (26.4)	76.7 (25.8)	75.7 (23.4)	21.1 (19.3)	6.5 (24.5)	8.46 (–2.57 to 19.50)	.040 (.13)
VPA[d] (min/day)	8.9 (5.8)	11.2 (6.6)	13.2 (8.7)	14.7 (8.8)	4.1 (5.8)	3.5 (7.8)	0.20 (–3.47 to 3.88)	.000 (.91)
TPA[e] (min/day)	309.5 (56.8)	328.8 (62.9)	358.9 (50.6)	344.1 (5.0)	58.7 (49.7)	10.4 (64.4)	30.88 (4.36 to 57.41)	.087 (.02)
SB[f] (min/day)	612.5 (87.0)	599.2 (78.2)	550.9 (66.5)	559.5 (71.4)	–69.8 (82.1)	–40.9 (76.6)	–13.61 (–45.98 to 18.76)	.012 (.40)
Sleep (hours/day)	8.6 (0.9)	8.5 (0.8)	8.8 (0.6)	8.9 (0.6)	0.2 (0.9)	0.5 (0.8)	–0.22 (–0.49 to 0.06)	.043 (.12)
HGS[g] (kg)	7.4 (2.1)	8.0 (2.1)	7.4 (2.0)	8.0 (2.3)	0.3 (2.4)	–0.1 (1.8)	–0.80 (–1.69 to 0.09)	.000 (.90)
SLJ[h] (cm)	87.6 (21.8)	98.2 (15.3)	99.0 (23.0)	93.9 (14.4)	10.8 (10.8)	–4.2 (14.3)	12.26 (6.23 to 18.28)	.205 (<.001)
4 × 10-meter test (s)	16.0 (2.0)	15.5 (1.6)	15.2 (1.8)	15.1 (1.2)	–1.0 (1.4)	–0.4 (1.2)	–0.34 (–0.87 to 0.19)	.027 (.20)
CRF[i] (laps)	24.2 (10.7)	22.4 (10.8)	25.6 (11.8)	24.0 (11.1)	1.4 (5.9)	2.1 (9.4)	0.29 (–3.61 to 4.20)	.000 (.88)
PF[j] (z-score)	–0.0 (0.5)	0.0 (0.4)	0.0 (0.4)	0.0 (0.4)	0.1 (0.3)	–0.1 (0.3)	0.15 (0.01 to 0.29)	.076 (.04)
Mobak OM[k] (points)	2.6 (1.9)	3.2 (2.23)	3.4 (2.4)	3.5 (2.3)	0.7 (1.9)	0.3 (1.7)	0.24 (–0.61 to 1.09)	.005 (.57)
Mobak SM[l] (points)	3.9 (2.4)	3.7 (2.4)	4.5 (2.2)	3.8 (2.3)	0.2 (1.8)	0.1 (1.8)	0.27 (–0.78 to 1.33)	.003 (.67)
Mobak total (points)	6.4 (3.8)	6.9 (3.8)	7.6 (4.2)	7.3 (4.1)	0.9 (2.9)	0.4 (2.2)	0.42 (–0.80 to 1.64)	.007 (.49)
Mr. Ant	2.3 (0.9)	2.0 (0.8)	2.7 (1.4)	2.2 (0.9)	0.5 (1.6)	0.2 (0.8)	0.46 (–0.24 to 1.16)	.036 (.19)
Go/no-go	0.8 (0.2)	0.7 (0.2)	0.8 (0.1)	0.8 (0.2)	–0.0 (0.2)	–0.0 (0.2)	0.02 (–0.07 to 0.10)	.311 (.73)
Not this	2.5 (0.6)	2.5 (0.7)	2.6 (0.7)	2.6 (0.8)	0.1 (0.8)	0.3 (0.8)	–0.13 (–0.55 to 0.28)	.003 (.92)
Card sorting	8.5 (3.0)	7.8 (3.5)	9.1 (2.7)	8.2 (2.8)	1.0 (2.3)	–0.2 (2.4)	1.08 (–0.16 to 2.32)	.066 (.08)

[a]MVPA: moderate to vigorous physical activity.

[b]LPA: light physical activity.

[c]MPA: moderate physical activity.

[d]VPA: vigorous physical activity.

XSL•FO
RenderX

[e]TPA: total physical activity.
[f]SB: sedentary behavior.
[g]HGS: handgrip strength.
[h]SLJ: standing long jump.
[i]CRF: cardiorespiratory fitness.
[j]PF: physical fitness.
[k]OM: object movement.
[l]SM: self-movement.

Discussion

Principal Findings

The major findings of this study showed that a gamified family-based exercise could be an effective intervention to affect preschool children's adherence to 24-hour movement behavior guidelines. The intervention successfully increased PA levels and decreased SB within the target population. In fact, this program optimized both engagement and adherence through gamification, hence proving again that young children enjoy active play and interactive approaches. Point accumulation, immediate rewards, and progress tracking were some of the features integrated to enhance positive behaviors and encourage consistent participation. The program also provided a supportive environment where the inclusion of family members in the process of developing healthier habits should help in reinforcing bonds between family members. This novel approach encouraged participation and also supported the establishment of lifelong PA habits in young children, which is crucial for achieving enduring health benefits.

The PF assessments in the exercise program included measures such as HGS, SLJ, a 4 × 10-meter speed/agility test, and the 20-meter shuttle run test, alongside accelerometer-based PA domains (LPA, MPA, VPA, MVPA, TPA, and SB). Average daily MVPA, assessed via GENEActiv accelerometers, was higher by 8.62 minutes in the exercise group than in the routine care group over the follow-up period. Although the routine care group showed improvement, the gamification group demonstrated notable progress, particularly in MVPA, which achieved the study's primary objective despite the absence of significant between-group differences. Indeed, previous studies have shown that promoting high-intensity PA at an early age may have positive effects on body composition and PF levels in children over the long term, particularly in enhancing muscular strength [41]. Research on preschool PF by Lu et al [42] further emphasized that PA intensity plays a crucial role in fostering PF at this age. The benefits observed may be attributed to the specificity of lower resistance and endurance exercises used in the study (eg, body weight exercises rather than weightlifting), which likely contributed to improvements in muscular fitness as measured by the SLJ tests. Replacing LPA or SB time with MVPA appears to be an effective strategy for enhancing the muscular fitness of preschool-aged children [43].

The secondary outcomes, including anthropometric measures and body composition, did not exhibit significant changes; however, improvements in PF, as evidenced by gains in the SLJ and overall fitness (z-score), were noted. Additionally, enhancements in various PA metrics (LPA, MPA, TPA, and MVPA) highlighted the intervention's effectiveness, with LPA and TPA showing statistically significant improvements. These findings align with longitudinal studies, such as those by Leppänen et al [43], that identified a strong association between higher levels of VPA/MVPA and improved CRF and muscular strength in young children. Similarly, our results mirror those of Fang et al's [44], who reported positive correlations between MVPA and the SLJ test, as well as negative correlations between MVPA and the 4 × 10-meter speed/agility test. These findings underscore the critical role of PA in promoting fitness. Furthermore, Ha et al [45], in their study "Active 1+ FUN," suggested that the "Active 1+ FUN" program was effective in improving the BMC of children. However, they concluded that further research is needed to explore how family-based initiatives could also effectively enhance PA behaviors.

Our results suggest that incorporating high-intensity PA at a young age offers lasting benefits, particularly in terms of CRF and muscular strength. These results are consistent with prior research highlighting the long-term positive effects of early PA on overall health [46,47]. However, sustaining PA over time requires the use of both intrinsic and extrinsic motivation strategies. Although gamification can effectively increase PA engagement, maintaining long-term adherence remains a significant challenge. Intrinsic motivation through enjoyment, a sense of mastery, and positive reinforcement can help children maintain an interest in PA as they age.

Gamification literature shows that engaging and interactive experiences have a stronger influence on PA than regular exercise methods among younger people who enjoy play and are at ease with digital technology [48]. Research on game-like interventions has found that children who take part in these activities show bigger gains in PA and PF than those following nongamified approaches [49]. By creating a playful yet structured environment, gamified programs increase motivation and encourage consistent engagement, which leads to better physical engagement [26]. For instance, Corepal et al [50] found that for some participants, gamification motivated adolescents to join in physically rewarding activities.

Our intervention distributed rewards like virtual accessories and sports-related items. This let children customize their avatars, which made them feel more connected to the program [49]. This reward system kept children coming back by setting up a clear path where they could open new "worlds" and move up levels to reach the "expert level." Our program also involved families, which helped with factors that affect how well children stick with such programs. These include support from parents and children believing in themselves and having fun. All of

XSL•FO
RenderX

these work together to create a good exercise experience that leads to lasting changes in behavior [43]. This shows how mixing game-like elements with family involvement is key. It meets the special motivational needs of young children, which in the end helps them stick with it and develop good health habits.

No previous research has looked at how gamification programs affect executive function outcomes, making it hard to compare results. We believe our study's strong point is measuring the impact on executive function, which shows skills that help children focus, plan, prioritize, work toward goals, and control their behaviors and emotions, and serves as a better indicator of children's mental health. What makes our findings even more convincing is that we looked at how many children in both the control and gamification groups had meaningful changes. We did not see big changes in the MOBAK KG battery [37]. However, the exercise group showed a clear trend in their favor. We think the groups did not differ much because children this age grow, develop, and mature.

One of the primary strengths of our study is the young age of the participants (mean age 5.0, SD 0.5 years), which allows for early intervention and potentially long-lasting impact on PA behaviors. Additionally, using objective PA measurements through accelerometers minimizes the subjective biases often encountered in questionnaire-based assessments. Accelerometers provide detailed insights into PA domains, capturing more precise data than self-reports. Another strength is the inclusion of the PREFIT fitness test [35], which is a validated and reliable measure of PF in this age group. This test enhances the robustness of our findings by ensuring that PF outcomes are accurately assessed in early childhood.

Our findings might be limited by the relatively small sample size (nonrepresentative), which may render some statistical analyses underpowered to detect significant differences. Additionally, some evaluators were not blinded to group allocation, which could introduce bias. The sample was drawn from a single primary care center in Pamplona, which may not be generalizable to the broader Spanish population due to potential regional variations in demographic characteristics and school environments. To improve representativeness, future studies should include multiple centers across diverse regions. Another limitation pertains to the use of accelerometers for measuring PA. Although accelerometers provide objective data, they may fail to capture some physical activities, such as cycling, swimming, or stair climbing, potentially leading to underestimation of overall PA levels. Additionally, the interpretation of PA intensity levels depends on predefined thresholds, which may vary by study and influence the consistency of results across different research contexts.

Parental involvement may have also influenced the outcomes, as young children required assistance from their parents to use the intervention platform. Variability in parental engagement likely impacted children's adherence and motivation, contributing to variability in the results. To address these limitations, future research could incorporate objective tracking methods, such as wearable devices, to monitor PA. In this study, participants were required to complete at least 50% of the training sessions (12 of 24 sessions), with failure to meet this criterion resulting in the discontinuation of the study for that participant. Evidence from prior research supports a 50% adherence rate, which has been shown to improve PF, PA levels, and BMC. It is also important to note that preschoolers rely on parental supervision and may participate in other extracurricular activities. Tools like wearable devices could provide a more accurate representation of daily activities, reducing reliance on parental reporting. Furthermore, a multisite study design and the inclusion of diverse PA modalities could strengthen the generalizability of the findings.

Conclusions

In summary, our study highlights that gamified family-based interventions have the potential to establish foundational PA behaviors in preschool children. By combining gamification with structured family support, these interventions can create a motivationally rich environment that encourages consistent engagement in PA. Incorporating such gamified strategies into school curricula and community programs may cultivate a culture of enjoyment around PA from an early age, helping to counter the natural tendency towards SB observed in later childhood stages. This approach could play a crucial role in the broader public health objectives of mitigating the prevalence of childhood obesity and promoting lifelong health.

Acknowledgments

Gobierno de Navarra provided funding for this study (Departamento de Salud) and cofinanced it with the European Regional Development Fund through the FEDER Operational Program 2014-2020 (Navarra).

Data Availability

Data are available from the corresponding author on reasonable request from January 1, 2026.

Conflicts of Interest

None declared.

Checklist 1
CONSORT-eHEALTH checklist (V 1.6.1).

XSL•FO
RenderX

[PDF File, 3994 KB - games_v13i1e60185_app1.pdf]

References

1. Law M, King G, King S, et al. Patterns of participation in recreational and leisure activities among children with complex physical disabilities. Develop Med Child Neuro 2006 May;48(5):337-342. [doi: 10.1017/S0012162206000740]
2. Gunter KB, Almstedt HC, Janz KF. Physical activity in childhood may be the key to optimizing lifespan skeletal health. Exerc Sport Sci Rev 2012 Jan;40(1):13-21. [doi: 10.1097/JES.0b013e318236e5ee] [Medline: 21918458]
3. Janssen I, Leblanc AG. Systematic review of the health benefits of physical activity and fitness in school-aged children and youth. Int J Behav Nutr Phys Act 2010 May 11;7:40. [doi: 10.1186/1479-5868-7-40] [Medline: 20459784]
4. Kriemler S, Meyer U, Martin E, van Sluijs EMF, Andersen LB, Martin BW. Effect of school-based interventions on physical activity and fitness in children and adolescents: a review of reviews and systematic update. Br J Sports Med 2011 Sep;45(11):923-930. [doi: 10.1136/bjsports-2011-090186] [Medline: 21836176]
5. Raboteg-Saric Z, Sakic M. Relations of parenting styles and friendship quality to self-esteem, life satisfaction and happiness in adolescents. Applied Research Quality Life 2014 Sep;9(3):749-765. [doi: 10.1007/s11482-013-9268-0]
6. Samaha M, Hawi NS. Relationships among smartphone addiction, stress, academic performance, and satisfaction with life. Comput Human Behav 2016 Apr;57:321-325. [doi: 10.1016/j.chb.2015.12.045]
7. WHO Guidelines on Physical Activity and Sedentary Behaviour: World Health Organization; 2020.
8. Moreno-Contreras MI, Pino-Ortega J, Martínez-Santos R. Leisure time physical activity and its relationships with some mental health indicators in Spain through the National Health Survey. Salud mental febrero de 2011;34(1):45-52 [FREE Full text]
9. Brosnahan J, Steffen LM, Lytle L, Patterson J, Boostrom A. The relation between physical activity and mental health among Hispanic and non-Hispanic white adolescents. Arch Pediatr Adolesc Med 2004 Aug;158(8):818-823. [doi: 10.1001/archpedi.158.8.818] [Medline: 15289257]
10. Hobza V, Maracek M, Hamrik Z. Organized sport activities of 11 to 15-year-old adolescents: trends from 2010-2018 and socioeconomic context. Int J Environ Res Public Health 2022 May 17;19(10):6074. [doi: 10.3390/ijerph19106074] [Medline: 35627611]
11. Tremblay MS, LeBlanc AG, Kho ME, et al. Systematic review of sedentary behaviour and health indicators in school-aged children and youth. Int J Behav Nutr Phys Act 2011 Sep 21;8:98. [doi: 10.1186/1479-5868-8-98] [Medline: 21936895]
12. Tucker P. The physical activity levels of preschool-aged children: a systematic review. Early Child Res Q 2008 Oct;23(4):547-558. [doi: 10.1016/j.ecresq.2008.08.005]
13. Dalene KE, Anderssen SA, Andersen LB, et al. Secular and longitudinal physical activity changes in population‐based samples of children and adolescents. Scandinavian Med Sci Sports 2018 Jan;28(1):161-171 [FREE Full text] [doi: 10.1111/sms.12876]
14. Hallal PC, Andersen LB, Bull FC, Guthold R, Haskell W, Ekelund U. Global physical activity levels: surveillance progress, pitfalls, and prospects. The Lancet 2012 Jul;380(9838):247-257. [doi: 10.1016/S0140-6736(12)60646-1]
15. Armstrong N, Tomkinson G, Ekelund U. Aerobic fitness and its relationship to sport, exercise training and habitual physical activity during youth. Br J Sports Med 2011 Sep;45(11):849-858. [doi: 10.1136/bjsports-2011-090200] [Medline: 21836169]
16. Sahoo K, Sahoo B, Choudhury AK, Sofi NY, Kumar R, Bhadoria AS. Childhood obesity: causes and consequences. J Family Med Prim Care 2015;4(2):187-192. [doi: 10.4103/2249-4863.154628] [Medline: 25949965]
17. Knuth AG, Hallal PC. Temporal trends in physical activity: a systematic review. J Phys Act Health 2009 Sep;6(5):548-559. [doi: 10.1123/jpah.6.5.548] [Medline: 19953831]
18. Utesch T, Dreiskämper D, Strauss B, Naul R. The development of the physical fitness construct across childhood. Scandinavian Med Sci Sports 2018 Jan;28(1):212-219 [FREE Full text] [doi: 10.1111/sms.12889]
19. Andersen RE, Crespo CJ, Bartlett SJ, Cheskin LJ, Pratt M. Relationship of physical activity and television watching with body weight and level of fatness among children: results from the Third National Health and Nutrition Examination Survey. JAMA 1998 Mar 25;279(12):938-942. [doi: 10.1001/jama.279.12.938] [Medline: 9544768]
20. Ratner GR, Durán AS, Garrido L MJ, Balmaceda HS, Jadue HL, Atalah S E. Impacto de una intervención en alimentación y actividad física sobre la prevalencia de obesidad en escolares [Article in Spanish]. Nutrición Hospitalaria 2013 Oct;28(5):1508-1514. [doi: 10.3305/nh.2013.28.5.6644]
21. Robinson LE, Stodden DF, Barnett LM, et al. Motor competence and its effect on positive developmental trajectories of health. Sports Med 2015 Sep;45(9):1273-1284. [doi: 10.1007/s40279-015-0351-6] [Medline: 26201678]
22. Sardi L, Idri A, Fernández-Alemán JL. A systematic review of gamification in e-Health. J Biomed Inform 2017 Jul;71:31-48. [doi: 10.1016/j.jbi.2017.05.011] [Medline: 28536062]
23. Ahola R, Pyky R, Jämsä T, et al. Gamified physical activation of young men--a multidisciplinary population-based randomized controlled trial (MOPO study). BMC Public Health 2013 Jan 14;13:32. [doi: 10.1186/1471-2458-13-32] [Medline: 23311678]
24. Hightow-Weidman L, Muessig K, Knudtson K, et al. A gamified smartphone app to support engagement in care and medication adherence for HIV-positive young men who have sex with men (AllyQuest): development and pilot study. JMIR Public Health Surveill 2018 Apr 30;4(2):e34. [doi: 10.2196/publichealth.8923] [Medline: 29712626]

XSL•FO
RenderX

25. Matallaoui A, Koivisto J, Hamari J, Zarnekow R. How effective is “exergamification”? A systematic review on the effectiveness of gamification features in exergames. Presented at: Hawaii International Conference on System Sciences 2017; Jan 4-7, 2017; Hawaii p. 3316-3325. [doi: 10.24251/HICSS.2017.402]
26. Mazeas A, Duclos M, Pereira B, Chalabaev A. Evaluating the effectiveness of gamification on physical activity: systematic review and meta-analysis of randomized controlled trials. J Med Internet Res 2022 Jan 4;24(1):e26779. [doi: 10.2196/26779] [Medline: 34982715]
27. Roshanpour R, Hazegh M. Investigating the impact of virtual reality and gamification on improving physical activities in school. J Dep Anxiety Sci 2022;1(1):01-08.
28. Edwards EA, Lumsden J, Rivas C, et al. Gamification for health promotion: systematic review of behaviour change techniques in smartphone apps. BMJ Open 2016 Oct 4;6(10):e012447. [doi: 10.1136/bmjopen-2016-012447] [Medline: 27707829]
29. Deterding S, Dixon D, Khaled R, Nacke L. From game design elements to gamefulness: defining “gamification”. Presented at: Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments (MindTrek ’11); Sep 28-30, 2011; Tampere, Finland p. 9-15. [doi: 10.1145/2181037.2181040]
30. Sailer M, Hense JU, Mayr SK, Mandl H. How gamification motivates: an experimental study of the effects of specific game design elements on psychological need satisfaction. Comput Human Behav 2017 Apr;69:371-380. [doi: 10.1016/j.chb.2016.12.033]
31. Alonso-Martínez AM, Legarra-Gorgoñon G, García-Alonso Y, et al. Gamified family-based health exercise intervention to improve adherence to 24-h movement behaviors recommendations in children: '3, 2, 1 Move on Study'. Trials 2023 Aug 14;24(1):531. [doi: 10.1186/s13063-023-07494-8] [Medline: 37580788]
32. Migueles JH, Rowlands AV, Huber F, Sabia S, van Hees VT. GGIR: a research community–driven open source R package for generating physical activity and sleep outcomes from multi-day raw accelerometer data. Journal for the Measurement of Physical Behaviour 2019;2(3):188-196. [doi: 10.1123/jmpb.2018-0063]
33. van Hees VT, Sabia S, Anderson KN, et al. A novel, open access method to assess sleep duration using a wrist-worn accelerometer. PLoS One 2015;10(11):e0142533. [doi: 10.1371/journal.pone.0142533] [Medline: 26569414]
34. Ortega FB, Cadenas-Sánchez C, Sánchez-Delgado G, et al. Systematic review and proposal of a field-based physical fitness-test battery in preschool children: the PREFIT battery. Sports Med 2015 Apr;45(4):533-555. [doi: 10.1007/s40279-014-0281-8] [Medline: 25370201]
35. Ahmaidi SB, Varray AL, Savy-Pacaux AM, Prefaut CG. Cardiorespiratory fitness evaluation by the shuttle test in asthmatic subjects during aerobic training. Chest 1993 Apr;103(4):1135-1141. [doi: 10.1378/chest.103.4.1135] [Medline: 8131453]
36. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126. [doi: 10.2196/jmir.1923] [Medline: 22209829]
37. Herrmann C, Ferrari Ehrensberger I, Wälti M, Wacker S, Kühnis J. MOBAK-KG: Basic Motor Competencies in Kindergarten – Test Manual, 3rd edition: Zenodo; 2020. [doi: 10.5281/zenodo.3774438]
38. Howard SJ, Melhuish E. An early years toolbox for assessing early executive function, language, self-regulation, and social development: validity, reliability, and preliminary norms. J Psychoeduc Assess 2017 Jun;35(3):255-275. [doi: 10.1177/0734282916633009] [Medline: 28503022]
39. Anthropometry Procedures Manual: National Health and Nutrition Examination Survey; 2007. URL: https://wwwn.cdc.gov/nchs/data/nhanes/public/2007/manuals/manual_an.pdf [accessed 2025-02-23]
40. Cohen J. Statistical Power Analysis for the Behavioral Sciences: Routledge Academic; 1988. [doi: 10.1016/C2013-0-10517-X]
41. Migueles JH, Delisle Nyström C, Leppänen MH, Henriksson P, Löf M. Revisiting the cross-sectional and prospective association of physical activity with body composition and physical fitness in preschoolers: a compositional data approach. Pediatr Obes 2022 Aug;17(8):e12909. [doi: 10.1111/ijpo.12909] [Medline: 35212168]
42. Lu Z, Guo J, Liu C, et al. Reallocation of time to moderate-to-vigorous physical activity and estimated changes in physical fitness among preschoolers: a compositional data analysis. BMC Public Health 2024 Oct 14;24(1):2823. [doi: 10.1186/s12889-024-20290-6] [Medline: 39402478]
43. Leppänen MH, Henriksson P, Delisle Nyström C, et al. Longitudinal physical activity, body composition, and physical fitness in preschoolers. Medicine & Science in Sports & Exercise 2017;49(10):2078-2085. [doi: 10.1249/MSS.0000000000001313]
44. Fang H, Quan M, Zhou T, et al. Relationship between physical activity and physical fitness in preschool children: a cross-sectional study. Biomed Res Int 2017;2017:9314026. [doi: 10.1155/2017/9314026] [Medline: 29359160]
45. Ha AS, Ng JYY, Lonsdale C, Lubans DR, Ng FF. Promoting physical activity in children through family-based intervention: protocol of the 'Active 1 + FUN' randomized controlled trial. BMC Public Health 2019 Feb 20;19(1):218. [doi: 10.1186/s12889-019-6537-3] [Medline: 30786902]
46. Leppänen MH, Nyström CD, Henriksson P, et al. Physical activity intensity, sedentary behavior, body composition and physical fitness in 4-year-old children: results from the ministop trial. Int J Obes 2016 Jul;40(7):1126-1133. [doi: 10.1038/ijo.2016.54]

XSL•FO
RenderX

47. Hodges EA, Smith C, Tidwell S, Berry D. Promoting physical activity in preschoolers to prevent obesity: a review of the literature. J Pediatr Nurs 2013 Jan;28(1):3-19. [doi: 10.1016/j.pedn.2012.01.002] [Medline: 22326400]
48. Pakarinen A, Parisod H, Smed J, Salanterä S. Health game interventions to enhance physical activity self-efficacy of children: a quantitative systematic review. J Adv Nurs 2017 Apr;73(4):794-811. [doi: 10.1111/jan.13160] [Medline: 27688056]
49. Camacho-Sánchez R, Manzano-León A, Rodríguez-Ferrer JM, Serna J, Lavega-Burgués P. Game-based learning and gamification in physical education: a systematic review. Education Sciences 2023;13(2):183. [doi: 10.3390/educsci13020183]
50. Corepal R, Best P, O'Neill R, et al. Exploring the use of a gamified intervention for encouraging physical activity in adolescents: a qualitative longitudinal study in Northern Ireland. BMJ Open 2018 Apr 20;8(4):e019663. [doi: 10.1136/bmjopen-2017-019663] [Medline: 29678971]

Abbreviations

ANCOVA: analysis of covariance
BMC: basic motor competencies
CDC-NHANES: Centers for Disease Control and Prevention's National Health and Nutrition Examination Survey
CONSORT: Consolidated Standards of Reporting Trials
CRF: cardiorespiratory fitness
HGS : handgrip strength
LPA: light physical activity
MPA: moderate physical activity
MVPA: moderate to vigorous physical activity
OM: object movement
PA: physical activity
PF: physical fitness
SB: sedentary behavior
SLJ: standing long jump
SM: self-movement
TPA: total physical activity
VPA: vigorous physical activity
WC: waist circumference
WHO: World Health Organization

Edited by R Baranyi; submitted 03.05.24; peer-reviewed by B Lai, D Lee; revised version received 11.12.24; accepted 16.12.24; published 04.03.25.

Please cite as:
Legarra-Gorgoñon G, García-Alonso Y, Ramírez-Vélez R, Alonso-Martínez L, Izquierdo M, Alonso-Martínez AM
Effect of a Gamified Family-Based Exercise Intervention on Adherence to 24-Hour Movement Behavior Recommendations in Preschool Children: Single-Center Pragmatic Trial
JMIR Serious Games 2025;13:e60185
URL: https://games.jmir.org/2025/1/e60185
doi:10.2196/60185

© Gaizka Legarra-Gorgoñon, Yesenia García-Alonso, Robinson Ramírez-Vélez, Loreto Alonso-Martínez, Mikel Izquierdo, Alicia M Alonso-Martínez. Originally published in JMIR Serious Games (https://games.jmir.org), 4.3.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Comparing In-Person, Standard Telehealth, and Remote Musculoskeletal Examination With a Novel Augmented Reality Exercise Game System: Pilot Cross-Sectional Comparison Study

Richard Wu[1,2], MD; Keerthana Chakka[1,3], MD; Sara Belko[4], MS; Ninad Khargonkar[5], MS; Kevin Desai[6], PhD; Balakrishnan Prabhakaran[5], PhD; Thiru Annaswamy[7], MD, MA

[1]The University of Texas Southwestern Medical Center, Dallas, TX, United States

[2]Johns Hopkins Hospital, Baltimore, MD, United States

[3]Northwestern Medicine, Chicago, IL, United States

[4]Sidney Kimmel Medical College at Thomas Jefferson University, Philadelphia, PA, United States

[5]Department of Computer Science, The University of Texas at Dallas, Richardson, TX, United States

[6]Department of Computer Science, The University of Texas at San Antonio, San Antonio, TX, United States

[7]Penn State Health Milton S. Hershey Medical Center, Penn State University College of Medicine, Hershey, PA, United States

Corresponding Author:
Thiru Annaswamy, MD, MA
Penn State Health Milton S. Hershey Medical Center, Penn State University College of Medicine, Hershey, PA, United States

Abstract

Background: Current telemedicine technologies are not fully optimized for conducting physical examinations. The Virtual Remote Tele-Physical Examination (VIRTEPEX) system, a novel proprietary technology platform using a Microsoft Kinect-based augmented reality game system to track motion and estimate force, has the potential to assist with conducting asynchronous, remote musculoskeletal examinations.

Objective: This pilot study evaluated the feasibility of the VIRTEPEX system as a supplement to telehealth musculoskeletal strength assessments.

Methods: In this cross-sectional pilot study, 12 study participants with upper extremity pain and/or weakness underwent strength evaluations for four upper extremity movements using in-person, telehealth, VIRTEPEX, and composite (telehealth plus VIRTEPEX) assessments. The evaluators were blinded to each other's assessments. The primary outcome was feasibility, as determined by participant recruitment, study completion, and safety. The secondary outcome was preliminary evaluation of inter-rater agreement between in-person, telehealth, and VIRTEPEX strength assessments, including κ statistics.

Results: This pilot study had an 80% recruitment rate, a 100% completion rate, and reported no adverse events. In-person and telehealth evaluations achieved highest overall agreement (85.71%), followed by agreements between in-person and composite (75%), in-person and VIRTEPEX (62.5%), and telehealth and VIRTEPEX (62.5%) evaluations. However, for shoulder flexion, agreement between in-person and VIRTEPEX evaluations (78.57%; κ=0.571, 95% CI 0.183 to 0.960) and in-person and composite evaluations (78.57%; κ=0.571, 95% CI 0.183 to 0.960) was higher than that between in-person and telehealth evaluations (71.43%; κ=0.429, 95% CI −0.025 to 0.882).

Conclusions: This study demonstrates the feasibility of asynchronous VIRTEPEX examinations and supports the potential for VIRTEPEX to supplement and add value to standard telehealth platforms. Further studies with an additional development of VIRTEPEX and larger sample sizes for adequate power are warranted.

(JMIR Serious Games 2025;13:e57443) doi:10.2196/57443

KEYWORDS

physical examination; telemedicine; tele-health; virtual care; virtual health; telerehabilitation; augmented reality; AR; game; simulation; digital world; virtual environment; motion; strength; force; musculoskeletal; remote examination; exercise; physical examinations; telehealth; cross-sectional; VIRTEPEX; patient; exergame

Introduction

Telemedicine has seen increased utilization in recent years, especially during the COVID-19 pandemic [1]. An important advantage of telemedicine is flexibility, which has led to its incorporation into various medical specialties, including radiology, psychiatry, dermatology, neurology, and cardiology [2]. Telemedicine can be useful for providing health care in

XSL•FO
RenderX

difficult-to-reach locations, such as global health settings [3], rural or wilderness areas [4], disaster scenarios [5], and even outer space [6]. Many of these applications use video conferencing technology combined with peripheral examination devices, such as electronic stethoscopes, tele-opthalmoscopy cameras, video otoscopes, tele-dermatoscopes, digital endoscopes, electronic scales, smartphones, and wearable devices [2]. In the field of rehabilitation medicine, telecommunications technology has enabled telerehabilitation facilitating remote patient interactions [7]. Examples of telerehabilitation applications include replacing in-person visits for chronic lower back pain evaluation and management [8], providing remote synchronous treatment interventions for musculoskeletal conditions [9], and implementing asynchronous rehabilitation programs following total knee replacement surgery [10].

However, existing telemedicine and telerehabilitation platforms have limitations. One major limitation is that physical examinations are difficult to conduct through virtual platforms, as most telemedicine systems are limited to video and audio transmission. Consequently, many physical examination components, including auscultation [11], palpation [2], and manual strength assessments [12] cannot be easily performed through video or audio alone. These assessments may require either another provider at the remote site to assist with physical examination [2] or a peripheral device to collect biometric data. For musculoskeletal examinations, health care providers using existing telemedicine systems may attempt to remotely evaluate patients' range of motion and concentric and eccentric strength by asking them to move a joint or extremity across its full range of motion, raise a limb against gravity, or lift objects of known weight [12]. However, these approaches are limited compared to in-person musculoskeletal examinations and may not always provide reliable information for a comprehensive tele-physical assessment (TelePA) by health care providers.

In recent years, emerging technologies have shown promise in improving reliability and accuracy of TelePA. Wearable biometric sensors can facilitate remote evaluation by transmitting spatiotemporal position, speed, acceleration, gait, force, and haptic data in real time [13-16]. Virtual and augmented reality technologies can immerse patients in simulated environments [13] and facilitate body ownership in virtual settings [17], thereby leading to greater patient motivation, engagement, and psychosocial benefits during telerehabilitation [18]. Some TelePA systems utilize motion-sensing 3D sensors, such as the Microsoft Kinect system [19], which is a portable, low-cost motion analysis system equipped with RGB+Depth (RGB-D) camera technology [20]. The Kinect system has been studied for various medical applications that involve examining upper and lower extremity function, including assessment of patient movement and function [21], detecting patient walking [22], performing gait analysis [23], assessing balance [24], and mobility monitoring in patients with Parkinson's disease [25].

Methods

Overview

This pilot study aimed to expand the scope of telemedicine and telerehabilitation by assessing the feasibility of the novel Virtual Remote Tele-Physical Examination (VIRTEPEX) system, which combines an augmented reality computer game environment with the Microsoft Kinect system to track motion and estimate force during gameplay [26]. The VIRTEPEX system can be operated asynchronously to evaluate motion and strength and has the potential to augment TelePA during or between telemedicine visits.

VIRTEPEX Game System

VIRTEPEX is a proprietary technology platform that uses the Microsoft Kinect (version 2) RGB-D camera [20] and machine learning software [27]. The technical specifications of VIRTEPEX have been previously published [26]. VIRTEPEX uses noninvasive motion tracking and inverse dynamics to estimate forces for 4 joint movements: shoulder abduction, shoulder flexion, elbow flexion, and wrist extension. As part of the user experience, patients perform the 4 joint movements while playing an augmented reality bowling computer game (Figure 1), which was rendered using Unity LTS (version 2018.4; Unity Technologies). For each joint movement completed, the patient's strength determines the momentum of a virtual bowling ball for each completed joint movement. VIRTEPEX records and transmits force estimates for each joint movement to a health care provider, who can asynchronously assess the patient's strength. During asynchronous evaluation, the health care provider performs the same 4 joint movements through VIRTEPEX; for each joint movement performed by both the patient and provider, VIRTEPEX synthesizes a comparative animation with two virtual bowling balls colliding into one another (Figure 2). In the animation, the momentum of one ball corresponds to the patient's strength for the joint movement, while the other corresponds to the provider's strength. A virtual midline marks the location where the balls would collide if both users applied equal strength; the provider visually estimates the distance between the bowling balls' collision point and the midline to subjectively estimate the difference in strength between a patient and the provider. In other words, if the balls collide further away from the midline and closer to the patient, the patient's joint movement strength is judged to be weaker compared to the provider.

XSL•FO
RenderX

Figure 1. The simulated bowling computer game interface [26]. PM&R: Physical Medicine and Rehabilitation. ROM: Range of Motion

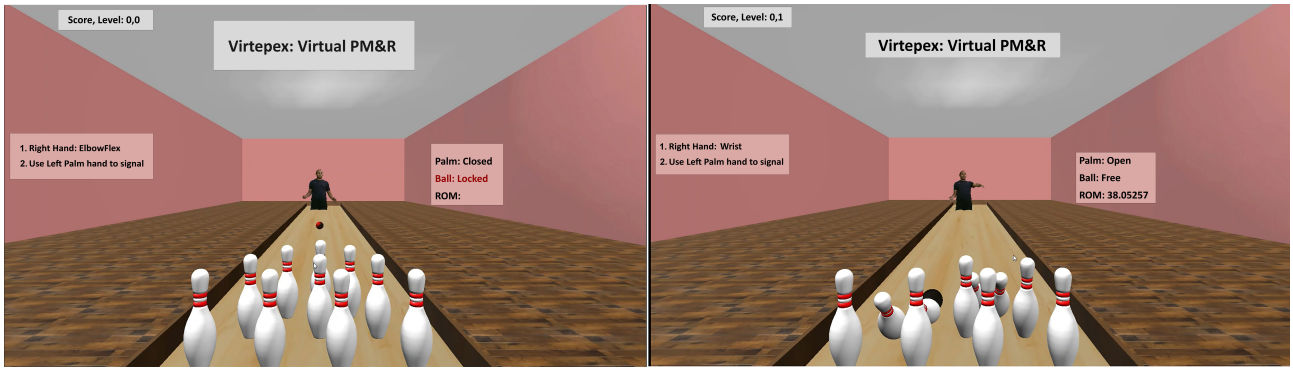
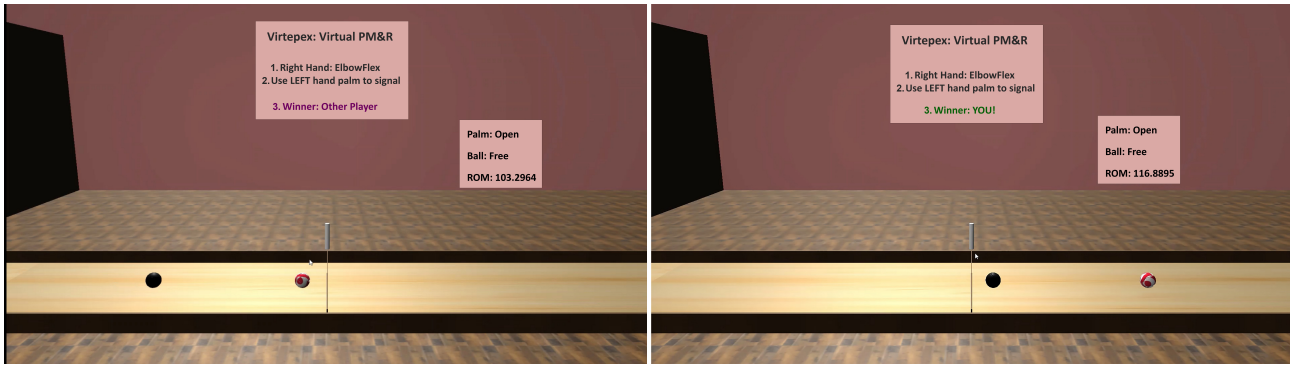


Figure 2. The simulated bowling ball collision animation as viewed by the health care provider for asynchronous remote evaluation of patient's strength [26]. PM&R: Physical Medicine and Rehabilitation. ROM: Range of Motion



Study Sample

This cross-sectional pilot study included patients with impaired upper extremity movements due to any condition causing shoulder, elbow, or wrist pain and/or weakness, with the aim of evaluating movement strength. Adult patients with history of pain and/or weakness in these areas were recruited from a Veterans Affairs hospital's outpatient physical medicine and rehabilitation clinic over one month. Patients scheduled for a therapy (ie, physical, occupational, or kinesiological) for any of the above symptoms were approached for participation; of 15 patients who nitially agreed to participate, 3 patients did not show up for their scheduled research visit, leaving 12 patients who were finally recruited. Of these, 10 participants had unilateral upper extremity pain and/or weakness while the other 2 participants had bilateral issues, resulting in a total of 14 evaluations. This number is similar to that previously reported for comparable clinical studies of telerehabilitation technology [7]. Before formal evaluations, participants and evaluators received training on VIRTEPEX and were allowed to interact with VIRTEPEX for some time to acclimate to the system.

Clinical Evaluation

Each participant's strength for 4 upper extremity movements—shoulder abduction, shoulder flexion, elbow flexion, and wrist extension—was assessed by 3 different evaluators blinded to each other's assessments. The first evaluator (a physiatrist with over 20 years of experience) conducted an in-person physical examination; the second evaluator (a medical student adequately trained in physical assessment and strength examination) conducted a synchronous standard telehealth assessment via real-time audio and video of the patient's physical movements; and the third (another trained medical student) conducted an asynchronous VIRTEPEX assessment using the augmented reality bowling computer game alone, without corresponding audio or video of the participants' movements. For every assessment of different joint movements, each evaluator subjectively rated patient strength on a binary scale (normal or impaired) to simplify the evaluation process and subsequent statistical analyses. In addition, the telehealth and VIRTEPEX evaluations were incorporated together into a composite evaluation that aimed to simulate a typical use-case scenario for VIRTEPEX in order to augment a typical telehealth assessment. The composite evaluation included data for wrist extension from telehealth assessments and data for shoulder abduction, shoulder flexion, and elbow flexion from VIRTEPEX assessments. This approach was formulated by the evaluator using clinical judgement to decide between data from VIRTEPEX and telehealth assessments, a process similar to using clinical judgement to extrapolate information from multiple in-person manual muscle testing data points. During evaluators' training and interactions with VIRTEPEX, it was noted that wrist extension was identified as the most difficult upper extremity movement to capture with the Kinect camera. Therefore, the composite evaluation defaulted to telehealth assessment data for all patients' wrist extension movements.

Study Outcomes

Given this was a pilot study, our primary outcome was to assess for feasibility, based on participant recruitment, study completion, and safety. Recruitment rate was defined as the number of individuals who successfully enrolled and completed the study out of the total patients who were initially approached. Participants completed the study after undergoing each of the

XSL•FO
RenderX

4 upper extremity movement evaluations with each of the 3 evaluators. Safety was measured using the number of adverse events reported during VIRTEPEX evaluation.

The secondary outcome was preliminary assessment of inter-rater agreements, which were calculated for each of the 4 joint movements between (1) in-person and telehealth evaluations, (2) in-person and VIRTEPEX evaluations, (3) in-person and composite evaluations, and (4) telehealth and VIRTEPEX evaluations. Raw percent agreement between two evaluators was calculated as the number of evaluations with agreement divided by the total number of evaluations. κ statistics with 95% CI were calculated using Microsoft Excel for each joint movement to further quantify agreement between in-person and telehealth evaluations, in-person and VIRTEPEX evaluations, in-person and composite evaluations, and telehealth and VIRTEPEX evaluations. κ values were categorized on the following scale used in previous literature [28]: 0.81-1.00 (almost perfect agreement), 0.61-0.80 (excellent agreement), 0.41-0.60 (moderate agreement), 0.21-0.41 (fair agreement), 0.0-0.20 (slight agreement), and <0.00 (no agreement). Overall raw percent agreement across all joint movements was also calculated for each participant.

Ethical Considerations

This study was approved by our institutional research ethics committee (VA Hospital Institutional Review Board, study 1643604). All participants provided written informed consent. Once evaluations were completed, data was deidentified. Participants in this project were not provided any compensation.

Results

User Statistics

Among 15 patients who were initially approached and agreed to participate in this study, 12 completed the evaluation, while 3 did not attend their scheduled visit. As 2 participants had bilateral upper extremity deficits and were separately evaluated for each side, there were a total of 14 completed evaluations. The study recruitment rate was 80% (N=12), with 100% of the 12 participants completing the study, and no adverse events reported.

For each evaluation method (in-person, telehealth, VIRTEPEX, and composite evaluations), 14 evaluations were conducted. This included assessments of 4 individual joint movements per evaluation, yielding 56 joint assessments per evaluation method and a total of 224 separate joint assessments across all methods. Among in-person evaluations, 38 joint assessments were rated as normal and 18 as impaired. Telehealth evaluations rated 40 joint assessments as normal and 16 as impaired. VIRTEPEX evaluations identified 37 joint assessments as normal and 19 as impaired, while composite evaluations identified 44 joint assessments as normal and 12 as impaired.

Evaluation Outcomes

In-Person Versus Telehealth

Overall raw agreements for individual joint movements ranged from 71.43% to 100%, with 85.71% for shoulder abduction, 71.43% for shoulder flexion, 100% for elbow flexion, and 85.71% for wrist extension during comparison between in-person and telehealth evaluations (Table 1). κ values exhibited moderate to almost perfect agreement, ranging from κ=0.429 to κ=1.000. Substantial agreement (κ=0.720, 95% CI 0.375 to 1.000) was observed for shoulder abduction, moderate agreement (κ=0.429, 95% CI −0.025 to 0.882) for shoulder flexion, almost perfect agreement (κ=1.000, 95% CI 1.000 to 1.000) for elbow flexion, and moderate agreement (κ=0.440, 95% CI −0.155 to 1.000) for wrist extension. Based on κ values, the highest agreement was observed for elbow flexion (almost perfect with 100.00% raw agreement; κ=1.000, 95% CI 1.000 to 1.000), and the lowest for shoulder flexion (moderate agreement with 71.43% raw agreement, κ=0.429, 95% CI −0.025 to 0.882). Overall agreement between in-person and telehealth assessments was 85.71%, varying from 50% to 100% per participant.

XSL•FO
RenderX

Table . Raw percent agreements, κ statistics, and CI for in-person versus telehealth evaluations.

Variables	Joints and movements				
	Shoulder abduction	Shoulder flexion	Elbow flexion	Wrist extension	Raw agreement (%)
Research evaluation[a]					
1	A[b]	A	A	A	100
2	D[c]	A	A	A	75
3	A	A	A	A	100
4	A	A	A	A	100
5	D	D	A	A	50
6	A	D	A	D	50
7	A	A	A	A	100
8	A	D	A	A	75
9	A	A	A	A	100
10	A	A	A	D	75
11	A	A	A	A	100
12	A	A	A	A	100
13	A	A	A	A	100
14	A	D	A	A	75
Overall agreement	85.71	71.43	100	85.71	85.71
κ statistic (95% CI)	0.720 (0.375 to 1.000)	0.429 (–0.025 to 0.882)	1.000 (1.000 to 1.000)	0.440 (–0.155 to 1.000)	—[d]

[a]Individual participants' responses.

[b]A: Agree.

[c]D: Disagree.

[d]Not applicable.

In-Person Versus VIRTEPEX

Overall raw agreements for individual joint movements ranged from 35.71% to 78.57%, with 57.14% for shoulder abduction, 78.57% for shoulder flexion, 78.57% for elbow flexion, and 35.71% for wrist extension during comparison between in-person evaluations and VIRTEPEX evaluations (Table 2). κ values exhibited slight to moderate agreement, ranging from κ=0.060 to κ=0.571. Fair agreement was observed (κ=0.222, 95% CI −0.073 to 0.518) for shoulder abduction, moderate agreement (κ=0.571, 95% CI 0.183 to 0.960) for shoulder flexion, fair agreement (κ=0.276, 95% CI −0.332 to 0.883) for elbow flexion, and slight agreement (κ=0.060, 95% CI −0.065 to 0.185) for wrist extension. Based on κ values, the highest agreement was observed for shoulder flexion (moderate agreement with 78.57% raw agreement; κ=0.571, 95% CI 0.183 to 0.960), and lowest for wrist extension (slight agreement with 35.71% raw agreement; κ=0.060, 95% CI −0.065 to 0.185). Overall agreement between in-person evaluations and VIRTEPEX was 62.5% , varying from 50% to 75% per participant.

XSL•FO
RenderX

Table . Raw percent agreements, κ statistics, and confidence intervals (CI) for in-person versus VIRTEPEX evaluations.

Variables	Joints and movements				
	Shoulder abduction	Shoulder flexion	Elbow flexion	Wrist extension	Raw agreement (%)
Research evaluation[a]					
1	D[b]	A[c]	A	A	75
2	D	A	A	D	50
3	A	A	A	D	75
4	A	D	A	A	75
5	D	D	A	A	50
6	D	A	D	A	50
7	A	D	A	D	50
8	D	A	A	A	75
9	A	A	D	D	50
10	A	A	A	D	75
11	A	A	A	D	75
12	A	A	A	D	75
13	A	A	D	D	50
14	D	A	A	D	50
Overall agreement (%)	57.14	78.57	78.57	35.71	62.50
κ statistic (95% CI)	0.222 (−0.073 to 0.518)	0.571 (0.183 to 0.960)	0.276 (−0.332 to 0.883)	0.060 (−0.065 to 0.185)	—[d]

[a]Individual participants' responses.

[b]D: Disagree.

[c]A: Agree.

[d]Not applicable.

In-Person Versus Composite Evaluations

Overall raw agreements for individual joint movements ranged from 57.14% to 85.71%, with 57.14% for shoulder abduction, 78.57% for shoulder flexion, 78.57% for elbow flexion, and 85.71% for wrist extension during comparison between in-person and composite evaluations (Table 3). The κ values exhibited fair to moderate agreement, ranging from κ=0.222 to κ=0.571. Fair agreement (κ=0.222, 95% CI −0.073 to 0.518) was observed for shoulder abduction, moderate agreement (κ=0.571, 95% CI 0.183 to 0.960) for shoulder flexion, fair agreement (κ=0.276, 95% CI −0.332 to 0.883) for elbow flexion, and moderate agreement (κ=0.440, 95% CI −0.155 to 1.000 for wrist extension. Based on κ values, the highest agreement was observed for shoulder flexion (moderate agreement with 78.57% raw agreement; κ=0.571, 95% CI 0.183 to 0.960), and lowest for shoulder abduction (fair agreement with 57.14% raw agreement; κ=0.222, 95% CI −0.073 to 0.518). Overall agreement between in-person and composite evaluations was 75%, varying from 25% to 100% per participant.

XSL•FO
RenderX

Table . Raw percent agreements, κ statistics, and CI) for in-person versus composite evaluations.

Variables	Joints and movements				
	Shoulder abduction	Shoulder flexion	Elbow flexion	Wrist extension	Raw agreement (%)
Research evaluation[a]					
1	D[b]	A[c]	A	A	75
2	D	A	A	A	75
3	A	A	A	A	100
4	A	D	A	A	75
5	D	D	A	A	50
6	D	A	D	D	25
7	A	D	A	A	75
8	D	A	A	A	75
9	A	A	D	A	75
10	A	A	A	D	75
11	A	A	A	A	100
12	A	A	A	A	100
13	A	A	D	A	75
14	D	A	A	A	75
Overall agreement (%)	57.14	78.57	78.57	85.71	75
κ statistic (95% CI)	0.222 (−0.073 to 0.518)	0.571 (0.183 to 0.960)	0.276 (−0.332 to 0.883)	0.440 (−0.155 to 1.000)	_[d]

[a]Individual participants' responses.

[b]D: Disagree.

[c]A: Agree.

[d]Not applicable.

Telehealth Versus VIRTEPEX

Overall raw agreements for individual joint movements ranged from 35.71% to 78.57%, with 71.43% for shoulder abduction, 64.29% for shoulder flexion, 78.57% for elbow flexion, and 35.71% for wrist extension during the comparison between telehealth and VIRTEPEX (Table 4). κ values exhibited poor to fair agreement, ranging from κ=−0.033 to κ=0.364. Fair agreement (κ=0.364, 95% CI −0.043 to 0.770) was observed for shoulder abduction; slight agreement (κ=0.186, 95% CI −0.342 to 0.714 for shoulder flexion, fair agreement (κ=0.276, 95% CI −0.332 to 0.883) for elbow flexion, and poor agreement (κ=−0.033, 95% CI −0.356 to 0.291) for wrist extension. Based on κ values, the highest agreement was observed for shoulder abduction (fair agreement with 71.43% raw agreement; κ=0.364, 95% CI −0.043 to 0.770), and lowest for wrist extension (poor agreement with 35.71% raw agreement; κ=−0.033, 95% CI −0.356 to 0.291). Overall agreement was 62.5%, varying from 0% to 100% per participant.

XSL•FO
RenderX

Table . Raw percent agreements, κ statistics, and CI for telehealth versus VIRTEPEX evaluations.

Variables	Joints and movements				
	Shoulder abduction	Shoulder flexion	Elbow flexion	Wrist extension	Raw agreement (%)
Research evaluation[a]					
1	D[b]	A[c]	A	A	75
2	A	A	A	D	75
3	A	A	A	D	75
4	A	D	A	A	75
5	A	A	A	A	100
6	D	D	D	D	0
7	A	D	A	D	50
8	D	D	A	A	50
9	A	A	D	D	50
10	A	A	A	A	100
11	A	A	A	D	75
12	A	A	A	D	75
13	A	A	D	D	50
14	D	D	A	D	25
Overall agreement (%)	71.43	64.29	78.57	35.71	62.5
κ statistic (95% CI)	0.364 (−0.043 to 0.770)	0.186 (−0.342 to 0.714)	0.276 (−0.332 to 0.883)	−0.033 (−0.356 to 0.291)	_[d]

[a]Individual participants' responses.

[b]D: Disagree.

[c]A: Agree.

[d]Not applicable.

Discussion

Principal Results

This pilot study suggests that using VIRTEPEX to supplement telehealth strength assessment is feasible and safe. Further, VIRTEPEX demonstrates sufficiently acceptable levels of inter-rater agreement with in-person examination to warrant further evaluation for clinical use.

The inclusion of the composite assessment in this study illustrates how VIRTEPEX could be integrated into an existing clinical workflow. VIRTEPEX enables asynchronous collection of strength assessment data, which can help inform subsequent synchronous telehealth examinations. These synchronous examinations could be more efficient by focusing on movements identified as deficient via VIRTEPEX. In clinical practice, the providers could use clinical judgment to create a comprehensive assessment based on datapoints from different examination platforms. The composite evaluation in this study aims to represent this process and evaluate the utility of VIRTEPEX, and findings suggest that VIRTEPEX has the potential to augment remote strength assessments.

Among telehealth, VIRTEPEX, and composite evaluations, telehealth evaluation showed the highest overall raw agreement with in-person evaluation (85.71%). This was higher than raw percent agreements between in-person and VIRTEPEX evaluations (62.5%), in-person and composite evaluations (75%), and telehealth and VIRTEPEX evaluations (62.5%). Previous studies on inter-rater reliability for upper extremity musculoskeletal examinations have shown a wide but comparable range of agreement values. For instance, one study demonstrated raw inter-rater agreement values ranging from 66.67% to 98.9% for in-person upper extremity examinations performed by different evaluators [28], while another study comparing telemedicine-based shoulder examinations with in-person evaluations exhibited raw agreement values ranging from 46.7% to 83.7% [29].

Statistically significant agreement was observed between in-person and telehealth evaluations for shoulder abduction and elbow flexion (substantial agreement: κ=0.720, 95% CI 0.375 to 1.000 for shoulder abduction; almost perfect agreement: κ=1.000, 95% CI 1.000 to 1.000 for elbow flexion). Statistically significant, moderate agreement was observed between in-person and VIRTEPEX (κ=0.571, 95% CI 0.183 to 0.960) and in-person and composite evaluations (κ=0.571, 95% CI 0.183 to 0.960) for shoulder flexion. However, the agreement between in-person and telehealth evaluations for shoulder flexion was not statistically significant (κ=0.429, 95% CI −0.025 to 0.882). In contrast, the statistical significance of the agreements for shoulder flexion between in-person and VIRTEPEX evaluations,

XSL•FO
RenderX

as well as between in-person and composite evaluations was particularly notable. These results suggest that while telehealth may have greater overall agreement with in-person evaluation and be better-suited for evaluating shoulder abduction and elbow flexion strength, VIRTEPEX and composite evaluations may be superior for assessing shoulder flexion strength in remote settings.

Interestingly, none of the remote evaluation methods demonstrated statistically significant agreement with in-person evaluations for wrist extension strength assessment. This suggests that further modifications to both existing telehealth technology and VIRTEPEX are needed to improve remote evaluation of patients with wrist weakness and/or pain. Additionally, while telehealth and VIRTEPEX evaluations showed an overall agreement of 62.5%, the lack of statistically significant κ values for the 4 joint movements suggests that the observed agreement may have been due to chance. Alternatively, it may indicate that telehealth and VIRTEPEX are better suited for assessing joint movements differently, resulting in distinct data patterns across these evaluation platforms.

Limitations

This pilot study had several limitations. First, given that using VIRTEPEX requires reliable internet connection, this system would have limited utility for patients without internet access. Second, the sample size was small (14 evaluations) and all participants were veterans, reducing generalizability of findings to nonveteran populations. As this was a pilot study focused on feasibility, the small sample size provided less than 80% power—corresponding to approximately 22-30 evaluations—to distinguish moderate inter-rater agreement (κ=0.6) in a two-sided test [30]. A larger, fully powered study would be necessary to better quantify agreement across different evaluation methods. Additionally, VIRTEPEX is currently still under development and is not yet available for clinical practice, as further work is needed to assess and improve the system's ease of use, cost-effectiveness, clinical outcomes, and user satisfaction. Similarly, the subjective binary strength grading used in this study was applied for simplicity and convenience during the evaluation process and statistical analyses; however, it is not applied as a standard in clinical practice.

Future Directions

In this pilot study, VIRTEPEX was only used to remotely assess upper extremity movement strength. However, future development could enable VIRTEPEX to remotely assess strength for joint movements in other areas of the body, such as hips, knees, ankles, torso, and neck, facilitating a more comprehensive TelePA. Additional development could potentially enable VIRTEPEX to evaluate multi-joint movements, gait, and/or balance. Additionally, given that the VIRTEPEX system did not evaluate pain experienced by patients during movements, another potential feature to implement would be pain assessment.

Other potential steps could help improve appeal and adoption of VIRTEPEX. Although VIRTEPEX is currently optimized for use with Microsoft Kinect, adapting it to other devices could facilitate wider adoption. Further, streamlining the user interface and incorporating more game design elements could simplify the user experience and improve user engagement during telerehabilitation. Finally, VIRTEPEX could be expanded beyond traditional telehealth settings to other clinical applications, such as remote physiological monitoring for patients needing recurrent physical assessments.

Conclusions

This study supports the feasibility of VIRTEPEX as a supplement to telehealth and demonstrates that VIRTEPEX can achieve moderate agreement with in-person evaluations. Notably, VIRTEPEX had greater agreement with in-person evaluations than telehealth for shoulder flexion, suggesting its potential to enhance existing telehealth technologies. Further technological developments to VIRTEPEX, combined with more adequately powered studies, can better evaluate the effectiveness and accuracy of VIRTEPEX-supplemented TelePA.

Acknowledgments

Research from this study was previously presented as poster presentations and published as abstracts for the following conferences: (1) the 2022 Association for Computing Machinery Designing Interactive Systems Conference 26, (2) 2022 American Academy of Physical Medicine & Rehabilitation Annual Assembly 31, (3) 2023 Association of Academic Physiatrists Annual Meeting 32

Disclaimer

VIRTEPEX is a proprietary device/product that has not yet been approved for use by the FDA for any purpose, is not labeled for use, and is still under investigation.

Data Availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: RW, KC, SB, NK, KD, BP, TA
Data curation: RW, KC, TA
Development of VIRTEPEX: NK, KD, BP, TA

XSL•FO
RenderX

Formal analysis: RW
Methodology: RW, KC, SB, NK, KD, BP, TA
Supervision: TA
Writing – original draft: RW, KC, SB
Writing – review & editing: TA

Conflicts of Interest

None declared.

References

1. Tenforde AS, Iaccarino MA, Borgstrom H, et al. Telemedicine during COVID-19 for outpatient sports and musculoskeletal medicine physicians. PM R 2020 Sep;12(9):926-932. [doi: 10.1002/pmrj.12422] [Medline: 32424977]
2. Weinstein RS, Krupinski EA, Doarn CR. Clinical examination component of telemedicine, telehealth, mHealth, and connected health medical practices. Med Clin North Am 2018 May;102(3):533-544. [doi: 10.1016/j.mcna.2018.01.002] [Medline: 29650074]
3. Dietrich D, Dekova R, Davy S, Fahrni G, Geissbühler A. Applications of space technologies to global health: scoping review. J Med Internet Res 2018 Jun 27;20(6):e230. [doi: 10.2196/jmir.9458] [Medline: 29950289]
4. Ting L, Wilkes M. Telemedicine for patient management on expeditions in remote and austere environments: a systematic review. Wilderness Environ Med 2021 Mar;32(1):102-111. [doi: 10.1016/j.wem.2020.09.009] [Medline: 33423896]
5. Merrell RC, Cone SW, Rafiq A. Telemedicine in extreme conditions: disasters, war, remote sites. Stud Health Technol Inform 2008;131:99-116. [Medline: 18305326]
6. Grigoriev AI, Orlov OI. Telemedicine and spaceflight. Aviat Space Environ Med 2002 Jul;73(7):688-693. [Medline: 12137107]
7. Schutte J, Gales S, Filippone A, Saptono A, Parmanto B, McCue M. Evaluation of a telerehabilitation system for community-based rehabilitation. Int J Telerehabil 2012;4(1):15-24. [doi: 10.5195/ijt.2012.6092] [Medline: 25945193]
8. Jansen-Kosterink S, Huis in 't Veld R, Wever D, Hermens H, Vollenbroek-Hutten M. Introducing remote physical rehabilitation for patients with chronic disorders by means of telemedicine. Health Technol (Berl) 2015 Jul 1;5(2):83-90. [doi: 10.1007/s12553-015-0111-5]
9. Cottrell MA, Galea OA, O'Leary SP, Hill AJ, Russell TG. Real-time telerehabilitation for the treatment of musculoskeletal conditions is effective and comparable to standard practice: a systematic review and meta-analysis. Clin Rehabil 2017 May;31(5):625-638. [doi: 10.1177/0269215516645148] [Medline: 27141087]
10. Bini SA, Mahajan J. Clinical outcomes of remote asynchronous telerehabilitation are equivalent to traditional therapy following total knee arthroplasty: a randomized control study. J Telemed Telecare 2017 Feb;23(2):239-247. [doi: 10.1177/1357633X16634518] [Medline: 26940798]
11. Kang SH, Joe B, Yoon Y, Cho GY, Shin I, Suh JW. Cardiac auscultation using smartphones: pilot study. JMIR Mhealth Uhealth 2018 Feb 28;6(2):e49. [doi: 10.2196/mhealth.8946] [Medline: 29490899]
12. Tanaka MJ, Oh LS, Martin SD, Berkson EM. Telemedicine in the era of COVID-19: the virtual orthopaedic examination. J Bone Joint Surg Am 2020 Jun 17;102(12). [doi: 10.2106/JBJS.20.00609] [Medline: 32341311]
13. Borresen A, Wolfe C, Lin CK, et al. Usability of an immersive augmented reality based telerehabilitation system with haptics (ARTESH) for synchronous remote musculoskeletal examination. Int J Telerehabil 2019;11(1):23-32. [doi: 10.5195/ijt.2019.6275] [Medline: 31341544]
14. Rodríguez-Martín D, Pérez-López C, Samà A, et al. A waist-worn inertial measurement unit for long-term monitoring of Parkinson's disease patients. Sensors (Basel) 2017 Apr 11;17(4):827. [doi: 10.3390/s17040827] [Medline: 28398265]
15. Zhang S, Guo S, Gao B, Hirata H, Ishihara H. Design of a novel telerehabilitation system with a force-sensing mechanism. Sensors (Basel) 2015 May 19;15(5):11511-11527. [doi: 10.3390/s150511511] [Medline: 25996511]
16. Piqueras M, Marco E, Coll M, et al. Effectiveness of an interactive virtual telerehabilitation system in patients after total knee arthoplasty: a randomized controlled trial. J Rehabil Med 2013 Apr;45(4):392-396. [doi: 10.2340/16501977-1119] [Medline: 23474735]
17. Matamala-Gomez M, Bottiroli S, Realdon O, et al. Telemedicine and virtual reality at time of COVID-19 pandemic: an overview for future perspectives in neurorehabilitation. Front Neurol 2021;12:646902. [doi: 10.3389/fneur.2021.646902] [Medline: 33841313]
18. Howard MC. A meta-analysis and systematic literature review of virtual reality rehabilitation programs. Comput Human Behav 2017 May 1;70:317-327. [doi: 10.1016/j.chb.2017.01.013]
19. Timmi A, Coates G, Fortin K, et al. Accuracy of a novel marker tracking approach based on the low-cost Microsoft Kinect v2 sensor. Med Eng Phys 2018 Sep;59:63-69. [doi: 10.1016/j.medengphy.2018.04.020] [Medline: 29983277]
20. Gaddam SPR, Chippa MK, Sastry S, Ange A, Berki V, Davis BL. Estimating forces during exercise activity using non-invasive Kinect camera. Presented at: 2015 International Conference on Computational Science and Computational Intelligence (CSCI); Dec 7-9, 2015; Las Vegas, NV, USA. [doi: 10.1109/CSCI.2015.126]

XSL•FO
RenderX

21. Capecci M, Ceravolo MG, Ferracuti F, et al. Accuracy evaluation of the Kinect v2 sensor during dynamic movements in a rehabilitation scenario. Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) 2016 ;2016:5409-5412. [doi: 10.1109/EMBC.2016.7591950] [Medline: 28269481]
22. Liu L, Mehrotra S. Patient walk detection in hospital room using Microsoft Kinect v2. Presented at: 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); Aug 16-20, 2016; Orlando, FL, USA p. 4395-4398. [doi: 10.1109/EMBC.2016.7591701]
23. Latorre J, Colomer C, Alcañiz M, Llorens R. Gait analysis with the Kinect v2: normative study with healthy individuals and comprehensive study of its sensitivity, validity, and reliability in individuals with stroke. J Neuroeng Rehabil 2019 Jul 26;16(1):97. [doi: 10.1186/s12984-019-0568-y] [Medline: 31349868]
24. Ayed I, Jaume-i-Capó A, Martínez-Bueso P, Mir A, Moyà-Alcover G. Balance measurement using Microsoft Kinect v2: towards remote evaluation of patient with the functional reach test. Appl Sci (Basel) 2021;11(13):6073. [doi: 10.3390/app11136073] [Medline: 6073]
25. Amini A, Banitsas K, Young WR. Kinect4FOG: monitoring and improving mobility in people with Parkinson's using a novel system incorporating the Microsoft Kinect v2. Disabil Rehabil Assist Technol 2019 Aug;14(6):566-573. [doi: 10.1080/17483107.2018.1467975] [Medline: 29790385]
26. Khargonkar N, Desai K, Prabhakaran B, Annaswamy T. Virtepex: virtual remote tele-physical examination system. Presented at: Designing Interactive Systems Conference Proceedings; Jun 13-17, 2022; Virtual Event Australia. [doi: 10.1145/3532106.3533486]
27. Shotton J, Fitzgibbon A, Cook M, et al. Real-time human pose recognition in parts from single depth images. 20-25 June, 2011 Presented at: Conference on Computer Vision and Pattern Recognition Proceedings. [doi: 10.1109/cvpr.2011.5995316]
28. Smith CK, Bonauto DK, Silverstein BA, Wilcox D. Inter-rater reliability of physical examinations in a prospective study of upper extremity musculoskeletal disorders. J Occup Environ Med 2010 Oct;52(10):1014-1018. [doi: 10.1097/JOM.0b013e3181f4396b] [Medline: 20881623]
29. Wang G, Fiedler AK, Warth RJ, Bailey L, Shupe PG, Gregory JM. Reliability and accuracy of telemedicine-based shoulder examinations. J Shoulder Elbow Surg 2022 Aug 1;31(8):e369-e375. [doi: 10.1016/j.jse.2022.04.005]
30. Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. Phys Ther 2005 Mar;85(3):257-268. [Medline: 15733050]
31. Wu R, Chakka K, Belko S, 2022 American Academy of Physical Medicine & Rehabilitation Annual Assembly. Virtual remote tele-physical examination (VIRTEPEX) system user validation. PM R 2022 Oct;14(S1). [doi: 10.1002/pmrj.12913]
32. Chakka K, Wu R, Belko S. User validation of the virtual remote tele-physical examination (VIRTEPEX) system. Am J Phys Med Rehabil 2023;102(4S):A1-A234. [doi: 10.1097/PHM.0000000000002184]

Abbreviations

RGB-D: RGB+Depth
TelePA: tele-physical assessment
VA: Veterans Affairs
VIRTEPEX: Virtual Remote Tele-Physical Examination

Edited by M Schmid; submitted 17.02.24; peer-reviewed by L Mosca, T Davidson; revised version received 09.12.24; accepted 22.12.24; published 05.02.25.

Please cite as:
Wu R, Chakka K, Belko S, Khargonkar N, Desai K, Prabhakaran B, Annaswamy T
Comparing In-Person, Standard Telehealth, and Remote Musculoskeletal Examination With a Novel Augmented Reality Exercise Game System: Pilot Cross-Sectional Comparison Study
JMIR Serious Games 2025;13:e57443
URL: https://games.jmir.org/2025/1/e57443
doi:10.2196/57443

© Richard Wu, Keerthana Chakka, Sara Belko, Ninad Khargonkar, Kevin Desai, Balakrishnan Prabhakaran, Thiru Annaswamy. Originally published in JMIR Serious Games (https://games.jmir.org), 5.2.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Virtual Reality High-Intensity Interval Training Exergaming Compared to Traditional High-Intensity Circuit Training Among Medical Students: Pilot Crossover Study

Pietro Merola[1,2,3], PhD; Marcos Barros Cardoso[4]; Gabriel Barreto[5], PhD; Matheus Carvalho Chagas[1,6]; Luana Farias Oliveira Saunders[5], PhD; Bryan Saunders[5,7,8], PhD; Danilo Cortozi Berton[2,3], MD, PhD

[1]Move Sapiens, Hospital de Clínicas de Porto Alegre, Porto Alegre, Brazil

[2]Hospital de Clínicas de Porto Alegre, Porto Alegre, Brazil

[3]Postgraduate Program in Pulmonary Sciences, Federal University of Rio Grande do Sul, Porto Alegre, Brazil

[4]Medical School, Federal University of Rio Grande do Sul, Porto Alegre, Brazil

[5]Applied Physiology and Nutrition Research Group, Faculty of Medicine, University of São Paulo, São Paulo, Brazil

[6]Biomedical Engineering, Federal University of ABC, São Bernardo do Campo, Brazil

[7]Center of Lifestyle Medicine, Faculty of Medicine, University of São Paulo, São Paulo, Brazil

[8]Applied Physiology and Nutrition Research Group, School of Physical Education and Sport and Faculdade de Medicina, Universidade de São Paulo, São Paulo, Brazil

Corresponding Author:
Pietro Merola, PhD
Move Sapiens, Hospital de Clínicas de Porto Alegre, Porto Alegre, Brazil

Abstract

Background: This study evaluated the effectiveness of a virtual reality (VR) high-intensity interval training (HIIT) boxing protocol compared to traditional high-intensity circuit training (HICT) in improving exercise motivation, engagement, and physiological responses among 30 healthy medical students.

Objective: The purpose was to compare the VR HIIT protocol, which involved using an Oculus Quest 2 for a futuristic exoskeleton game experience, with a traditional 12-exercise HICT.

Methods: In total, 30 medical students engaged in both VR HIIT, using an Oculus Quest 2 for a futuristic exoskeleton game experience, and a traditional 12-exercise HICT. Metrics included heart rate (HR) and blood lactate levels before and after exercise alongside ratings of perceived exertion and the Situational Motivation Scale.

Results: VR HIIT showed significantly higher mean HR (mean 161, SD 15 vs mean 144, SD 11 bpm; d=1.5; P<.001), peak HR (mean 182, SD 15 vs mean 176, SD 11 bpm; d=0.8; P=.001), and ratings of perceived exertion (mean 16, SD 2 vs mean 15, SD 2; d=0.4; P=.03). Postexercise lactate levels were higher in HICT (mean 8.8, SD 4.5 vs mean 10.6, SD 3.0 mmol/L; d=0.6; P=.006). Intrinsic motivation and other psychological measures showed no significant differences, except for lower fatigue in HICT (d=0.5; P=.02).

Conclusions: VR HIIT significantly enhances physiological parameters while maintaining intrinsic motivation, making it a viable alternative to traditional HICT. However, the short-term nature of this study is a limitation, and future research should explore the long-term engagement and therapeutic impacts of VR exercise in diverse and clinical populations.

(JMIR Serious Games 2025;13:e63461) doi:10.2196/63461

KEYWORDS

virtual reality; VR; high-intensity interval training; exercise motivation; exergame; physical activity; exercise; heart rate

Introduction

Exercise and physical activity contribute to enhanced health outcomes, leading to a reduction in the risk of chronic diseases such as cardiovascular disease, diabetes, and various types of cancer [1]. Despite the well-documented benefits of regular moderate- and high-intensity exercise in reducing health risk factors, a significant portion of the global population remains physically inactive. According to the World Health Organization [2], approximately 28% of adults worldwide do not engage in sufficient physical activity, as detailed in the "Global Action Plan on Physical Activity 2018‐2030" report. Physical inactivity has been identified as the fourth leading cause of mortality globally, accounting for approximately 6% of deaths, while obesity contributes to about 5% of mortality [1]. This lack of physical activity significantly increases the risk of heart

XSL•FO
RenderX

disease and diabetes across various countries and social groups regardless of income level [1]. Furthermore, it has been demonstrated that physical exercise is fundamentally important for energy balance and body mass control [3] and for mental health and sleep quality [4]. One of the most cited barriers to exercising is the lack of time and motivation [5].

High-intensity interval training (HIIT) offers a solution to these barriers by providing a time-efficient and engaging exercise modality. HIIT consists of alternating short "bursts" of intense exercise with passive rest or active periods of low-intensity exercise [6,7]. Studies show that HIIT is equally beneficial, or perhaps even superior, to traditional continuous aerobic exercise in many variables related to health and fitness, such as cardiovascular endurance, metabolic rate, and muscle strength [7]. Specifically, for exercise to be considered true HIIT, it must include periods of exercise reaching 85% - 95% of the maximum heart rate (HR) during the high-intensity intervals [8].

Additionally, HIIT's design facilitates sustaining high-intensity activity peaks during exercise sessions. The high-intensity nature of HIIT protocols means that a complete exercise session can be performed in a shorter period (7-minute exercise) compared to classic continuous or endurance exercise protocols, making it a practical solution for time-restricted individuals [8]. However, sustaining the high intensity of HIIT can be challenging and uncomfortable, which can potentially decrease motivation for individuals [9].

In recent years, exergames (a combination of video games and exercise) have been proposed as a solution to improve motivation and engagement in physical exercise practice [9,10], and it has already demonstrated that exergames can bring benefits to the physical and mental health of players of different ages [11]. For example, exergames have been shown to improve physical fitness parameters such as cardiovascular endurance and muscle strength as well as enhance mental health outcomes such as motivation, affect, and mood restoration [12,13]. The advancement of virtual reality (VR) technologies, which allow greater sensory immersion, has provided an evolution of these devices from mere entertainment tools to potential serious games with significant health benefits [14,15]. A study suggested that playing VR exergames helps to promote enhancements in mood in young adults [16]. In addition, VR games using stationary bikes have incorporated HIIT protocols, effectively achieving the intensity required for cardiovascular and metabolic benefits. Studies have shown that VR enhances performance during HIIT [14], improves motivation [5], and maintains the necessary exercise intensity for health outcomes [17].

VR boxing may be a suitable exercise activity, given that it is feasible and effective with an HIIT protocol and logistically compatible without the need for specialized equipment. Boxing has been demonstrated to be feasible in VR [18], and high-intensity boxing training has been shown to be effective for improving fitness, making it suitable for HIIT [19]. This is endorsed by a recent study, which suggests that engaging in VR fitness boxing games can lead to vigorous physical activity with high energy expenditure comparable to traditional forms of exercise [20].

To distinguish an HIIT VR exergame from existing models [20], we emphasize a design that prioritizes greater freedom of movement, enabling players to engage in high-intensity exercises without the constraints of rigid gameplay mechanics [21,22]. Unlike traditional VR games that synchronize movement to specific patterns or beats, where players must ever dodge or hit approaching orbs and objects [20], the new approach allows for dynamic and unrestricted physical activity within a clear HIIT time structure [6,23]. During each bout, players are encouraged to deliver as many shots as possible, reinforcing the all-out effort characteristic of HIIT [6], while maintaining a more immersive and intense workout experience [21,22]. In this study, we focused on the acute effects of a single session of the game to analyze exercise intensity using HR and blood lactate concentration, standard metrics in HIIT for assessing exercise intensity and physiological stress [6,23], and the immediate motivation to perform at higher intensities. This approach was chosen over a chronic study to directly assess how the game influences the intensity and motivation during high-intensity exercise sessions.

The objective was to evaluate how a VR exergame, Move Sapiens, influenced acute physiological responses and exercise motivation in comparison to a traditional high-intensity circuit training (HICT) model. The primary outcome of this study was the physiological response to VR HIIT compared to the control, a traditional HICT, measured by HR and blood lactate levels. Secondary outcomes included psychological measures such as intrinsic motivation, identified regulation, external regulation, and amotivation, assessed using the Situational Motivation Scale (SIMS). Exploratory outcomes included ratings of perceived exertion (RPE) and symptoms of simulator sickness.

Methods

Participants

The sample size calculation was based on the exergame study of Martin-Niedecken et al [9], which, although not identical in design, also focused on an HIIT exergame. With an effect size of d=0.73, guided by cardiac responses reported in their study, we aimed for 80% power and a 5% significance level. This necessitated at least 11 participants per condition or group to provide objective, quantifiable data critical for evaluating exercise intensity and effectiveness. G*Power software (Heinrich-Heine-Universität Düsseldorf) was used for this transparent and reproducible calculation. However, due to the convenience of the sample, this study increased the participant pool to 30 healthy individuals aged 18 to 30 years, consisting of both male and female medical students from the Federal University of Rio Grande do Sul School of Medicine (Table 1). Recruitment was conducted through advertising on social networks and within the university community. Participation in the research was entirely voluntary, with students given the option to freely choose whether to take part. Nonparticipation did not result in any detriment to their university activities, ensuring that the right to choose was fully respected without any prejudice or consequence. Maximal HR was calculated using the 220−age formula [24].

XSL•FO
RenderX

Participants were required to be healthy and free of major health issues, including severe psychiatric disorders, cardiovascular diseases such as congenital heart defects or arrhythmias, serious chronic conditions like uncontrolled type 1 diabetes, binocular vision anomalies, or upper and lower limb neuromuscular restrictions. Additionally, individuals with recent muscular injuries, flu-like symptoms, or any infectious conditions that could hinder HIIT performance were also excluded. Participants were classified as gamers or nongamers based on their self-reported video gaming habits. Nongamers were defined as those who reported playing less than 1 hour of video games per week over the past 2 years [25], a criterion that did not include experience with VR gaming, given that everyone reported having no previous experience with VR exergames.

Table . Demographic and physiological characteristics (N=30).

Characteristic	Values
Age (years), mean (SD)	24 (3)
Body mass (kg), mean (SD)	69.0 (11.3)
Height (m), mean (SD)	1.69 (0.09)
BMI (kg/m^2), mean (SD)	24.1 (2.8)
Maximum heart rate (bpm), mean (SD)	196 (3)
Sex, n (%)	
Female	12 (40)
Male	18 (60)
Gamer, n (%)	19 (63)
Nongamer, n (%)	11 (37)
IPAQ SF[a], n (%)	
High	16 (53)
Moderate	6 (20)
Low	8 (27)

[a]IPAQ SF: International Physical Activity Questionnaire—Short Form.

Ethical Considerations

All participants provided written informed consent prior to participation in the study. The study protocol was reviewed and approved by the Federal University of Rio Grande do Sul Institutional Review Board (approval 59636722.1.0000.5327). Data collected during the study were anonymized to ensure participant confidentiality, and all privacy measures adhered to institutional and legal requirements. No financial compensation was provided to participants, as participation was entirely voluntary and without any expectation of remuneration.

Experimental Design

The study was conducted using a crossover design consisting of 3 visits. During the first visit, participants were introduced to the Move Sapiens exergame on the Oculus Quest 2 VR device (Meta) and the 12-exercise HICT. The familiarization session involved participants completing a half session of the exergame, which included 6 blocks of 16 seconds of exercise, followed by 20-second pauses. Additionally, participants performed 10 seconds in each of the HICT exercises. For the randomization of activities, the Research Randomizer tool (Social Psychology Network) [26] was used to generate a random sequence for each participant. This tool is specifically designed for research purposes and provides a reliable method for randomization [26], involved either the Move Sapiens exergame or the control exercise condition (HICT). All sessions occurred within the School of Medicine at Clinics Hospital of Porto Alegre, performed 48 hours apart, and at consistent times of the day to control for circadian variation. On test days, participants were instructed to abstain from other exercises and to avoid alcohol and caffeine for 12 hours before testing. These restrictions were emphasized during the study briefing and reinforced through reminders sent to participants 24 hours before each test session. Compliance was self-reported by participants upon arrival on test days.

Mint chewing gum was provided as a preventive measure to mitigate any potential initial discomfort related to motion sickness during the first visit [27], which served as a familiarization with the VR equipment and protocol. This was intended to help participants acclimate to the VR environment. No formal scale was used during the familiarization session; only anecdotal records about discomfort with VR were kept, and no participants reported any symptoms of motion sickness. Consequently, chewing gum was not provided in subsequent VR sessions, as the initial preventive measure appeared sufficient to alleviate any discomfort during the first exposure.

Procedures

VR Headset and Game Setup

Participants interacted with Move Sapiens using the Oculus Quest 2 VR headset, which was selected because the game was specifically developed for the Oculus platform. The Oculus Quest 2 provides an immersive VR experience through its

XSL•FO
RenderX

head-mounted display and 2 handheld controllers, which allow players to engage fully in the game's mechanics. Before beginning the session, a 5-foot by 5-foot (1.5 m by 1.5 m) play area was calibrated for each participant to ensure safe movement within the virtual environment. The floor level was also adjusted, and the headset straps were customized to fit each participant comfortably, ensuring optimal performance during gameplay.

High-Intensity Interval Protocols

In Move Sapiens, players are immersed by VR in a futuristic laboratory while equipped with advanced exoskeleton armor that enhances their physical abilities set within a narrative of futuristic human augmentation via virtual exoskeletons [28]. This setting serves as the backdrop for the first mechanic of the prototype, where the objective is to punch a drone as many times as possible during timed intervals. The armor features a heads-up display that provides players with essential information, including round time, rest periods, the number of completed rounds, and the total punches thrown. Visual and auditory cues guide the players through each phase, ensuring that they follow the high-intensity interval structure. This "hypercasual" prototype focuses on a simple, effective mechanic—delivering rapid punches—while a ranking system tracks the best sessions to motivate continued performance improvement.

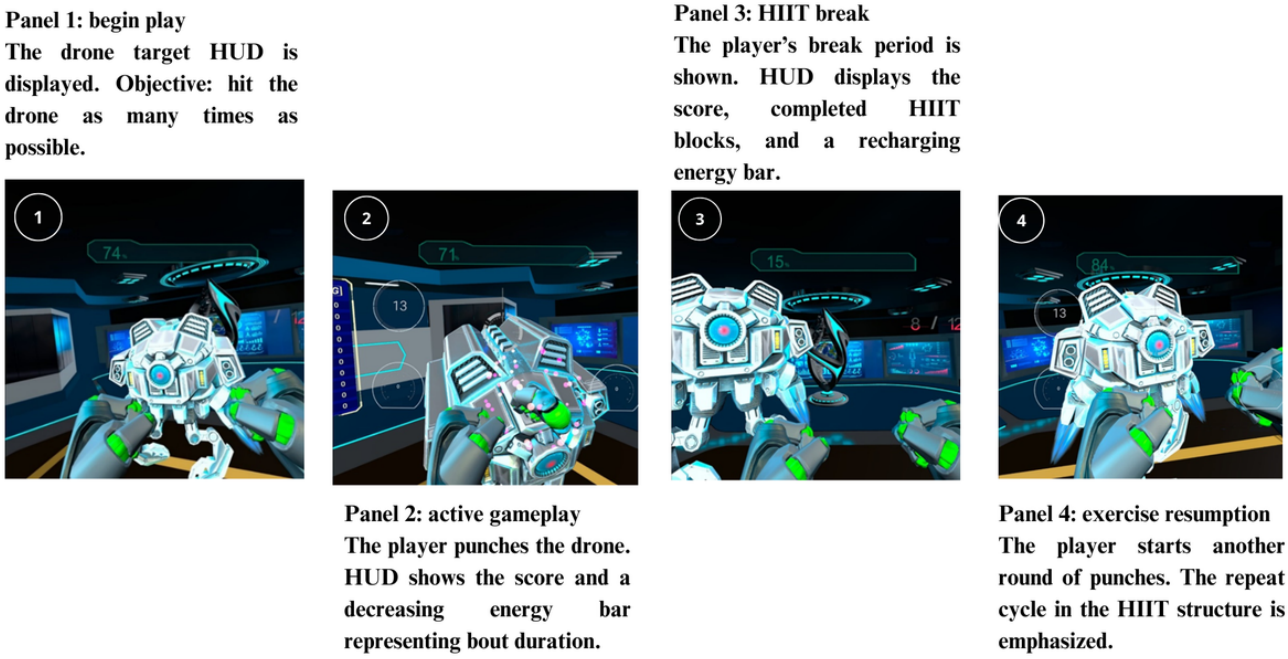
We used a low-volume HIIT protocol to mitigate typical VR usability issues such as dizziness [15], in addition to sweat and discomfort [29]. The following short HIIT shadow boxing all-out protocol was used: 12 sets of 16 seconds with 20 seconds of passive rest, lasting approximately 7 minutes, concluding with 1 minute of cooldown [6,23]. The HIIT protocol, including exact times for bouts and rest periods, was custom-integrated into the game. The bouts were controlled by the game itself, which provided cues to start and stop each bout, ensuring that participants followed the protocol accurately. During bouts, players were required to strike the virtual "dummy drone" as many times as possible with punches and scored points each time they hit the target.

The game mechanics are designed to reward the volume and speed of punches, encouraging players to deliver as many blows as possible at a high intensity. This approach motivates players to sustain a high level of effort, resulting in consistently vigorous exercise. In developing the simulation, we focused on creating a fast-punching experience by incorporating essential factors from existing VR boxing games [18,20] but with a more minimalist design [22,30]. This minimalist approach allows players greater freedom in their movements without being constrained by mechanics that require specific, predefined actions [22,31] (Figure 1). The game was developed using the Unreal Engine 4 platform (Epic Games).

As a comparator condition, we used an HICT protocol consisting of 12 exercises including jumping jacks, wall sits, push-ups, abdominal crunches, step-ups onto a chair, squats, triceps dips on a chair, planks, high knees running in place, lunges, push-up and rotation, and side planks [32]. Each exercise was performed for 30 seconds, with 10 seconds of transition time between sets. The total time for the entire circuit training was approximately 7 minutes.

The exercise and rest durations between the HIIT VR protocol and the HICT protocol differ, aimed at assessing external validity. The HICT protocol, with its 7-minute exercise duration, is widely used in home workouts and popularized by smartphone apps, making it a relevant comparison [32]. The objective was to compare 2 existing and widely available do-it-yourself exercise modalities.

Figure 1. Overview of Move Sapiens HIIT VR exergame mechanics. HIIT: high-intensity interval training; HUD: heads-up display; VR: virtual reality.



XSL•FO
RenderX

Subjective Measures

After each session, participants completed the Simulator Sickness Questionnaire, which uses a 0‐ to 3-point scale to measure simulator sickness symptoms across nausea, oculomotor, and disorientation subscales. This 16-item questionnaire rates symptoms from 0=none to 3=severe, enabling the evaluation of specific issues like stomach awareness, eyestrain, and dizziness. Responses provide subscale and total scores, quantifying the overall impact of simulator sickness [33]. Subjective perception of effort was recorded at the end of the test using a 6‐ to 20-point RPE scale [34]. The SIMS was used to evaluate both intrinsic and extrinsic motivation among participants, featuring 16 distinct items: (1) I think this activity is interesting, (2) I am doing it for my own good, (3) I am supposed to do it, (4) I do not see any good reasons for doing this activity, (5) I think this activity is pleasant, (6) this activity is good for me, (7) it is something I have to do, (8) I do this activity but do not see its value, (9) this activity is fun, (10) it aligns with how I choose to live my life, (11) I feel obligated to do it, (12) I do this activity but do not know what I gain from it, (13) I enjoy this activity, (14) this activity is important for me, (15) I feel forced to do it, and (16) I do not see what this activity brings me [35]. This comprehensive questionnaire is segmented into 4 motivational factors: intrinsic motivation, identified regulation, external regulation, and amotivation. Each factor is represented by specific items that participants respond to, rating their level of agreement or how applicable each statement feels to them on a nuanced 7-point scale. This scale ranges from 1=does not apply at all to 7=completely applies, facilitating a detailed exploration of participants' motivational states across various situations. The 16 items are designed to capture a wide range of motivational attitudes, from personal interest and enjoyment (intrinsic motivation) to compliance with external demands (external regulation) and lack of motivation (amotivation).

Objective Measures

HR was measured continuously during each 7-minute training session using an HR monitor (Polar H10) and was used to calculate average and peak HR. Blood samples were collected from the fingertip and immediately analyzed on a radiometer (ABL 800 flex, Radiometer; Radiometer Medical ApS) to determine blood lactate. Blood collections were performed twice: before the exercise and 5 minutes after the end of the exercise [36].

Data were analyzed using the RStudio software (version 2023.12.1, Build 402; RStudio PBC), and the significance level was defined at $P<.05$. Linear mixed models were used for all the analyses with condition (2 levels: HIIT VR vs HICT) as a fixed factor and participant ID as a random factor. Mean and peak HR as well as postexercise blood lactate were adjusted for baseline values with their addition to the model as a covariate. This was done to account for initial individual differences. This model was used due to its robustness to data missing at random. Data are represented as estimated means (emmeans) and 95% CIs, except stated otherwise. Cohen *d* was used to calculate effect sizes for objective and subjective data, offering a standardized way to evaluate the practical significance of the observed effects. For Cohen *d* calculation, a transformation from Student *t* values to Cohen *d* was performed:

$$d=2*t \div dferror$$

Cohen *d* classifications are interpreted as follows: a value below 0.2 indicates a very small effect, above 0.2 indicates a small effect, above 0.5 indicates a medium effect, and above 0.8 indicates a large effect.

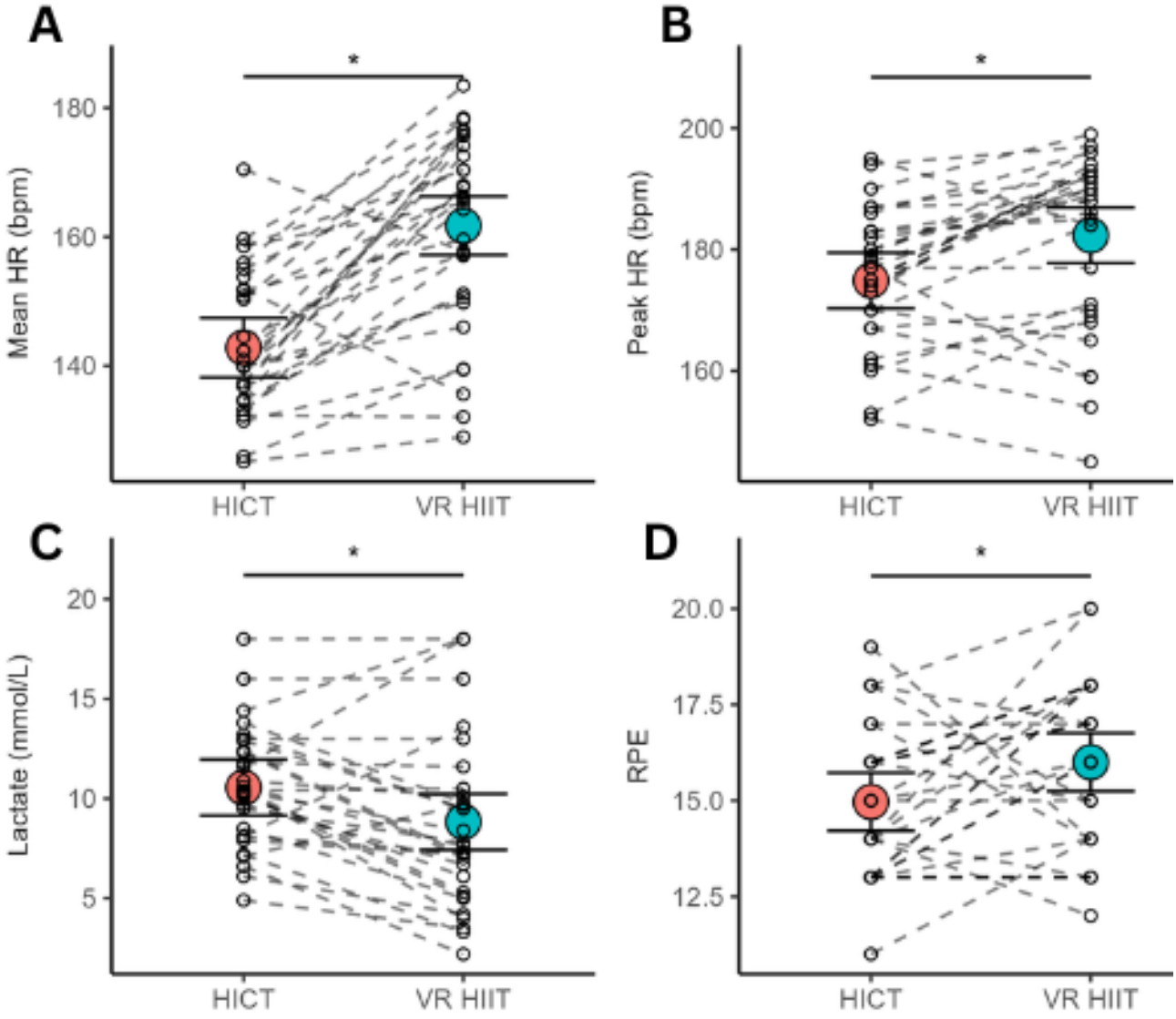
Results

The results revealed significant differences in several metrics between VR HIIT and HICT. For mean HR, VR HIIT exhibited higher values (emmean 162 bpm, 95% CI 157-166) compared to HICT (emmean 143 bpm, 95% CI 138-147; $P<.001$; $d=2.07$; 95% CI 1.40-2.72; Figure 2A). Peak HR was also higher in VR HIIT (emmean 182 bpm, 95% CI 178-187) compared to HICT (emmean 175 bpm, 95% CI 170-180; $P=.001$; $d=1.20$; 95% CI 0.61-1.77; Figure 2B). Postexercise lactate concentration was higher following HICT (emmean 10.6 mmol/L, 95% CI 9.14-12.0) compared to VR HIIT (emmean 8.83 mmol/L, 95% CI 7.43-10.2; $P=.006$; $d=0.80$; 95% CI 0.25-1.34; Figure 2C). RPE were higher for VR HIIT (emmean 16, 95% CI 15-17) compared to HICT (emmean 15, 95% CI 14-16; $P=.03$; $d=0.61$; 95% CI 0.07-1.14).

Regarding reported symptoms, perceived fatigue (not RPE) was higher for VR HIIT (emmean 1.8, 95% CI 1.6-2.0) compared to HICT (emmean 1.5, 95% CI 1.3-1.7; $P=.02$; $d=0.67$; 95% CI 0.13-1.20). No other differences in symptoms were observed between conditions (all $P\geq.14$; Figure 3). There were no significant differences between VR HIIT and HICT for intrinsic motivation ($P=.06$; $d=0.53$), identified regulation ($P=.70$; $d=0.10$), external regulation ($P=.10$; $d=0.32$), or amotivation ($P=.35$; $d=0.26$).

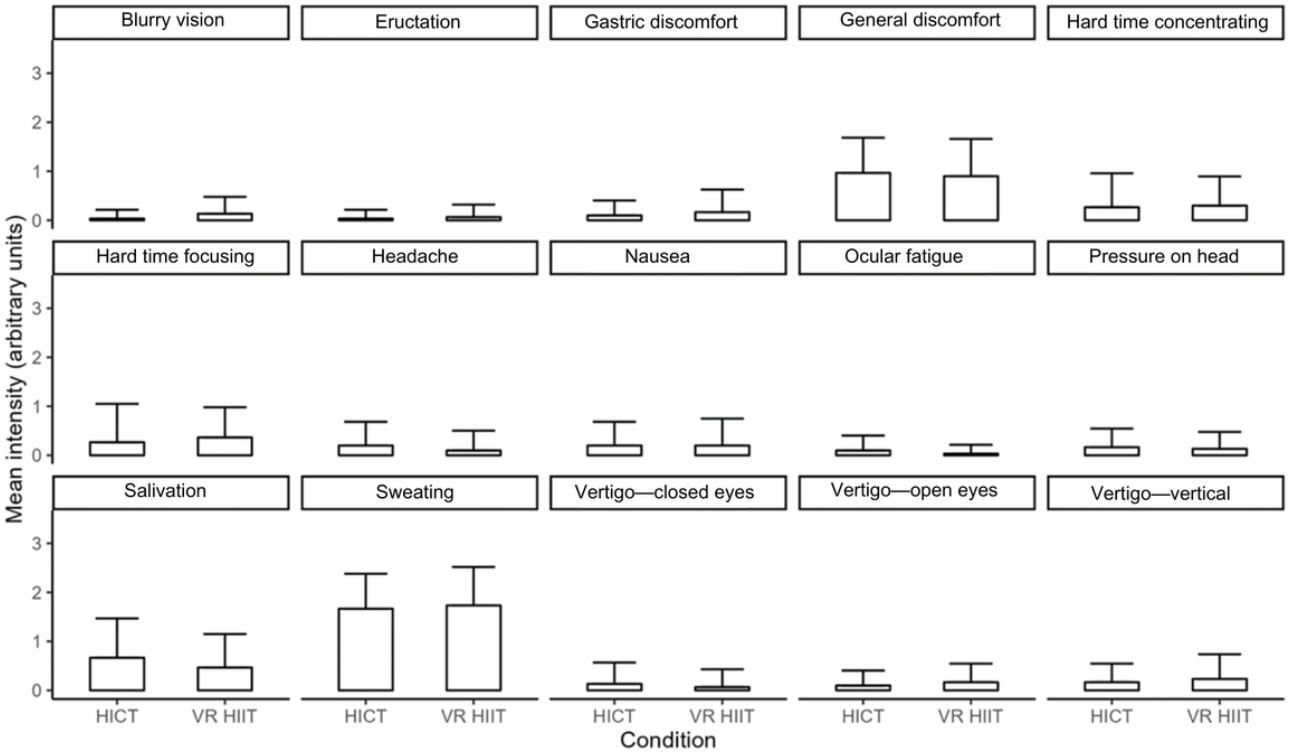
XSL•FO
RenderX

Figure 2. Mean values of (A) mean HR, (B) peak HR, (C) postexercise lactate, and (D) RPE, along with their respective 95% CIs for the standard HIIT protocol (pink circles) and VR HIIT (blue circles). Individual participant data are represented by small dots, while a line connects their values between conditions. HICT: high-intensity circuit training; HIIT: high-intensity interval training; HR: heart rate; RPE: ratings of perceived exertion; VR: virtual reality. *Significant differences between protocols.



XSL•FO
RenderX

Figure 3. Summary of mean values for symptoms as reported by the participants. Values are represented as mean (SD). HICT: high-intensity circuit training; HIIT: high-intensity interval training; VR: virtual reality.



Discussion

Principal Findings

The findings from this study underscore the efficacy of VR HIIT exergaming to increase physiological measures such as HR and increase perceived exertion without reducing intrinsic motivation when compared to traditional HICT. This demonstrates that VR HIIT is not just effective in providing physical exercise benefits akin to those of HICT but also has the potential to maintain acute motivation for exercise. Despite the greater perceived exercise intensity, intrinsic motivation did not decrease, given the sensory stimuli of the VR environment. The immersive nature of VR HIIT, marked by interactive and engaging elements, may contribute to these positive outcomes [21,22]. This aligns with previous studies showing that VR exergaming effectively increases enjoyment during a single bout of HIIT in untrained individuals [14] and improves mood [16].

Lactate levels were substantially increased following both exercise sessions but were significantly higher in the HICT condition compared to the VR HIIT. This can be attributed to the nature of HICT, which typically incorporates a higher volume of strength exercises and calisthenics using body weight [32]. Such activities are known to facilitate a greater accumulation of metabolites due to the anaerobic nature of the exertion, leading to higher lactate production [37]. Despite this, the VR HIIT condition also achieved substantial blood lactate concentrations, indicative of significant metabolic stress [23,38]. Notably, this was achieved alongside higher values in HR during the exercise, suggesting that this VR HIIT boxing protocol effectively stimulates cardiovascular and metabolic responses even in the absence of traditional strength and calisthenic exercises [8]. Our HR results demonstrated that the VR HIIT boxing protocol achieved vigorous intensity levels comparable to the “Supernatural” VR fitness game’s flow and boxing modes. These modes are associated with significant caloric expenditure, ranging from approximately 12.01 to 13.11 kilocalories per minute, with metabolic equivalent of task values of 11.44 for flow and a peak of 12.49 for boxing at higher intensities. [20].

Importantly, the VR game, although a distinct exercise modality from HICT, induced metabolic stress like that of an exercise model validated to be performed independently [32]. This finding underscores the potential of VR HIIT to offer a comparable physiological challenge to HICT, leveraging the immersive and engaging qualities of VR technology to simulate a validated exercise environment effectively.

The exercise intensity within the VR setting is inherently self-selected, despite the game design being crafted to encourage engagement at the highest possible intensity levels [9,21]. Similarly, exercise intensity is also self-selected for HICT, as every type of exercise involves motivation, volition, and intensity self-regulation. The incentive to increase intensity in the game was a better score achieved via the greatest number of punches within the HIIT blocks [21,31]. The stimulus of the VR game may have led to increased effort during the activity, leading to increased HR. Whether this leads to greater health benefits or increased engagement in exercise over time remains to be investigated. Evidently, how to increase motivation response in VR is a considerable question ahead [39]. The foremost challenge lies in the evolution of game design, where the objective is to increasingly leverage game mechanics and sensory stimuli to foster higher motivation among users [21,22]. However, it is important to acknowledge that the novelty of VR may initially boost motivation and engagement, potentially

XSL•FO
RenderX

skewing performance in the short term [40]. This effect underscores the need for creating more compelling and immersive experiences that not only draw participants in [39] but also encourage them to exert themselves more vigorously during the exercise [9,14]. This endeavor requires a nuanced understanding of human motivation and behavior, alongside a mastery of VR technology, to craft experiences that are both engaging and physically demanding [41].

Future Directions

Integrating HIIT with VR in the study demonstrated a notable safety profile, with no adverse effects reported, particularly concerning motion sickness. The intersection of HIIT and VR with continuous innovation [42] represents a promising approach, particularly for improving engagement and motivation in exercise routines [21,43]. By combining the cognitive and physiological benefits of HIIT with the immersive qualities of VR, this approach holds the potential for a holistic method that addresses physical health. While this study did not specifically examine mental health special populations, the immersive and engaging nature of VR HIIT suggests potential applicability for these groups. This is particularly relevant for youth populations, such as patients with attention-deficit/hyperactivity disorder [44,45], who may benefit from such a different and multifaceted approach to exercise [21,46]. While the study confirmed safety among 30 medical students with heterogeneous physical activity levels, the findings are specific to this group. More research is needed to determine the effectiveness of VR HIIT exergames for these kinds of patients.

Limitations

This study acknowledges certain limitations, primarily its reliance on physiological markers, such as HR and blood lactate levels, and subjective motivation assessments [47]. The study uniformly administered the Simulator Sickness Questionnaire after the exercise across both experimental and control conditions. While this approach maintains comparability, it does not capture baseline symptom levels [48]. We chose to compare an HIIT VR protocol to a traditional HICT protocol, which may have led to some of the differences observed.

The study uniformly administered the SIMS after the exercise across both experimental and control conditions. While this approach maintains comparability, it does not capture baseline motivation or symptom levels [48], limiting our ability to assess changes in motivation due to the exercise protocols themselves. Additionally, the choice of highly active participants, who likely had high baseline motivation, may have introduced a ceiling effect, making it difficult to detect significant changes in motivation between the VR HIIT and HICT conditions. This limitation should be considered when interpreting the null findings in motivation, as the participants' pre-existing motivation levels could have constrained the potential for further increases. Importantly, we do not infer any superiority of either exercise type despite some differences in physiological responses, and further research should make comparisons between our VR HIIT protocol and other non-VR HIIT protocols.

The data derived from this study do not necessarily suggest that VR HIIT will be better adhered to over the long term nor that it will generate similar or better results when applied in a prolonged context. This points to a significant area for future investigation, emphasizing the need to assess the long-term adherence to, and effectiveness of, VR HIIT programs [16,39]. Despite these constraints, the findings contribute valuable information on the physiological responses to VR HIIT. Although the study did not show significant changes in intrinsic motivation, it demonstrated that VR HIIT does not reduce intrinsic motivation compared to HICT. Since subjective motivation plays a crucial role in determining whether participants will continue to engage in an activity [39,49], future research should aim to enhance subjective motivation and examine its impact on long-term adherence and engagement.

Conclusions

VR HIIT achieves acute significant increases in key physiological measures, affirming its effectiveness as an exercise modality comparable to traditional HICT in terms of likely long-term physical benefits. The combination of VR and HIIT has proven to be safe, with no adverse effects, and has maintained intrinsic motivation despite greater perceived exercise intensity due to the sensory stimuli provided by the VR environment. Additionally, VR HIIT's capacity to deliver immersive and tailored exercise experiences presents promising applications in therapeutic contexts, particularly for populations with specific needs where conventional exercise methods may fall short. While this study demonstrates the immediate benefits of VR HIIT, future research is essential to evaluate the sustained engagement and long-term health outcomes associated with this modality. Investigating its impact over extended periods will be crucial to fully understand the breadth of VR HIIT's benefits and to optimize its application for various exercise and therapeutic needs, particularly in clinical populations.

Acknowledgments

All authors declared that they had insufficient funding to support open access publication of this manuscript, including from affiliated organizations or institutions, funding agencies, or other organizations. JMIR Publications provided article processing fee (APF) support for the publication of this article. The project "Move Sapiens" received funding from the "Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS)" in partnership with the "Serviço Brasileiro de Apoio às Micro e Pequenas Empresas (SEBRAE)" (grant 22/2551-0001041-5) as part of the "Doutor Empreendedor FAPERGS/SEBRAE 03/2022" initiative. The grant was awarded to PM, acknowledging the project's potential to contribute significantly to the field of science and sports exercise through innovative research and development activities. This study was supported by the "Fundo de Incentivo à Pesquisa e Eventos do Hospital de Clínicas de Porto Alegre." GB (2020/12036-3) and BS (2021/06836-0) acknowledge the receipt of personal research grants from the São Paulo Research Foundation. BS also acknowledges a scholarship granted by the São Paulo

XSL•FO
RenderX

Research Foundation and the Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo—Laboratório de Investigação Médica “Adopt a Researcher” program.

Conflicts of Interest

PM is the chief executive officer and MCC is the chief technology officer of the startup Move Sapiens, and both are creators of the exergame used in this study.

References

1. Murray CJL, Aravkin AY, Zheng P, et al. Global burden of 87 risk factors in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. The Lancet 2020 Oct;396(10258):1223-1249. [doi: 10.1016/S0140-6736(20)30752-2]
2. Global action plan on physical activity 2018–2030: more active people for a healthier world. World Health Organization. 2018. URL: https://www.who.int/publications/i/item/9789241514187 [accessed 2024-12-05]
3. Westerterp KR. Exercise, energy balance and body composition. Eur J Clin Nutr 2018 Sep;72(9):1246-1250. [doi: 10.1038/s41430-018-0180-4] [Medline: 30185845]
4. Elbe AM, Lyhne SN, Madsen EE, Krustrup P. Is regular physical activity a key to mental health? Commentary on “Association between physical exercise and mental health in 1.2 million individuals in the USA between 2011 and 2015: a cross-sectional study”, by Chekroud et al., published in Lancet Psychiatry. J Sport Health Sci 2019 Jan;8(1):6-7. [doi: 10.1016/j.jshs.2018.11.005]
5. Barathi SC, Finnegan DJ, Farrow M, et al. Interactive feedforward for improving performance and maintaining intrinsic motivation in VR exergaming. Presented at: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems; Apr 21-26, 2018; Montreal, QC, Canada p. 1-14. [doi: 10.1145/3173574.3173982]
6. Buchheit M, Laursen PB. High-intensity interval training, solutions to the programming 424 puzzle: part I: cardiopulmonary emphasis. Sports Med 2013 May;43(5):313-338. [doi: 10.1007/s40279-013-0029-x]
7. Gibala MJ. Physiological basis of interval training for performance enhancement. Exp Physiol 2021 Dec;106(12):2324-2327. [doi: 10.1113/EP088190] [Medline: 32362039]
8. Gibala MJ, MacInnis MJ. Physiological basis of brief, intense interval training to enhance maximal oxygen uptake: a mini-review. Am J Physiol Cell Physiol 2022 Nov 1;323(5):C1410-C1416. [doi: 10.1152/ajpcell.00143.2022]
9. Martin-Niedecken AL, Mahrer A, Rogers K, de Bruin ED, Schättin A. “HIIT” the ExerCube: comparing the effectiveness of functional high-intensity interval training in conventional vs. exergame-based training. Front Comput Sci 2020 Oct 23;2. [doi: 10.3389/fcomp.2020.00033]
10. Siegel SR, L Haddock B, Dubois AM, Wilkin LD. Active video/arcade games (exergaming) and energy expenditure in college students. Int J Exerc Sci 2009;2(3):165-174. [Medline: 20407622]
11. Marques LM, Uchida PM, Barbosa SP. The impact of exergames on emotional experience: a systematic review. Front Public Health 2023 Sep 7;11. [doi: 10.3389/fpubh.2023.1209520]
12. Keller N, Whittle RS, McHenry N, et al. Virtual reality “exergames”: a promising countermeasure to improve motivation and restorative effects during long duration spaceflight missions. Front Physiol 2022;13:932425. [doi: 10.3389/fphys.2022.932425] [Medline: 36304582]
13. Feodoroff B, Konstantinidis I, Froböse I. Effects of full body exergaming in virtual reality on cardiovascular and muscular parameters: cross-sectional experiment. JMIR Serious Games 2019 Aug 28;7(3):e12324. [doi: 10.2196/12324] [Medline: 31464194]
14. Farrow M, Lutteroth C, Rouse PC, Bilzon JLJ. Virtual-reality exergaming improves performance during high-intensity interval training. Eur J Sport Sci 2019 Jul;19(6):719-727. [doi: 10.1080/17461391.2018.1542459] [Medline: 30403927]
15. Xu W, Liang HN, Zhang Z, Baghaei N. Studying the effect of display type and viewing perspective on user experience in virtual reality exergames. Games Health J 2020 Dec;9(6):405-414. [doi: 10.1089/g4h.2019.0102] [Medline: 32074463]
16. Ochi G, Kuwamizu R, Fujimoto T, Ikarashi K, Yamashiro K, Sato D. The effects of acute virtual reality exergaming on mood and executive function: exploratory crossover trial. JMIR Serious Games 2022 Sep 28;10(3):e38200. [doi: 10.2196/38200] [Medline: 36169992]
17. Haller JC, Jang YH, Haller J, Shaw L, Wünsche BC. HIIT the road. Presented at: ACSW 2019: Proceedings of the Australasian Computer Science Week Multiconference; Jan 29 to Feb 1, 2019; Sydney, Australia p. 1-9. [doi: 10.1145/3290688.3290752]
18. Xu W, Liang HN, Ma X, Li X. VirusBoxing: a HIIT-based VR boxing game. Presented at: CHI PLAY ’20: Extended Abstracts of the 2020 Annual Symposium on Computer-Human Interaction in PLAY; Nov 2-4, 2020 p. 98-102. [doi: 10.1145/3383668.3419958]
19. Cheema BS, Davies TB, Stewart M, Papalia S, Atlantis E. The feasibility and effectiveness of high-intensity boxing training versus moderate-intensity brisk walking in adults with abdominal obesity: a pilot study. BMC Sports Sci Med Rehabil 2015;7(1):3. [doi: 10.1186/2052-1847-7-3] [Medline: 25973207]

XSL•FO
RenderX

20. Craig TV, Rhodes RE, Sui W. Examining and comparing the energy expenditure of two modes of a virtual reality fitness game (Supernatural): indirect calorimetry study. JMIR Serious Games 2024 Jun 4;12:e53999. [doi: 10.2196/53999] [Medline: 38833285]
21. Kojić T, Spang R, Vergari M, Meier L, Möller S, Voigt-Antons JN. Effects of user factors on user experience in virtual reality: age, gender, and VR experience as influencing factors for VR exergames. Qual User Exp 2023 Dec;8(1):3. [doi: 10.1007/s41233-023-00056-5]
22. Li Y, Muñoz J, Mehrabi S, Middleton L, Cao S, Boger J. Multidisciplinary Iterative Design of Exergames (MIDE): a framework for supporting the design, development, and evaluation of exergames for health. In: Fang X, editor. HCI in Games: Springer, Cham; 2020:128-147. [doi: 10.1007/978-3-030-50164-8_9]
23. Buchheit M, Laursen PB. High-intensity interval training, solutions to the programming puzzle. Part II: anaerobic energy, neuromuscular load and practical applications. Sports Med 2013 Oct;43(10):927-954. [doi: 10.1007/s40279-013-0066-5] [Medline: 23832851]
24. Fox SM, Naughton JP, Haskell WL. Physical activity and the prevention of coronary heart disease. Ann Clin Res 1971 Dec;3(6):404-432. [Medline: 4945367]
25. Boot WR, Kramer AF, Simons DJ, Fabiani M, Gratton G. The effects of video game playing on attention, memory, and executive control. Acta Psychol (Amst) 2008 Nov;129(3):387-398. [doi: 10.1016/j.actpsy.2008.09.005] [Medline: 18929349]
26. Urbaniak GC, Plous S. Research Randomizer (Version 4.0) [computer software]. 2013. URL: https://www.randomizer.org/ [accessed 2025-01-02]
27. Kaufeld M, De Coninck K, Schmidt J, Hecht H. Chewing gum reduces visually induced motion sickness. Exp Brain Res 2022 Feb;240(2):651-663. [doi: 10.1007/s00221-021-06303-5]
28. McGonigal J. Reality Is Broken: Why Games Make Us Better and How They Can Change the World: Penguin; 2011.
29. Shaw LA, Wünsche BC, Lutteroth C, et al. Challenges in virtual reality exergame design. Presented at: Australasian User Interface Conference (AUIC); Jan 27-30, 2015; Sydney, Australia.
30. Gubin TA, Iyer HP, Liew SN, et al. A systems approach to healthcare innovation using the MIT hacking medicine model. Cell Syst 2017 Jul 26;5(1):6-10. [doi: 10.1016/j.cels.2017.02.012] [Medline: 28750199]
31. Barisch-Fritz B, Nigg CR, Barisch M, Woll A. App development in a sports science setting: a systematic review and lessons learned from an exemplary setting to generate recommendations for the app development process. Front Sports Act Living 2022;4:1012239. [doi: 10.3389/fspor.2022.1012239] [Medline: 36685062]
32. Klika B, Jordan C. High-intensity circuit training using body weight. ACSMs Health Fit J 2013;17(3):8-13. [doi: 10.1249/FIT.0b013e31828cb1e8]
33. Kennedy RS, Lane NE, Berbaum KS, Lilienthal MG. Simulator Sickness Questionnaire: an enhanced method for quantifying simulator sickness. Int J Aviat Psychol 1993 Jul;3(3):203-220. [doi: 10.1207/s15327108ijap0303_3]
34. Borg GAV. Psychophysical bases of perceived exertion. Med Sci Sports Exerc 1982 May;14(5):377. [doi: 10.1249/00005768-198205000-00012]
35. Guay F, Vallerand RJ, Blanchard C. On the assessment of situational intrinsic and extrinsic motivation: the Situational Motivation Scale (SIMS). Motiv Emot 2000;24(3):175-213. [doi: 10.1023/A:1005614228250]
36. Warr-di Piero D, Valverde-Esteve T, Redondo-Castán JC, Pablos-Abella C, Sánchez-Alarcos Díaz-Pintado JV. Effects of work-interval duration and sport specificity on blood lactate concentration, heart rate and perceptual responses during high intensity interval training. PLoS One 2018;13(7):e0200690. [doi: 10.1371/journal.pone.0200690] [Medline: 30011320]
37. Bellissimo GF, Ducharme J, Mang Z, et al. The acute physiological and perceptual responses between bodyweight and treadmill running high-intensity interval exercises. Front Physiol 2022;13:824154. [doi: 10.3389/fphys.2022.824154] [Medline: 35360255]
38. Gastin PB. Energy system interaction and relative contribution during maximal exercise. Sports Med 2001;31(10):725-741. [doi: 10.2165/00007256-200131100-00003] [Medline: 11547894]
39. Mouatt B, Smith AE, Mellow ML, Parfitt G, Smith RT, Stanton TR. The use of virtual reality to influence motivation, affect, enjoyment, and engagement during exercise: a scoping review. Front Virtual Real 2020 Dec 23;1. [doi: 10.3389/frvir.2020.564664]
40. Lemmens JS, von Münchhausen CF. Let the beat flow: how game difficulty in virtual reality affects flow. Acta Psychol 2023 Feb;232:103812. [doi: 10.1016/j.actpsy.2022.103812] [Medline: 36566612]
41. Ali SG, Wang X, Li P, et al. A systematic review: virtual-reality-based techniques for human exercises and health improvement. Front Public Health 2023 Mar 23;11. [doi: 10.3389/fpubh.2023.1143947]
42. Mejtoft T, Lindahl O, Öhberg F, et al. Medtech innovation guide: an empiric model to support medical technology innovation. Health Technol 2022 Sep;12(5):911-922. [doi: 10.1007/s12553-022-00689-0]
43. Li Pira G, Aquilini B, Davoli A, Grandi S, Ruini C. The use of virtual reality interventions to promote positive mental health: systematic literature review. JMIR Ment Health 2023 Jul 6;10:e44998. [doi: 10.2196/44998] [Medline: 37410520]
44. Goharinejad S, Goharinejad S, Hajesmaeel-Gohari S, Bahaadinbeigy K. The usefulness of virtual, augmented, and mixed reality technologies in the diagnosis and treatment of attention deficit hyperactivity disorder in children: an overview of relevant studies. BMC Psychiatry 2022 Jan 4;22(1):4. [doi: 10.1186/s12888-021-03632-1] [Medline: 34983446]

XSL•FO
RenderX

45. Ji H, Wu S, Won J, et al. The effects of exergaming on attention in children with attention deficit/hyperactivity disorder: randomized controlled trial. JMIR Serious Games 2023 May 9;11:e40438. [doi: 10.2196/40438] [Medline: 37159253]
46. Shah SHH, Karlsen AST, Solberg M, Hameed IA. A social VR-based collaborative exergame for rehabilitation: codesign, development and user study. Virtual Real 2022 Nov 28:1-18. [doi: 10.1007/s10055-022-00721-8] [Medline: 36465891]
47. Laursen P, Buchheit M. Science and Application of High-Intensity Interval Training: Solutions to the Programming Puzzle: Human Kinetics; 2019. [doi: 10.5040/9781492595830]
48. Brown P, Spronck P, Powell W. The Simulator Sickness Questionnaire, and the erroneous zero baseline assumption. Front Virtual Real 2022 Sep 6;3. [doi: 10.3389/frvir.2022.945800]
49. Morris LS, Grehl MM, Rutter SB, Mehta M, Westwater ML. On what motivates us: a detailed review of intrinsic v. extrinsic motivation. Psychol Med 2022 Jul;52(10):1801-1816. [doi: 10.1017/S0033291722001611] [Medline: 35796023]

Abbreviations

emmean: estimated mean
HICT: high-intensity circuit training
HIIT: high-intensity interval training
HR: heart rate
RPE: ratings of perceived exertion
SIMS: Situational Motivation Scale
VR: virtual reality

Edited by R Baranyi; submitted 04.07.24; peer-reviewed by C Wender, W Dhahbi; revised version received 02.10.24; accepted 09.10.24; published 07.01.25.

Please cite as:
Merola P, Cardoso MB, Barreto G, Chagas MC, Farias Oliveira Saunders L, Saunders B, Cortozi Berton D
Virtual Reality High-Intensity Interval Training Exergaming Compared to Traditional High-Intensity Circuit Training Among Medical Students: Pilot Crossover Study
JMIR Serious Games 2025;13:e63461
URL: https://games.jmir.org/2025/1/e63461
doi:10.2196/63461

© Pietro Merola, Marcos Barros Cardoso, Gabriel Barreto, Matheus Carvalho Chagas, Luana Farias Oliveira Saunders, Bryan Saunders, Danilo Cortozi Berton. Originally published in JMIR Serious Games (https://games.jmir.org), 7.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Tongue Muscle Training App for Middle-Aged and Older Adults Incorporating Flow-Based Gameplay: Design and Feasibility Pilot Study

Kuan-Chu Su[1], MA; Ko-Chiu Wu[1], PhD; Kuei-Ru Chou[2], PhD; Chia-Hsu Huang[1], MDes

[1]Department of Interaction Design, National Taipei University of Technology, Rm.701-4, Design Building, No.1, Sec.3, Chung-hsiao E. Rd, Taipei, Taiwan

[2]School of Nursing, Taipei Medical University, Taipei, Taiwan

Corresponding Author:
Ko-Chiu Wu, PhD
Department of Interaction Design, National Taipei University of Technology, Rm.701-4, Design Building, No.1, Sec.3, Chung-hsiao E. Rd, Taipei, Taiwan

Abstract

Background: Complications due to dysphagia are increasingly prevalent among older adults; however, the tediousness and complexity of conventional tongue rehabilitation treatments affect their willingness to rehabilitate. It is unclear whether integrating gameplay into a tongue training app is a feasible approach to rehabilitation.

Objective: Tongue training has been proven helpful for dysphagia treatment. Following the development of a tongue training app, a feasibility trial aimed to identify physiological and psychological factors that affect user and flow experience and explored whether training specialized muscles could produce a flow experience for optimal immersion. We aimed to provide a useful tool for medical rehabilitation so that older adults could retain tongue muscle flexibility.

Methods: After consulting professional nurses, we developed a mobile gaming app for middle-aged and older adults to train their tongue muscles. This pilot study used an image recognition system to detect the tongue movements of 32 healthy middle-aged and older adults (7 males, 21.9%; 25 females, 78.1%) during 3 game training tasks, each requiring different reaction speeds. Their physiological and psychological signals, as well as the results of the Flow State Scale 2 (FSS2) questionnaire, were used for correlation analysis regarding relevant flow experiences to establish and evaluate the feasibility of our method.

Results: Through exploratory factor analyses, a 2-factor (operation and immersion) structure was confirmed to have an adequate model fit (χ^2_{36}=448.478; P<.001; Kaiser-Meyer-Olkin=0.757) and internal consistency reliability (Cronbach α=0.802). The slow, medium, and fast levels all significantly affected the FSS2 score for operation (P=.001), the National Aeronautics and Space Administration Task Load Index (P<.001), and flow distance (P<.001). K-means clustering revealed that participants could be further categorized into 3 groups. Through the analysis of changes in the participants' physiological and psychological signals for each given task, Pearson correlation indicated that changes were primarily related to flow distance. For the 12 indicators measured in this study, the low, medium, and high operation groups showed significance in 58% (7/12), 50% (6/12), and 25% (3/12) of the indicators, respectively. Similarly, the low, medium, and high immersion groups had changes in 50% (6/12), 33% (4/12), and 17% (2/12) of indicators, respectively.

Conclusions: Our research supports the further development of a gaming app to aid older adults with tongue muscle training and measure flow using physiological and psychological signals to enhance training accuracy and feasibility. Next, we aim to conduct a randomized pilot trial, improve app functions, offer alternative rehabilitation options, and encourage long-term participation. Future goals include enhancing long-term efficacy, diversifying training modes, and adding a multiuser interactive option for an added challenge.

(JMIR Serious Games 2025;13:e53045) doi:10.2196/53045

KEYWORDS

exergame; mobile app; flow; self-care; feasibility; older adults; dysphagia; tongue exercises

XSL•FO
RenderX

Introduction

Background

Compared with desktop-based platforms, mobile phones can provide on-demand leisure and game time [1], in addition to being a handy tool for daily life [2]. Mobile health has become a promising tool to assist older adults with self-care through health and personal care education, medication compliance support, dietary restriction support, setting exercise goals, stress reduction strategies, calling for assistance, and gaining skills for self-management [3]. Mobile health further provides patients with the ability to share data with caregivers and establishes a simple but proactive framework for health management [4].

Dysphagia can lead to rapid aging, disease, and weakened oral expression skills. In older adults, dysphagia is accompanied with a risk of aspiration pneumonia [5-8]. These effects and symptoms affect social activities as well as reduce the dignity and self-esteem of patients [9]. Therefore, the prevention of dysphagia and reduction of potential symptoms are critical.

Disease Prevention Through Tongue Exercise

Studies have used oral diadochokinesis assessments together with the 10-item Eating Assessment Tool (EAT-10) to evaluate swallowing function [10]. To reduce swallowing impairments in older individuals, the use of speech, tongue-resistance exercises, and head-raising exercises have been noted [11-13]. Tongue-strengthening training devices or exercises [14] are intended to improve swallowing [15-17]. In general, there are 3 different tongue-training methods: stimulus-response therapeutic tongue exercises, playing computer games with the tongue using a tongue drive system, and tongue-protrusion tasks. Findings derived from these tongue-training methods suggest a differential effect of tongue-training paradigms on training-induced cortical plasticity and subject-based scores of fun, motivation, and pain in healthy participants [18]. In addition, exergame training can result in neuroplasticity and cognitive improvement for older adults who are institutionalized [19], while increasing cognitive and physical function in healthy individuals. Exergame training incorporating both cognitive engagement and physical activity exerts greater benefits than cognitively engaging video game training alone [20].

Incorporating gameplay into rehabilitation and training can make such regimens more interesting. Most exergames have been used to improve balance, reduce the effects of Parkinson disease [21], alleviate depressive symptoms in adults [22], and also improve overall fitness [23]. To date, however, there have been very few studies regarding tongue muscle training.

Roles of Physiological Signals

Having a higher heart rate variability (HRV) is a biomarker reflecting autonomic function and is associated with a greater emotional well-being [24]. It is known that the effect of transcutaneous auricular vagus nerve stimulation on HRV is not regulated by the duration of stimulation. In fact, changes in HRV occur most substantially at the beginning of stimulation [25]. Physiological signals such as pulse, respiration, blood pressure, heart rate, and body temperature can be collected by physiological sensors [26,27]. For example, Garmin device monitors can collect accurate data on heart rate and the number of steps taken during the day [28]. There are also biosensors, which can recognize audio features [29], facial expressions [30], body gestures [31], and even touches on sensitive screens [32]. Furthermore, biosensors can be useful for the detection of emotions by monitoring autonomic nervous system activity [33,34]. Heart rate is a common measurement for cardiovascular strain during training and can provide a visualization of a participants's training data as well as inform detailed recommendations for training. Heart rate recordings can be performed precisely and in a noninvasive way, which reduces the need for specialized equipment. Exercising according to defined heart rate zones is already well established in professional and recreational endurance training [35]. To a large extent, HRV is modulated by stimulating sympathetic and repressing parasympathetic influences of the autonomic nervous system [36]. It is also known that age and sex have varying effects on heart rate [37].

Several important indicators of variability include heart rate, the root mean square of successive differences (RMSSD), the natural logarithm of the RMSSD (lnRMSSD), the SD of the NN interval (SDNN), low frequency power (LFP), high frequency power, the low frequency to high frequency (LF/HF) ratio, the number of pairs of successive NN intervals that differ by more than 50 milliseconds (NN50), and the proportion of the NN50 divided by the total number of NN intervals (PNN50) [38,39]. However, the complex interactions among physiological and psychological signals mean that they are rarely discussed and studied together in game mechanics.

Flow Definition and Measurement Questionnaire

Serious games, which are games designed for a purpose other than pure entertainment, can increase patient motivation. Meeting the needs of an individual is often a precursor of a flow state, which is a crucial yet often overlooked feature of serious games [40,41]. Flow is an immersive experience characterized by an optimal balance between one's current skills and the level of challenge [42-44]. It enables an individual to become highly absorbed, forget the passing of time, and enter a psychological state of ecstatic happiness and contentment. Flow encompasses the following eight dimensions: (1) a combination of challenge and skill, (2) the merging of actions with awareness, (3) clear task goals, (4) direct and immediate feedback, (5) concentration on the task at hand, (6) a sense of control, (7) a loss of self-consciousness, and (8) a perception of the transformation of time [45]. Subsequent studies have added the dimensions of learning and positive subjective experiences [46]. Skills and the degree of learning increase with experience, whereas attention, challenge, the sense of presence, flow, and exploratory behavior decrease with experience [47]. We found that generating feedback, challenge, and reward mechanisms in games can increase user interest. The qualities of the user experience generated in a game, such as hedonic quality and pragmatic quality, which can be determined using the net promoter score (NPS) and the User Experience Questionnaire (UEQ), can lead to the exploration of potential market opportunities [48,49].

XSL•FO
RenderX

The National Aeronautics and Space Administration Task Load Index (NASA-TLX) questionnaire is used to assess perceived workloads and determine the effectiveness of gaming tasks [50]. For the measurement of flow experience, the effectiveness of the Flow State Scale 2 (FSS2), the Dispositional Flow Scale 2, flow distance (FD), EGameFlow, and the Game Experience Questionnaire have been confirmed [24,51]. Among them, FD can be divided into 3 levels to distinguish the current state of experience: anxiety, flow, and boredom [52].

For flow during exercise, an electroencephalogram is often used to investigate the influence of music tempo (fast, slow, or no music control) on flow. An analysis of data collected using the short FSS2 revealed that music tempo exerts a significant impact on subjective experiences and objective physiological characteristics. Higher subjective flow levels have been observed in participants listening to fast-tempo music while walking briskly, showing that fast tempos are conducive to movement flow. These findings demonstrate the benefits of music during sports training to improve training effectiveness [53]. The evidence above highlights that physiological signals can be measured during exercise, and the results of flow questionnaires can reveal whether the user has entered a state of immersion. However, flow states during tongue muscle training have rarely been studied.

Study Aims

The purpose of this study was to evaluate our mobile health app, which incorporates gameplay for specialized tongue muscle training. The goal of the app is to enable older adults to rehabilitate unassisted and immerse themselves in the process, optimizing their flow. To determine the feasibility of integrating gameplay, we conducted a small, randomized pilot trial by recruiting 32 healthy middle-aged and older users and set up 3 training tasks at different speeds. Since the analysis of patients with dysphagia may be related to more complex variables, this study included healthy individuals for simplicity in the pilot run. This app, by combining tongue exercises with interactive games, could make the process of rehabilitating tongue muscles more interesting and engaging, emphasizing skill training rather than strength training and focusing on retaining users instead of short-term entertainment. The findings derived from our study could serve as a valuable reference for the next stage of development and evaluation. Our aim is to created an app that can help to alleviate and mitigate dysphagia-related conditions and enable older adults to maintain a healthy life.

Methods

Overview

The study was conducted in two phases. In the first phase, the study team designed and integrated the game into the tongue app. In the second phase, the study team evaluated the feasibility and acceptability of the integrated game using a single-arm trial.

Phase 1: Game Design and Development

Registered professional nurses at Taipei Medical University designed the training movements used in the game based on the tongue exercises in the Guidance Manual for Care and Guidance of Eating and Swallowing Difficulties [54], including 6 tongue stretching exercises (Table 1). Once experts had confirmed the accuracy of the tongue exercises, we designed several game icons for the tongue movements, including up, down, left, right, close, and open.

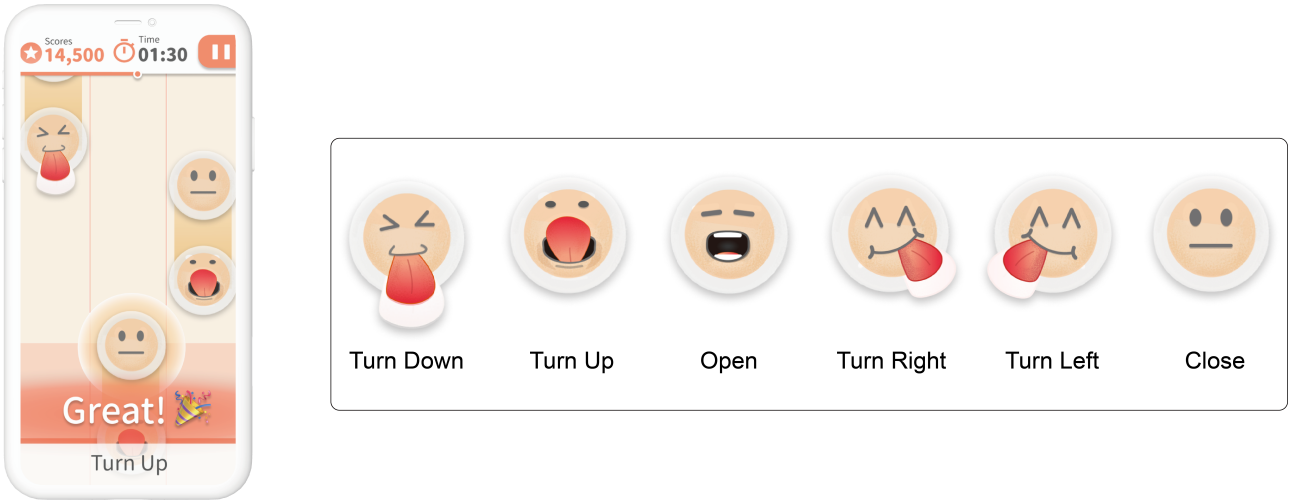
Table . Descriptions of training movements designed by experts for a mobile game app.

No.	Explanation	Movement segments (image recognition definition)
A1	Extend tongue out of the mouth as far as possible and then retract. Repeat 5 times.	Extend tongue, retract, and close mouth.
A2-1, A2-2	Extend tongue to right corner of mouth as far as possible and then extend to left corner. Repeat 10 times.	Extend tongue to the right, retract, extend tongue to the left, retract, and close mouth.
A3-1, A3-2	Open mouth as much as possible, and circle tongue along lips clockwise. Repeat 10 times.	Open mouth, extend tongue, make a circle, and close mouth.
A4	Push tongue hard against upper front teeth and hold for 10 seconds. Repeat 10 times.	Extend tongue, tongue up, and close mouth.

We first photographed 60 participants and removed blurry sections. Teachable Machine version 1.0.1 (Google) was used for initial model training. We imported 200 images for each of the 6 tongue training movements (ie, a total of 1200 training images) to verify that the different movements could be accurately recognized by the camera. The recognition accuracy rates for up, down, left, right, close, and open were 100%, 82%, 96%, 83%, 90%, and 97%, respectively. The trained model was output in the Keras vesion 2.2.0 (Oneiros) and Tensorflow Lite version 1.9 (Google Brain Team) formats. We used the Unity Beats Detection module of Unity Engine version 2020.3.48f1 (Unity Technologies) to analyze the rhythm tempo, and we also used lively and fast-paced music, which began playing at the beginning of the game. In the program design, we generated training levels using the training movements to give the participants a gaming experience, thereby producing an exergame (Figure 1).

XSL•FO
RenderX

Figure 1. Mobile game interface and the schematic diagram of tongue training actions in the game.



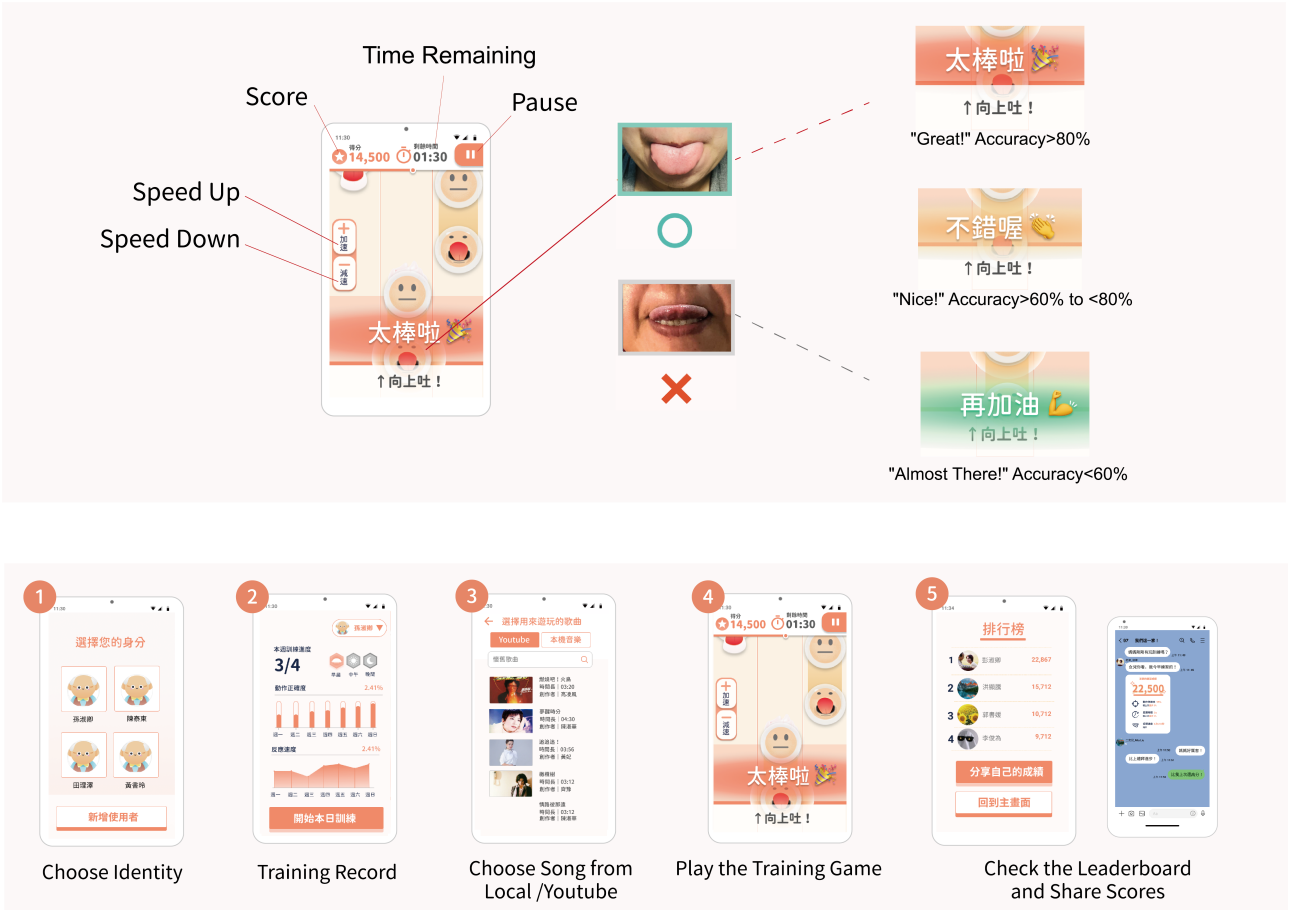
Furthermore, to ensure proper hygiene, image recognition methods that did not require invasion into the oral cavity were used. Images of tongue movements were detected, and suitable feedback and reward mechanisms were also provided. We set up 3 levels with varying speeds (slow, medium, and fast) and skill difficulty requirements to test the mental flow state of the participants under different conditions.

A smartphone was used to record the participant's heart rate (Garmin heart rate chest belt) and mental flow score. To set up the Garmin HRM-PRO activity tracker belt, we created an anonymous user account with the manufacturer that was not associated with any of the patients. Before beginning the game, the participants were separated into the 3 training modes (M1: slow speed; M2: medium speed; or M3: fast speed) by a Latin square method. They then filled out their personal information, read and reviewed their rights and interests, and gave their informed consent. They then put on the Garmin HRM-PRO belt and were asked to rest quietly for 5 minutes before beginning the game tasks. They were given instructions and allowed to practice for 3 minutes, and then played in assigned modes of different difficulties. Subsequently, they rested for 3 minutes and then played again. After each game, they were given a questionnaire (including the FSS2, FD, the NASA-TLX, the UEQ, pragmatic quality, hedonic quality, and NPS), and at the end, an interview was conducted. Using physiological and psychological signals, we explored and analyzed whether the tongue muscle training app could produce a positive flow experience.

After logging in to the daily training page, the homepage was shown. The subject could then select previous records or enter the song menu to choose their favorite music. By tapping the play icon, they could begin training, and after completion, the screen showed the resulting score and their ranking among friends and relatives, which could be shared on social networks (Figure 2). By emulating the correct tongue movements, the system presented encouraging sounds and words such as great, nice, and almost there to the subject. This study explored whether psychological and physiological signals influenced each other, thereby leading to optimal flow states in participants, with the purpose of maintaining the training motivation of middle-aged and older adults. The study integrated big data that could also be used for future smart oral-medical research and analysis.

XSL•FO
RenderX

Figure 2. Development of the mobile app interface, the feedback diagram, and the game step flow chart.



Phase 2: Feasibility Testing

This study included a single-arm, unblinded evaluation of the feasibility and acceptability of the tongue app, in preparation for a planned randomized pilot trial. All participants in the feasibility study were given free access to the tongue app. Participants' use of the app was passively monitored for the pilot trial.

In the early stages, Figma version 95.7 (Figma Inc) was used to create a preliminary model, and the Wizard of Oz method was used to conduct a usability test on 5 middle-aged and older adults aged 55-70 years. After iterative improvements, the final version of the gaming app was developed. We officially started testing on October 5, 2021. All testers were either recruited from Facebook or from friends and relatives of the researchers.

A Google Form published on Facebook served as our web-based survey platform. Recruitment remained open until the prespecified sample size of 32 (7 males, 21.9%; 25 females, 78.1%) healthy middle-aged and older participants was met. We confirmed the eligibility of the test takers based on a questionnaire (open between September 30, 2021, and October 4, 2021) and ensured that participants consented to the feasibility trial. Note that there was a 100% (32/32) response rate for the baseline questionnaire and that this was the denominator used for most of the analyses.

All participants filled in the informed consent form and swallowing function test form (EAT-10). People with specific illnesses or health problems (eg, those with pre-existing dysphagia or cancer) were excluded from this study, and only settings dedicated to home or informal care for older adults were included. In this study, the application was only functional on Android phones, tablets, and computers. We aim to upload the app to the Google Play Store in the near future.

Ethical Considerations

Taipei Medical University-Joint Institutional Review Board approved this study (N202109022), which was planned, carried out, analyzed, and interpreted independently of any industrial partners. Details about procedures, potential risks, confidentiality measures, data storage protocols, and benefits were shared with all eligible participants. Signed informed consent was collected from each participant. To protect confidentiality and ensure anonymity, participant names were replaced with unique identification codes. Digital data were securely stored on encrypted and password-protected systems. Participants in this study received US $15 as a transport subsidy. The specifics of the compensation were clearly communicated during the consent process.

Data Collection

This study was based on the baseline model game experience. After answering the questionnaire items regarding demographic background, the participants began to play. During the experiment, each participant played the game for 3 minutes with a song, playing 3 times in each of 3 difficulty levels (M1: slow; M2: medium; M3: fast). Physiological signals (heart rate, low frequency HRV, high frequency HRV, LF/HF ratio, lnRMSSD,

XSL•FO
RenderX

RMSSD, NN50, PNN50, and SDNN) were collected by the Garmin HRM-PRO. After the game, participants filled out the rest of the questionnaire, which comprised the following scales: FD, FSS2, NPS, UEQ, System Usability Scale, and NASA-TLX. In the questionnaire, there were measures for the challenge level of each exercise and measures of the skill level of the participant.

Subtracting the challenge score from the skill score and dividing it by 4 resulted in the FD, a score that fell between 1 and −1. A score of 0 meant that the skill and challenge levels were balanced, which was considered to be the state of flow. If a score was between 0 and 1, the subject was classified as bored (ie, the person's skill levels were higher than the challenge level). If a score was between −1 and 0, the person was classified as anxious. We integrated the questionnaire data and physiological data, which were downloaded from Elitehrv's paid platform (version 5.5.8, Elite HRV Inc) and organized in an Excel (version 1808, Microsoft) spreadsheet.

Statistical Analysis

In this study, SPSS version 30 (IBM Corp) was used as the main analysis software, and ANOVA tests were used to analyze whether differences among individual FDs in the M1 (slow), M2 (medium), and M3 (fast) levels were significant. Paired 2-tailed sample *t* tests were used to further analyze whether the training in the 3 levels affected the learning results in terms of operability. In addition, we also investigated whether there were differences between sexes, and the Pearson correlation coefficient was used to explore the relationship between the FSS2 scores and physiological signals. Exploratory factor analysis was used to extract the two factors (operation and immersion), and cluster analysis and k-means clustering were also used to divide the population into 3 groups based on speed, before reperforming the Pearson analysis.

Results

The reliability of the FSS2 questionnaire, as measured by the Cronbach α, was 0.802. An exploratory factor analysis was conducted on questions 1 to 9 of the FSS2 questionnaire, yielding a χ^2_{36} of 448.478 (P<.001; Kaiser-Meyer-Olkin=0.757). The FSS2 questionnaire was divided into two factors that confirmed the reliability of the structure: the FSS2-01 (FSS2 items 5, 7, 8, and 9) for immersion and FSS2-02 (FSS2 items 1, 2, 3, 4, and 6) for operation. An ANOVA test found that the F values were >0.05, indicating that the M1, M2, and M3 training modes all significantly affected the FSS2-02 (P=.001), NASA-TLX (P<.001), and FD (P<.001).

Within these 3 modes, there were significant differences in load level, flow, and FD (P<.001; Table 2). A Scheffe post hoc test confirmed that there was a significant difference in the NASA-TLX score between the M1 and M3 modes and between the M2 and M3 modes. There was a significant difference in the FSS2-02 score between the M1 and M3 modes. For FD, there were significant differences between the M1 and M2 modes, as well as the M1 and M3 modes (Multimedia Appendix 1).

Table . Relationship between the 3 training modes (M1, M2, and M3) and flow outcomes.

Variable	Sum of squares	Mean square	F test (df)	P value[a]
NASA-TLX[b]	1116823.000	558411.500	7.887 (2, 93)	.001
FSS2-02[c]	147.896	73.948	8.542 (2, 93)	<.001
Flow distance	27513.021	13756.510	12.198 (2, 93)	<.001

[a]The α level was .01 (2-tailed).

[b]NASA-TLX: National Aeronautics and Space Administration Task Load Index.

[c]FSS2-02: Flow State Scale 2, items 1, 2, 3, 4, and 6.

In order to determine the degree of influence of the FSS2-02 score on flow between each training mode pair, paired 2-tailed sample *t* tests were used. Significant differences were found among the 3 modes, indicating that participants were in the state of operation and demonstrating that the task design of this study was effective (Table 3). However, for the FSS2-01 questionnaire, only M1 and M3 exhibited a significant difference in immersion (P=.03). This could indicate that, in the M2 mode, the ease or difficulty of tasks was biased toward the M1 or M3 modes ().

XSL•FO
RenderX

Table . Valid operations by participants in the 3 modes (M1, M2, and M3).

Mode[a] comparisons	Mean difference (SD)	*t* test (*df*)	*P* value[b]
FSS2-02[c] (operation)			
M1 to M2	–1.71875 (2.61798)	–3.714 (31)	.001[d]
M2 to M3	–1.31250 (2.57077)	–2.888 (31)	.007[d]
M1 to M3	–3.03125 (4.18511)	–4.097 (31)	<.001[d]
FSS2-01[e] (immersion)			
M1 to M2	0.40625 (2.56351)	0.896 (31)	.38[f]
M2 to M3	0.68750 (3.71950)	1.046 (31)	.30[f]
M1 to M3	1.09375 (2.64404)	2.340 (31)	.03[f]

[a]Mode: training modes slow (M1), medium (M2), and fast (M3).

[b]Significance was determined using a one-way ANOVA.

[c]FSS2-02: Flow State Scale 2, items 1, 2, 3, 4, and 6.

[d]The α level was .01 (2-tailed).

[e]FSS2-01: Flow State Scale 2, items 5, 7, 8, and 9.

[f]The α level was .05 (2-tailed).

In addition, the results of the *t* tests (n=32) showed a significant difference in heart rate between men and women in the M1 mode (t_{28}=3.697; *P*=.001) and a significant difference in load level indicated by the NASA-TLX (t_{30}=2.276; *P*=.03). For the M2 mode, only the heart rate was significantly different between men and women (t_{30}=4.791; *P*<.001). Finally, there were significant differences in flow based on the FSS2-01 (t_{28}=5.431; *P*<.001), FSS2-02 (t_{29}=3.379; *P*=.002), FD (t_{30}=2.043; *P*=.05), as well as a significant difference in heart rate (t_{30}=3.693; *P*=.001) between men and women in the M3 mode (Table 4).

Table . Relationship between sex and physiological signals in the 3 mobile game modes (M1, M2, and M3).

Measurements	*F* test (*df*)	*P* value	*t* test (*df*)	*P* value
Slow training mode (M1)				
Heart rate[a]	5.228 (30, 27.5)	.03	3.697 (28)	.001[b]
NASA-TLX[c]	1.720 (30, 7.9)	.20	2.276 (30)	.03[d]
Medium training mode (M2)				
Heart rate	1.233 (30, 7.8)	.28	4.791 (30)	<.001[b]
Fast training mode (M3)				
FSS2-01[e]	5.496 (30, 28)	.03	5.431 (28)	<.001[b]
FSS2-02[f]	5.921(30, 29)	.02	3.379 (29)	.002[b]
Flow distance	0.601 (30, 8.5)	.44	2.043 (30)	.050[d]
Heart rate	0.832 (30, 8.3)	.37	3.693 (30)	.001[b]

[a]Measured in beats per minute.

[b]The α level was .01 (2-tailed).

[c]NASA-TLX: National Aeronautics and Space Administration Task Load Index.

[d]The α level was .05 (2-tailed).

[e]FSS2-01: Flow State Scale 2, items 5, 7, 8, and 9.

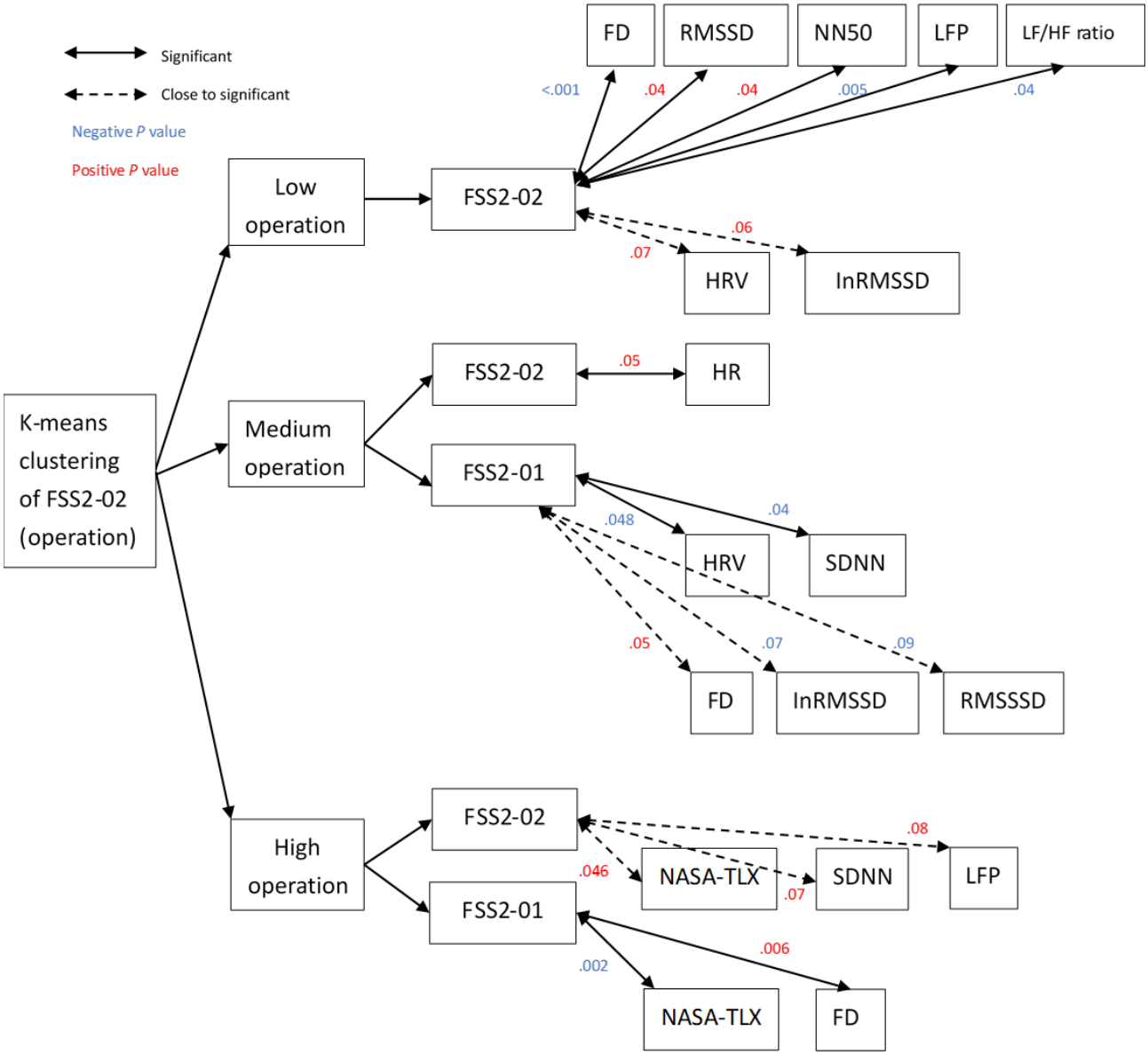
[f]FSS2-02: Flow State Scale 2, items 1, 2, 3, 4, and 6.

A Pearson correlation for physiological and psychological signals was performed for each training task mode and we found that the FSS2-01 score was significantly related to age (*r*=0.39; *P*<.001), pragmatic quality (*r*=0.21; *P*=.02), NASA-TLX score (*r*=–0.30; *P*=.003), FSS2-02 score (*r*=0.32; *P*=.001), and FD (*r*=0.33; *P*=.001). The FSS2-02 score was significantly correlated with age (*r*=0.34; *P*=.001), FSS2-01 score (*r*=0.32; *P*=.001), and FD (*r*=–0.34; *P*=.001). In addition, the FSS2-02 score was nonsignificantly correlated with the LF/HF ratio (*r*=–0.20; *P*=.054).

XSL•FO
RenderX

In order to explore whether certain training groups correlated with physiological and psychological signals, we divided the data into 3 groups through k-means clustering, partitioning the FSS2-02 scores into low, medium, and high operation groups. Correlation analysis between the 3 groups and physiological signals found that the FSS2-02 score for the low operation group was significantly correlated to FD (r=–0.79; P<.001), RMSSD (r=0.46; P=.04), NN50 (r=0.46; P=.04), LFP (r=–0.60; P=.005), and LF/HF ratio (r=–0.47; P=.04). There was a low correlation between the FSS2-02 score and HRV (r=0.42; P=.07) and lnRMSSD (r=0.43; P=.06). In the medium operation group, the FSS2-02 score correlated with heart rate (r=0.41; P=.050), while the FSS2-01 score correlated with age (r=0.58; P=.003), pragmatic quality (r=0.56; P=.005), UEQ score (r=0.48; P=.02), HRV (r=–0.41; P=.048), and the SDNN (r=–0.43; P=.04) and nonsignificantly correlated to hedonic quality (r=0.37; P=.07), FD (r=0.40; P=.053), lnRMSSD (r=–0.38; P=.07), and RMSSD (r=–0.35; P=.09). In the high operation group, the FSS2-02 score was correlated with the NASA-TLX score (r=0.28; P=.046) and was nonsignificantly correlated with SDNN (r=0.07; P=.07) and LFP (r=0.24; P=.08), while the FSS2-01 score was correlated with the NASA-TLX score (r=–0.41; P=.002) and FD (r=0.38; P=.006; Figure 3).

Figure 3. K-means clustering of FSS2-02 divided the relationship between mental flow and various physiological signals into 3 groups. Positive P values indicate a positive correlation, and negative P values indicate a negative correlation. FD: flow distance; FSS2-01: Flow State Scale 2, items 5, 7, 8, and 9; FSS2-02: Flow State Scale 2, items 1, 2, 3, 4, and 6; HR: heart rate; HRV: heart rate variability; LF/HF: low frequency to high frequency; LFP: low frequency power; lnRMSSD: natural logarithm of the root mean square of the successive differences; NN50: number of pairs of successive NN intervals that differ by more than 50 milliseconds; RMSSD: root mean square of the successive differences; SDNN: SD of the NN interval.

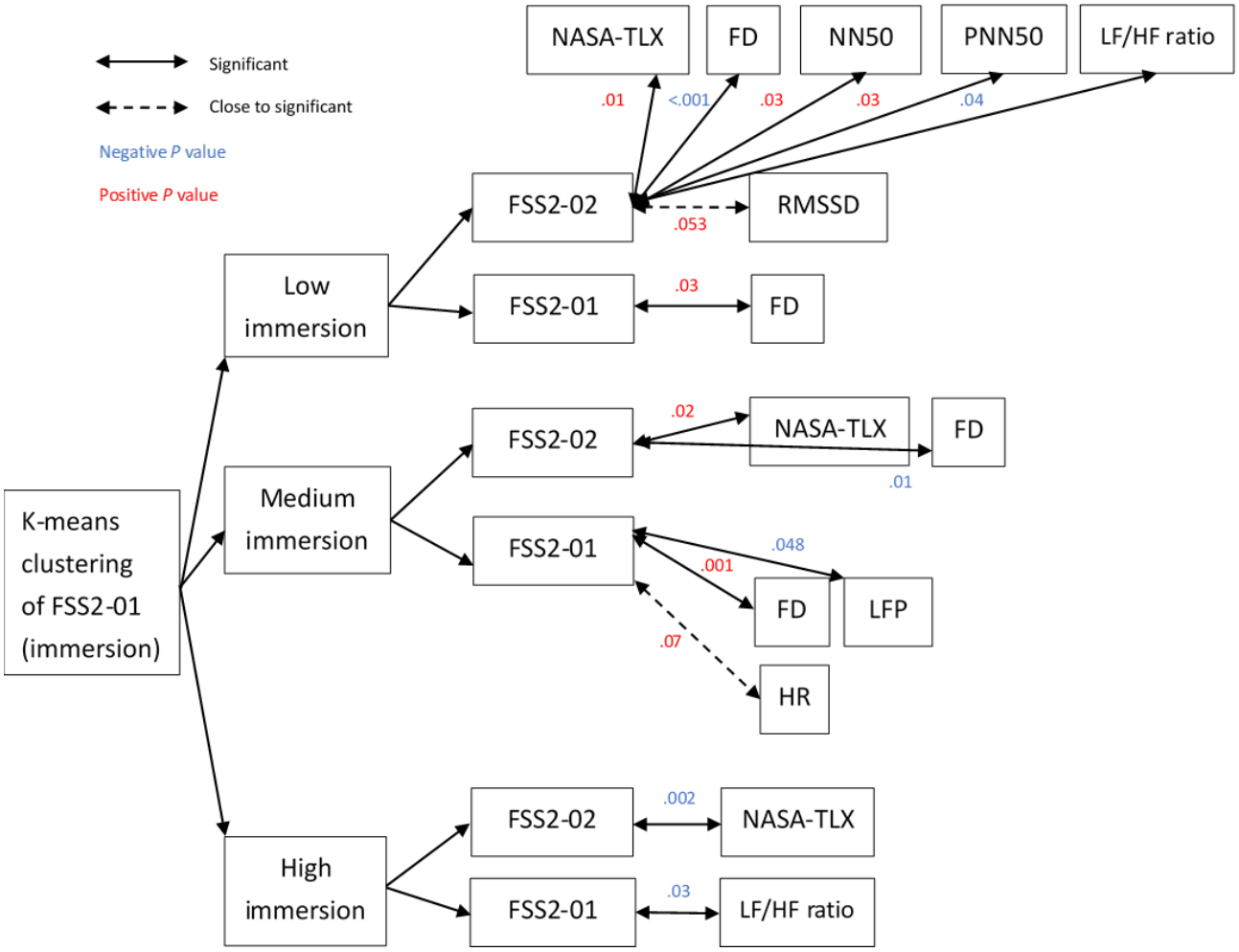


In addition, the research data for the FSS2-01 was divided into low, medium, and high immersion groups through k-means. Correlation analysis with the physiological signals found that the FSS2-02 score in the low immersion group was related to the NASA-TLX score (r=0.46; P=.009), FD (r=–0.69; P<.001), NN50 (r=0.39; P=.03), PNN50 (r=0.38; P=.03), and LF/HF ratio (r=–0.37; P=.04), while the FSS2-02 score had a low correlation with age (r=0.33; P=.07) and RMSSD (r=0.35;

XSL•FO
RenderX

P=.053), and the FSS2-01 score had a positive correlation with FD (*r*=0.39; *P*=.03). On the other hand, in the medium immersion group, the FSS2-02 score was correlated with the NASA-TLX score (*r*=0.39; *P*=.02) and FD (*r*=–0.42; *P*=.01), while the FSS2-01 score was correlated with hedonic quality (*r*=0.35; *P*=.04), UEQ score (*r*=0.36; *P*=.04), FD (*r*=0.53; *P*=.001), and LFP (*r*=–0.34; *P*=.048) and was nonsignificantly correlated with heart rate (*r*=0.31; *P*=.07). Finally, in the high immersion group, it was found that the FSS2-01 score exhibited a negative correlation with LF/HF ratio (*r*=–0.38; *P*=.03) and a low correlation with pragmatic quality (*r*=0.31; *P*=.08), and there was a negative correlation between the FSS2-02 and NASA-TLX scores (*r*=–0.51; *P*=.002; Figure 4).

Figure 4. K-means clustering of FSS2-01 divided the relationship between heart flow and various physiological signals into 3 groups. Positive *P* values indicate a positive correlation, and negative *P* values indicate a negative correlation. FD: flow distance FSS2-01: Flow State Scale 2, items 5, 7, 8, and 9; FSS2-02: Flow State Scale 2, items 1, 2, 3, 4, and 6; HR: heart rate; HRV: heart rate variability; LF/HF: low frequency to high frequency; LFP: low frequency power; NASA-TLX: National Aeronautics and Space Administration Task Load Index; NN50: number of pairs of successive NN intervals that differ by more than 50 milliseconds; PNN50: proportion of NN50 divided by the total number of NN intervals; RMSSD: root mean square of the successive differences.



Discussion

Principal Findings

We successfully identified 6 different methods for tongue training and developed a mobile game app for this purpose, aiming for users to experience flow through the training of specialized muscles (ie, the tongue). Most studies focuses on psychological and physiological signals (shown in Figure 3 and Figure 4), and this study emphasized 12 of these indicators (ie, NASA-TLX score, FD, heart rate, HRV, lnRMSSD, RMSSD, NN50, PNN50, SDNN, LFP, high frequency power, and LF/HF ratio). Among the psychological and physiological signals, flow was measured using FD, SDNN, RMSSD, NN50, HRV, and LFP. The LF/HF ratio was also used, which indicated if the findings were related to FD and if these factors generated flow. We found that our participants had effective learning operations and immersive experiences during the tasks and achieved optimal flow, confirming the effectiveness of the app. Importantly, flow generated during gameplay is known to be affected by age and sex, which in turn affects physiological signals, such as HRV [37].

In our mobile health app, interactive games were designed to train tongue muscles, which could be conducted anytime and anywhere, reducing the burden of conventional training. These games can also increase task repetition [55-57], which influences learning effectiveness [58,59]. The tasks were designed with 3 different levels of difficulty. Research has confirmed that different game designs can help participants to improve

XSL•FO
RenderX

operability and that introducing gamification into a task mechanism can intensify the balance of physiological and psychological signals [18]. When the balance between challenge and skill is achieved, flow optimization can be achieved [60]. Our findings demonstrate that the participants were immersed in training during the game, confirming the effectiveness of the tongue training app. By changing its speed and rhythm, we could modulate a user's flow experience. The tempo of the music was linked to the speed at which the tongue instruction icons moved down the smartphone screen, and training levels were generated. Thus, the M3 mode tasks were difficult and the most fast-paced. The participants had to sync the speed of the tongue training motions to the rhythm of the music. This study combined image recognition and music rhythm stimulation to explore and analyze complex physiological and psychological signals. The results show that such methods could generate an ideal flow experience.

Strengths and Weaknesses of the Study

Our research has successfully simplified complicated rehabilitation procedures, using a simple game app combined with audio effects to facilitate tongue training. To verify the effectiveness of the mobile app games, the FSS2 questionnaire for the game was divided into two groups by exploratory factor analyses: operation and immersion. An analysis of the results for the tasks performed at different speeds revealed that training outcomes were highly related to FD. This demonstrates that through this tongue training app, we can both facilitate tongue muscle rehabilitation and generate flow in users. Flow includes experiencing complete concentration, the desire to challenge oneself, and the generation of feedback reward mechanisms [61], supporting that physiological signals are affected by task stimulation [62].

The image recognition system had certain limitations. Data collection during experiments was also a time-consuming process. In our study, the participant had to wear a heart rate detector during the test, which may have caused some inconvenience and led to minor variations in heart rhythm or physiological signals. Second, research has found that there are indeed differences in flow and physiological signals between men and women, indicating differences regarding the influence of personal characteristics and environmental factors [37]. Under the same difficulty settings, results showed that there were fewer psychological signals in the M3 training mode. In the future, we can appropriately adjust the task difficulty of the M3 mode to achieve optimal training. In addition, despite the small sample size and the skewed sex ratio of the participants in this study, our image recognition system achieved a high level of accuracy, paving the way for scaling up the recruitment of participants in future experiments.

The model of this complete mobile app game was trained using a limited dataset and thus only recognized the mouths and tongues of Asian individuals. Therefore, if the mobile app was applied to non-Asian individuals, it may not perform well, leading to inaccurate experiments. Moreover, these mobile app games were only used by middle-aged and older adults, and the findings may not apply to users in other age groups. In the future, we aim to collect more diverse data to improve the accuracy and breadth of the platform.

Implications and Future Research

We expect to include the addition of a 2-player competition mode to the interface to enable not only interactions with family or friends but also to include a wider range of advanced training functions. The interface for middle-aged and older adults could be optimized with more options, such as realistic characters and diverse oral and tongue exercises, to enhance the experience of the app. The interface only allowed users to change basic settings such as age, music speed, and mode. Since this study was a pilot study, there were no experiment cases without accompanying soundtracks (ie, during the trial, there were corresponding soundtracks in all 3 modes).

In this study, the generation of a soundtrack required an internet connection. An offline training mode is planned for subsequent iterations of the app so that full training can be achieved anytime and anywhere, even in places with poor signal reception, such as rural areas, elevators, or basements. Furthermore, the soundtrack could be adjusted based on pitch or volume, or it could be muted in future experiments. In this study, our main goal was to provide either an alternative to clinical rehabilitation or a more relaxed training setting for healthy people. When used in actual clinical settings, complying with medical standards and regulations could require the demonstration of efficacy across multiple conditions. For example, the app may be further modified to include oral training. Moreover, the development of the image recognition system was susceptible to slight errors depending on variations in ambient lighting. In the future, we hope to enhance the accuracy of the image recognition system and collect more lighting data to reduce errors under different scenes and levels of light.

Conclusions

Despite the many ways to train tongue muscles, most studies have been concerned with the treatment and rehabilitation of patients with oral cavity diseases. In addition, equipment expenses or a lack of continuous professional medical services are associated with declines in patient compliance for rehabilitation. There are few gamified apps targeting health care professions and education and even fewer considering factors that may increase efficacy. Following the completion of the pilot trial and subsequent analysis, the participant cohort could be categorized into two groups based on their physiological and psychological signals: operation and immersion. Furthermore, tongue muscle training using our app could produce a superior flow experience for users. This was determined using several variables including FD, SDNN, RMSSD, NN50, HRV, LFP, and LF/HF ratio and could be repeated in different training groups, which confirmed the effectiveness of this app. Compared to strength training and full-body movement training, our method was more engaging and produced a better flow experience. We hope that the optimal flow generated by this app will encourage users to train independently and mitigate or prevent dysphagia. This study strengthens the potential for incorporating game training into mobile health.

XSL•FO
RenderX

Acknowledgments

The findings and conclusions are those of the authors, who are responsible for its contents. We thank Taipei Medical University and the University System of Taipei Joint Research Program (USTP-NTUT-TMU-110-03) for their assistance.

Data Availability

The datasets generated and/or analyzed during this study are not publicly available due to security purposes requested by the academic institution but are available from the corresponding author on reasonable request.

Authors' Contributions

KCS, KCW, KRC, and CHH conceptualized the trial. KCS, KCW, and CHH collected and analyzed the data. KCS wrote the manuscript, to which all authors gave critical feedback. All authors approved the final draft of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Multiple comparisons from Scheffe test.
[PDF File, 100 KB - games_v13i1e53045_app1.pdf]

Checklist 1
Reporting a pilot and feasibility trial checklist.
[PDF File, 285 KB - games_v13i1e53045_app2.pdf]

References

1. Busch PA, Hausvik GI, Ropstad OK, Pettersen D. Smartphone usage among older adults. Comput Hum Behav 2021 Aug;121:106783. [doi: 10.1016/j.chb.2021.106783]
2. Turel O, Serenko A. The benefits and dangers of enjoyment with social networking websites. Eur J Inf Syst 2012 Sep;21(5):512-528. [doi: 10.1057/ejis.2012.1]
3. Lee JA, Nguyen AL, Berg J, et al. Attitudes and preferences on the use of mobile health technology and health games for self-management: interviews with older adults on anticoagulation therapy. JMIR Mhealth Uhealth 2014 Jul 23;2(3):e32. [doi: 10.2196/mhealth.3196] [Medline: 25098413]
4. Foster M, Xiong W, Quintiliani L, Hartmann CW, Gaehde S. Preferences of older adult veterans with heart failure for engaging with mobile health technology to support self-care: qualitative interview study among patients with heart failure and content analysis. JMIR Form Res 2022 Dec 20;6(12):e41317. [doi: 10.2196/41317] [Medline: 36538348]
5. Sheikhany AR, Hady AFA, Farag S. Oropharyngeal dysphagia profile in early versus late stage dementia. Egypt J Otolaryngol 2019 Jan;35(1):103-109. [doi: 10.4103/ejo.ejo_98_18]
6. Singh S, Hamdy S. Dysphagia in stroke patients. Postgrad Med J 2006 Jun;82(968):383-391. [doi: 10.1136/pgmj.2005.043281] [Medline: 16754707]
7. Tjaden K. Speech and swallowing in Parkinson's disease. Top Geriatr Rehabil 2008;24(2):115-126. [doi: 10.1097/01.TGR.0000318899.87690.44] [Medline: 19946386]
8. Manabe T, Teramoto S, Tamiya N, Okochi J, Hizawa N. Risk factors for aspiration pneumonia in older adults. PLoS ONE 2015;10(10):e0140060. [doi: 10.1371/journal.pone.0140060] [Medline: 26444916]
9. Ekberg O, Hamdy S, Woisard V, Wuttge-Hannig A, Ortega P. Social and psychological burden of dysphagia: its impact on diagnosis and treatment. Dysphagia 2002;17(2):139-146. [doi: 10.1007/s00455-001-0113-5] [Medline: 11956839]
10. Takeuchi N, Sawada N, Ekuni D, Morita M. Oral diadochokinesis is related to decline in swallowing function among community-dwelling Japanese elderly: a cross-sectional study. Aging Clin Exp Res 2021 Feb;33(2):399-405. [doi: 10.1007/s40520-020-01547-7] [Medline: 32306370]
11. Robbins J, Kays SA, Gangnon RE, et al. The effects of lingual exercise in stroke patients with dysphagia. Arch Phys Med Rehabil 2007 Feb;88(2):150-158. [doi: 10.1016/j.apmr.2006.11.002] [Medline: 17270511]
12. Antunes EB, Lunet N. Effects of the head lift exercise on the swallow function: a systematic review. Gerodontology 2012 Dec;29(4):247-257. [doi: 10.1111/j.1741-2358.2012.00638.x] [Medline: 22612867]
13. Shaker R, Kern M, Bardan E, et al. Augmentation of deglutitive upper esophageal sphincter opening in the elderly by exercise. Am J Physiol 1997 Jun;272(6 Pt 1):G1518-G1522. [doi: 10.1152/ajpgi.1997.272.6.G1518] [Medline: 9227489]

XSL•FO
RenderX

14. Fukuoka T, Ono T, Hori K, Kariyasu M. Effects of tongue-strengthening exercise on tongue strength and effortful swallowing pressure in young healthy adults: a pilot study. J Speech Lang Hear Res 2022 May 11;65(5):1686-1696. [doi: 10.1044/2022_JSLHR-21-00331] [Medline: 35363546]
15. Carnaby-Mann GD, Crary MA. McNeill dysphagia therapy program: a case-control study. Arch Phys Med Rehabil 2010 May;91(5):743-749. [doi: 10.1016/j.apmr.2010.01.013] [Medline: 20434612]
16. Carnaby-Mann G, Crary MA, Schmalfuss I, Amdur R. "Pharyngocise": randomized controlled trial of preventative exercises to maintain muscle structure and swallowing function during head-and-neck chemoradiotherapy. Int J Radiat Oncol Biol Phys 2012 May 1;83(1):210-219. [doi: 10.1016/j.ijrobp.2011.06.1954] [Medline: 22014959]
17. Miller N, Noble E, Jones D, Burn D. Hard to swallow: dysphagia in Parkinson's disease. Age Ageing 2006 Nov;35(6):614-618. [doi: 10.1093/ageing/afl105] [Medline: 17047007]
18. Kothari M, Svensson P, Jensen J, et al. Training-induced cortical plasticity compared between three tongue-training paradigms. Neuroscience 2013 Aug 29;246:1-12. [doi: 10.1016/j.neuroscience.2013.04.040] [Medline: 23632170]
19. Monteblanco Cavalcante M, Fraga I, Dalbosco B, et al. Exergame training-induced neuroplasticity and cognitive improvement in institutionalized older adults: a preliminary investigation. Physiol Behav 2021 Nov 1;241:113589. [doi: 10.1016/j.physbeh.2021.113589] [Medline: 34509470]
20. Hou HY, Li HJ. Effects of exergame and video game training on cognitive and physical function in older adults: a randomized controlled trial. Appl Ergon 2022 May;101:103690. [doi: 10.1016/j.apergo.2022.103690] [Medline: 35066398]
21. Chuang CS, Chen YW, Zeng BY, et al. Effects of modern technology (exergame and virtual reality) assisted rehabilitation vs conventional rehabilitation in patients with Parkinson's disease: a network meta-analysis of randomised controlled trials. Physiotherapy 2022 Dec;117:35-42. [doi: 10.1016/j.physio.2022.07.001] [Medline: 36242929]
22. Huang K, Zhao Y, He R, et al. Exergame-based exercise training for depressive symptoms in adults: a systematic review and meta-analysis. Psychol Sport Exerc 2022 Nov;63:102266. [doi: 10.1016/j.psychsport.2022.102266]
23. Huang HC, Wong MK, Lu J, Huang WF, Teng CI. Can using exergames improve physical fitness? A 12-week randomized controlled trial. Comput Hum Behav 2017 May;70:310-316. [doi: 10.1016/j.chb.2016.12.086]
24. Jackson SA, Eklund RC. Assessing flow in physical activity: the flow state scale–2 and dispositional flow scale–2. J Sport Exerc Psychol 2002;24(2):133-150. [doi: 10.1123/jsep.24.2.133]
25. Geng D, Yang K, Fu Z, Zhang Y, Wang C, An H. Circadian stage-dependent and stimulation duration effects of transcutaneous auricular vagus nerve stimulation on heart rate variability. PLoS ONE 2022;17(11):e0277090. [doi: 10.1371/journal.pone.0277090] [Medline: 36327249]
26. Li D, Gao W. Physiological state assessment and prediction based on multi-sensor fusion in body area network. Biomed Signal Process Control 2021 Mar;65:102340. [doi: 10.1016/j.bspc.2020.102340]
27. Liu Z, Wei W, Fu W. Amplified acquisition of physiological signal in human body communication. J Internet Technol 2021 Dec;22(7):1649-1659. [doi: 10.53106/160792642021122207017]
28. Collins T, Woolley SI, Oniani S, et al. Version reporting and assessment approaches for new and updated activity and heart rate monitors. Sensors (Basel) 2019 Apr 10;19(7):1705. [doi: 10.3390/s19071705] [Medline: 30974755]
29. Patel N, Patel S, Mankad SH. Impact of autoencoder based compact representation on emotion detection from audio. J Ambient Intell Humaniz Comput 2022;13(2):867-885. [doi: 10.1007/s12652-021-02979-3] [Medline: 33686349]
30. Gantayat SS, Lenka S. Study of algorithms and methods on emotion detection from facial expressions: a review from past research. Comm Softw Netw 2021:231-244. [doi: 10.1007/978-981-15-5397-4_24]
31. Wu J, Zhang Y, Sun S, Li Q, Zhao X. Generalized zero-shot emotion recognition from body gestures. Appl Intell 2022 Jun;52(8):8616-8634. [doi: 10.1007/s10489-021-02927-w]
32. Yang K, Wang C, Gu Y, et al. Behavioral and physiological signals-based deep multimodal approach for mobile emotion recognition. IEEE Trans Affect Comput 2021;14(2):1082-1097. [doi: 10.1109/TAFFC.2021.3100868]
33. Levenson RW, Carstensen LL, Friesen WV, Ekman P. Emotion, physiology, and expression in old age. Psychol Aging 1991 Mar;6(1):28-35. [doi: 10.1037//0882-7974.6.1.28] [Medline: 2029364]
34. Levenson RW. The autonomic nervous system and emotion. Emot Rev 2014 Apr;6(2):100-112. [doi: 10.1177/1754073913512003]
35. Ludwig M, Hoffmann K, Endler S, Asteroth A, Wiemeyer J. Measurement, prediction, and control of individual heart rate responses to exercise-basics and options for wearable devices. Front Physiol 2018;9:778. [doi: 10.3389/fphys.2018.00778] [Medline: 29988588]
36. Stauss HM. Heart rate variability. Am J Physiol Regul Integr Comp Physiol 2003 Nov;285(5):R927-R931. [doi: 10.1152/ajpregu.00452.2003] [Medline: 14557228]
37. Zhang J. Effect of age and sex on heart rate variability in healthy subjects. J Manipul Physiol Ther 2007 Jun;30(5):374-379. [doi: 10.1016/j.jmpt.2007.04.001] [Medline: 17574955]
38. Bourdillon N, Yazdani S, Vesin JM, Schmitt L, Millet GP. RMSSD is more sensitive to artifacts than frequency-domain parameters: implication in athletes' monitoring. J Sports Sci Med 2022 Jun;21(2):260-266. [doi: 10.52082/jssm.2022.260] [Medline: 35719238]

XSL•FO
RenderX

39. Garg V, Verma S, Connelly KA, et al. Does empagliflozin modulate the autonomic nervous system among individuals with type 2 diabetes and coronary artery disease? The EMPA-HEART CardioLink-6 Holter analysis. Metabol Open 2020 Sep;7:100039. [doi: 10.1016/j.metop.2020.100039] [Medline: 32812924]
40. Von Bargen T, Zientz C, Haux R. Gamification for mHealth - a review of playful mobile healthcare. Stud Health Technol Inform 2014;202:225-228. [Medline: 25000057]
41. Miller AS, Cafazzo JA, Seto E. A game plan: gamification design principles in mHealth applications for chronic disease management. Health Informatics J 2016 Jun;22(2):184-193. [doi: 10.1177/1460458214537511] [Medline: 24986104]
42. Csikszentmihalyi M. Creativity: Flow and the Psychology of Discovery and Invention: Harper Perennial; 1996.
43. Csikszentmihalyi M. Flow: The Psychology of Optimal Experience: Harper Perennial; 1990. [doi: 10.1080/00222216.1992.11969876]
44. Novak TP, Hoffman DL, Yung YF. Measuring the customer experience in online environments: a structural modeling approach. Mktg Sci 2000 Feb;19(1):22-42. [doi: 10.1287/mksc.19.1.22.15184]
45. Csikszentmihalyi M, Hunter J. Happiness and creativity. Fut 2003;4(2):185-199. [doi: 10.1023/A:1024409732742]
46. Hoffman DL, Novak TP. Marketing in hypermedia computer-mediated environments: conceptual foundations. J Mark 1996 Jul;60(3):50-68. [doi: 10.1177/002224299606000304]
47. Novak TP, Hoffman DL, Duhachek A. The influence of goal-directed and experiential activities on online flow experiences. J Consum Psychol 2003 Mar 1;13(1):3-16. [doi: 10.1207/153276603768344744]
48. Tong Y, Liang Y, Spasic I, Hicks Y, Hu H, Liu Y. A data-driven approach for integrating hedonic quality and pragmatic quality in user experience modeling. J Comput Inf Sci Eng 2022 Dec 1;22(6). [doi: 10.1115/1.4054155]
49. Atoum I. Measurement of key performance indicators of user experience based on software requirements. Sci Comput Program 2023 Mar;226:102929. [doi: 10.1016/j.scico.2023.102929]
50. Bouchard MA, Bergeron-Boucher J, Chamberland C, Tremblay S, Jackson PL. Assessing Differences in Emotional Expressivity Between Expert and Non Expert Video Game Players Using Facial Electromyography. Neuroergonomics 2018:313-314. [doi: 10.1016/B978-0-12-811926-6.00091-9]
51. Choi DS, Kim HY, Kim JW. A cognitive and emotional strategy for computer game design. Asia Pac J Inf Syst 2000;10:165-187 [FREE Full text]
52. Chen LX, Sun CT. Self-regulation influence on game play flow state. Comput Hum Behav 2016 Jan;54:341-350. [doi: 10.1016/j.chb.2015.08.020]
53. Zhang J, Huang Y, Dong Y, Li J, Zhu L, Zhao M. The effect of music tempo on movement flow. Front Psychol 2024;15:1292516. [doi: 10.3389/fpsyg.2024.1292516] [Medline: 38348258]
54. Hwu YJ, Kuo CT, Chiang QG, et al. Guidance manual for care and guidance of eating and swallowing difficulties. : Health Promotion Administration, Ministry of Health and Welfare; 2019 Jun URL: https://www.mohw.gov.tw/dl-58241-7dacaed8-99a1-4053-aa6a-a7f106435fc6.html [accessed 2021-09-01]
55. Huo X, Wang J, Ghovanloo M. A wireless tongue-computer interface using stereo differential magnetic field measurement. Annu Int Conf IEEE Eng Med Biol Soc 2007;2007:5724-5727. [doi: 10.1109/IEMBS.2007.4353646] [Medline: 18003312]
56. Huo X, Wang J, Ghovanloo M. Introduction and preliminary evaluation of the Tongue Drive System: wireless tongue-operated assistive technology for people with little or no upper-limb function. J Rehabil Res Dev 2008;45(6):921-930. [doi: 10.1682/jrrd.2007.06.0096] [Medline: 19009478]
57. Huo X, Wang J, Ghovanloo M. A magneto-inductive sensor based wireless tongue-computer interface. IEEE Trans Neural Syst Rehabil Eng 2008 Oct;16(5):497-504. [doi: 10.1109/TNSRE.2008.2003375] [Medline: 18990653]
58. Hartmann T, Klimmt C. Gender and computer games: exploring females' dislikes. J Comp Mediated Comm 2006 Jul;11(4):910-931. [doi: 10.1111/j.1083-6101.2006.00301.x]
59. Hartman JM. Self-controlled use of a perceived physical assistance device during a balancing task. Percept Mot Skills 2007 Jun;104(3 Pt 1):1005-1016. [doi: 10.2466/pms.104.3.1005-1016] [Medline: 17688157]
60. Shin N. Online learner's 'flow' experience: an empirical study. Brit J Educ Tech 2006 Sep;37(5):705-720. [doi: 10.1111/j.1467-8535.2006.00641.x]
61. Bressler DM, Bodzin AM. A mixed methods assessment of students' flow experiences during a mobile augmented reality science game. Comp Assisted Learn 2013 Dec;29(6):505-517. [doi: 10.1111/jcal.12008]
62. Martin-Niedecken AL, Schwarz T, Schättin A. Comparing the impact of heart rate-based in-game adaptations in an exergame-based functional high-intensity interval training on training intensity and experience in healthy young adults. Front Psychol 2021;12:572877. [doi: 10.3389/fpsyg.2021.572877] [Medline: 34234705]

Abbreviations

EAT-10: 10-item Eating Assessment Tool
FD: flow distance
FSS2: Flow State Scale 2
FSS2-01: Flow State Scale 2, items 5, 7, 8, and 9
FSS2-02: Flow State Scale 2, items 1, 2, 3, 4, and 6

XSL•FO
RenderX

HRV: heart rate variability
LF/HF: low frequency to high frequency
LFP: low frequency power
lnRMSSD: natural logarithm of the root mean square of successive differences
NASA-TLX: National Aeronautics and Space Administration Task Load Index
NN50: number of pairs of successive NN intervals that differ by more than 50 milliseconds
NPS: net promoter score
PNN50: proportion of NN50 divided by the total number of NN intervals
RMSSD: root mean square of successive differences
SDNN: SD of the NN interval
UEQ: User Experience Questionnaire

Edited by A Coristine, T Leung; submitted 25.09.23; peer-reviewed by A Hadianfard, MT Irshad; revised version received 13.11.24; accepted 13.11.24; published 09.01.25.

Please cite as:
Su KC, Wu KC, Chou KR, Huang CH
Tongue Muscle Training App for Middle-Aged and Older Adults Incorporating Flow-Based Gameplay: Design and Feasibility Pilot Study
JMIR Serious Games 2025;13:e53045
URL: https://games.jmir.org/2025/1/e53045
doi:10.2196/53045

© Kuan-Chu Su, Ko-Chiu Wu, Kuei-Ru Chou, Chia-Hsu Huang. Originally published in JMIR Serious Games (https://games.jmir.org), 9.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

The Usability and Effect of a Novel Intelligent Rehabilitation Exergame System on Quality of Life in Frail Older Adults: Prospective Cohort Study

Chien-Hsiang Chang[1], PhD; Chun-Chun Wei[2], PhD; Wei-Chih Lien[3*], PhD; Tai-Hua Yang[4], PhD; Bo Liu[1], PhD; Yimo Lin[1], PhD; Poh Thong Tan[1], PhD; Yang-Cheng Lin[1*], PhD

[1]Department of Industrial Design, National Cheng Kung University, Tainan, Taiwan

[2]Department of Creative Design, National Yunlin University of Science and Technology, Yunlin, Taiwan

[3]Department of Physical Medicine and Rehabilitation, National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University, Tainan, Taiwan

[4]Department of Biomedical Engineering, National Cheng Kung University, Tainan, Taiwan

[*]these authors contributed equally

Corresponding Author:
Yang-Cheng Lin, PhD
Department of Industrial Design, National Cheng Kung University, Tainan, Taiwan

Abstract

Background: Aging in older adults results in a decline in physical function and quality of daily life. Due to the COVID-19 pandemic, the exercise frequency among older adults decreased, further contributing to frailty. Traditional rehabilitation using repetitive movements tends not to attract older adults to perform independently.

Objective: Intelligent Rehabilitation Exergame System (IRES), a novel retro interactive exergame that incorporates real-time surface electromyography, was developed and evaluated.

Methods: Frail older adults were invited to use the IRES for rehabilitation using lower limb training twice per week for 4 weeks. Participants were required to have no mobility or communication difficulties and be willing to complete the 4-week study. The enrolled cohort had baseline scores ranging from 1 to 5 on the Clinical Frailty Scale, as described by Rockwood et al. Three major lower limb movements (knee extension, plantar flexion, and dorsiflexion) were performed 20 times for each leg within 30 minutes. The surface electromyography collected and analyzed muscle potential signals for review by health care professionals to customize the protocol for the next training. The System Usability Scale (SUS) and Taiwanese version of the EuroQol-5 Dimensions (EQ-5D) were administered after completing the first (week 1, baseline) and last training (week 4, one-month follow-up) to evaluate the usability of the IRES and its effects on the quality of life of participants.

Results: A total of 49 frail older adults (mean age 74.6 years) were included in the analysis. The usability of the IRES improved according to the mean SUS score, from 82.09 (good) at baseline to 87.14 (good+) at 1-month follow-up. The willingness to use (t_{96}=−4.51; P<.001), learnability (t_{96}=−4.83; P<.001), and confidence (t_{96}=−2.27; P=.02) in working with the IRES increased. After using the IRES for 1 month, significant improvements were observed in the ease of use (t_{47}=2.05; P=.04) and confidence (t_{47}=2.68; P<.001) among participants without previous rehabilitation experience. No sex-based differences in the SUS scores were found at baseline or 1-month follow-up. The quality of life, as assessed by the EQ-5D, improved significantly after 1 month of IRES training compared to that at baseline (t_{96}=6.03; P<.001).

Conclusions: The novel IRES proposed in this study received positive feedback from frail older adults. Integrating retro-style exergame training into rehabilitation not only improved their rehabilitation motivation but also increased their learning, system operability, and willingness to continue rehabilitation. The IRES provides an essential tool for the new eHealth care service in the post–COVID-19 era.

(JMIR Serious Games 2025;13:e50669) doi:10.2196/50669

KEYWORDS

frailty; exergame; older adults; intelligent rehabilitation; reminiscence therapy; eHealth care

XSL•FO
RenderX

Introduction

Global populations are aging at an escalating speed. The aging-related decline in physical function and quality of life imposes a huge health care burden. Frailty is a clinical aging syndrome attributed to chronic disease interactions and other factors, including an unbalanced diet and adverse lifestyle habits. In Taiwan, the prevalence of frailty and prefrailty in older adults aged 65 years and above is 5.4% and 41.5%, respectively, with a higher prevalence among females than males [1-3]. The clinical presentation of frailty includes decreased activity, abnormal weight loss, reduced appetite, fatigue without apparent cause, muscle loss, gait and balance abnormalities, and cognitive function impairment [4]. Frailty affects the endocrine, cardiovascular, and musculoskeletal systems [5]. Age-related muscle atrophy, known as sarcopenia, correlates highly with frailty [6]. The gait speed of older adults gradually decreases with age and is directly related to the bone and muscle mass of the lower limbs [7]. At present, lower limb rehabilitation involves repetitive training and often does not encourage older adults to perform independently [8,9].

Globally, the predicted demand for health care professionals will reach 80 million in 2030, while the number of available professionals will be less than 18 million [10]. To address this need, the health care system is using innovative technology such as intelligent telemedicine, smart sensor technology, and wearable devices [11,12]. An intelligent health care ecosphere created by information and communications technology is helpful for preventive health care, fitness, and health care services [13]. Interactive technology devices, as effective training tools, improve mobility and balance in patients [14]. Previous studies report significant improvements in several domains of physical function via game-related rehabilitation [15,16]. These findings suggest the effectiveness of exergames, which integrate games into exercise training, to improve physical function during rehabilitation. In addition, mental health is also improved by exergames, which increase the willingness of older adults to participate in long-term exercise [17]. Using exergames for rehabilitation to restore balance in older adults and to shorten the rehabilitation treatment period is an economically beneficial option for home training. Furthermore, exergames with visual feedback provide more effective rehabilitation; such improvement is unlikely to be achieved by traditional manual rehabilitation training, which diminishes with age [18].

Reminiscence therapy, which incorporates life experiences using old photographs or scenarios into activities, is widely used in the care of older adults [19]. This method stimulates long-term memory and aims to maintain patients' physical and mental health and quality of life, regardless of their cognitive function. The benefits of reminiscence therapy include increasing self-esteem, decreasing depression, and promoting mental health, thereby increasing life satisfaction [20,21].

Older adults are at high risk of COVID-19 infection due to frailty. The COVID-19 pandemic has placed immense pressure on health care systems; in response, a variety of exergames with interactive technology were launched during the pandemic for home use. These exergames can be used as activity tools to maintain muscle strength and decrease physical and mental stress during measures for COVID-19 control [8,9]. In the post–COVID-19 era [22], exergames may help them to keep training and to reduce the burden on the health care system.

In this study, an Intelligent Rehabilitation Exergame System (IRES) that included a retro interactive exergame, real-time surface electromyography (sEMG) used to control the game, and reminiscence therapy was developed for use in the rehabilitation of frail older adults [19,23]. In addition, IRES records and provides data to health care professionals during rehabilitation to allow for the optimization of the rehabilitation protocol [19,23]. The System Usability Scale (SUS) was used to analyze the usability of IRES for rehabilitation [8,24], and the Taiwanese version of the EuroQol-5 Dimensions (EQ-5D) was used to evaluate changes in the quality of life after IRES training in frail older adults [25].

Methods

Intelligent Rehabilitation Exergame System

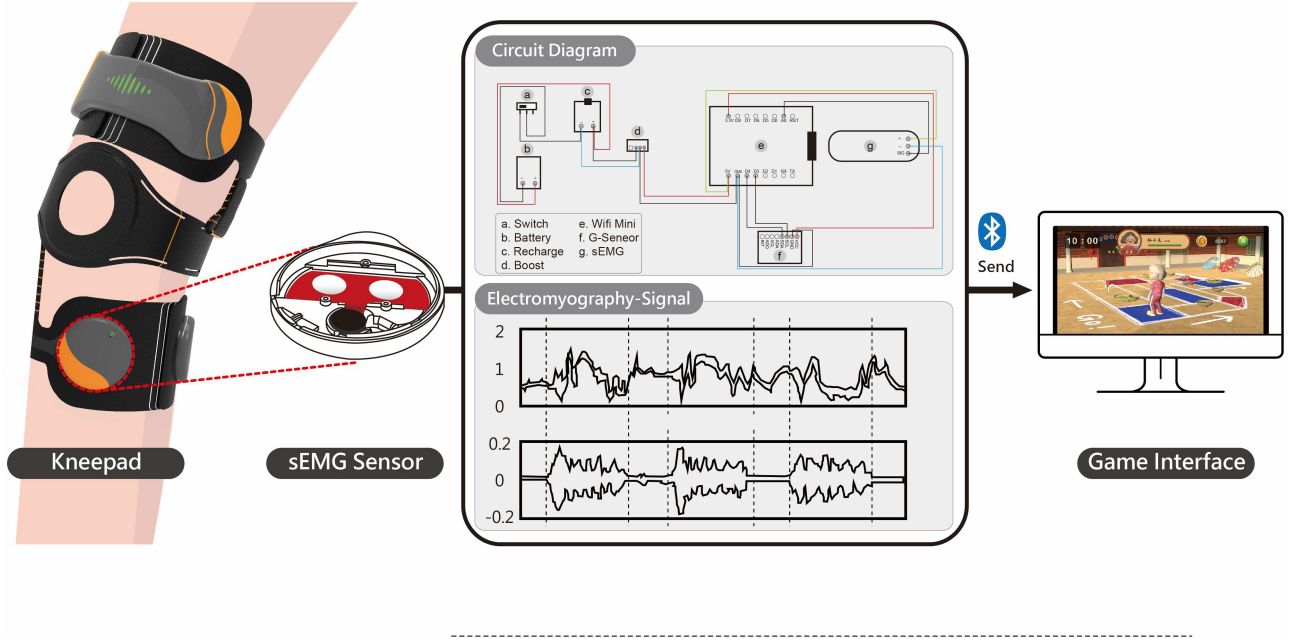
IRES comprises an exergame for rehabilitation that interfaces with a virtual physician, historical photographs with constructed scenarios for reminiscence therapy, and an sEMG for controlling the exergame and recording electrophysiological data for muscle activity during rehabilitation.

sEMG Signal-Sensing Technology

sEMG is composed of electrode patches and sensors that are placed on the surface of the skin to detect signals useful for human-machine interactions, gesture recognition, and muscle function evaluation, as previously described [26,27]. Touch fasteners were used to position a kneepad with 3 electrode patches at the bottom of the sensor. The changes in muscle potential signals caused by muscle contractions during limb exercise were measured [28]. The signals were amplified and transmitted by the wireless transmission module to the main unit of the sEMG, then displayed on the screen and stored [27]. The signals were also transmitted to the Unity game software (Unity Technologies), which serves as an interactive multimedia medium to provide visual feedback to the user (Figure 1).

XSL•FO
RenderX

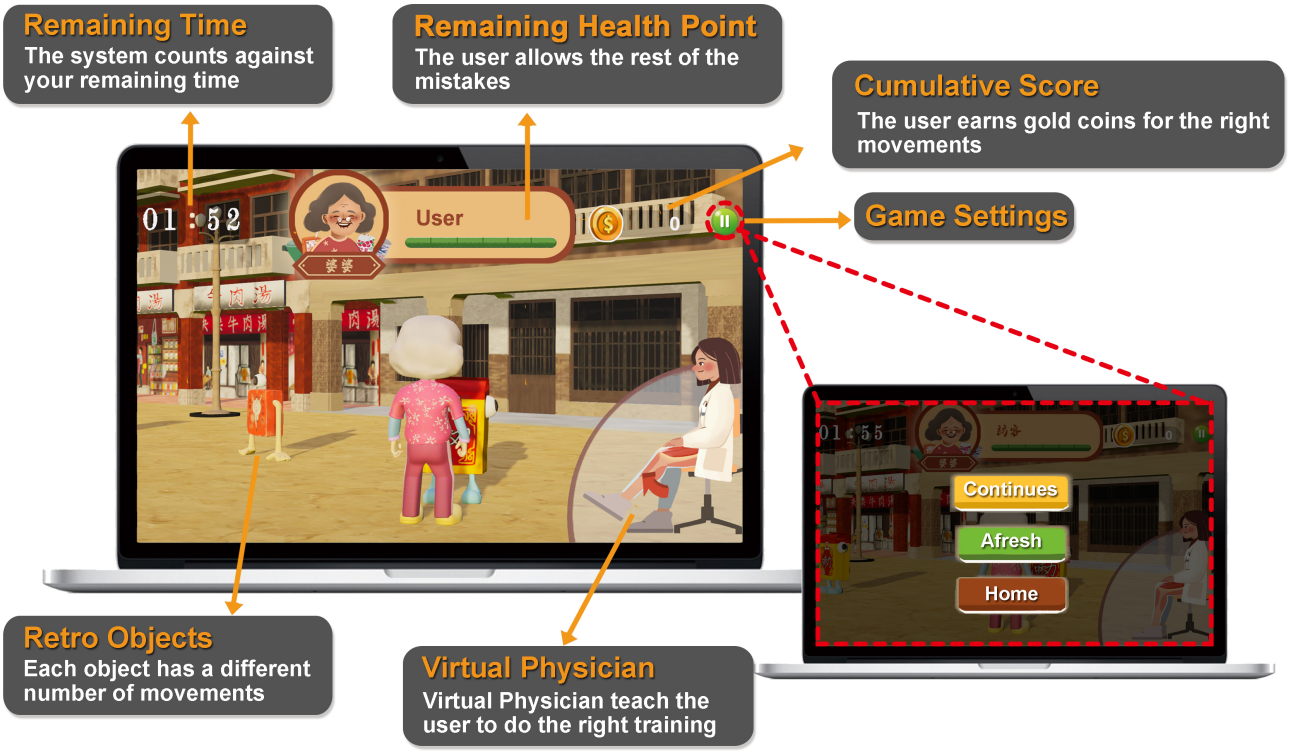
Figure 1. Illustration of the real-time surface electromyography, circuit diagram, and electromyography signal.



Design of the Exergame

The IRES system incorporates several innovative features: (1) it is designed to accommodate both medical professionals and frail older adults, (2) it records the sEMG data of muscle group movement during rehabilitation training, (3) this data is then sent for analysis, (4) feedback is provided to medical professionals to assess and adjust rehabilitation parameters, and (5) it recommends the most suitable rehabilitation treatment for the frail older user. Representative historical and cultural images were used for reminiscence therapy that targeted older adults. For example, photographs of the Hayashi Department Store in Tainan in the 1960s were used as the inspiration to draw the interface background, and the interactive components were also designed based on items that were well-known in Taiwan in the 1960s (Figure 2).

Figure 2. The interface design with the retro style of the IRES (Intelligent Rehabilitation Exergame System).

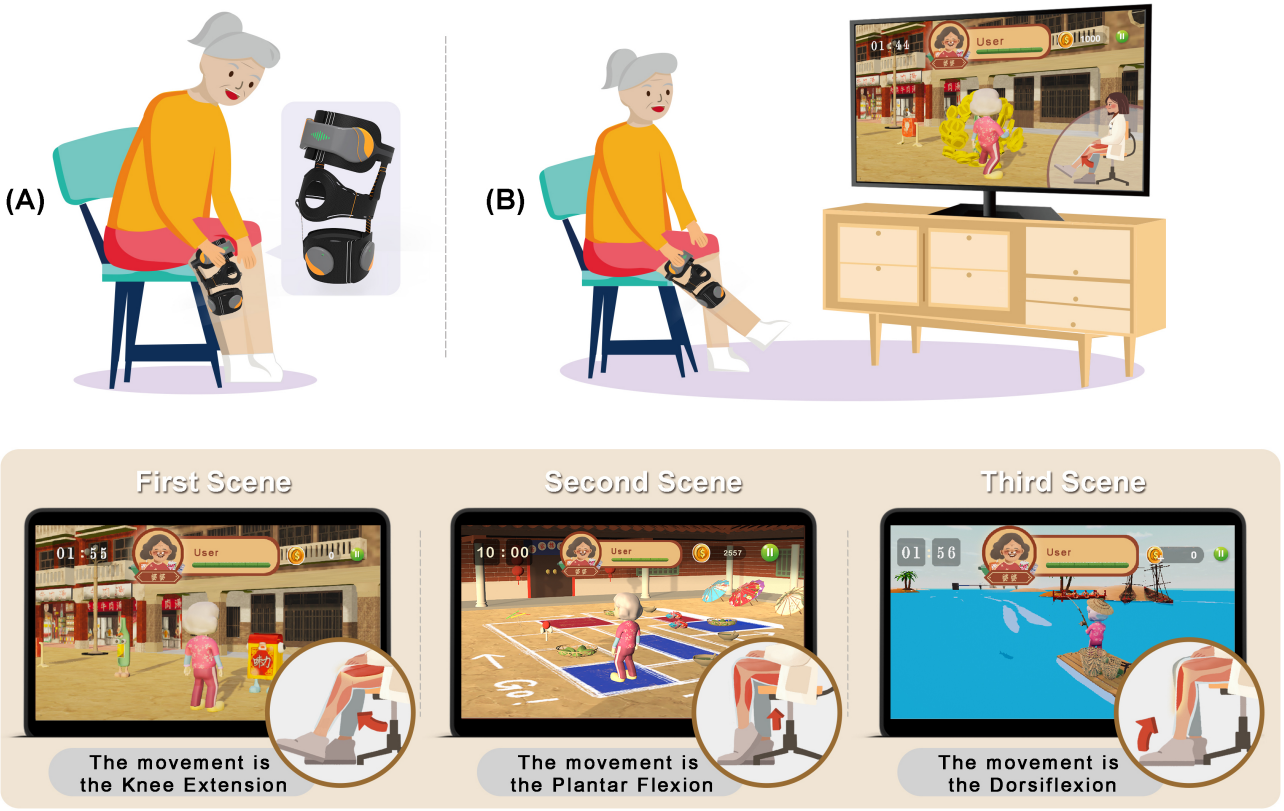


Three scenes were constructed for 3 major lower limb rehabilitation movements—knee extension (Figure 3), ankle plantar flexion, and ankle dorsiflexion for the training of the quadriceps, tibialis anterior, and gastrocnemius muscles, respectively. Participants were able to check the remaining time, remaining health points, and cumulative score on the screen and freely set the exergame to return to the main menu, retry, or continue. The virtual physician was also shown on the screen

XSL•FO
RenderX

as a coach to guide and encourage the participants to perform correct movements during training. Misplacement of the kneepad was alerted by the virtual physician (Figure S1 in Multimedia Appendix 1).

Figure 3. The operation procedure of the IRES (Intelligent Rehabilitation Exergame System): (A) frail older adults wearing the kneepad and placing the surface electromyography sensors and (B) entering the exergame system to perform 3 major rehabilitation movements (ie, knee extension, plantar flexion, and dorsiflexion).

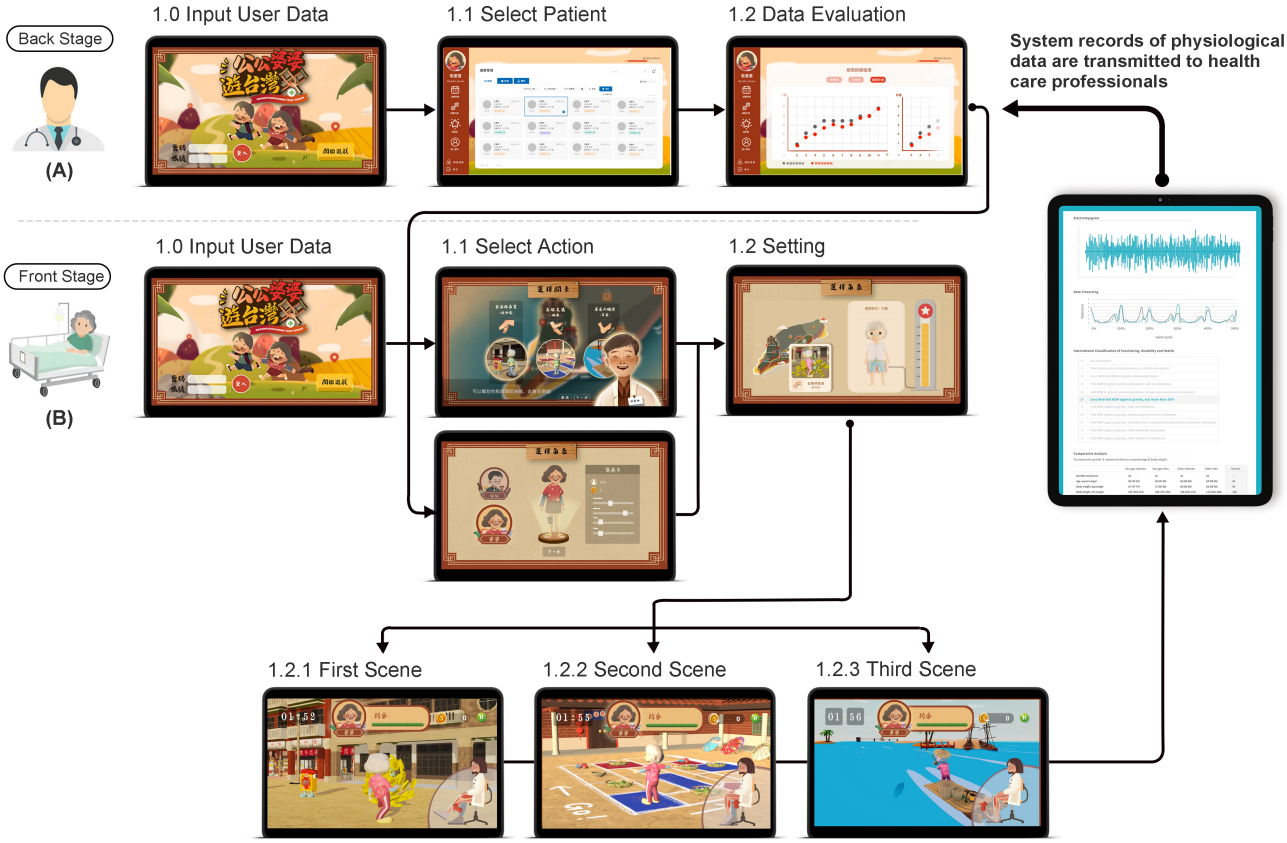


Operation of the Exergame

Before the start of the experiment, the researcher explained how the IRES operates and how to accurately wear the kneepad and calibrate the placement of the device. After kneepad placement, the participants logged in to the IRES and selected one scene for training. The first try with maximum strength was used to calibrate the baseline, and then the training began. Each movement was repeated 20 times for each leg within 30 minutes. Training was performed 2 times per week (Monday and Wednesday) for 4 weeks. The training record was analyzed for frequency, intensity, time, and type by the IRES and health care professionals to customize the content for the next training (Figure 4). The environment of the experimental setup is shown in Figure S2 in Multimedia Appendix 1.

XSL•FO
RenderX

Figure 4. Flowchart for operating the IRES (Intelligent Rehabilitation Exergame System) for (A) health care professionals and (B) frail older adults.



Participant Recruitment

This prospective cohort study was advertised in 2 community centers (Hesing Village and Liren Village, Tainan, Taiwan). Older adults over 60 years old with no mobility and communication difficulties who were willing to complete the 4-week study were enrolled as the inclusion criteria. The exclusion criteria were as follows: (1) unable to walk independently, (2) unable to understand or follow simple instructions, and (3) unwilling to complete the 4-week training.

Participant Assessment

The Clinical Frailty Scale from 1 (very fit) to 9 (terminally ill) was used to classify the health status of older adults by the physicians at the initial assessment (Table S1 in Multimedia Appendix 1) [29]. Participants assessment was performed after completing the first (week 1, baseline) and the last (week 4, one-month follow-up) training using the SUS for IRES usability [8,24] and the Taiwanese version of the EQ-5D questionnaire for quality of life [25]. The items included in the SUS are listed in Table S2 in Multimedia Appendix 1.

Ethical Considerations

This study was approved by the National Cheng Kung University Human Research Ethics Committee (NCKU HREC-F-109-497-2), and the participants provided their written informed consent to participate in this study after receiving a comprehensive explanation of the study from the investigators. All data were deidentified before analysis.

Statistical Analysis

As previously described by Chang et al [11], at a 95% power and 5% 2-tailed significance level, complete data from at least 15 participants are required to accurately detect differences in the usability analysis. We analyzed power using the G*Power 3.1 software (University of Kiel) [30]. Descriptive statistical data are reported as the mean and SD. Distributions of continuous variables were evaluated using the Shapiro-Wilk test. To assess statistically significant differences in the studied parameters, the independent samples *t* test or Mann-Whitney *U* test was used to compare continuous variables based on the results of the normality tests. The *t* test was used to verify the correlation between usability or EQ-5D and sex or previous rehabilitation experience at baseline and 1-month follow-up. Statistical significance was defined as $P<.05$.

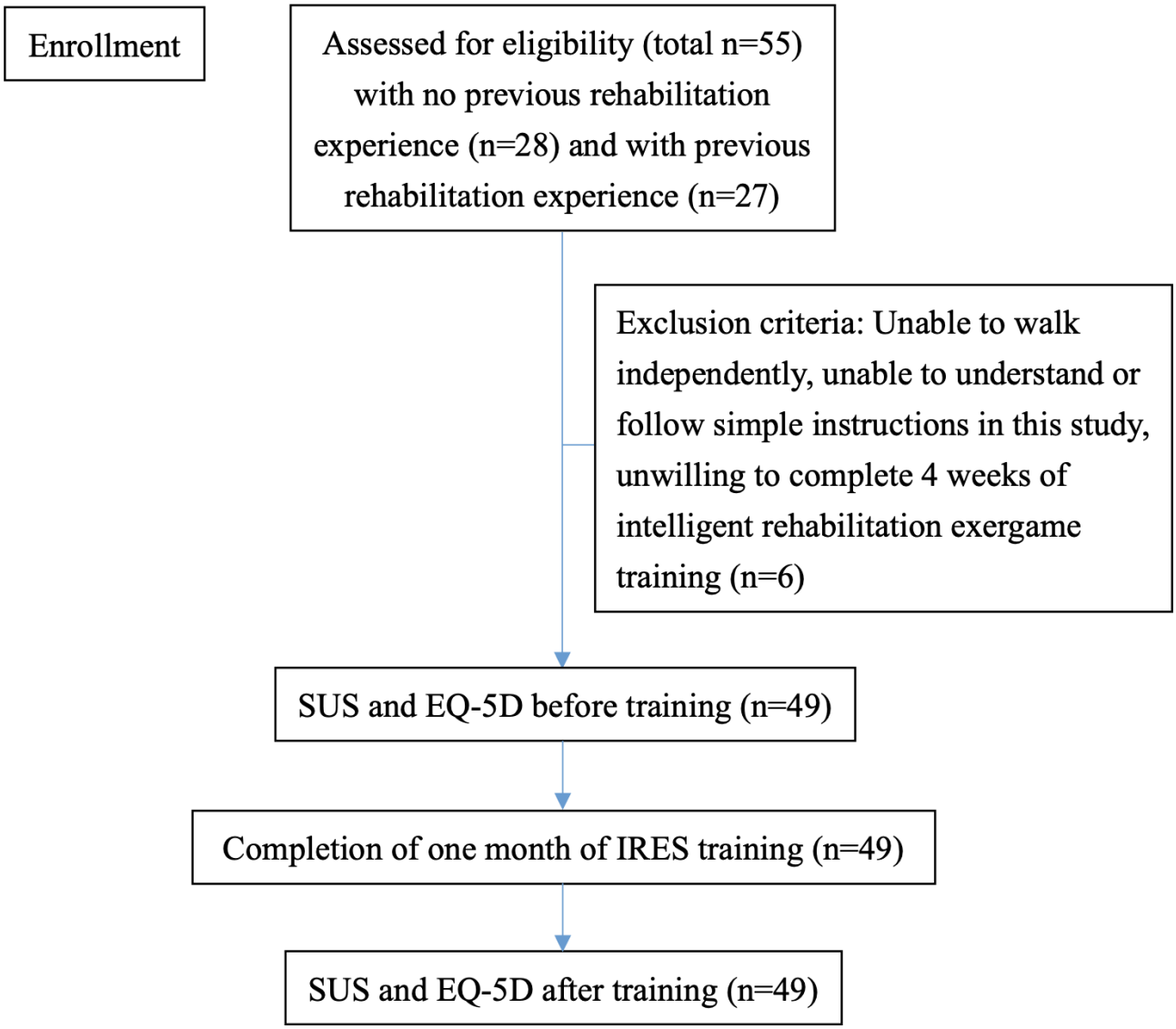
Results

Participant Demographics and Change in SUS After IRES Rehabilitation

A total of 55 older adults were recruited, including 27 with previous rehabilitation experience and 28 without. Of these participants, 6 were excluded according to the exclusion criteria, leaving 49 participants in the final cohort (Figure 5). The frailty scale scores of the 49 participants ranged from 1 to 5 (score of 1, four participants; score of 2, ten participants; score of 3, twenty-five participants; score of 4, seven participants; and score of 5, three participants). The mean age of the 49 enrolled participants was 74.6 years.

XSL•FO
RenderX

Figure 5. Flowchart for participant recruitment into the study cohort. IRES: Intelligent Rehabilitation Exergame System; SUS: System Usability Scale; EQ-5D: EuroQol-5 Dimensions.



The SUS scores for 10 selected participants at baseline and 1-month follow-up are shown in Table 1. The mean SUS score at baseline was 82.09, with a rating of "good." At 1-month follow-up, the mean SUS score increased to 87.14 with a rating of "good+." We observed a significant difference in overall usability between the baseline and the 1-month follow-up (t_{96}=−2.19; P=.03), indicating that the participants felt that the IRES was easy to operate and that its usability increased after 1 month of use.

XSL•FO
RenderX

Table . SUS[a] scores for 10 selected participants at baseline (mean 82.09, SD 10.62) and 1-month follow-up (mean 87.14, SD 12.17). A significant difference is observed in overall usability between the baseline and the 1-month follow-up (t_{96}=−2.19; P=.03).

Participant	Sex	CFS[b]	Previous rehabilitation experience	SUS scores	
				Baseline	1-month follow-up
1	Female	5	No	85	72.5
2	Male	3	No	95	80
3	Male	3	No	90	90
4	Male	3	No	92.5	87.5
5	Female	3	No	82.5	97.5
20	Female	4	No	70	95
21	Female	1	No	70	90
22	Female	2	No	87.5	100
23	Female	3	No	100	100
24	Female	4	Yes	80	90
25	Female	2	Yes	92.5	90

[a]SUS: System Usability Scale.

[b]CFS: Clinical Frailty Score.

Comparisons of the SUS score between baseline and 1-month follow-up for individual items are shown in Table 2. At baseline, only two items, item 6 (inconsistency) and item 7 (learnability), were “average”; all others achieved “good.” One month after completing the IRES training, item 6 (inconsistency) remained “average,” and the others achieved “good.” The scores increased from “average−” to “average+” for item 6 and from “average” to “good+” for item 7.

XSL•FO
RenderX

Table . SUS[a] component and total scores at baseline and 1-month follow-up stratified by previous rehabilitation experience and sex.

Item	Previous rehabilitation experience						Sex						Total (N=49)			Item Benchmark
	Yes (n=26)			No (n=23)			Male (n=12)			Female (n=37)			Mean (SD)	Max	Min	
	Mean (SD)	Max	Min	Mean (SD)	Max	Min	Mean (SD)	Max	Min	Mean (SD)	Max	Min				
1. Willingness																
Baseline	4.38 (0.69)	5	3	4.52 (0.84)	5	3	4.66 (0.65)	5	3	4.37 (0.79)	5	3	4.44 (0.76)	5	3	Good+
1-month follow-up	4.92 (0.27)	5	4	5.00 (0.00)	5	5	4.91 (0.28)	5	4	4.97 (0.16)	5	4	4.95 (0.19)	5	4	Good+
2. Complexity																
Baseline	1.84 (1.04)	4	1	1.52 (0.79)	3	1	1.33 (0.88)	4	1	1.81 (0.93)	4	1	1.69 (0.93)	4	1	Good+
1-month follow-up	1.80 (1.41)	5	1	1.43 (1.07)	5	1	1.50 (1.24)	5	1	1.67 (1.29)	5	1	1.63 (1.26)	5	1	Good+
3. Convenience																
Baseline	4.53 (0.64)	5	3	4.43 (0.84)	5	3	4.50 (0.79)	5	3	4.48 (0.73)	5	3	4.48 (0.73)	5	3	Good+
1-month follow-up	4.34 (0.89)	5	3	4.78 (0.51)	5	3	4.58 (0.79)	5	3	4.54 (0.76)	5	3	4.55 (0.76)	5	3	Good+
4. Stress																
Baseline	1.11 (0.32)	2	1	1.04 (0.20)	2	1	1.08 (0.28)	2	1	1.08 (0.27)	2	1	1.08 (0.27)	2	1	Good+
1-month follow-up	1.19 (0.80)	5	1	1.34 (1.02)	5	1	1.33 (1.15)	5	1	1.24 (0.83)	5	1	1.26 (0.90)	5	1	Good+
5. Integration																
Baseline	4.50 (0.64)	5	3	4.56 (0.72)	5	3	4.41 (0.79)	5	3	4.56 (0.64)	5	3	4.53 (0.68)	5	3	Good+
1-month follow-up	4.26 (1.28)	5	1	4.56 (0.99)	5	1	4.75 (0.62)	5	3	4.29 (1.26)	5	1	4.41 (1.15)	5	1	Good+
6. Inconsistency																
Baseline	2.03 (1.31)	5	1	2.82 (1.52)	5	1	2.25 (1.35)	5	1	2.45 (1.50)	5	1	2.40 (1.45)	5	1	Average–
1-month follow-up	2.07 (1.29)	5	1	2.08 (1.20)	5	1	2.16 (1.19)	5	1	2.05 (1.26)	5	1	2.08 (1.23)	5	1	Average+
7. Learnability																
Baseline	3.57 (1.33)	5	1	3.86 (1.32)	5	1	3.83 (1.26)	5	2	3.67 (1.35)	5	1	3.71 (1.32)	5	1	Average
1-month follow-up	4.81 (0.49)	5	3	4.69 (0.92)	5	1	4.66 (1.15)	5	1	4.78 (0.53)	5	3	4.75 (0.72)	5	1	Good+
8. Cumbersomeness																
Baseline	1.69 (0.92)	4	1	1.52 (1.20)	5	1	1.33 (0.88)	4	1	1.70 (1.10)	5	1	1.61 (1.05)	5	1	Good
1-month follow-up	2.11 (1.68)	5	1	1.82 (1.40)	5	1	2.41 (1.83)	5	1	1.83 (1.44)	5	1	1.97 (1.54)	5	1	Good–
9. Confidence																

XSL•FO
RenderX

Item	Previous rehabilitation experience						Sex						Total (N=49)			Item Benchmark
	Yes (n=26)			No (n=23)			Male (n=12)			Female (n=37)			Mean (SD)	Max	Min	
	Mean (SD)	Max	Min	Mean (SD)	Max	Min	Mean (SD)	Max	Min	Mean (SD)	Max	Min				
Baseline	3.96 (1.03)	5	2	4.30 (1.06)	5	1	4.08 (1.24)	5	1	4.13 (1.00)	5	2	4.12 (1.05)	5	1	Good–
1-month follow-up	4.26 (1.11)	5	1	4.91 (0.28)	5	4	4.50 (1.24)	5	1	4.59 (0.76)	5	2	4.57 (0.88)	5	1	Good+
10. Difficulty																
Baseline	1.80 (1.13)	5	1	1.56 (0.89)	4	1	1.83 (1.19)	4	1	1.64 (0.97)	5	1	1.69 (1.02)	5	1	Good
1-month follow-up	1.65 (1.19)	5	1	1.26 (0.75)	4	1	1.66 (1.30)	5	1	1.41 (0.92)	5	1	1.46 (1.02)	5	1	Good+

[a]SUS: System Usability Scale.

The Effect of Sex and Rehabilitation History on SUS

The effect of sex on differences in SUS scores between baseline and 1-month follow-up was evaluated (Table 3). No significant differences were observed between the sexes for any of the 10 items evaluated (P>.05), indicating that the usability of the IRES was consistent between male and female participants.

Table . Comparison of SUS[a] scores before and after training according to sex (N=49; 12 men and 37 women).

Item	Baseline		1-month follow-up	
	t test (*df*)	*P* value	*t* test (*df*)	*P* value
1. Willingness	−1.13 (47)	.26	0.84 (47)	.4
2. Complexity	1.55 (47)	.12	0.41 (47)	.68
3. Convenience	−0.05 (47)	.95	−0.16 (47)	.86
4. Stress	−0.02 (47)	.98	−0.29 (47)	.76
5. Integration	0.66 (47)	.5	−1.18 (47)	.24
6. Inconsistency	0.42 (47)	.66	−0.27 (47)	.78
7. Learnability	−0.35 (47)	.72	0.48 (47)	.63
8. Cumbersomeness	1.05 (47)	.29	−1.12 (47)	.26
9. Confidence	0.14 (47)	.88	0.31 (47)	.75
10. Difficulty	−0.53 (47)	.59	−0.76 (47)	.44

[a]SUS: System Usability Scale.

Whether previous rehabilitation experience affected the change in SUS scores between baseline and 1-month follow-up was evaluated (Table 4). At baseline, no significant differences were found between participants with and without rehabilitation experience for any of the 10 items (P>.05), indicating that experience does not affect usability. A significant difference between baseline and follow-up was observed for item 3 (convenience: t_{47}=2.05; P=.04) and item 9 (confidence: t_{47}=2.68; P<.001), indicating that IRES use can increase convenience and confidence in older adults without previous rehabilitation experience.

XSL•FO
RenderX

Table . SUS[a] component scores before and after training according to previous rehabilitation experience (N=49; yes=26, no=23).

Item	Baseline		1-month follow-up	
	t test (*df*)	*P* value	*t* test (*df*)	*P* value
1. Willingness	0.62 (47)	.53	1.35 (47)	.18
2. Complexity	−1.21 (47)	.23	−1.02 (47)	.31
3. Convenience	−0.48 (47)	.62	2.05 (47)	.04[b]
4. Stress	−0.90 (47)	.36	0.59 (47)	.55
5. Integration	0.33 (47)	.74	0.89 (47)	.37
6. Inconsistency	1.94 (47)	.06	0.02 (47)	.97
7. Learnability	0.76 (47)	.44	−0.53 (47)	.59
8. Cumbersomeness	−0.55 (47)	.57	−0.64 (47)	.51
9. Confidence	1.14 (47)	.25	2.68 (47)	<.001[b]
10. Difficulty	−0.82 (47)	.41	−1.35 (47)	.18

[a]SUS: System Usability Scale.

[b]Significant difference between groups.

The effect of previous rehabilitation experience on the difference between baseline and 1-month follow-up on each component of the SUS score was analyzed. A significant difference was observed for item 1 (willingness: t_{96}=−4.51; *P*<.001), item 7 (learnability: t_{96}=−4.83; *P*<.001), and item 9 (confidence: t_{96}=−2.27; *P*=.02). Among older adults with previous rehabilitation experience, we observed significant differences in item 1 (willingness: t_{50}=−3.66; *P*<.001) and item 7 (learnability: t_{50}=−4.42; *P*<.001). Among older adults without previous rehabilitation experience, significant differences were observed in item 1 (willingness: t_{44}=−2.71; *P*<.001), item 7 (learnability: t_{44}=−2.45; *P*=.01), and item 9 (confidence: t_{44}=−2.65; *P*=.021) (Table 5).

Table . Changes in SUS[a] component scores between baseline and 1-month follow-up according to previous rehabilitation experience.

Item	Baseline vs 1-month follow-up					
	Total (N=49)		Previous rehabilitation experience (n=26)		No rehabilitation experience (n=23)	
	t test (*df*)	*P* value	*t* test (*df*)	*P* value	*t* test (*df*)	*P* value
1. Willingness	−4.51 (96)	<.001[b]	−3.66 (50)	<.001[b]	−2.71 (44)	<.001[b]
2. Complexity	0.27 (96)	.78	0.11 (50)	.91	0.31 (44)	.75
3. Convenience	−0.40 (96)	.68	0.89 (50)	.37	−1.68 (44)	.09
4. Stress	−1.35 (96)	.17	−0.45 (50)	.65	−1.39 (44)	.17
5. Integration	0.64 (96)	.52	0.81 (50)	.41	0.00 (44)	>.99
6. Inconsistency	1.19 (96)	.23	−0.11 (50)	.91	1.82 (44)	.07
7. Learnability	−4.83 (96)	<.001[b]	−4.42 (50)	<.001[b]	−2.45 (44)	.01[b]
8. Cumbersome-ness	−1.37 (96)	.17	−1.12 (50)	.26	−0.79 (44)	.43
9. Confidence	−2.27 (96)	.02[b]	−1.02 (50)	.3	−2.65 (44)	.01[b]
10. Difficulty	1.08 (96)	.28	0.47 (50)	.63	1.24 (44)	.21

[a]SUS: System Usability Scale.

[b]Significant difference between baseline and 1-month follow-up.

The Change in Quality of Life After IRES Rehabilitation

The change in participant quality of life between baseline and 1-month follow-up was evaluated using the Taiwanese version of the EQ-5D questionnaire (Table 6). For the total cohort, all males and all females quality of life improved significantly after the IRES training program (all *P*<.05), indicating a positive effect on the quality of life of older adults regardless of sex.

XSL•FO
RenderX

Table . Differences in the EQ-5D[a] between baseline and 1-month follow-up according to sex.

EQ-5D	Total (N=49)		Male (n=12)		Female (n=37)	
	t test (*df*)	*P* value	*t* test (*df*)	*P* value	*t* test (*df*)	*P* value
Baseline	0.64 (96)	0.12	0.65 (22)	0.13	0.64 (72)	0.12
1-Month Follow-Up	0.74 (96)	0.1	0.74 (22)	0.11	0.74 (72)	0.1
Total	6.03 (96)	<.001[b]	2.47 (22)	0.03[b]	5.49 (72)	<.001[b]

[a]EQ-5D: EuroQol-5 Dimensions.

[b]Significant difference between baseline and 1-month follow-up.

Discussion

Principal Findings

We examined the usability of IRES and its effects on the quality of life of frail older adults. The results indicate that IRES increased the willingness of frail older adults to undergo continuous long-term rehabilitation training. The learnability and confidence (ie, self-affirmation) were also higher after 1 month of training compared to those at the baseline. This result indicates that IRES that combines gamified rehabilitation content and movements instructed by a virtual coach can increase the user's ability to operate and understand the system, thereby decreasing their uncertainty and increasing their confidence in using this novel technology product. In addition, no gender differences were found in the SUS scores between baseline and 1-month follow-up, indicating that both male and female older adults had consistent subjective willingness to use the IRES for the duration of the training. A significant difference was observed in the quality of life between baseline and 1 month after IRES training as assessed by the EQ-5D, further indicating that the quality of life of the participants improved, regardless of their sex.

The Evolution and Current Trends of Digital Rehabilitation Aids

Rehabilitation aids that combine exercise and games are effective training methods for alleviating frailty and maintaining the health and quality of life of older adults. Such aids also encourage older adults to participate in long-term rehabilitation training, effectively improving physical and mental function while decreasing societal health care burdens [18]. With the maturation and popularization of digital technology, digital games have been used in emerging cognitive training models. Such games, known as serious games, differ from normal entertainment games, as they are purposeful and beneficial to the players. Serious games can encourage older adults to exercise and then increase their physical activity level, further improving muscle strength and cardiopulmonary fitness. Serious games also overcome the low adherence that is generally observed in traditional rehabilitation training, thereby improving the motivation for long-term training [31]. By improving mental focus and increasing muscle mass, exergames can increase the balance, gait, and activity capacity of older adults. Moreover, the integration feature of exergames for older adults provides psychological health benefits [32-34]. The benefit of IRES on the self-esteem and quality of life of the older adults observed in this study likely results from the exergame-based training.

Indeed, information and communications technology and auxiliary life services can effectively improve the life satisfaction of older adults and may increase their willingness to use technology products or services. Considering the larger health care needs of this population, their travel limitations, and health care personnel shortages, intelligent telemedicine care models should be integrated into current health care systems as new tools or services to maintain patient health and prevent disease [35,36]. The following 3 major features need to be incorporated into future health care systems: (1) information and data platform improvements, including the collection and analysis of patient data, immediate provision of individual assistance, and data sharing among medical professionals; (2) health and care services, including the use of virtual and physical communities to provide customized products and care services for patients; and (3) care service support, including supporting and assisting patient care services and promoting health and welfare [36,37]. The appropriate incorporation of intelligent telemedicine can promote the reformation of traditional health care. In the UK National Health Service, the long-term plan is to make intelligent medicine a mainstream practice [38]. However, customized training modules and diverse service content are critical factors proven to affect user acceptance of intelligent telemedicine systems for both patients and health care professionals [39].

Features and Advantages of the IRES Program Design

The IRES uses gamified training content to increase the pleasure of frail older adults during training, thereby increasing their motivation to continue long-term rehabilitation. In addition, the IRES effectively records, transmits, and analyzes data regarding lower limb muscle movement by sEMG. sEMG measures changes in muscle potential signals that occur when muscles in different motor units are stimulated by motor neurons during limb exercise. The contraction of individual muscles during movements and their potential differences were used to indirectly evaluate individual muscle strength [28]. sEMG is a clinical method widely used in human-machine interactions, hand gesture recognition, and muscle function evaluation. This technology provides data that can be used to evaluate muscle activity and joint movement [26]. These data can be reviewed and studied by health care professionals to evaluate and adjust exercise regimens to provide the most suitable customized rehabilitation treatment for frail older adults.

XSL•FO
RenderX

This feature is another advantage of the IRES over other exergames that are only designed for patients; IRES serves and provides benefits for both patients and health care professionals. Health care professionals could use IRES for continuous interaction and telemedicine services with patients in the post-COVID-19 era [37,40]. Optimization may increase older adults' acceptance and willingness to continue using this novel product and service, as reported in previous studies [2,6,40]. In the future, IRES-based muscle strength training modules and systems should be developed in medical institutions to help manage health conditions common in frail older adults.

Reminiscence therapy is a people-centered nursing activity that uses patient memories, related past experiences, or specific events [41,42]. Wu et al [43] used old photographs for reminiscence therapy in older adults and found that the photographs significantly evoked autobiographical memories and semantic memory in older adults. Furthermore, reminiscence therapy as a nursing activity improves the sense of self-worth of patients, especially in older adults [44]. Recollecting meaningful people, events, and objects and guiding older adults to discuss past experiences helps them to find meaning in their lives, improves their self-esteem and confidence, alleviates depression, and improves their overall quality of life [45]. Old public photographs are a good format for use in the early care phase of older adults [43]. Integrating these historical images into exergame design can help prevent dementia, apart from providing entertainment. Tsao et al [46] integrated virtual and augmented reality to stimulate memories in older adults while immersing them in retro videos and music environments. These effects are also seen in the results of this study.

Reception and Effects of the IRES Program

We used the SUS and EQ-5D to assess the effects of IRES. The SUS is a widely used tool for evaluating the usability of products and systems [24]. After using a product, users are asked to rate their agreement with a series of statements, which are then used to calculate a usability score. The SUS is helpful for developers looking to quickly assess the usability of their product [47]. To correlate clinical information with specific usability criteria in this study, the EQ-5D was used to evaluate older adults' quality of life. This assessment provides a use value based on the 5D health state classifications: mobility, self-care, usual activities, pain or discomfort, and anxiety or depression. The SUS and EQ-5D questionnaire results revealed that using a retro style in the exergame design made it very interesting and invoked memories and shared experiences among older adults. This finding is consistent with the results of related studies [11,43,46]. A strong correlation should exist between clinical improvement and the willingness to continue using the rehabilitation system.

A semi-structured interview was also used to gather the participants' thoughts, willingness to use the IRES, and problems encountered when using it. The results indicate that the retro-style content resonated with older adults, thus improving their willingness to continue using the system. In other words, integrating retro-style content into rehabilitation training could increase the rehabilitation motivation of older adults. The participants reported positive feedback for the IRES system. The functions and treatment course arrangements did not cause adverse events. The IRES system may be suitable for general use among older adults. The human-machine interface design of intuitive interaction and the instructions provided by the virtual coach enabled the system to have better learnability and operability, thus increasing the willingness of older adults to continue rehabilitation.

Limitations

First, the sample size in this study is relatively small, which limits the generalizability of the results. Further studies using a larger sample size are needed to comprehensively evaluate the usability of the IRES system training in frail older adults. Second, this study has no direct physical function measurement for older adults. In future studies, we would evaluate additional clinical measures such as the strength of the muscle groups used and well-established scales like gait speed, and the timed up-and-go test to investigate the effect on physical function comprehensively. Third, a 1-month training and follow-up may need to be extended in order to achieve greater functional improvement. However, we found a significant improvement in the quality of life for older adults after IRES use. Further long-term studies may be needed to evaluate the long-term effects of IRES system training on frail older adults. Lastly, despite the longitudinal data presented here, the absence of a control group limited a direct comparison to no treatment and quantification of the benefits of IRES, particularly in mitigating physical or cognitive decline over time. The inclusion of a control group should be considered in future long-term studies to validate and extend our findings.

Conclusion

Population aging and the new normal brought about by the COVID-19 pandemic have increased the importance of frailty and sarcopenia prevention in older adults. Exercise effectively improves fitness, and integrating retro-style exergame training encourages older adults to participate in long-term rehabilitation training to prevent these conditions and decrease the burden on clinics. The results of this study show significant improvements in user willingness and learnability after a 1-month training using IRES. Furthermore, IRES use significantly improved confidence in the use of technology products in frail older adults without previous rehabilitation experience. Overall, IRES not only increased the rehabilitation motivation of frail older adults but also increased their willingness to continue rehabilitation.

Acknowledgments

This study was supported by the Ministry of Education (MOE) Teaching Practice Research Program titled "Post-Pandemic Curriculum-Redesign for Flipped Learning in Rehabilitation Medicine: Introducing 'AI-Based Functional Mobility Assessment'

XSL•FO
RenderX

Technology Into the Training Curriculum to Improve the Mobility Assessment Skills of Medical Students" (grant PMN1120522) and the National Science and Technology Council (grants NSTC 111-2314-B-006-064-MY3 and NSTC 111-2923-E-006 -002 -MY3, for a 3-year project). The financial sponsors played no role in the study design and execution, the analysis and interpretation of data, or in writing the manuscript. WCL and YCL are co-corresponding authors (WCL can be reached at lwclwhab@ms8.hinet.net for correspondence).

Data Availability

The datasets generated during and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: CHC, CCW, WCL, YCL
Data curation: CHC
Formal analysis: CHC, CCW, WCL, YCL
Funding acquisition: CCW, WCL, YCL
Investigation: CHC, CCW, THY, WCL, BL, YL, PTT, YCL
Methodology: CHC, CCW, WCL, YCL
Project administration: CHC, CCW, WCL, YCL
Resources: CCW, WCL, YCL
Software: CHC, CCW, WCL, YCL
Supervision: CCW, WCL, YCL
Validation: CHC, CCW, WCL, YCL
Visualization: CHC, CCW, WCL, BL, YCL
Writing – original draft: CHC, CCW, THY, WCL, YCL
Writing – review & editing: CHC, CCW, THY, WCL, YCL

Conflicts of Interest

None declared.

Multimedia Appendix 1
The illustration of IRES and CFS. IRES: Intelligent Rehabilitation Exergame System; CFS: Clinical Frailty Score.
[DOCX File, 280 KB - games_v13i1e50669_app1.docx]

References

1. Chang CC, Wu CS, Tseng HY, et al. Assessment of incident frailty hazard associated with depressive symptoms in a Taiwanese longitudinal study. Int Psychogeriatr 2022 Jan;34(1):61-70. [doi: 10.1017/S1041610221000806] [Medline: 34275507]
2. Gabrovec B, Antoniadou E, Soleymani D, et al. Need for comprehensive management of frailty at an individual level: European perspective from the advantage joint action on frailty. J Rehabil Med 2020 Jun 30;52(6):jrm00075. [doi: 10.2340/16501977-2687] [Medline: 32399576]
3. O'Caoimh R, Sezgin D, O'Donovan MR, et al. Prevalence of frailty in 62 countries across the world: a systematic review and meta-analysis of population-level studies. Age Ageing 2021 Jan 8;50(1):96-104. [doi: 10.1093/ageing/afaa219] [Medline: 33068107]
4. Rohrmann S. Epidemiology of frailty in older people. In: Veronese N, editor. Frailty and Cardiovascular Diseases: Springer; 2020, Vol. 1216:21-27. [doi: 10.1007/978-3-030-33330-0_3]
5. Angulo J, El Assar M, Álvarez-Bustos A, Rodríguez-Mañas L. Physical activity and exercise: strategies to manage frailty. Redox Biol 2020 Aug;35:101513. [doi: 10.1016/j.redox.2020.101513] [Medline: 32234291]
6. Cruz-Jentoft AJ, Bahat G, Bauer J, et al. Sarcopenia: revised European consensus on definition and diagnosis. Age Ageing 2019 Jan 1;48(1):16-31. [doi: 10.1093/ageing/afy169] [Medline: 30312372]
7. Kang MG, Kang SJ, Roh HK, et al. Accuracy and diversity of wearable device-based gait speed measurement among older men: observational study. J Med Internet Res 2021 Oct 11;23(10):e29884. [doi: 10.2196/29884] [Medline: 34633293]
8. Chang CH, Liu KH, Kajihara H, et al. Designing a somatosensory interactive game of lower extremity muscle rehabilitation for the elderly. In: Gao Q, Zhou J, editors. Human Aspects of IT for the Aged Population: Supporting Everyday Life Activities: Springer; 2021:23-27. [doi: 10.1007/978-3-030-78111-8_2]
9. Chen PJ, Chang CH, Kuo YL, Lin YC. Designing and evaluating a wearable sEMG device for the elderly. Presented at: 14th International Conference on ICT, Society, and Human Beings, ICT 2021, 18th International Conference on Web Based Communities and Social Media, WBC 2021 and 13th International Conference on e-Health, EH 2021 - Held at the 15th Multi-Conference on Computer Science and Information Systems, MCCSIS 2021; Jul 20-23, 2021.

XSL•FO
RenderX

10. New global medical crisis—medical labor shortage: global health human resources crisis and response strategies. KPMG. 2019 Jul 16. URL: https://kpmg.com/tw/zh/home/insights/2019/07/kpmg-tw-2019-human-solving-the-global-workforce-crisis-in-healthcare.html [accessed 2024-11-27]
11. Chang CH, Yeh CH, Chang CC, Lin YC. Interactive somatosensory games in rehabilitation training for older adults with mild cognitive impairment: usability study. JMIR Serious Games 2022 Jul 14;10(3):e38465. [doi: 10.2196/38465] [Medline: 35834303]
12. Lu L, Zhang J, Xie Y, et al. Wearable health devices in health care: narrative systematic review. JMIR Mhealth Uhealth 2020 Nov 9;8(11):e18907. [doi: 10.2196/18907] [Medline: 33164904]
13. Zhang Q, Guo X, Vogel D. Information and communication technology use for life satisfaction among the elderly: a motivation perspective. Am J Health Behav 2021 Jul 26;45(4):701-710. [doi: 10.5993/AJHB.45.4.9] [Medline: 34340737]
14. Ong DSM, Weibin MZ, Vallabhajosyula R. Serious games as rehabilitation tools in neurological conditions: a comprehensive review. Technol Health Care 2021;29(1):15-31. [doi: 10.3233/THC-202333] [Medline: 32804107]
15. Gomes GCV, Simões MDS, Lin SM, et al. Feasibility, safety, acceptability, and functional outcomes of playing Nintendo Wii Fit PlusTM for frail older adults: a randomized feasibility clinical trial. Maturitas 2018 Dec;118:20-28. [doi: 10.1016/j.maturitas.2018.10.002] [Medline: 30415751]
16. Mollà-Casanova S, Muñoz-Gómez E, Sempere-Rubio N, et al. Effect of virtual running with exercise on functionality in pre-frail and frail elderly people: randomized clinical trial. Aging Clin Exp Res 2023 Jul;35(7):1459-1467. [doi: 10.1007/s40520-023-02414-x] [Medline: 37188994]
17. Liu YH, Liu JC, Lin MT, Chen WC. Participation of senior citizens in somatosensory games: a correlation between the willingness to exercise and happiness. J Ambient Intell Human Comput 2020. [doi: 10.1007/s12652-020-01918-y]
18. Gueye T, Dedkova M, Rogalewicz V, Grunerova-Lippertova M, Angerova Y. Early post-stroke rehabilitation for upper limb motor function using virtual reality and exoskeleton: equally efficient in older patients. Neurol Neurochir Pol 2021;55(1):91-96. [doi: 10.5603/PJNNS.a2020.0096] [Medline: 33314016]
19. Lu Z, Wang W, Yan W, Kew CL, Seo JH, Ory M. The application of fully immersive virtual reality on reminiscence interventions for older adults: scoping review. JMIR Serious Games 2023 Oct 6;11:e45539. [doi: 10.2196/45539] [Medline: 37801360]
20. Lök N, Bademli K, Selçuk-Tosun A. The effect of reminiscence therapy on cognitive functions, depression, and quality of life in Alzheimer patients: randomized controlled trial. Int J Geriatr Psychiatry 2019 Jan;34(1):47-53. [doi: 10.1002/gps.4980] [Medline: 30246408]
21. Tam W, Poon SN, Mahendran R, Kua EH, Wu XV. The effectiveness of reminiscence-based intervention on improving psychological well-being in cognitively intact older adults: s systematic review and meta-analysis. Int J Nurs Stud 2021 Feb;114:103847. [doi: 10.1016/j.ijnurstu.2020.103847] [Medline: 33352435]
22. Yamada M, Kimura Y, Ishiyama D, et al. Effect of the COVID-19 epidemic on physical activity in community-dwelling older adults in Japan: a cross-sectional online survey. J Nutr Health Aging 2020;24(9):948-950. [doi: 10.1007/s12603-020-1424-2] [Medline: 33155619]
23. Al-Ayyad M, Owida HA, De Fazio R, Al-Naami B, Visconti P. Electromyography monitoring systems in rehabilitation: a review of clinical applications, wearable devices and signal acquisition methodologies. Electronics (Basel) 2023;12(7):1520. [doi: 10.3390/electronics12071520]
24. Brooke J. SUS—a quick and dirty usability scale. In: Usability Evaluation in Industry: CRC Press; 1996:189-194.
25. Lee HY, Hung MC, Hu FC, Chang YY, Hsieh CL, Wang JD. Estimating quality weights for EQ-5D (EuroQol-5 dimensions) health states with the time trade-off method in Taiwan. J Formos Med Assoc 2013 Nov;112(11):699-706. [doi: 10.1016/j.jfma.2012.12.015] [Medline: 24183199]
26. Wang C, Cai M, Hao Z, et al. Stretchable, multifunctional epidermal sensor patch for surface electromyography and strain measurements. Adv Intell Syst 2021 Nov;3(11):2100031. [doi: 10.1002/aisy.202100031] [Medline: 35663639]
27. Chang CH, Lien WC, Chiu TP, et al. A novel smart somatosensory wearable assistive device for older adults' home rehabilitation during the COVID-19 pandemic. Front Public Health 2023;11:1026662. [doi: 10.3389/fpubh.2023.1026662] [Medline: 37790724]
28. Kawase T, Tagami T, Miyazaki T, Kanno T, Kawashima K. Effect of electromyography signals on single joint motion forecasting. Sens Mater 2020;32(3):967-980. [doi: 10.18494/SAM.2020.2654]
29. Rockwood K, Song X, MacKnight C, et al. A global clinical measure of fitness and frailty in elderly people. CMAJ 2005 Aug 30;173(5):489-495. [doi: 10.1503/cmaj.050051] [Medline: 16129869]
30. Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007 May;39(2):175-191. [doi: 10.3758/bf03193146] [Medline: 17695343]
31. Moizer J, Lean J, Dell'Aquila E, et al. An approach to evaluating the user experience of serious games. Comput Educ 2019 Jul;136:141-151. [doi: 10.1016/j.compedu.2019.04.006]
32. Santamaría KG, Salicetti Fonseca A, Jiménez JM, Mora LCS. Balance, attention and concentration improvements following an exergame training program in elderly [Article in Spanish]. Retos 2018;33:102-105. [doi: 10.47197/retos.v0i33.43574]
33. Zheng L, Li G, Wang X, et al. Effect of exergames on physical outcomes in frail elderly: a systematic review. Aging Clin Exp Res 2020 Nov;32(11):2187-2200. [doi: 10.1007/s40520-019-01344-x]

XSL•FO
RenderX

34. Xu W, Liang HN, Baghaei N, Wu Berberich B, Yue Y. Health benefits of digital videogames for the aging population: a systematic review. Games Health J 2020 Dec 1;9(6):389-404. [doi: 10.1089/g4h.2019.0130]
35. Spadaro B, Martin-Key NA, Bahn S. Building the digital mental health ecosystem: opportunities and challenges for mobile health innovators. J Med Internet Res 2021 Oct 13;23(10):e27507. [doi: 10.2196/27507] [Medline: 34643537]
36. Chang KV, Chen KH, Chen YH, et al. A multicenter study to compare the effectiveness of the inpatient post acute care program versus traditional rehabilitation for stroke survivors. Sci Rep 2022 Jul 27;12(1):12811. [doi: 10.1038/s41598-022-16984-9] [Medline: 35896786]
37. 2020 global health care outlook: laying a foundation for the future. Deloitte Insights. 2020. URL: https://www2.deloitte.com/content/dam/Deloitte/global/Images/infographics/lifesciences-healthcare/gx-lshc-2020-global-healthcare-infographic.pdf [accessed 2024-12-23]
38. Dorsey ER, Topol EJ. Telemedicine 2020 and the next decade. Lancet 2020 Mar 14;395(10227):859. [doi: 10.1016/S0140-6736(20)30424-4] [Medline: 32171399]
39. Liao X, Wu D, Zhang Q, Han G. How to improve users' loyalty to smart health devices? The perspective of compatibility. Sustainability 2021;13(19):10722. [doi: 10.3390/su131910722]
40. Hong Z, Li N, Li D, et al. Telemedicine during the COVID-19 pandemic: experiences from western China. J Med Internet Res 2020 May 8;22(5):e19577. [doi: 10.2196/19577] [Medline: 32349962]
41. Woods RT, Bruce E, Edwards RT, et al. REMCARE: reminiscence groups for people with dementia and their family caregivers: effectiveness and cost-effectiveness pragmatic multicentre randomised trial. Health Technol Assess 2012;16(48):v-xv. [doi: 10.3310/hta16480] [Medline: 23211271]
42. Grilli MD, Wank AA, Bercel JJ, Ryan L. Evidence for reduced autobiographical memory episodic specificity in cognitively normal middle-aged and older individuals at increased risk for Alzheimer's disease dementia. J Int Neuropsychol Soc 2018 Nov;24(10):1073-1083. [doi: 10.1017/S1355617718000577] [Medline: 30136918]
43. Wu D, Chen T, Huang X, et al. The role of old photos in reminiscence therapy in elderly women with depressive symptoms: a functional magnetic resonance imaging study. Biol Res Nurs 2020 Apr;22(2):234-246. [doi: 10.1177/1099800420908002] [Medline: 32174161]
44. Kaplan T, Keser İ. The effect of individual reminiscence therapy on adaptation difficulties of the elderly: a randomized clinical trial. Psychogeriatrics 2021 Nov;21(6):869-880. [doi: 10.1111/psyg.12761] [Medline: 34530495]
45. Abu Khait A, Shellman J. Reminiscence in the care of people with mild dementia: a concept analysis. Nurs Forum 2020 Apr;55(2):282-293. [doi: 10.1111/nuf.12427] [Medline: 31919849]
46. Tsao YC, Shu CC, Lan TS. Development of a reminiscence therapy system for the elderly using the integration of virtual reality and augmented reality. Sustainability 2019 Sep;11(17):4792. [doi: 10.3390/su11174792]
47. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. Int J Hum Comput Interact 2008 Jul 29;24(6):574-594. [doi: 10.1080/10447310802205776]

ABBREVIATIONS

EQ-5D: EuroQol-5 Dimensions
IRES: Intelligent Rehabilitation Exergame System
sEMG: surface electromyography
SUS: System Usability Scale

Edited by A Coristine, T Leung; submitted 15.07.23; peer-reviewed by F Noveletto, M Jordan-Marsh; revised version received 12.10.24; accepted 16.10.24; published 21.01.25.

Please cite as:
Chang CH, Wei CC, Lien WC, Yang TH, Liu B, Lin Y, Tan PT, Lin YC
The Usability and Effect of a Novel Intelligent Rehabilitation Exergame System on Quality of Life in Frail Older Adults: Prospective Cohort Study
JMIR Serious Games 2025;13:e50669
URL: https://games.jmir.org/2025/1/e50669
doi:10.2196/50669

© Chien-Hsiang Chang, Chun-Chun Wei, Wei-Chih Lien, Tai-Hua Yang, Bo Liu, Yimo Lin, Poh Thong Tan, Yang-Cheng Lin. Originally published in JMIR Serious Games (https://games.jmir.org), 21.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly

XSL•FO
RenderX

cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Exploring Virtual Reality and Exercise Simulator Interventions in Patients With Attention Deficit Hyperactivity Disorder: Comprehensive Literature Review

Gurdeep Sarai[1*], BA; Prem Prakash Jayaraman[2*], PhD; Oren Tirosh[3*], PhD; Nilmini Wickramasinghe[4*], MBA, PhD

[1]School of Health Sciences, Swinburne University of Technology, Melbourne, Australia

[2]School of Science, Computing and Engineering Technologies, Factory of the Future and Digital Innovation Lab, Swinburne University of Technology, Melbourne, Australia

[3]Department of Health and Biomedical Sciences, Clinical Biomechanics and Rehabilitation, RMIT University, Melbourne, Australia

[4]School of Computing, Engineering and Mathematical Sciences, Optus Chair Digital Health, La Trobe University, Melbourne, Australia

[*]all authors contributed equally

Corresponding Author:
Gurdeep Sarai, BA
School of Health Sciences
Swinburne University of Technology
1 Almeida Cresent
Melbourne, 3141
Australia
Phone: 61 411362177
Email: gurdeep.sarai@gmail.com

Abstract

Background: This review explores virtual reality (VR) and exercise simulator–based interventions for individuals with attention-deficit/hyperactivity disorder (ADHD). Past research indicates that both VR and simulator-based interventions enhance cognitive functions, such as executive function and memory, though their impacts on attention vary.

Objective: This study aimed to contribute to the ongoing scientific discourse on integrating technology-driven interventions into the management and evaluation of ADHD. It specifically seeks to consolidate findings on how VR and exercise simulators may support individuals with ADHD, acknowledging associated challenges and implications inherent in both technological approaches.

Methods: This research looks at existing literature to examine the potential efficacy of VR and exercise simulator–based interventions for individuals with ADHD. It evaluates the capacity of these interventions to address specific challenges along with an emphasis on the adjustments for accommodating unique user behaviors. Additionally, it underscores the limited exploration of user perceptions of exercise simulator–based interventions and the undervalued role of motor function in both ADHD assessment and symptom management.

Results: The findings of this scoping review reveal that, while these interventions enhance user motivation and enjoyment, certain challenges resist modification through technology. Furthermore, this study explores the intricate complexities involved in customizing these technologies to accommodate the diverse aspects of user behavior and highlights the potential limitations in the use of VR.

Conclusions: This scoping review contributes to the ongoing research on enhancing interventions to support individuals with ADHD. It advocates for participant-centric approaches that aim to optimize both cognitive and motor outcomes while prioritizing the enhancement of user experiences. This study emphasizes the need for a comprehensive approach to interventions, recognizing the relationship between cognitive and motor abilities, and calls for improving technological interventions to address the varied needs of individuals with ADHD.

(JMIR Serious Games 2025;13:e57297) doi:10.2196/57297

KEYWORDS

exercise-based simulator; exergame; virtual reality; physical activity; attention-deficit/hyperactivity disorder

XSL•FO
RenderX

Introduction

Attention-deficit/hyperactivity disorder (ADHD) continues to be a prevalent neurodevelopmental condition, marked by persistent patterns of inattention or hyperactivity-impulsivity that interfere with an individual's functioning and development [1]. Traditional approaches toward managing ADHD involve a combination of therapeutic strategies and medication. However, achieving sustained patient focus and commitment to these treatments can be challenging. Physical activity and exercise have been acknowledged as beneficial nonpharmacological adjuncts to traditional ADHD therapies, contributing to enhancing cognitive function, behavior, and physical well-being [2]. However, it can be especially difficult for people with ADHD to maintain consistent and interesting physical activity due to the symptoms they face, such as impulsivity, short attention span, and hyperactivity. Thus, there is a growing demand for new and captivating interventions. In various health care sectors, inventive strategies involving the use of exercise-based simulators and virtual reality (VR) are being progressively examined and employed [3]. These technological advancements provide engaging and immersive environments, which may raise adherence and active participation in physical activity.

In addition, it is noteworthy that the timing deficits and motor skill impairments commonly associated with ADHD could be addressed through these interactive interventions. Noreika et al [4] provided evidence suggesting that dysfunctions in the neural networks responsible for timing tasks are integral to ADHD, affecting a wide range of cognitive operations, including motor timing, time perception, and temporal foresight. Therefore, therapeutic strategies incorporating physical activities facilitated by simulators and VR could not only capture attention but also help improve motor proficiency and timing deficits. By promoting movement coordination and rhythm, these technologies might assist in reinforcing the neural circuits that underpin time processing and motor skills, presenting a complementary approach to traditional ADHD interventions.

Physical exercise is a key factor for children with ADHD, playing a crucial role in developing brain structure and function, which in turn influences the developmental trajectory of the disorder [5]. Regular participation in physical activity improves the flow of oxygen to the brain and enhances connections between neurons, particularly supporting development of the frontal lobe [5,6]. This area is important for cognitive processing and is often underdeveloped in individuals with ADHD. Exercise may improve executive functions like inhibitory control, working memory, and cognitive flexibility, by increasing the essential neurotransmitters dopamine and norepinephrine, which are pivotal for attention and concentration [7]. Moreover, physical activity enhances the processing of external information, improves coordination, and can reduce impulsivity and aggressive behaviors by establishing structured behavioral patterns [8]. Without these benefits, children with ADHD risk continued struggles with attention, learning difficulties, motor skills deficits, and disruptive behavior, which can seriously inhibit their cognitive and social development [5]. Compounding the importance of exercise, the systematic review conducted by Cortese et al [9] identified a noteworthy link between obesity and ADHD, suggesting shared underlying factors, including dopaminergic dysregulation and oxidative stress, and underscoring the necessity of addressing obesity risk in ADHD management and vice versa. This association highlights the urgency for continued research into the complex interrelations among ADHD, exercise, obesity, and the potential long-term effects of ADHD medications on weight, advocating for an integrated and multifaceted approach to treatment [9]. In parallel with this perspective, VR and exercise-based simulators have emerged as progressive tools in various fields, including fitness, health, and rehabilitation, offering new avenues for such an approach in the management of ADHD.

VR and exercise-based simulators have grown in popularity over the last few years, especially in various fields of fitness, health, and rehabilitation. These innovative technologies offer interactive and immersive environments that provide users with multisensory feedback and real-time performance tracking. The main purpose of these systems is to enhance user experience while optimizing the effectiveness of physical training interventions [10]. While both VR and exercise-based simulators aim to improve physical training outcomes, they do differ in certain aspects. For instance, VR typically involves the use of a VR headset that immerses users in computer-generated environments [11]. Exercise-based simulators are advanced physical devices designed to imitate and recreate the movements and activities associated with exercise routines or sports [12]. These sophisticated simulators aim to replicate the complex motion, resistance, and environment of real-world exercises to enhance skill acquisition and promote physical development. By closely mimicking the intricacies of actual exercise movements, these simulators provide users with a realistic training experience that can improve their overall fitness level [12]. Furthermore, these devices offer opportunities for individuals to engage in targeted practice sessions that focus on specific skills or aspects of physical performance. The immersive nature of exercise simulators creates an optimal setting for individuals to develop proficiency while also enjoying a virtual yet highly tangible workout experience [13]. Beyond their physical benefits, these simulators are increasingly harnessed for cognitive training, offering a diverse range of exercises that are designed to stimulate mental processes, enhance neuroplasticity, and contribute to cognitive rehabilitation.

However, it is only through this extensive scoping review that we can begin to uncover the potential effects, applications, limitations, and future opportunities associated with these cutting-edge technologies. Specifically, we have applied the PICO (Population, Intervention, Comparison, Outcome) framework to structure our research question and methodology [14]. We aim to address the following research question: In individuals with ADHD (Population), can immersive forms of technology (Intervention), such as VR and exercise-based simulators (Comparison), lead to improvements in cognitive and motor function (Outcome)? This comprehensive analysis of the literature will explore the effectiveness, benefits, and potential drawbacks of these interventions, ultimately informing future research and clinical practice.

XSL•FO
RenderX

Methods

Design

We performed a scoping review following the methodology outlined in the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) guidelines [15]. The PRISMA-ScR checklist is provided in Multimedia Appendix 1. The PRISMA-ScR statement includes a 27-item checklist and a 4-phase flow diagram to ensure transparency and complete reporting in systematic reviews. Our protocol was drafted in accordance with these guidelines and was revised by the research team. The final protocol was registered prospectively with the Open Science Framework on July 19, 2024 [16].

Eligibility Criteria

To be included in the review, papers needed to focus on the use of VR or exercise-based simulators for managing the symptoms of ADHD. Peer-reviewed journal papers were included if they met the following criteria: published between 2019 and 2023; written in English; involved human participants diagnosed with ADHD; and described an intervention using VR, a motion-based device, or exercise-based simulators. Studies were included if they reported on outcomes related to cognitive and motor functions, such as attention, inhibitory control, cognitive flexibility, processing speed, motor skills, and physical fitness. Quantitative, qualitative, and mixed-method studies were considered to capture various aspects of the effectiveness of the interventions. Papers were excluded if they did not specifically focus on ADHD; involved nonhuman participants; or described interventions that did not use VR, a motion-based device, or exercise-based simulators. Studies focusing on other neurodevelopmental disorders or specific learning disabilities were also excluded unless they included a distinct analysis for ADHD participants.

Information Sources

To identify potentially relevant documents, the following bibliographic databases were searched from 2019 to April 2023: PubMed, Scopus, PsycINFO, and Web of Science. The search strategies were drafted and further refined through team discussions. The final search results were exported into EndNote, and duplicates were removed. The electronic database search was supplemented by scanning relevant reviews and reference lists of included studies to identify additional sources.

Literature Search Strategy

The literature search was conducted across the following major electronic databases: PubMed, Scopus, PsycINFO, and Web of Science. The search covered years from 2020 up until July 2023. The search strategy employed in our review consistently followed a structured approach. The base search string included the population terms "ADHD" and "attention-deficit/hyperactivity disorder," combined with intervention-related keywords such as "virtual reality," "VR," "exercise simulator," "exergaming," and "serious games." To refine the focus on relevant outcomes, additional terms were incorporated, including "cognitive function," "motor skills," "executive function," and "physical activity." This search string was systematically applied across all databases to ensure comprehensive retrieval of studies addressing the intersection of ADHD, VR, and exercise simulator interventions. We used Boolean operators such as "AND" to combine different concepts (eg, exercise simulators AND VR) and "OR" to include alternative terms (eg, ADHD OR attention-deficit/hyperactivity disorder). Articles were included if they focused on ADHD interventions using VR or exercise simulators. Studies were excluded if they were not in English, focused on non-ADHD populations, or were review articles without original data. This comprehensive search strategy ensured a thorough examination of relevant literature.

Selection Process

To ensure a rigorous and standardized selection process for the included sources of evidence, we developed a detailed selection spreadsheet. This included specific questions to evaluate the relevance and quality of the sources, such as the focus on VR or exercise-based simulators for ADHD, type of study, reported outcomes, and keywords. The spreadsheet was initially used by 1 reviewer on a random sample of 979 full-text articles. The results were then discussed with another reviewer, and the form was refined to resolve any inconsistencies and to ensure clarity. Key changes included clarifying the criteria for inclusion and exclusion and adding more specific questions related to the interventions and outcomes.

To initiate the full screening process, duplicate articles were first eliminated from consideration. Subsequently, titles and then abstracts of potential articles underwent a thorough assessment to exclude any irrelevant publications. Publications were then once again crosschecked with our inclusion and exclusion criteria. Any disagreements that arose during the screening process were resolved through collaborative decision-making. If the reviewers were unable to reach a consensus, a third reviewer was brought in to help determine the appropriate course of action. This approach ensured that all included studies met the predefined eligibility criteria. Microsoft Excel was used for managing and charting records during the data charting process, while the software EndNote was used for managing selected references and tracking records after the eligibility stage of the flowchart.

Data Charting and Items

A data charting form was jointly developed to determine which variables to extract for the review. The initial form included categories such as "search typed in/keyword, title, database/search engine/top-level source, platform/publisher, date, range of coverage from source, link, methodology, strengths, weaknesses, data sets, data analysis, sample size, target population and area, theory, and personal interest." This comprehensive set of categories was designed to capture all relevant aspects of each source and to ensure consistency and thoroughness in data extraction.

The data charting was conducted using Microsoft Excel, which allowed for efficient organization and manipulation of the extracted data. Calibration was carried out by having 2 supervisors review and provide feedback on the initial charting form. This feedback was used to refine the form iteratively.

XSL•FO
RenderX

Each reviewer independently charted data from a sample set of sources to identify any discrepancies or areas for improvement in the form. The results of these independent charting exercises were discussed in detail, and the form was continuously updated to address any inconsistencies and to ensure clarity and comprehensiveness.

Three individuals were involved in the charting process: GS served as the primary charting agent, while 2 supervisors reviewed the charted data. The primary charting was done independently, and the supervisors provided periodic reviews to ensure accuracy and consistency. Any inconsistencies identified during these reviews were resolved through discussion and consensus, ensuring that all data entries were accurate and aligned with the review's objectives.

Throughout the process, any items that required interpretation, such as subjective assessments of strengths and weaknesses or methodological nuances, were carefully documented. These interpretative items were discussed among the team to reach a consensus, ensuring that the final data entries were as objective and consistent as possible.

Revisions to the charting form were made iteratively based on the feedback and findings during the initial calibration exercises. For instance, some categories were refined to provide clearer definitions or to better capture the nuances of the data. The rationale for these changes was to enhance the reliability and validity of the data extraction process.

Data Synthesis

We synthesized the range of evidence included in this review to answer our research questions and objectives by grouping the studies based on their intervention type: VR, motion-based devices, and exercise-based simulators. For each group, we summarized the study characteristics in terms of sample, focus of the study, trials and duration, and outcomes. This structured approach allowed us to systematically compare the various interventions and their effectiveness.

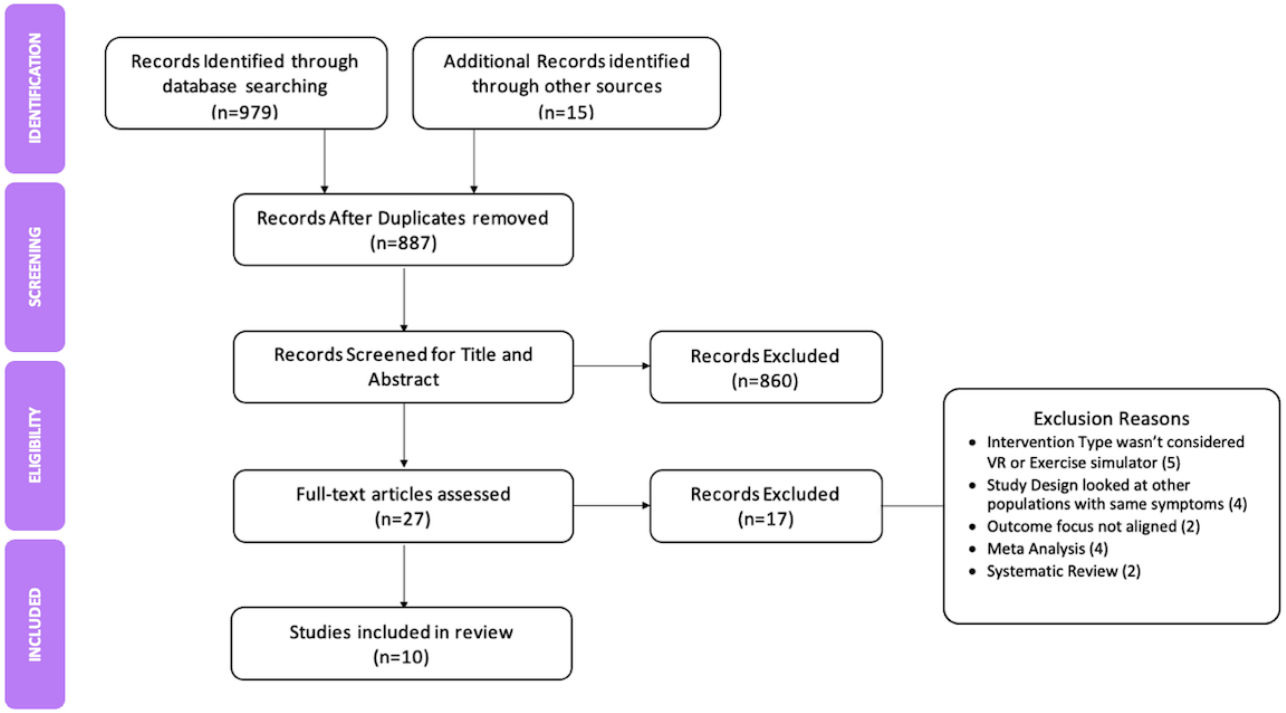
The evidence from the included studies has been presented in tables for clarity and ease of comparison. The tables include columns for sample, focus of the study, trials and duration, and outcomes, allowing for a quick visual representation of the key aspects of each intervention type. This approach facilitates a comprehensive understanding of the strengths, weaknesses, and overall effectiveness of the different interventions.

Results

Included Studies

Out of the initial 994 articles that were identified, a total of 10 were selected and re-evaluated for eligibility (Figure 1). The research designs used in these studies included randomized controlled trial, comparative study, pilot study, and exploratory research study. Throughout this process, a recurring emphasis on cognitive function and attention, as well as some aspects of motor function, in relation to the use of both simulators and VR, became apparent.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram.



VR Interventions for ADHD

A summary of VR interventions for ADHD is presented in Table 1. The study by Rodrigo-Yanguas et al [17] used "The Secret Trail of Moon," a VR game specifically designed for ADHD treatment, comprising various mini-games aimed at enhancing cognitive functions such as inhibitory control, selective attention, cognitive flexibility, and processing speed. Participants reported high levels of enjoyment, suggesting the potential of VR games as engaging therapeutic tools for ADHD management. Similarly, Bernardelli et al [18] assessed a VR-based serious game

XSL•FO
RenderX

designed to enhance attention and cognitive skills in children with ADHD, confirming the high feasibility and usability of the application. This study indicates potential future improvements in attention and cognitive skills among children with ADHD.

Table 1. Summary of study traits for virtual reality.

Study	Sample	Focus	Trials and duration	Outcome
Rodrigo-Yanguas et al [17], 2021	N=37 (all subtypes of ADHD[a]); 68% male; age: 12-22 years	To enhance cognitive abilities such as inhibitory control, selective attention, cognitive flexibility, and processing speed	Single session: patients tested each mini-game as long as desired within 1 session	86% reported enjoying playing TSTM[b] game; 83% expressed interest in continuing to play it; 14% reported dizziness or motion sickness
Stokes et al [19], 2022	N=20; 90% male; age: 10 years (range 8-12 years)	To evaluate whether environmental distractors in the VR[c] setting would result in lower response rates during behavioral tasks	Single session; 3 blocks of Stroop, Math, and AX-CPT[d] in a virtual classroom with 200 self-paced trials each; recording responses and eye movement data	Significant decrease in the time spent looking at the whiteboard from the pre- to postdistraction period
Ou et al [20], 2020	N=3; 66% female; age: 9.67 years	To test VR rehabilitation games designed specifically for children with ADHD	36 sessions, 3 times a week for 3 months; 40 minutes per session, divided into four 10-minute game sessions with 5-minute breaks	VR games helped improve attention, impulse control, and defiance in ADHD rehabilitation
Bernardelli et al [18], 2021	The panel involved in this study consisted of 4 experts who all specialized in bioengineering	To develop and test a VR serious game for children with ADHD	Single 30-minute session; task completion times vary by the game level	The study confirmed the feasibility of the VR application for remote performance validation and provided input for possible future improvements
Wiebe et al [21], 2023	N=32; 65.63% female; age: 23.03 years	Feasibility study of VR for assessing attention, impulsivity, and hyperactivity	Only 2 sessions; each of these sessions or blocks had a duration of 24 minutes	Showed the feasibility of the virtual seminar room approach; consistent CPT[e] performance and no distraction effect; linked inattention and impulsivity

[a]ADHD: attention-deficit/hyperactivity disorder.

[b]TSTM: The Secret Trail of Moon.

[c]VR: virtual reality.

[d]AX-CPT: AX-continuous performance task.

[e]CPT: continuous performance task.

Stokes et al [19] used VR technology integrated with eye tracking to understand attentional distractions in children with ADHD. Participants performed behavioral tasks in a VR classroom while their eye movements were tracked, providing insights into how environmental distractors affect attention. This setup highlighted the feasibility and utility of VR in assessing and understanding attentional processes in ADHD. Similarly, Wiebe et al [21] used a VR seminar room to simulate a classroom environment for assessing attention, impulsivity, and hyperactivity in children with ADHD. The study confirmed the feasibility of such VR setups, demonstrating their potential for detailed assessments of ADHD-related symptoms.

Ou et al [20] employed VR rehabilitation games targeting improvements in attention, impulse inhibition, and oppositional defiance in children with ADHD. Over multiple sessions spanning 3 months, significant positive changes in cognitive and behavioral parameters were observed, demonstrating the effectiveness of VR-based rehabilitation games in managing ADHD symptoms. These studies collectively underscore the effectiveness and feasibility of VR interventions in enhancing cognitive functions and managing ADHD symptoms, highlighting the promise of VR technology as a therapeutic tool for ADHD.

Motion Sense Devices for ADHD

A summary of motion sense devices for ADHD is presented in Table 2. Gao et al [22] investigated the use of exergaming with motion sense controllers like Xbox Kinect and Wii systems among preschool children. Preschool children engaged in these exergames, which significantly increased their moderate-to-vigorous physical activity (MVPA) levels. This study highlights the potential of exergaming to enhance physical activity engagement in young children.

Benzing and Schmidt [23] focused on children with ADHD, who participated in exergaming sessions aimed at improving executive functions and motor abilities. The intervention led to faster reaction times and improved motor skills, demonstrating the potential benefits of exergaming in enhancing cognitive and motor functions in children with ADHD.

XSL•FO
RenderX

Table 2. Summary of study traits for motion sense devices.

Study	Sample	Focus	Trials and duration	Outcome
Gao et al [22], 2020	N=65; 55% female; age: 4.46 years	To assess the effects of exergaming on the physical skills and activity of preschoolers	40 sessions, 5 times a week for 8 weeks; each session: 30 minutes total (20 minutes exergaming and 10 minutes warmup/cooldown)	The exergaming group showed more moderate-to-vigorous physical activity than standard care
Benzing and Schmidt [23], 2019	N=51; 82.4% male; age: 10.43 years	To investigate the effects of cognitively and physically demanding exergaming on executive functions, ADHD[a] symptoms, and motor abilities	24 sessions, 3 times a week for 8 weeks; each session: 3 hours; participants logged training duration in diaries	The exergaming group had faster reaction times in inhibition and switching tasks and better motor performance after the intervention compared to controls

[a]ADHD: attention-deficit/hyperactivity disorder.

Exercise-Based Simulators for ADHD

A summary of exercise-based simulators for ADHD is presented in Table 3. Shema-Shiratzky et al [24] combined a treadmill with a VR environment to enhance cognitive and behavioral functions in children with ADHD. Significant improvements were observed in executive function, memory, and gait regularity, alongside reductions in social problems and psychosomatic behavior. This study indicates the comprehensive benefits of combining physical and cognitive training through VR-based interventions. Ko et al [25] explored the use of a ski simulator for VR and non-VR exercises, finding that VR exercises resulted in enhanced exercise capacity and concentration levels, with greater engagement and concentration observed in the VR group compared to the non-VR group. This study underscores the added benefits of VR in exercise-based interventions.

Ji et al [26] evaluated the effects of both exergaming and traditional bicycle exercises in children with ADHD. Both interventions led to improvements in selective attention and self-control, with exergaming showing superior outcomes in cognitive control processes. This study highlights the comparative effectiveness of exergaming over traditional exercises in managing ADHD symptoms, emphasizing the potential of exergaming as a more effective intervention for enhancing cognitive and behavioral functions in children with ADHD.

Table 3. Summary of study traits for exercise-based simulators.

Study	Sample	Focus	Trials and duration	Outcome
Shema-Shiratzky et al [24], 2019	N=14; age: 8-12 years	To assess the effects of a VR[a] training program on executive function and attention abilities in children diagnosed with ADHD[b]	18 sessions, 3 times a week for 6 weeks; each session: 30 minutes to 1 hour	Reduced social problems and psychosomatic behavior; improved executive function, memory skills, and dual-task walking
Ko et al [25], 2020	N=10; mean age 28.50 years (SD 6.11 years)	To evaluate the effects of VR and non-VR exercises on both the exercise capacity and concentration levels when using a ski exergame	Single session; 10 minutes of free stretching, 2 minutes of rest, and 2 sets of 1-minute exercises each, with a 30-minute rest between sets	The results showed that there were significant differences in the VR and non-VR exercises in terms of the range of motion of the ankle and the user's concentration
Ji et al [26], 2023	N=42; 30% dropout; 87% male; age: 9 years	To investigate the impact of exergaming on attention and self-control in children with ADHD	12 sessions, 3 times a week for 4 weeks; each session: 50 minutes, 60% to 80% of heart rate; 10-minute warmup, 30-minute main exercise, 10-minute cooldown	Both the exergaming group (EXG) and bicycle exercise group had increased selective attention and self-control after exercise; only the EXG showed improved sustained attention and response control

[a]VR: virtual reality.

[b]ADHD: attention-deficit/hyperactivity disorder.

Areas of Application

The studies reviewed demonstrate the diverse applications of VR and exergaming interventions in managing ADHD symptoms. VR interventions typically focus on cognitive and educational outcomes, employing immersive and interactive environments that engage children dynamically. They are also applied as assessment tools in identifying cognitive traits related to those with ADHD. These interventions have shown significant improvements in attention, memory, and academic skills, as well as high levels of feasibility and acceptance among participants.

In contrast, exergaming interventions primarily aim to enhance physical activity and motor skills through interactive gaming. Studies like those by Gao et al [22] and Benzing and Schmidt

XSL•FO
RenderX

[23] highlight the substantial benefits of these interventions in increasing physical activity engagement and improving motor abilities in children with ADHD.

Moreover, studies combining physical and cognitive training, such as those by Shema-Shiratzky et al [24] and Ji et al [26], demonstrate the potential of integrated approaches in providing comprehensive benefits. These interventions not only address cognitive deficits but also enhance physical health and overall well-being, offering a holistic approach to ADHD management.

Discussion

Overview

This scoping review aimed to investigate and critically analyze the use, effectiveness, benefits, and potential drawbacks of VR and exercise simulator–based interventions used for individuals diagnosed with ADHD. The core findings indicate that VR-based games and exercise simulators improve attention, inhibitory control, memory, and motor functions, targeting both cognitive and physical deficits in individuals with ADHD. These interventions also significantly increase user motivation and engagement, making them more appealing compared to traditional methods. Participants reported high levels of enjoyment and adherence, which are critical for long-term treatment success.

The review also highlights the importance of tailoring interventions to individual needs, such as adjusting the virtual environments and difficulty levels to ensure their effectiveness across varying symptom severities. The study advances knowledge by demonstrating the dual role of VR and simulators in enhancing cognitive and motor outcomes, advocating for integrated approaches that address the multifaceted nature of ADHD.

In relation to previous literature, the findings are consistent with research on the benefits of physical exercise for ADHD. However, this paper expands the scope by introducing VR and simulator interventions as more engaging and potentially more effective alternatives to conventional exercise and cognitive training programs. Additionally, the review identifies gaps, particularly regarding long-term effects and optimizing user experience to mitigate challenges like motion sickness in VR environments. This comprehensive analysis offers new insights into the role of immersive technology in ADHD treatment while emphasizing the need for further research.

Impact of VR and Exercise Simulator–Based Interventions

Cognitive Function

The selected literature primarily focuses on how VR systems and exercise simulator–based systems can enhance cognitive function in individuals diagnosed with ADHD. For example, VR serious mini-games focused on enhancing cognitive capacity and information processing were implemented in a recent study [17]. This was done through a VR headset using a video game called The Secret Trail of Moon designed through 5 mini-games, with each involving different aspects of cognitive function. For instance, a game called “Enigma” enhances working memory and cognitive flexibility by requiring the user to match elements within a time limit.

Research studies investigating the use of motion-based controllers in cognitive function typically use VR applications. The study by Ou et al [20] focused on ADHD children, employing a VR intervention and evaluating the outcomes through the Test of Nonverbal Intelligence (TONI-4) and the Wisconsin Card Sorting Test (WCST). The TONI-4 results suggested cognitive advancements after the intervention, while WCST’s outcomes highlighted a mixed response. These findings suggest enhancing cognitive functions using VR warranted exploration. Meanwhile, the study by Benzing and Schmidt [23] employed computer-based tasks, such as the Simon task for inhibition, Flanker task for switching, and N-back task for updating. Most notably, the intervention group had improved reaction times, indicating an enhancement in executive functions. However, no significant changes to the updating abilities were found, leading to suggestions of exploring more sensitive tests for detecting exercise-triggered cognitive changes.

Another study indicated that interventions can effectively target cognitive functions such as working memory and executive function [27]. The study by Shema-Shiratzky et al [24] involving a treadmill with a simulated program evaluated cognitive function using the NeuroTrax computerized neuropsychological battery. This battery includes 5 tasks that encompass various aspects of cognitive function. The tasks include the Stroop test, the Go-No-Go task, verbal and nonverbal memory tasks, and a “Catch game” (measuring set-shifting, adaptation, and planning). The testing for cognitive function was conducted before and after the intervention, as well as during a 6-week follow-up period. The study showed that these interventions had a measured impact on cognitive functions such as working memory and executive function. However, it should be noted that the study did not find significant improvements in the attention score.

Attention

Attention, as articulated by Wu [28] in “We know what attention is,” can be defined as the mental mechanism by which an individual selectively focuses on certain stimuli while disregarding other perceivable information to effectively guide behaviors and responses. This selective process can often be sustained over a period, maintaining a consistent behavioral orientation toward relevant targets; this aspect is termed sustained attention. It involves the ability to persist in cognitive engagement with a single target or task over time, a critical function when continuous monitoring or prolonged concentration is required [28]. Attention can manifest in various forms, from single-mindedly concentrating on one aspect, a concept known as selective attention, to distributing one’s focus across multiple stimuli or tasks, a process referred to as divided attention [28].

This concept forms the basis of a few studies, each offering unique insights into ADHD-related attention processes through various interventions. Bernardelli et al [18] explored selective attention in children with ADHD using a VR application designed to enhance “top-down” and “bottom-up” attention skills, as well as working memory, inhibition, and distributed

XSL•FO
RenderX

attention components. The study's findings indicate that children showed improved ability to filter out distractions and focus on specific tasks within the VR environment, highlighting the potential of VR in enhancing selective attention.

Ji et al [26] focused on sustained attention in children with ADHD through exergaming. This study examined how continuous physical engagement can help children maintain focus over extended periods. The researchers used tools like the Frankfurt Attention Inventory (FAIR) test and the Go-No-Go task, along with monitoring N2 amplitudes associated with conflict monitoring and attentional allocation. The results suggested that sustained physical activity through exergaming significantly improved children's ability to maintain attention over time.

Stokes et al [19] addressed attentional distraction and selective attention by using VR technology coupled with eye tracking to measure how virtual environments can reduce distractions and enhance selective attention in children with ADHD. Unlike other studies that used VR as an intervention, this study employed it as an assessment tool, demonstrating VR's potential in evaluating attention difficulties. The study found that children were better able to focus on tasks and reduce distractions when engaging in VR-based activities.

Wiebe et al [21] examined divided attention in adults with ADHD using a multimodal VR-based approach. This feasibility study explored how individuals manage to distribute their focus across multiple stimuli or tasks within a virtual environment, measuring factors such as omission errors, reported distraction levels, and electroencephalography (EEG) metrics like the theta/beta ratio. The findings indicated that VR could effectively help individuals with ADHD improve their ability to manage divided attention.

These studies collectively contribute to understanding the multifaceted nature of attention in individuals with ADHD, providing in-depth insights into how VR and exercise-based simulators can be used to assess and enhance different aspects of attention.

Motor Function

A key theme that emerged from the respective findings is the positive impact of these interventions on cognitive functions, specifically on executive function, in children diagnosed with ADHD. This suggests that interventions using VR and exercise simulator–based systems could potentially address cognitive deficits associated with ADHD. However, only a couple of papers focused on both the cognitive and motor function aspects of these technologies. The impact of a ski exercise–based simulator on exercise capacity and concentration levels was investigated in a study conducted by Ko et al [25]. To assess concentration, the authors used an EEG system and measured motor components such as ankle range of motion and rate of perceived exertion. Benzing and Schmidt [23] used a subset of 6 test items from the German Motor Test to capture various dimensions of motor ability [23]. These dimensions encompass endurance, speed, strength, coordination, and flexibility. Following the intervention, the exergaming group displayed significantly superior overall motor performance compared to the control group. Specifically, improvements in strength and coordination were observed.

This represents a significant void in research as assessing both motor function and cognitive performance is essential for gaining a comprehensive understanding of how VR and exercise-based simulators relate to ADHD. Research has provided unique insights into the importance of assessing motor function in children with ADHD. It underscores significant deficits in motor function in children with ADHD [29,30]. The ambiguity regarding whether motor impairments are inherent to ADHD or triggered by co-occurring developmental coordination disorder has been highlighted, emphasizing the need for more research and the inclusion of motor skills assessment in ADHD care [31]. Children with ADHD were found to have worse dynamic balance performance, another indicator of motor function [32]. In children with ADHD, gross motor function was significantly impaired [33]. Collectively, these studies build a convincing case for exploring motor skills as part of an integrated approach to ADHD diagnosis and treatment. They argue for the necessity to consider these motor impairments when devising treatment plans, to look beyond the evident cognitive aspects of ADHD, and to understand the broader impacts that motor function deficits could have.

Perceptions and Experiences

Presenting users in a first-person perspective, VR can create a sense of presence and immersion that enhances engagement and motivation during exercise [34]. VR and exercise simulator–based interventions have shown promise in improving health outcomes and increasing physical activity levels [35]. This was a common theme explored through the literature that was also able to identify similar results with ADHD. The studies by Rodrigo-Yanguas et al [17], Stokes et al [19], and Shema-Shiratzky et al [24] present the recurring theme of interest and enjoyment associated with the use of VR and exercise simulator–based systems. All 3 studies emphasize the role such systems play in enhancing engagement in tasks, fostering better cognitive outcomes, and enhancing the ability to persevere in accomplishing therapy.

Additionally, a study looking at a VR headset focused on the perception of children and adolescents with ADHD [17]. The study considered mental health professionals, such as psychologists and neuropsychologists, and educators, including teachers, pedagogues, counsellors, and others, and found that 40 out of 56 (71%) either provide treatment and education or had provided these services to individuals with ADHD [17]. The perception of participants toward VR was gathered using an ad hoc questionnaire. The questionnaire had specific questions about each game's mechanics, the VR system's usability, and the overall experience. The examination of exercise-based simulators and their influence on perception in individuals has not been the focus of any specific studies yet. More direct qualitative responses from participants about their experience and preferences would provide a clearer understanding of their perception toward exercise-based simulators.

XSL•FO
RenderX

Implications of VR and Exercise Simulator–Based Interventions

Positive Changes in Cognitive Domains

The discussed studies provide intriguing implications for the effectiveness of VR and exercise simulator–based interventions for individuals with ADHD. The research studies identified that these interventions bring about positive changes in various cognitive domains. Specifically, improvements have been observed in working memory and executive function, as well as increased levels of concentration.

Efficacy of Exercise-Based Simulators

A study emphasizing selective attention and self-control through exergaming and bicycle exercises demonstrated effectiveness in enhancing cognitive function in children with ADHD [26]. This is evident as both the exergaming group and bicycle exercise group showed significant increases in performance, quality, and continuity values. The mean performance value increased from 190.44 to 269.06 in the exergaming group and from 231.15 to 303.92 in the bicycle exercise group after the intervention, denoting significant improvements (P<.001). Similarly, the mean quality value increased in both groups (exergaming group: from 0.90 to 0.95; P=.02; bicycle exercise group: from 0.92 to 0.97; P=.005). Continuity values also increased significantly in both groups (exergaming group: from 171.33 to 262.44; P<.001; bicycle exercise group: from 214.66 to 294.70; P<.001). The statistical data implied that the interventions effectively enhanced attention and self-control in children, with slight but greater improvements observed in the exergaming group than in the bicycle exercise group.

Significant changes in both behavior and cognitive function were observed [24], contributing to existing knowledge in this area. The study used Cohen *d* as a measure of effect size, which is often used to indicate the standardized difference between 2 means [24]. Values are interpreted as follows: 0.2, small effect size; 0.5, medium effect size; and ≥0.8, large effect size. Behavior and social problem measurements had notable effect sizes, with Cohen *d* values of 0.75 and 0.66 for improvements in psychosomatic behavior after training and at follow-up, respectively. Cognitive function showed considerable improvements, especially in executive function and memory. Cohen *d* was reported as 0.63 for the executive function index score and remained >1.01 after a 6-week follow-up for the color trail tests, demonstrating large effect sizes. However, studies on the effects of exercise-based simulators on attention have presented mixed results.

An analysis of sensorimotor rhythm waves associated with concentration revealed that VR exercise resulted in a more pronounced EEG signal compared to non-VR exercise [25]. A more pronounced sensorimotor rhythm wave, often measured with an EEG device, indicates heightened brain activity in the sensorimotor cortex. This area of the brain is associated with processing sensory input and controlling motor functions. A pronounced wave can suggest increased concentration, attention, and engagement in a task. On the other hand, Shema-Shiratzky et al [24] found that VR-based exercise did not significantly improve attention as the attention index scores remained unchanged between testing points. The authors speculated that this could be due to the multifaceted nature of attention, as there are various types of attention, which may not be sensitive enough to reveal changes in different aspects of attention. Some forms of attention may be resistant to VR training effects, and it is plausible that other modalities did not exhibit a response. Furthermore, it has been reported that the attention index score of individuals with ADHD showed no response to either a placebo or methylphenidate, suggesting that some measures of attention may not show changes with various treatments or interventions [36].

Design and Implementation Considerations

When designing and implementing interventions using VR systems and exercise-based simulators, it is crucial to consider the specific features and requirements that can maximize their effectiveness. Differences in design and implementation across various studies significantly impact the outcomes of these interventions. For instance, a study used a ski exergame sports simulator, while another focused on an exercise-centered treadmill. These variations highlight the importance of carefully considering the type of VR and exercise equipment used when analyzing the effectiveness of these interventions.

Case Studies and Examples

Concrete examples of successful serious games and exergaming applications for ADHD provide valuable insights for guiding future designs. For instance, “The Secret Trail of Moon,” a VR game specifically designed for ADHD treatment, effectively enhances cognitive functions such as inhibitory control, selective attention, cognitive flexibility, and processing speed [17]. This aligns with broader findings that engaging game mechanics are crucial for therapeutic success, as they boost user motivation and adherence [37,38]. Additionally, games based on motion-based devices like “Ring Fit Adventure” demonstrate how commercial exergaming can motivate physical activity through fun interactive gameplay, a finding supported by studies highlighting the importance of enjoyment in exergaming for sustained engagement [39]. By analyzing these successful examples, it becomes clear that incorporating immersive and enjoyable elements is essential for maintaining user engagement and achieving desired therapeutic outcomes in ADHD interventions.

Exergaming Integration

The integration of exergaming into ADHD interventions is supported by a growing body of literature emphasizing the benefits of physical activity for cognitive and motor development [26,40,41]. Effective integration involves the use of hardware, such as motion sensors and VR headsets, to create interactive and engaging experiences. Studies show that motion sense controllers like Xbox Kinect or Nintendo Switch significantly increase MVPA levels, particularly in young populations [22]. For ADHD interventions, game mechanics should blend physical activity with cognitive challenges. Tasks like obstacle courses requiring quick decision-making or following complex instructions while physically active can enhance both motor skills and executive functions [23]. By incorporating these elements, interventions align with guidelines

XSL•FO
RenderX

on promoting physical activity as a core component of ADHD treatment [9].

Accessibility and Inclusivity

Ensuring accessibility and inclusivity is a critical aspect of designing effective VR and exergaming interventions, as highlighted in existing guidelines on technology-assisted therapies [42]. Adjustability in difficulty levels and customizable controls are recommended to cater to users with diverse needs and preferences [21]. Studies have emphasized that settings allowing users to modify task intensity and complexity ensure that the intervention remains appropriately challenging for individuals with varying skill levels [43]. Moreover, diverse representations in games, including characters and scenarios reflecting different cultures and backgrounds, are crucial for increasing engagement across broad user groups [44]. Accessibility should also consider physical adaptations, such as seated play options or alternative control schemes, ensuring that users with physical disabilities can benefit from the interventions. These considerations reflect best practices in inclusive game design, promoting broader access and greater adherence to treatment programs.

Strengths and Weaknesses of Exercise-Based Simulators and VR

Exercise simulators offer a robust solution for managing ADHD by integrating physical activity with cognitive training. As detailed in Table 4, the design features identified through various studies, such as immersive VR environments, motion sense controllers, and customizable settings, can be effectively applied to the exercise simulator. By combining cognitive training elements with realistic and engaging VR experiences, the sailing simulator can provide a comprehensive therapeutic tool for individuals with ADHD. Emphasizing user safety, engagement, and adaptability ensures that the intervention is both effective and accessible, ultimately improving cognitive and behavioral outcomes for users.

While the studies revolving around exercise-based simulators had not focused on the perception of the participants, all of them did mention a few positive points related to the experience. For example, 1 study identified that the game-like properties of the exercise simulator potentially made the training regimen more engaging, likely promoting participants' adherence to the program and being effectively used by individuals who generally do not prefer traditional exercise but enjoy gaming [24]. The 3 studies found no adverse effects linked to the use of exercise-based simulators. However, a potential limitation that was identified is the lack of variability in the types of games or exercises that can be played on these devices. Since most simulators are designed around specific sports or activities, there may be limited options for different types of exercises or games to be performed on these devices.

Motion-based devices coupled with serious games present a compelling avenue for interventions, particularly for individuals with ADHD. The consensus points toward the engaging nature of these tools, using movement and interactivity to capture attention and promote short-term participation, a known challenge in traditional ADHD interventions. This form of technology can potentially enhance motivation and short-term adherence to a therapeutic program due to its active style. While research is ongoing, preliminary findings suggest that motion-based serious games can yield improvements in areas like attention, executive function, and motor coordination, all of which are commonly affected in ADHD. However, a further investigation is needed to solidify these findings and determine the long-term efficacy. Despite the promising outlook, considerations regarding the inherent structure of games, with defined levels and goals, could lead to a plateau in learning and engagement once those targets are achieved. This inherent limitation might result in diminished motivation over time, particularly for individuals with ADHD who thrive on novelty and challenge. Alternatively, exercise-based simulators, grounded in real-life sports and activities, offer a more open-ended experience. The pursuit of mastery in activities like cycling or sailing, with their nuanced techniques and physical demands, presents a continuous challenge that surpasses the limited nature of game-based scenarios. The open-ended nature of exercise-based simulators, which allows for continuous skill development and application in real-world scenarios, aligns well with the need for sustained engagement in ADHD interventions. This approach may lead to greater long-term adherence and more impactful therapeutic outcomes compared to more constrained game-based scenarios.

In contrast, VR systems offer a wide range of environments and experiences, providing a unique platform for various interventions. Studies have highlighted the effectiveness of VR-based interventions in managing ADHD. For example, Rodrigo-Yanguas et al [17] reported high levels of user satisfaction, with 86% of participants enjoying the game, 83% wanting to replay it, 78% finding it easy to understand, and all participants feeling good after playing. This suggests that children with ADHD prefer VR-based cognitive tests over traditional ones, which has also been noted by Fang et al [45].

The VR-based "virtual seminar room" has proven to be effective in ADHD assessment by offering a standardized yet realistic setting. Wiebe et al [21] found that this environment enhances ecological validity and provides insights into participants' distractibility and impulsiveness, potentially improving ADHD evaluations. Such positive engagement in VR experiences may lead to increased task endurance among children with ADHD.

Bernardelli et al [18] conducted a comprehensive analysis of VR interventions for children with ADHD, emphasizing their dual benefits for cognitive training and physical activity. Their findings indicated improvements in both attention and motor skills, underscoring the versatility of VR systems in ADHD treatment. The study also highlighted the need for personalized VR settings to minimize discomfort and maximize engagement, reinforcing the importance of tailored approaches in VR interventions.

Despite the numerous benefits, VR interventions are not without drawbacks. Approximately 14% of users reported discomfort, and motion sickness was a significant concern [17]. Addressing these issues is crucial to optimize the effectiveness of VR-based therapies.

XSL•FO
RenderX

Table 4. Summary of the strengths and weaknesses found through exercise simulators and virtual reality.

Study	Technology used	Strengths	Weaknesses
Rodrigo-Yanguas et al [17], 2021	VR[a] headset, console (PlayStation 4), and serious game: The Secret Trail of Moon (TSTM)	High interest and enjoyment: 83% of children expressed interest in playing. Most found the instructions easy to understand and the games easy to play	14% reported discomfort or motion sickness from VR glasses, with dizziness being a potential side effect
Stokes et al [19], 2022	VR headset (HTC Vive), PC, eye tracking (SensoMotoric Instruments), and 3D-simulated classroom	Children with ADHD[b] found VR cognitive testing more enjoyable than traditional tests, potentially increasing persistence.	VR classrooms have shown worse behavioral performance in other studies
Ko et al [25], 2020	Ski exergame (Ski Fit 360)	Enhancement of exercise and concentration: VR increased exercise capacity, with wider ankle motion and higher SensoMotoric Rhythm (SMR) wave activity indicating improved concentration	The ski motion platform made it difficult for users to maintain balance when moving side to side
Ji et al [26], 2023	Exergame group: ExerHeart devices (D&J Humancare App Alchemist's Treasure); Nonexergame group: Stationary bike (Fit Elite-Whole body exerciser 100)	Improved attention and control: exergaming improved selective attention and self-control, with N2 amplitude changes indicating better cognitive control	Only 1 game could be played on the simulator, limiting variety
Shema-Shiratzky et al [24], 2019	Exergame: the system utilized projects a computer-generated simulation in front of a treadmill	Gait and cognitive improvements: VR improved gait regularity, executive function, and memory, and reduced social and psychosomatic issues in children with ADHD	No negatives or limitations were mentioned regarding the simulator setup
Benzing and Schmidt [23], 2019	Motion sense controller, console (Xbox Kinect), and serious game: Shape Up	Improved reaction times and motor skills: exergaming led to faster reaction times and better motor abilities; participants rated enjoyment highly (3.62/4)	High dropout rate; no significant improvements in some tasks like updating; general health benefits were seen, but not specific ADHD symptom improvements
Ou et al [20], 2020	VR headset (HTC Vive) and PC game: "Fishing Master, Fruit Train and Ocean Manager"	Cognitive and behavioral enhancements: participants showed cognitive improvements and reductions in ADHD symptoms, such as inattention and hyperactivity	Participants lost interest after mastering the games, and frequent breaks were needed to prevent discomfort from the VR headgear
Bernardelli et al [18], 2021	VR headset (Oculus Quest) and PC serious game: ADAD	User-friendly VR application: VR reduced impulsivity and improved attention, with a customizable intervention for individual needs	Need for automatic level selection to reduce errors; remote implementation complexities affected operations
Gao et al [22], 2020	Motion sense controller, console (Wii or Xbox Kinect) games: Just Dance for Kids (both consoles), and Wii Nickelodeon Fit	Engagement in physical activity: participants found the intervention engaging, spending 40% of school time in moderate-to-vigorous physical activity	No negatives or limitations were mentioned regarding the intervention setup
Wiebe et al [21], 2023	A virtual seminar room (VSR) was created using Unity 3D and featured a VR headset (HTC Vive) and PC, replicating a real-world seminar room with typical furniture, a front canvas, and virtual classmates	Feasibility of the VSR: the VSR was well-accepted, with successful behavioral and neurophysiological data recording and low cybersickness reports	Head movements and strain from the VR headset caused artifacts that interfered with electroencephalography signals; cybersickness was still reported

[a]VR: virtual reality.

[b]ADHD: attention-deficit/hyperactivity disorder.

Limitations

Despite the promising results of integrating VR and exercise-based simulators for managing ADHD, some limitations must be acknowledged.

Methodological Limitations

Despite employing a structured search across multiple databases, this review may be susceptible to publication bias, potentially favoring studies that report positive or significant findings. As a result, there is a risk of overestimating the effectiveness of VR and exercise-based interventions, as studies with neutral or negative outcomes might be underrepresented. Additionally,

XSL•FO
RenderX

the search strategy, while comprehensive, focused on specific terms such as "virtual reality," "exercise simulator," "exergaming," and "serious games." This narrow focus may have unintentionally excluded relevant studies that use different terminology or related technologies. Furthermore, the decision to limit the search to articles published between 2019 and 2023 may have omitted earlier foundational research, limiting the understanding of the field's broader development and historical context.

Another limitation is language and population bias. By including only English-language publications, the review may have excluded valuable research conducted in other languages, which could provide additional insights. Moreover, the review's focus on human participants diagnosed with ADHD may have inadvertently excluded research on other populations, such as those with cognitive impairments or neurodevelopmental disorders that share overlapping symptoms, potentially limiting the scope of findings that could otherwise be applicable to ADHD management.

Limitations Within the Included Studies

Many of the included studies featured relatively small sample sizes, which limits the generalizability of their findings to broader ADHD populations. Larger and more diverse samples are needed to validate these results and ensure their applicability to a wider demographic. Additionally, most studies lacked long-term follow-up data, making it difficult to assess whether the short-term benefits of VR and exercise-based interventions are sustained over time. Longitudinal studies are essential to evaluate the durability and potential long-term effects of these interventions. Furthermore, several of the studies lacked rigorous design frameworks, such as randomized controlled trials, which are critical for establishing clear causal relationships between the interventions and the observed improvements in ADHD symptoms. This absence of methodological rigor weakens the strength of the conclusions drawn from the available evidence. These limitations underscore the need for cautious interpretation of the review's findings. While the review offers valuable insights, it should be seen as a foundation for further investigation. Future research should prioritize larger and more diverse samples, long-term follow-up, and rigorous study designs, including randomized controlled trials, to assess the potential of VR and exercise-based simulators more fully for managing ADHD.

Conclusion

VR and exercise-based interventions show promise as potentially effective approaches for people with ADHD. Studies suggest that individuals with ADHD find VR cognitive testing and experiences more pleasurable compared to traditional tests, which may result in increased engagement and persistence in tasks. The use of VR in ADHD interventions offers a wide range of virtual environments and activities that can be tailored to individual needs, providing a more versatile and engaging experience for participants. Both forms of technology have also shown improvements in a variety of cognitive functions. However, it is crucial to recognize the potential drawbacks of VR interventions, including certain users experiencing discomfort and dizziness. These negative reactions can impact adherence to the program and add additional cognitive load during tasks, potentially compromising the effectiveness of the intervention. Previous research conducted on exercise-based simulators did not consider the perspectives of individuals with ADHD toward this type of intervention. This implies that there is a gap in understanding how people diagnosed with ADHD perceive and respond to exercise-based simulators. It is important to address such misunderstandings so that people with ADHD can benefit from the opportunities provided by VR and exercise-based interventions to support the challenges they face with ADHD, which will enable them to enjoy a higher quality of life and a normal life.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[PDF File (Adobe PDF File), 61 KB - games_v13i1e57297_app1.pdf]

References

1. Faraone S, Asherson P, Banaschewski T, Biederman J, Buitelaar J, Ramos-Quiroga J, et al. Attention-deficit/hyperactivity disorder. Nat Rev Dis Primers 2015 Aug 06;1:15020 [FREE Full text] [doi: 10.1038/nrdp.2015.20] [Medline: 27189265]
2. Sanabria-Mazo JP, Montero-Marin J, Feliu-Soler A, Gasión V, Navarro-Gil M, Morillo-Sarto H, et al. Mindfulness-based program plus amygdala and insula retraining (MAIR) for the treatment of women with fibromyalgia: a pilot randomized controlled trial. J Clin Med 2020 Oct 11;9(10):3246 [FREE Full text] [doi: 10.3390/jcm9103246] [Medline: 33050630]
3. Georgieva I, Georgiev G. Reconstructing personal stories in virtual reality sas a mechanism to recover the self. Int J Environ Res Public Health 2019 Dec 18;17(1):26 [FREE Full text] [doi: 10.3390/ijerph17010026] [Medline: 31861440]
4. Noreika V, Falter C, Rubia K. Timing deficits in attention-deficit/hyperactivity disorder (ADHD): evidence from neurocognitive and neuroimaging studies. Neuropsychologia 2013 Jan;51(2):235-266 [FREE Full text] [doi: 10.1016/j.neuropsychologia.2012.09.036] [Medline: 23022430]

XSL•FO
RenderX

5. Sun W, Yu M, Zhou X. Effects of physical exercise on attention deficit and other major symptoms in children with ADHD: A meta-analysis. Psychiatry Res 2022 May;311:114509 [FREE Full text] [doi: 10.1016/j.psychres.2022.114509] [Medline: 35305344]
6. Di Liegro C, Schiera G, Proia P, Di Liegro I. Physical activity and brain health. Genes (Basel) 2019 Sep 17;10(9):720 [FREE Full text] [doi: 10.3390/genes10090720] [Medline: 31533339]
7. Zheng J, Su X, Xu C. Effects of exercise intervention on executive function of middle-aged and elderly people: A systematic review of randomized controlled trials. Front Aging Neurosci 2022 Aug 12;14:960817 [FREE Full text] [doi: 10.3389/fnagi.2022.960817] [Medline: 36034137]
8. Kotulska M, Jędrzejewska B, Lichman M, Borycka A, Lubczyńska Z, Potocka Z, et al. The influence of physical activity on reducing ADHD symptoms – a systematic review. J Educ Health Sport 2023 May 16;31(1):46-54 [FREE Full text] [doi: 10.12775/jehs.2023.31.01.004]
9. Cortese S, Moreira-Maia C, St Fleur D, Morcillo-Peñalver C, Rohde L, Faraone S. Association between ADHD and obesity: a systematic review and meta-analysis. Am J Psychiatry 2016 Jan;173(1):34-43 [FREE Full text] [doi: 10.1176/appi.ajp.2015.15020266] [Medline: 26315982]
10. Hauer K, Litz E, Günther-Lange M, Ball C, de Bruin ED, Werner C. Effectiveness and sustainability of a motor-cognitive stepping exergame training on stepping performance in older adults: a randomized controlled trial. Eur Rev Aging Phys Act 2020 Sep 29;17(1):17 [FREE Full text] [doi: 10.1186/s11556-020-00248-4] [Medline: 33005256]
11. Maresca G, Maggio M, Buda A, La Rosa G, Manuli A, Bramanti P, et al. A novel use of virtual reality in the treatment of cognitive and motor deficit in spinal cord injury: A case report. Medicine (Baltimore) 2018 Dec;97(50):e13559 [FREE Full text] [doi: 10.1097/MD.0000000000013559] [Medline: 30558016]
12. Ortiz J, Palacios-Navarro G, Andaluz V, Guevara B. Virtual reality-based framework to simulate control algorithms for robotic assistance and rehabilitation tasks through a standing wheelchair. Sensors (Basel) 2021 Jul 27;21(15):5083 [FREE Full text] [doi: 10.3390/s21155083] [Medline: 34372320]
13. Masengo G, Zhang X, Dong R, Alhassan AB, Hamza K, Mudaheranwa E. Lower limb exoskeleton robot and its cooperative control: A review, trends, and challenges for future research. Front Neurorobot 2022 Jan 12;16:913748 [FREE Full text] [doi: 10.3389/fnbot.2022.913748] [Medline: 36714152]
14. Brown D. A review of the PubMed PICO tool: using evidence-based practice in health education. Health Promot Pract 2020 Jul;21(4):496-498 [FREE Full text] [doi: 10.1177/1524839919893361] [Medline: 31874567]
15. Tricco A, Lillie E, Zarin W, O'Brien K, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
16. Exploring Virtual Reality and Exercise Simulator Interventions in Attention Deficit Hyperactivity Disorder: A Comprehensive Literature Review. OSF. URL: https://osf.io/8xsy9 [accessed 2024-12-22]
17. Rodrigo-Yanguas M, Martin-Moratinos M, Menendez-Garcia A, Gonzalez-Tardon C, Royuela A, Blasco-Fontecilla H. A virtual reality game (The Secret Trail of Moon) for treating attention-deficit/hyperactivity disorder: development and usability study. JMIR Serious Games 2021 Sep 01;9(3):e26824 [FREE Full text] [doi: 10.2196/26824] [Medline: 34468332]
18. Bernardelli G, Flori V, Greci L, Scaglione A, Zangiacomi A. A Virtual Reality Based Application for Children with ADHD: Design and Usability Evaluation. In: De Paolis LT, Arpaia P, Bourdot P, editors. Augmented Reality, Virtual Reality, and Computer Graphics. AVR 2021. Lecture Notes in Computer Science, vol 12980. Cham: Springer; 2021:363-375.
19. Stokes J, Rizzo A, Geng J, Schweitzer J. Measuring attentional distraction in children with ADHD using virtual reality technology with eye-tracking. Front Virtual Real 2022 Mar;3:855895 [FREE Full text] [doi: 10.3389/frvir.2022.855895] [Medline: 35601272]
20. Ou Y, Wang Y, Chang H, Yen S, Zheng Y, Lee B. Development of virtual reality rehabilitation games for children with attention-deficit hyperactivity disorder. J Ambient Intell Human Comput 2020 Apr 10;11(11):5713-5720 [FREE Full text] [doi: 10.1007/s12652-020-01945-9]
21. Wiebe A, Kannen K, Li M, Aslan B, Anders D, Selaskowski B, et al. Multimodal virtual reality-based assessment of adult ADHD: a feasibility study in healthy subjects. Assessment 2023 Jul 18;30(5):1435-1453 [FREE Full text] [doi: 10.1177/10731911221089193] [Medline: 35435010]
22. Gao Z, Lee J, McDonough D, Albers C. Virtual reality exercise as a coping strategy for health and wellness promotion in older adults during the COVID-19 pandemic. J Clin Med 2020 Jun 25;9(6):1986 [FREE Full text] [doi: 10.3390/jcm9061986] [Medline: 32630367]
23. Benzing V, Schmidt M. The effect of exergaming on executive functions in children with ADHD: A randomized clinical trial. Scandinavian Med Sci Sports 2019 May 23;29(8):1243-1253 [FREE Full text] [doi: 10.1111/sms.13446]
24. Shema-Shiratzky S, Brozgol M, Cornejo-Thumm P, Geva-Dayan K, Rotstein M, Leitner Y, et al. Virtual reality training to enhance behavior and cognitive function among children with attention-deficit/hyperactivity disorder: brief report. Dev Neurorehabil 2019 Aug;22(6):431-436 [FREE Full text] [doi: 10.1080/17518423.2018.1476602] [Medline: 29771624]
25. Ko J, Jang S, Lee HT, Yun H, Kim YS. Effects of virtual reality and non-virtual reality exercises on the exercise capacity and concentration of users in a ski exergame: comparative study. JMIR Serious Games 2020 Oct 28;8(4):e16693 [FREE Full text] [doi: 10.2196/16693] [Medline: 33112240]

XSL•FO
RenderX

26. Ji H, Wu S, Won J, Weng S, Lee S, Seo S, et al. The effects of exergaming on attention in children with attention deficit/hyperactivity disorder: randomized controlled trial. JMIR Serious Games 2023 May 09;11:e40438 [FREE Full text] [doi: 10.2196/40438] [Medline: 37159253]
27. Shaw P, Gilliam M, Liverpool M, Weddle C, Malek M, Sharp W, et al. Cortical development in typically developing children with symptoms of hyperactivity and impulsivity: support for a dimensional view of attention deficit hyperactivity disorder. Am J Psychiatry 2011 Feb;168(2):143-151 [FREE Full text] [doi: 10.1176/appi.ajp.2010.10030385] [Medline: 21159727]
28. Wu W. We know what attention is!. Trends Cogn Sci 2024 Apr;28(4):304-318 [FREE Full text] [doi: 10.1016/j.tics.2023.11.007] [Medline: 38103983]
29. Cho H, Ji S, Chung S, Kim M, Joung Y. Motor function in school-aged children with attention-deficit/hyperactivity disorder in Korea. Psychiatry Investig 2014 Jul;11(3):223-227 [FREE Full text] [doi: 10.4306/pi.2014.11.3.223] [Medline: 25110492]
30. Mokobane M, Pillay B, Meyer A. Fine motor deficits and attention deficit hyperactivity disorder in primary school children. S Afr J Psychiatr 2019;25:1232 [FREE Full text] [doi: 10.4102/sajpsychiatry.v25i0.1232] [Medline: 30899581]
31. Goulardins J, Marques J, De Oliveira J. Attention deficit hyperactivity disorder and motor impairment. Percept Mot Skills 2017 Apr;124(2):425-440 [FREE Full text] [doi: 10.1177/0031512517690607] [Medline: 28361657]
32. Goetz M, Paulasova Schwabova J, Hlávka Z, Ptacek R, Surman C. Dynamic balance in children with attention-deficit hyperactivity disorder and its relationship with cognitive functions and cerebellum. NDT 2017 Mar;Volume 13:873-880 [FREE Full text] [doi: 10.2147/ndt.s125169]
33. Pila-Nemutandani R, Pillay B, Meyer A. Gross motor skills in children with Attention Deficit Hyperactivity Disorder. S. Afr. j. occup. ther 2018;48(3):19-23 [FREE Full text] [doi: 10.17159/2310-3833/20l7/vol48n3a4]
34. Blair C, Walsh C, Best P. Immersive 360° videos in health and social care education: a scoping review. BMC Med Educ 2021 Nov 24;21(1):590 [FREE Full text] [doi: 10.1186/s12909-021-03013-y]
35. Lee G. Effects of virtual reality exercise program on balance, emotion and quality of life in patients with cognitive decline. J Kor Phys Ther 2016 Dec 30;28(6):355-363 [FREE Full text] [doi: 10.18857/jkpt.2016.28.6.355]
36. Leitner Y, Doniger G, Barak R, Simon E, Hausdorff J. A novel multidomain computerized cognitive assessment for attention-deficit hyperactivity disorder: evidence for widespread and circumscribed cognitive deficits. J Child Neurol 2007 Mar;22(3):264-276 [FREE Full text] [doi: 10.1177/0883073807299859] [Medline: 17621495]
37. Dan B. Gamification of therapy: the fun factor in rehabilitation. Dev Med Child Neurol 2022 Mar;64(3):276 [FREE Full text] [doi: 10.1111/dmcn.15126] [Medline: 35120264]
38. Luckner N, Kayali F, Purgathofer P. Game mechanics to further engagement with physical activity. Interaction Design and Architecture 2018 Mar 10(36):43-57 [FREE Full text] [doi: 10.55612/s-5002-036-003]
39. Rüth M, Schmelzer M, Burtniak K, Kaspar K. Commercial exergames for rehabilitation of physical health and quality of life: a systematic review of randomized controlled trials with adults in unsupervised home environments. Front Psychol 2023 Jun 2;14:1155569 [FREE Full text] [doi: 10.3389/fpsyg.2023.1155569] [Medline: 37333591]
40. Chan Y, Jang J, Ho C. Effects of physical exercise on children with attention deficit hyperactivity disorder. Biomed J 2022 Apr;45(2):265-270 [FREE Full text] [doi: 10.1016/j.bj.2021.11.011] [Medline: 34856393]
41. Li D, Wang D, Cui W, Yan J, Zang W, Li C. Effects of different physical activity interventions on children with attention-deficit/hyperactivity disorder: A network meta-analysis of randomized controlled trials. Front Neurosci 2023 Mar 20;17:1139263 [FREE Full text] [doi: 10.3389/fnins.2023.1139263] [Medline: 37021131]
42. Erlandson R. Universal and Accessible Design for Products, Services, and Processes. Boca Raton: CRC Press; 2007.
43. Staiano A, Calvert S. Exergames for physical education courses: physical, social, and cognitive benefits. Child Dev Perspect 2011 Jun;5(2):93-98 [FREE Full text] [doi: 10.1111/j.1750-8606.2011.00162.x] [Medline: 22563349]
44. Brand J, Knight S, Majewski J. The Diverse Worlds of ComputerGames: A Content Analysis of Spaces,Populations, Styles and Narratives. In: Proceedings of DiGRA 2003 Conference: Level Up. 2003 Presented at: DiGRA 2003 Conference: Level Up; November 4-6, 2003; Utrecht, The Netherlands URL: https://dl.digra.org/index.php/dl/article/view/88/88
45. Fang Y, Han D, Luo H. A virtual reality application for assessment for attention deficit hyperactivity disorder in school-aged children. Neuropsychiatr Dis Treat 2019;15:1517-1523 [FREE Full text] [doi: 10.2147/NDT.S206742] [Medline: 31239686]

Abbreviations

ADHD: attention-deficit/hyperactivity disorder
EEG: electroencephalography
MVPA: moderate-to-vigorous physical activity
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews
TONI: Test of Nonverbal Intelligence
VR: virtual reality
WCST: Wisconsin Card Sorting Test

XSL•FO
RenderX

Edited by R Baranyi; submitted 14.02.24; peer-reviewed by E Vallefuoco, M Martín Moratinos, M Bella-Fernández; comments to author 29.04.24; revised version received 14.07.24; accepted 08.11.24; published 29.01.25.

Please cite as:
Sarai G, Jayaraman PP, Tirosh O, Wickramasinghe N
Exploring Virtual Reality and Exercise Simulator Interventions in Patients With Attention Deficit Hyperactivity Disorder: Comprehensive Literature Review
JMIR Serious Games 2025;13:e57297
URL: https://games.jmir.org/2025/1/e57297
doi:10.2196/57297
PMID:39879092

©Gurdeep Sarai, Prem Prakash Jayaraman, Oren Tirosh, Nilmini Wickramasinghe. Originally published in JMIR Serious Games (https://games.jmir.org), 29.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Playful Antisedentary Interactions for Online Meeting Scenarios: A Research Through Design Approach

Jiaqi Jiang[1*], BS; Shanghao Li[2*], BS, MS; Xian Li[1], BE; Yingxin Xu[1], BE; Jian Zhao[3], BS, MS, PhD; Pengcheng An[1], BS, MS, PhD

[1]School of Design, Southern University of Science and Technology, Shenzhen, Guangdong, China

[2]Department of Computer Science, University of Illinois Chicago, Chicago, IL, United States

[3]School of Computer Science, University of Waterloo, Waterloo, ON, Canada

[*]these authors contributed equally

Corresponding Author:
Pengcheng An, BS, MS, PhD
School of Design
Southern University of Science and Technology
No. 1088 Xueyuan Avenue
Nanshan District
Shenzhen, Guangdong, 518055
China
Phone: 86 13166272076
Email: anpengcheng88@gmail.com

Abstract

Background: Online meetings have become an integral part of daily life for many people. However, prolonged periods of sitting still in front of screens can lead to significant, long-term health risks. While previous studies have explored various interventions to address sedentary lifestyles, few have specifically focused on mitigating sedentary behavior during online meetings. Furthermore, design opportunities to address this issue in the context of online meetings remain underexplored.

Objective: This study aims to investigate the design of effective antisedentary interactions for online meeting scenarios and understand user experiences with gamified bodily interactions as an antisedentary measure during online meetings.

Methods: This study adopts a "research through design" approach to develop and explore user experiences of gamified bodily interactions as interventions to mitigate sedentary behavior during online meetings. In collaboration with 11 users, we co-designed and iterated 3 prototypes, which led to the development of the Bodily Interaction Gamification towards Anti-sedentary Online Meeting Environments (BIG-AOME) framework. Using these prototypes, we conducted user studies with 3 groups totaling 15 participants. During co-design and evaluation, all group semistructured interviews were transcribed into written format and analyzed using a conventional qualitative content analysis method.

Results: The findings demonstrate that gamified bodily interactions encourage users to engage in physical movement while reducing the awkwardness of doing so during online meetings. Seamless integration with meeting software and the inclusion of long-term reward mechanisms can further contribute to sustained use. In addition, such games can serve as online icebreakers or playful tools for decision-making. Drawing from 3 design prototypes, this study offers a comprehensive analysis of each design dimension within the BIG-AOME framework: bodily engagement, attention, bodily interplay, timeliness, and virtual and physical environments.

Conclusions: Our research findings indicate that antisedentary bodily interactions designed for online meetings have the potential to mitigate sedentary behaviors while enhancing social connections. Furthermore, the BIG-AOME framework that we propose explores the design space for antisedentary physical interactions in the context of online meetings, detailing pertinent design choices and considerations.

(JMIR Serious Games 2025;13:e62778) doi:10.2196/62778

KEYWORDS

gamification; sedentary behavior; videoconferencing; exertion games; embodied interaction; design research

XSL•FO
RenderX

Introduction

Background

Online meetings have become a significant aspect of modern work, introducing new daily health challenges. An increasing number of activities are being conducted primarily or entirely online [1]. As noted by Wu and Yu [2], individuals tend to prefer online meeting platforms due to their advantages such as flexibility and cost-effectiveness, even when working from home is not mandatory. However, increasing research evidence has revealed the presence of technostress in virtual conferencing, which can negatively impact mental and physical well-being, productivity, job satisfaction, and group commitment [1,3-5]. Among these challenges, prolonged sedentary behavior, with its potential for causing irreversible health damage over time, stands out as a significant threat [6-8].

Sedentary behavior occur in periods of wakefulness during which the body engages in minimal movement, such as sitting, lying, or reclining [9]. The negative effects of prolonged sedentary behavior—distinct from merely a lack of physical activity—cannot be fully mitigated by intermittent physical exercise [10]. According to current medical studies, no "gold standard" exists for measuring sedentary behavior, nor are there specific guidelines on how to effectively interrupt such behaviors [11]. The prevailing recommendation is to reduce excessive sedentary behavior by replacing it with mild to vigorous physical activity [12,13]. Seated activities that incorporate moderate movements of the lower or upper body can help reduce the negative effects of sedentary behavior [12]. This approach underscores the necessity to integrate more active routines in daily life to counteract the health risks associated with prolonged sedentary periods.

Many interventions have been developed to reduce sedentary time and encourage more active engagement in physical activities [14,15]. These include break reminder software for computers [16,17], mobile apps designed to interrupt prolonged sedentary behavior [18], dedicated devices equipped with LED displays [19], posture-adjusting smart chairs [20], and indoor location–based mobile games [21]. However, online meeting environments, as distinct sources of technostress due to remote social interactions, remain underexplored in design research. Therefore, there is a pressing need for research to collect a broader range of design examples and provide specific guidance for developing effective and engaging interventions to address sedentary behavior in online meetings.

In-person meetings typically feature habitual breaks, encouraged by social and environmental elements [22]. However, online meetings often lack these cues or social expectations, leading to fewer affordances for breaking sedentary routines [23]. This indicates the need for designing antisedentary activities that could fit into online meetings and engage users in a socially amusing way. Analysis of large-scale video meeting data shows that multitasking is a common behavior in online meetings. While multitasking can lead to negative outcomes, such as increased distraction [24], it may also have positive effects, such as improved efficiency in attention division or task shifting [25]. This highlights a design opportunity to harness the positive aspects of multitasking in online meetings to reduce sedentary behavior by incorporating physical exercises into the meeting process as secondary or parallel activities. Gamified bodily interactions offer a promising approach to integrating simple, brief physical exercises into meeting routines.

Gamification and Exertion Games

Gamification is defined as "the use of game elements in non-game contexts" [26]. It applies the principles of game design and elements such as storytelling, leaderboards, and winning rules to address real-world challenges in areas such as training, health care, and education [27]. Previous studies have demonstrated that gamification can facilitate behavior change [28-30] and make products more engaging [31]. Exertion games or exergames [32], which promote physical activity by incorporating exercise into digital games, have become an increasingly relevant area of research. According to Mueller et al [33], all computer systems that facilitate physical exertion as part of the interaction could be regarded as exertion games, regardless of whether their primary focus is on gaming or exercise.

Many studies have focused on motivating physical activity through gamification by developing software or using hardware [34-39]. Exertion games have demonstrated not only their effectiveness in promoting physical activity but also their ability to enhance social interaction among users [40,41]. Mandryk et al [42] analyzed exertion games from the lens of mitigation of sedentary behaviors and articulated 2 design principles: "providing an easy entry into play" and "motivating repeated play sessions throughout the day." Gamified bodily interactions have been applied in areas such as fitness, education, and training [30,43,44]. However, to the best of our knowledge, few studies have investigated how gamified bodily interactions can be integrated into videoconferencing scenarios to effectively promote physical activity and combat sedentary behavior.

At the intersection of remote education and exertion games, Shin et al [45] proposed Jumple, a virtual physical education classroom that leveraged artificial intelligence pose estimation technology to facilitate physical activity for students in online environments. By leveraging commonly available devices, Jumple aimed to address the challenges of remote learning and the negative impact of the COVID-19 pandemic on children's well-being. Sachan and Peiris [46] investigated the use of augmented reality–based micro–health interventions to mitigate "Zoom fatigue"—a term describing the somatic and cognitive exhaustion caused by the intensive or inappropriate use of videoconferencing tools [5]—among college students during virtual classes, exploring their effectiveness in reducing sedentary behavior and promoting physical activity. These studies offer design instances from online classroom settings, which is a similar context to online meetings. Building on this foundation, our exploration pertinently focuses on the online meeting context and aims to generate further knowledge through a design framework to systematically reveal the design space and implications.

XSL•FO
RenderX

Understanding the Online Meeting Environment

For decades, scholars have predicted that online meetings will reform the conventional routine of commuting to and from the workplace and the way people collaborate with others [47]. Nevertheless, for many people, the heightened reliance on online meetings has proven to be mentally and physically challenging. Rudnicka et al [48] found that break taking is a behavior mediated by the social norms among coworkers, while remote work environments often lead to extended periods of inactivity and excessive work due to the absence of social cues. Given the drawbacks and challenges associated with current videoconferencing software, we propose to add socially engaging bodily interactions to afford more antisedentary components in online meetings. Many people perceive breaks involving sedentary screen activities (eg, playing games, using social media, browsing the web, and watching videos) as unhelpful, while considering physical activities to be helpful breaks [49]. However, in practice, people tend to exceed the intended duration for digital and static breaks, whereas physical breaks, especially outdoor activities, are less likely to extend beyond the planned time [50]. To address this, we explore the potential of adding low-threshold physical movements as a playful antisedentary option. As a supplement to higher-threshold physical activities (eg, standing up or walking around), these games provide an additional way to promote helpful breaks during online meetings.

Multitasking is a common phenomenon during meetings. In face-to-face meetings, people engage in various physical activities such as pacing, standing, and stretching, which are beneficial to overall meeting performance [25]. During telephone calls, people also tend to perform minor movements while talking without disrupting the communication [47]. However, online meetings include more of digital multitasking than physical multitasking (eg, locomotion or other movements). On a Zoom call, maintaining a central position within the camera view with one's face clearly visible to others is regarded by cultural norms as professional and trustworthy [51]. Research often stresses the negative impacts of multitasking, including increased mental workload and reduced productivity, which could potentially contribute to burnout and depression [5]. However, according to a large-scale analysis of remote meeting multitasking behaviors, in-meeting multitasking can also lead to positive outcomes when participants are able to regulate their attention, making flexible use of time during less demanding or noncritical parts of a meeting [25]. The authors suggest we should allow "space for positive multitasking" and "shorten meeting duration and insert breaks" [25]. In the future, people's tendency to multitask is unlikely to diminish due to the growing prevalence of digital devices and the nature of the modern workplace [52]. Inspired by the idea of leveraging positive multitasking to mitigate sedentary routines, our design incorporates bodily game elements into the online meeting interface.

Objectives

Given the limited knowledge on *how to design gamified bodily interactions as antisedentary interventions in online meetings*, our study adopted a research through design methodology [53] to address 2 research questions (RQs):

- How can gamified bodily interactions be integrated into online meetings to reduce sedentary behavior?
- What are the relevant design options and considerations to properly design such gamified bodily interactions for online meeting contexts?

To address the first part of this research, we conducted 3 rounds of co-design activities with 11 participants. During these sessions, participants brainstormed, elaborated, and assessed initial bodily game ideas and low-fidelity prototypes considering their prior online meeting experiences. On the basis of the co-design insights and outcomes, we propose the initial Bodily Interaction Gamification towards Anti-sedentary Online Meeting Environments (BIG-AOME) framework that describes a design space to help designers and researchers consciously navigate multiple relevant design dimensions. Across these dimensions, the BIG-AOME framework reveals various design options that can be considered to make thoughtful design choices according to specific design goals and scenarios.

To further consolidate and contextualize the BIG-AOME framework, in the second part of this research, we implemented 3 game design ideas into functioning prototypes and evaluated them with 15 participants in online meeting settings. The 3 designs were chosen because they were hypothesized to represent distinct design options across various dimensions, enabling a concrete exploration of the design space and uncovering context-specific insights and implications underlying the BIG-AOME framework. These bodily game prototypes can be integrated into mainstream online meeting platforms (eg, Zoom or VooV Meeting) via virtual camera software such as OBS Studio (Open Broadcaster Software) [54]. The visual components of the games are overlaid on the camera view or video tiles of online meeting attendees. While the games are active, the attendees can interact with each other and the game objects via bodily movements adapted from existing beneficial physical exercises. This approach allows gamified exercises to be embedded flexibly into meeting sessions as a playful way to mitigate sedentary meeting routines, with the participants still being seamlessly engaged in the meeting and connected with other attendees.

Our study offers a design-oriented exploration of gamified bodily interactions as a novel approach to breaking sedentary online meeting routines. Our contributions are two-fold: (1) a total of 3 prototypes as design instances of antisedentary bodily gamification for online meetings, along with empirical findings regarding user experiences; and (2) a preliminary design framework for creating gamified bodily interactions for online meetings, highlighting a promising design space with relevant design options and considerations to inform and inspire future research.

XSL•FO
RenderX

Methods

Design Explorations and the Formulation of the BIG-AOME Framework

In this section, we present our iterative research through design approach [53], including our early design explorations, co-design workshops, and the initial formulation of the BIG-AOME design framework.

Early Design Explorations

Overview

In the initial stage, we explored the possibility of integrating gamified bodily interactions as antisedentary interventions in online meeting scenarios. To do so, we first studied existing literature [26-30] and online resources (eg, credible health information websites such as MedlinePlus [55], the Centers for Disease Control and Prevention [56], and the World Health Organization [57], as well as YouTube channels maintained by professional physical therapists or exercise coaches). The purpose was to accumulate suitable antisedentary physical exercises as design inputs. These exercises specifically target areas such as the shoulders, neck, and back and can be performed in typical online meeting settings. This corpus of verified healthy movements served as an inspiration resource for gamification design ideas to build upon. Subsequently, we generated a wide range of gamification ideas, implemented them in functional low-fidelity prototypes, and tested them for firsthand experience and as preparatory materials for later co-design workshops.

Insights From Early Design Explorations

The most important lesson we learned was that simply using bodily interactions to control a “classic” digital game would often fail to provide appropriate gaming experiences. Two unsuccessful early attempts—arm-controlled versions of Angry Birds and Snake—were perceived as blunt and tedious, with arm gestures leading to fatigue rather than alleviating the strain of sedentary behavior. As Mueller et al [58] noted, these designs treated the body as merely a controller for digital objects, failing to leverage the advantages of “body as play.” Therefore, we focused on creating gameplay experiences that could naturally integrate the players into online meeting environments to achieve in-depth and seamless bodily engagement.

The early design explorations laid the groundwork for our BIG-AOME framework. In the initial stages, we developed design considerations based on literature research, online resources, and insights from early prototyping. These considerations addressed various aspects and formed the initial elements of our framework. In the subsequent co-design sessions, we refined these considerations based on users’ needs and experiences. This process led us to the final version of the framework.

Co-Design Workshops

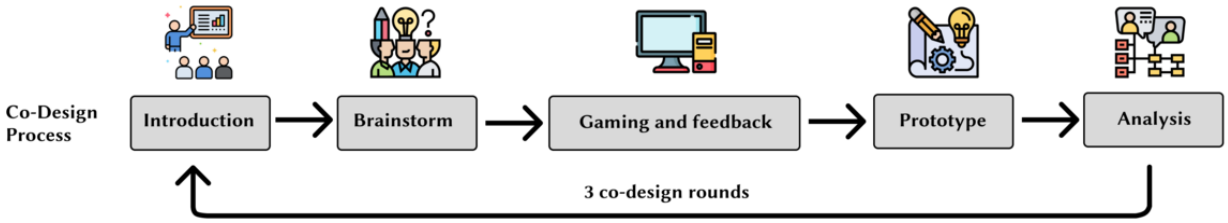
Overview

To address user needs in real-world contexts and refine the design space, we used a participatory design approach, inviting individuals with diverse online meeting experiences to contribute to the development of gamified bodily interactions. We organized 3 online co-design workshops to elicit participants’ meeting experiences, create new design ideas and functional low-fidelity prototypes, and gather design insights. We recruited 11 participants (n=3, 27% male; n=8, 73% female) with diverse educational and professional backgrounds, including business, design, and engineering. Of these 11 participants, 7 (64%) were aged 18 to 24 years, 2 (18%) were aged 25 to 34 years, and 2 (18%) were aged 45 to 54 years. We refer to these participants as P1, P2, and so on up to P11. Each workshop session lasted approximately 2 hours. To facilitate the co-design process, we prepared prompt cards to inspire the participants during their discussions and idea sketches. These cards were designed based on our literature study and firsthand experience. The participants were presented with 8 card sets, reflecting our initial design considerations as mentioned previously. Each set addressed a key consideration, including start time, duration, bodily movements as input, bodily interplay, action, game elements, attention, and exercise intensity; for example, the card set titled Action Cards included 21 illustrations, with each depicting a beneficial bodily exercise found in the literature and credible online sources. Similarly, the Attention Card set featured a spectrum ranging from games that allow participants to play while focusing on the meeting to those requiring full concentration.

The workshops were hosted on Zoom (Zoom Video Communications, Inc) to simulate an online meeting setting and allow the participants to test functional low-fidelity prototypes within the Zoom interface. Design activities were carried out using Figma (Figma, Inc) and its FigJam feature. As illustrated in Figure 1, the co-design workshop consisted of 5 steps: introduction, brainstorming, gaming and feedback, prototyping, and iterating. Participants were briefed on the research background, purpose, and workshop schedule in the introduction stage. During brainstorming, they shared their ideas, provided suggestions, and generated new concepts. The gaming and feedback stage involved playing gamification low-fidelity prototypes or acting out design ideas together and exchanging opinions in a group discussion. In the prototyping stage, participants sketched out their ideas based on insights from the previous stages. Finally, they refined their low-fidelity prototypes and shared their design considerations with others. After the workshops, we acquired a rich understanding of underlying real-world needs from the participants. We collected 2 types of data during this process. First, we amassed a significant number of ideas from the brainstorming stage of the co-design workshops. Second, we documented group discussion records. To further clarify the insights from the idea cards, we used an affinity diagram to identify common themes, patterns, and relationships among the data points. For the discussion content, we used the conventional qualitative content analysis method outlined by Hsieh and Shannon [59] to examine how the initial design considerations supported participants in formulating their ideas.

XSL•FO
RenderX

Figure 1. An illustration of our design process, highlighting the various stages involved.



Insights From Co-Design Workshops

The co-design workshops yielded rich insights. We identified 4 major themes in the affinity diagram (Textbox 1). First, the social aspect was identified as critical, with competition and cooperation influencing user engagement. Second, participants expressed a preference for games with low learning curves, which aligned with a design principle proposed by Mandryk et al [42]: "providing an easy entry into play." The third point concerned the physical fatigue and mental stress associated with online meetings, highlighting the need to differentiate between the intensity of physical activity and the attentional demands of games. Finally, we observed variations in the layout view and display order of video tiles among users, with tile sizes varying based on the number of meeting participants. Fluctuations in the number of attendees, speakers, and layout settings within the meeting platform can impact the meeting interfaces. These factors should be considered when designing games.

Textbox 1. User needs identified from workshops and representative responses from participants.

Consideration of social factors

- "When I see the name of one of my best friends on it, my desire to win is very strong." [P8]
- "I am more inclined toward something that can be played alone." [P6]

Desire for simplicity and familiarity

- "Everyone is already familiar with the rules of these games [a reference to games such as Fruit Ninja and Whack-a-Mole, which were the inspirations behind game ideas suggested by participants]." [P3]

Adaptability to physical and mental states

- "The mind is already very tired [during the meeting]." [P11]
- "If [the intensity of physical activity is] too gentle, I can't feel my body being relaxed." [P2]

Clear visibility and usability

- "I can't clearly see the sticker elements on the video." [P9]

The discussion content analysis revealed that the design consideration cards we supplied played a substantial role in shaping and refining participants' ideas. The analysis identified 3 key ways the cards were used: inspiring and guiding nonexperts in proposing designs, adapting designs to the online meeting context, and evaluating design choices. These findings also offer insights into the application of the proposed BIG-AOME framework.

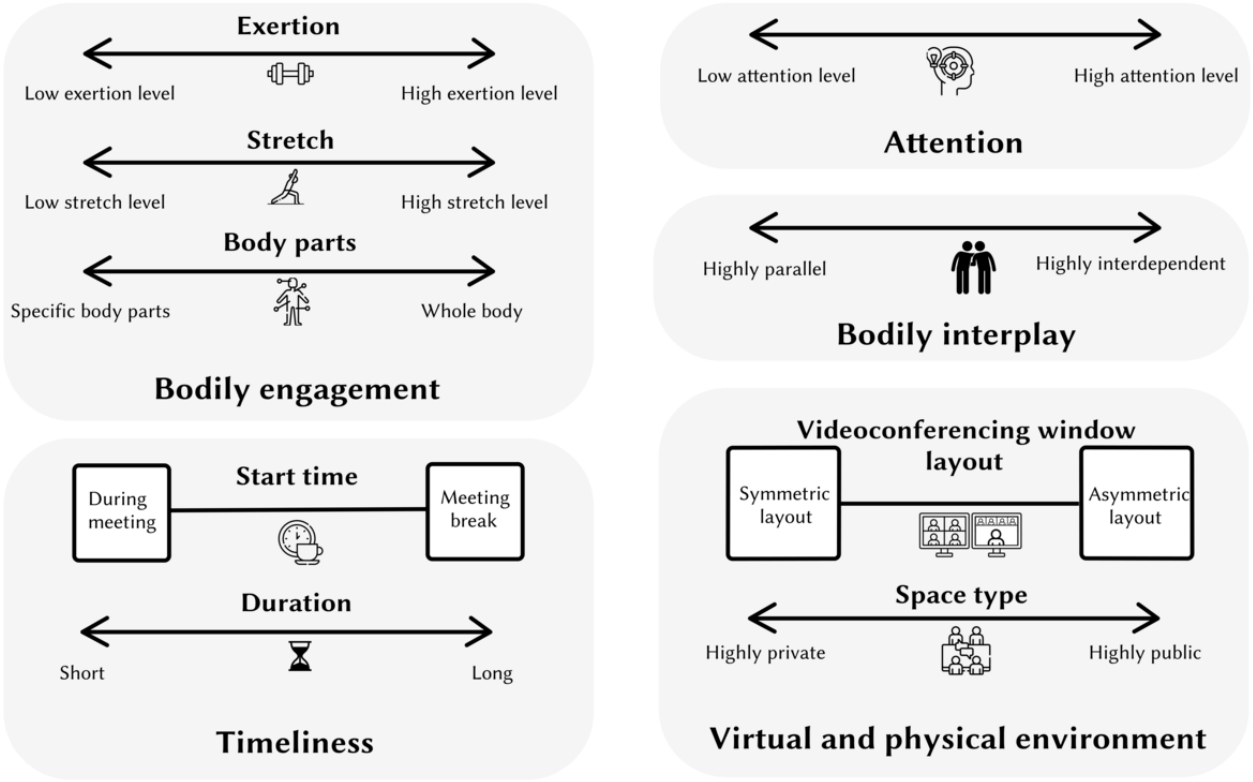
BIG-AOME Framework

Overview

In this section, we present the BIG-AOME framework. Initially formulated during early design explorations, the framework has been enriched and refined based on insights and design implications derived from the co-design sessions. As illustrated in Figure 2, the framework consists of 5 dimensions that specify a broad design space for creating gamified antisedentary bodily interactions for online meeting scenarios. Here, we briefly introduce each dimension, with a more concrete examination provided later using empirical data from the user evaluation.

XSL•FO
RenderX

Figure 2. The Bodily Interaction Gamification towards Anti-sedentary Online Meeting Environments framework.



Bodily Engagement

The *bodily engagement* dimension describes players' physical experience during gameplay. It includes 3 subdimensions: *exertion*, *stretch*, and *body parts*. *Exertion* represents the perceived intensity of the physical effort exerted during the game. *Stretch* refers to the perceived extent of physical reach and flexibility required by the game. The *body parts* subdimension indicates the perceived extent to which various parts of the body are engaged during gameplay: some games primarily involve a specific movement, such as head nodding, while others require full-body activities, such as dancing. Incorporating the 3 subdimensions allows us to design games that are more pertinent to the nuanced needs and design goals.

Attention

The *attention* dimension measures the level of concentration needed for gameplay and highlights positive multitasking during online meetings. Games requiring peripheral attention enable users to engage in movement without losing focus on the meeting. Conversely, games demanding greater focus provide high engagement and offer attendees a pause, promoting physical activity and fostering interactive play among participants. With numerous design nuances, the *attention* dimension spans a broad spectrum for game design, ranging from lightly engaging to fully immersive experiences.

Bodily Interplay

The *bodily interplay* dimension builds upon the concept proposed by Mueller et al [41]. It evaluates to what extent users' bodies interact with and influence each other during the gaming experience. Highly interdependent bodily interplay requires collective participation, fostering a shared gaming experience. By contrast, highly parallel bodily interplay allows participants to engage with the game independently, without requiring cooperation from others. This form of interaction is noninterfering and can be asynchronous. The level of bodily interplay is a vital design consideration, depending on the design objectives and the social dynamics of the meeting attendees. Therefore, it requires thoughtful deliberation in the design process.

Timeliness

The *timeliness* dimension comprises 2 subdimensions: *start time* and *duration*. The *start time* subdimension determines the most suitable moment to commence the game. Some games are designed to be engaged with during the meeting, whereas others might be better suited to ad hoc breaks. The *duration* subdimension focuses on the length of an episode of the gaming experience. Episodes can be brief, offering a quick gamified experience that ends once completed, or continual, allowing intermittent engagement over a longer period. The design choices for both subdimensions depend on the meeting context and attendees' preferences, requiring careful consideration.

Virtual and Physical Environments

This dimension addresses 2 crucial factors: the layout of the videoconferencing window and the nature of the user's physical space. Videoconferencing software typically provides either symmetrical (eg, thumbnail video tiles) or asymmetrical (eg, speaker view) layouts. In an asymmetrical layout, typically used when a participant shares their screen, the shared content occupies a larger portion of the display. This setup enables the shared screen to display gameplay elements to all participants, collectively drawing their attention to the game. The larger

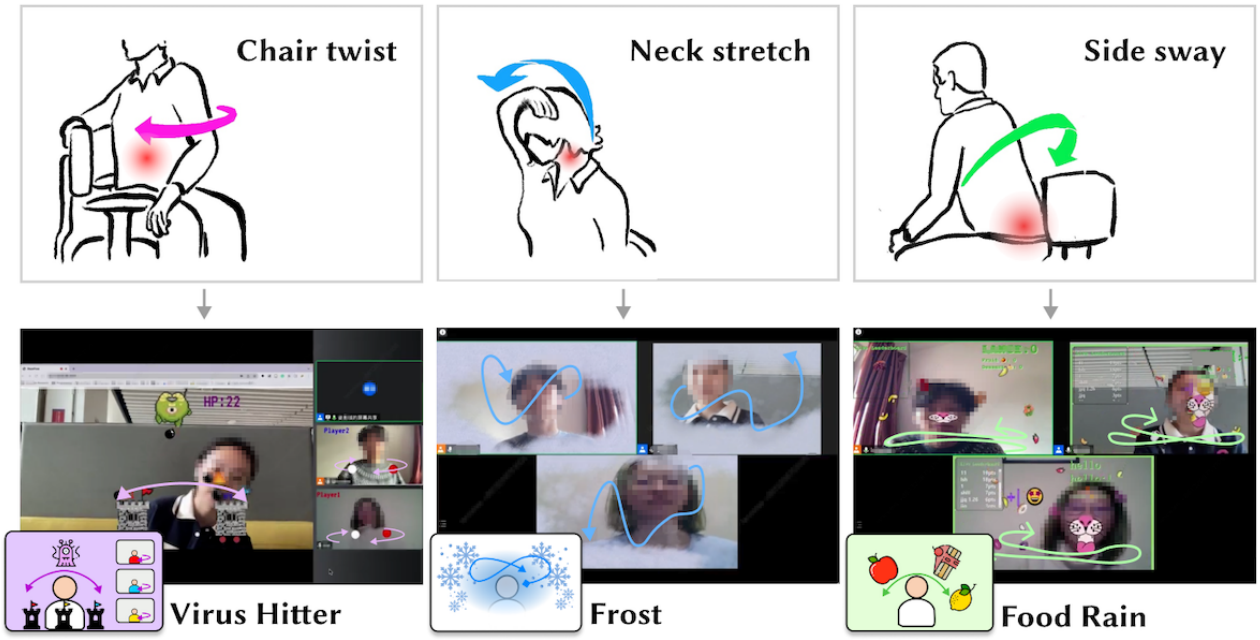
XSL•FO
RenderX

display area in this layout may accommodate more visual elements. Conversely, in a symmetrical layout, the screen is divided equally among participants. This layout allows users to see more faces but may result in smaller videoconferencing windows, making it difficult to discern finer visual details, particularly in large meetings. Designers should recognize the differences between symmetrical and asymmetrical layouts and consider the nature of the meeting. Turning to the physical environment, the *space type* pertains to the level of privacy and ownership of a user's surroundings. A highly private space, such as a personal study, enables users to move and speak freely without social concerns. Conversely, a highly public space, such as a library or an office, may limit movement and vocal interactions to avoid disturbing others or creating social awkwardness. These constraints can affect users' willingness and ability to engage in games during meetings.

Three Design Prototypes

To explore how gamified bodily interactions can be integrated into online meetings to reduce sedentary behavior (RQ1) and to contextually examine and refine the BIG-AOME framework for relevant design implications (RQ2), we developed 3 high-fidelity prototypes for user evaluation: Virus Hitter, Frost, and Food Rain (Figure 3 [60,61]; Multimedia Appendix 1).

Figure 3. The 3 high-fidelity prototypes developed to probe the design of antisedentary bodily gamification for online meetings.



Virus Hitter is a multiplayer cooperative game in which players are assigned roles: 1 participant is randomly selected as the "hitter," while the others serve as the hitter's assistants. The assistants' interaction is adapted from an existing exercise called "chair twist" [60]: participants, while seated, are asked to rotate their upper bodies side to side, facilitating a relaxing effect. The hitter, meanwhile, moves a torch that appears on the screen and is attached to their nose via motion tracking, using it to launch bombs at watchtowers. The hitter's interaction is derived from an exercise called "side sway" [61], which involves side-to-side sway while remaining seated. Every assistant corresponds to 1 watchtower matched by color. A brief animated tutorial is offered so that players can quickly and easily understand how to play the game. Virus Hitter is designed to promote a high level of bodily interactivity and requires the cooperation of all participants, thereby not only encouraging physical activity but also fostering social bonding among players.

Frost is a gamified interaction designed for use during online meetings. It simulates frost slowly forming on the surface of the videoconferencing window, gradually spreading from the edges toward the user's image. To prevent their personal window from becoming entirely obscured by frost, users must move any part of their body to "swipe" it away. The suggested movement for this activity is inspired by the"neck stretch" [60], which involves the simple action of raising and lowering the head to stretch the neck muscles. This specific interaction was chosen to provide a natural and inconspicuous reason for meeting participants to engage in light physical activity without feeling self-conscious. This game is characterized by low levels of exertion, stretch, and attention, making it ideal for relieving muscle and spine fatigue through gentle and unobtrusive movements. These movements do not significantly distract from meeting discussions or disrupt the flow of the meeting, whether in virtual or physical settings. Frost is designed to be played seamlessly as a secondary task, complementing the primary activities of an online meeting, thus enabling participants to stay active without interrupting the meeting dynamics.

Food Rain is a multiplayer competitive game. After inputting a nickname, players see fruits and desserts falling from the top of the screen. They move their bodies and open their mouths to catch the falling food items in their own window to score points. Catching a fruit increases the score by 1 point, while catching a dessert decreases it by 1 point. This movement is also derived from the side sway exercise [61]. During this activity, players' neck and waist muscles receive a moderate workout. As some users might feel awkward opening their mouths in front of

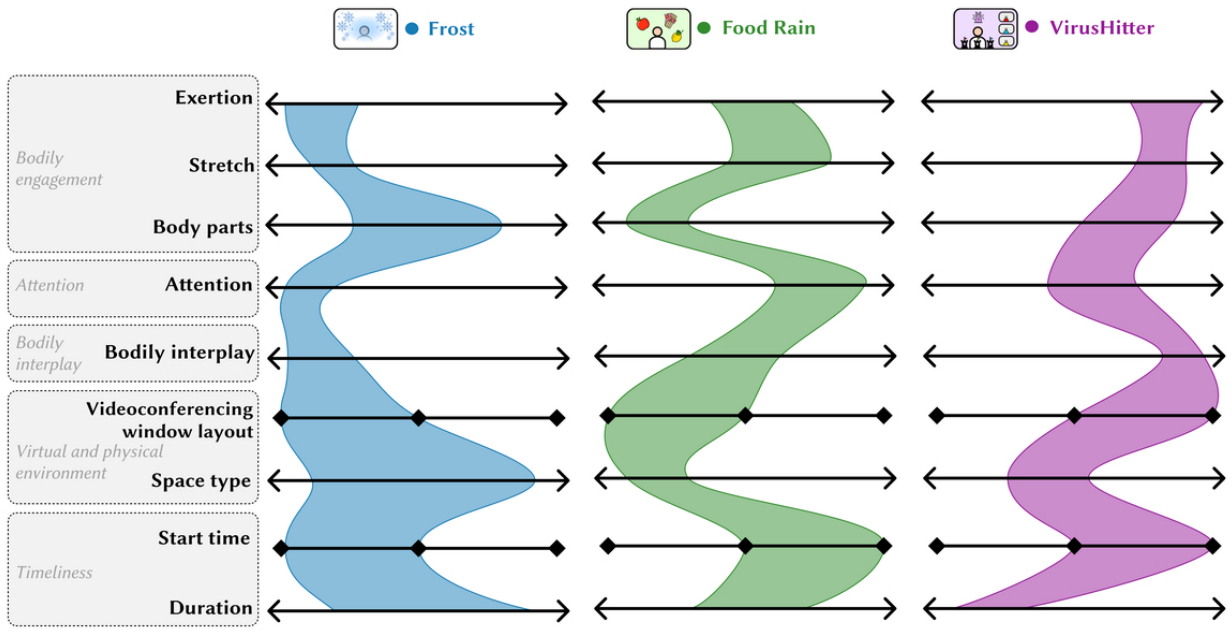
XSL•FO
RenderX

others, the design superimposes a cartoonish animal mouth that tracks users' mouth movements (opening and closing) in real time. A leaderboard in the top left corner of each user's window displays the rankings and scores of all players. The competition mechanism encourages long-term participation. Food Rain, which has a moderate duration, was designed with the assumption that it would trigger relatively high social engagement.

The 3 gamified bodily interactions are implemented on a web-based application and streamed via OBS Studio [54] as a virtual camera, which enables its integration with any online meeting platform (eg, Zoom and VooV Meeting). As online meeting platforms and their application ecosystems continue to evolve, gamified bodily interactions have the potential to emerge in various forms in the future. They could be available as add-on applications within the platforms, integrated as features of the meeting interface itself, or even implemented as virtual backdrops for meetings.

The prototypes presented here are representative examples chosen to illustrate and explore the design space. They were selected from a range of game designs developed during our early exploration and co-design phases. In the process of transforming ideas into prototypes, we considered various dimensions of the framework. As depicted in Figure 4, each prototype is intentionally situated in different regions within these dimensions. This strategic placement results in each prototype embodying unique design patterns. Each dimension contributes to the overall game design, and different combinations of dimensions can yield a diverse range of games.

Figure 4. Distribution of dimensions on the framework when designing the 3 prototypes.



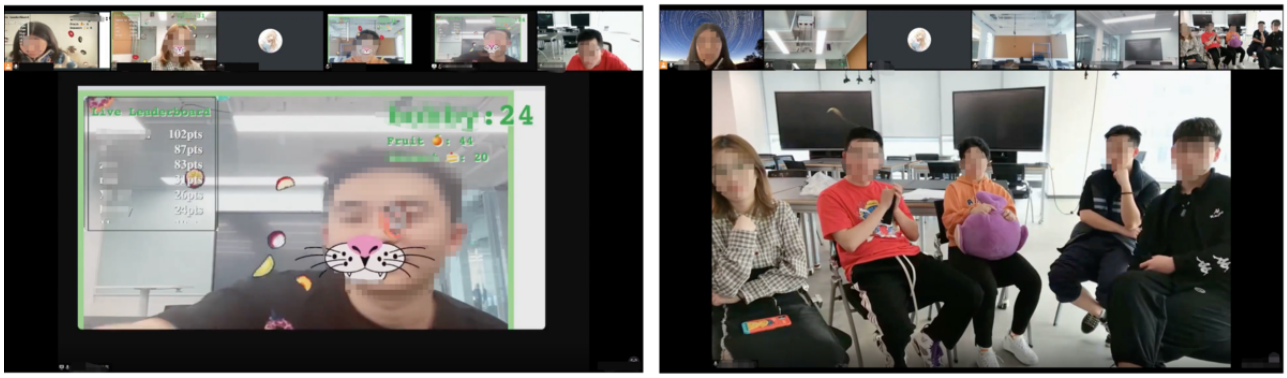
Evaluation Study

Participants and Setup

The study involved 3 rounds of sessions with 4, 5, and 6 participants, totaling 15 participants (n=10, 67% male; n=5, 33% female), referred to as P12, P13, and so on up to P26. Of these 15 participants, 11 (73%) were aged 18 to 24 years and 4 (27%) were aged 25 to 34 years. A prestudy survey was conducted to gather participants' basic information and online meeting habits. The results showed that 53% (8/15) of the participants attended ≥1 online meetings weekly, and 87% (13/15) spent >1 hour in each meeting. To simulate a realistic online meeting scenario, we organized 3 rounds of online seminars focused on ChatGPT. Each session included discussions and integrated gamified activities. The seminars were designed as collaborative learning experiences, where participants were introduced to foundational ChatGPT concepts through lecture videos and group discussions featuring open-ended questions. Fixed-interval breaks were incorporated into the meetings, during which participants engaged with gamified exercises. After the seminar, participants independently evaluated the games by completing an evaluation panel designed using Figma. This panel, based on the BIG-AOME framework, aimed to capture their feedback on the gaming experience and design elements (refer to Multimedia Appendix 2 for details). Each session concluded with a focus group interview, where participants discussed their gaming experiences and shared insights into their evaluation choices (refer to Multimedia Appendix 3 for details). To ensure a smooth process, all games were preconfigured on participants' browsers, and clear guidelines were provided for the evaluation and interview stages. The entire meeting process, including the focus group discussion, was recorded for subsequent analysis. As shown in Figure 5, screenshots illustrate the user evaluation process and interview process conducted on the online meeting platform.

XSL•FO
RenderX

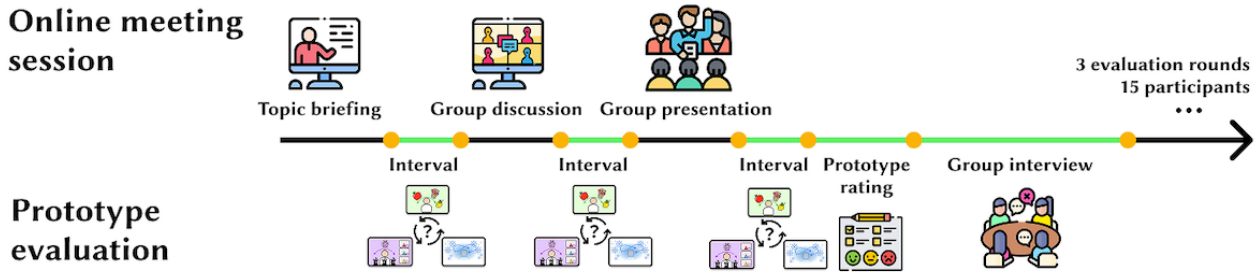
Figure 5. Screenshots showcasing: the user evaluation process (Left) and the focus group interview process (Right) conducted on the online meeting platform.



Procedure

As illustrated in Figure 6, the entire study spanned 105 minutes, comprising 45 minutes for learning activities, 30 minutes for gameplay, 10 minutes for completing the evaluation panel, and 20 minutes for a focus group interview. Each game was allocated 10 minutes, providing participants ample time to engage as they wished.

Figure 6. A visualization of the overall evaluation procedure.



The study began with the host introducing its objectives and process, after which participants watched a YouTube video providing an overview of ChatGPT and its underlying mechanism. They then took their first meeting break to play a randomly selected game. After the break, participants were divided into groups to discuss the risks and opportunities associated with using ChatGPT. Each group was assigned a team member to facilitate the discussion. Another game was then randomly selected for all participants to experience. After the gameplay, 1 representative from each group presented their opinions to the rest of the participants. The final game was played during the last meeting break.

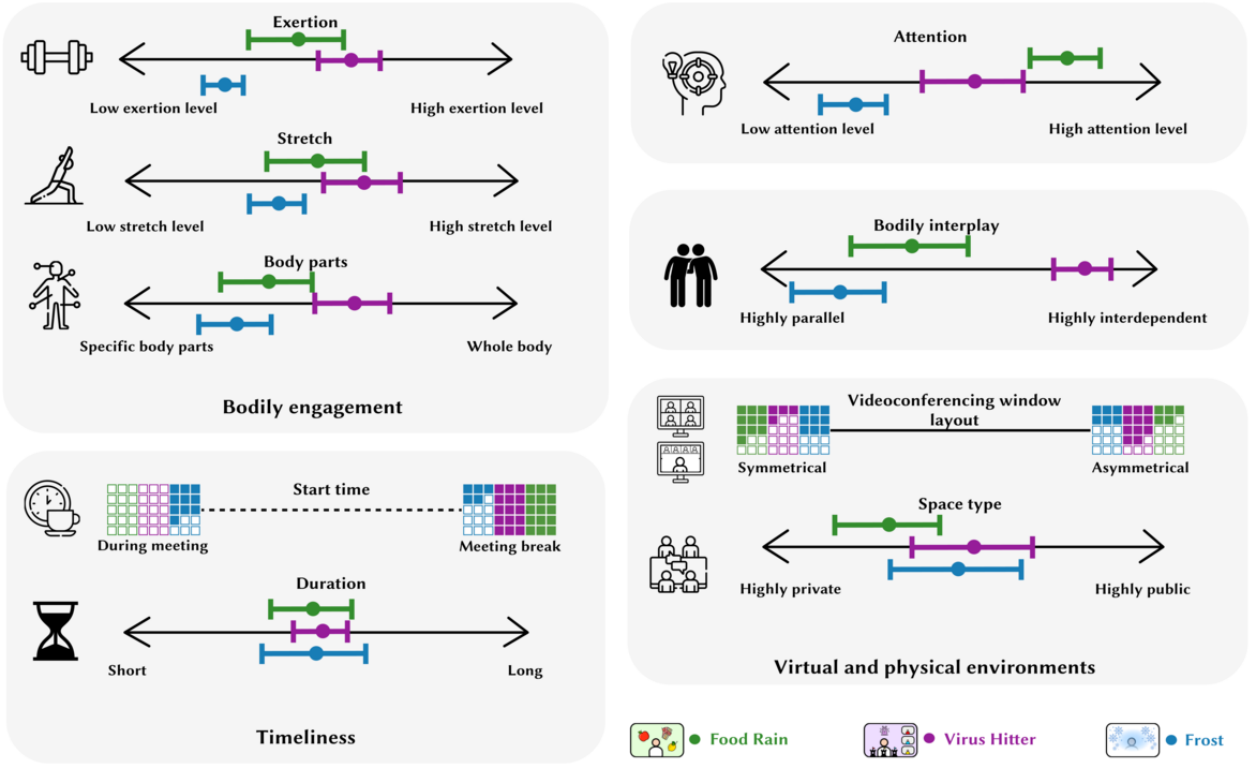
After completing the seminar simulation, participants were invited to evaluate the games and share their personal insights on how these games could be applied in real-life scenarios. Participants were given 10 minutes to complete the evaluation panel. We emphasized that they should provide responses based on their own experience and that all perspectives were valid. Finally, a 20-minute focus group interview was conducted to explore participants' experiences with gamified bodily interactions and discuss potential application scenarios. The group discussions followed a semistructured format, using open-ended questions to facilitate open and comprehensive conversations. First, we asked participants to recall their gaming experience, including physical sensations, mental responses, and interactions with others. Next, participants explained the reasons behind their ratings on the evaluation panel. Finally, we invited participants to describe how they would use the framework to create a similar game if given the opportunity.

Data Collection and Analysis

Data collection and analysis comprised 2 parts. First, we conducted focus group interviews to gather users' experiences with gamified bodily interactions during online meetings. All interview data were transcribed verbatim. We used the conventional qualitative content analysis method outlined by Hsieh and Shannon [59] for an in-depth understanding of participants' perceptions of and experiences with the games played during meeting breaks. This process led to the development of preliminary codes, which were iteratively refined and adjusted as our understanding of the data evolved. These codes were then organized into clusters, providing a clear outline of the major findings from our qualitative data. Second, participants rated each of the 3 games based on their experiences within the dimensions of the BIG-AOME framework. These ratings were instrumental in identifying the framework regions occupied by each prototype and in validating our initial design assumptions. The ratings were processed using Python (Python Software Foundation) and are descriptively visualized in Figure 7.

XSL•FO
RenderX

Figure 7. Results of users' evaluation regarding the distribution of the 3 games on the framework. The horizontal line for each game represents the IQR (Q1-Q3), defining the middle 50% of the data, which highlights the central distribution and concentration of the data values.



Ethical Considerations

This study has been approved by the Medical Ethics Committee of Southern University of Science and Technology (approval number: 20230044; application number: 2023PES044).

Informed consent was obtained from all participants before their involvement. The informed consent form provided a detailed explanation of the study's procedures, potential risks, and expected benefits. It also outlined the confidentiality of participants' personal data and emphasized that all ethical standards would be upheld throughout the research. Participants were made aware that their involvement was entirely voluntary, and they could withdraw from the study at any time without facing any penalties. To ensure privacy protection, all collected data were anonymized and securely stored in an online central database. Each participant received compensation of US $15.

Results

Integrating Gamified Bodily Interactions Into Online Meetings (RQ1)

The first aim of our study was to explore how gamified bodily interactions can be effectively integrated into online meeting settings (RQ1). This investigation seeks to provide empirical insights into this novel antisedentary approach, identifying potential opportunities and addressing challenges to inform future applications. Our findings highlight 3 key themes (Textbox 2): (1) perceived value in breaking sedentary behavior patterns, (2) potential factors influencing motivation and sustained use, and (3) anticipated use scenarios and additional benefits.

Textbox 2. Key themes and descriptions.

Perceived value in breaking sedentary behavior patterns

- Provides a reason to move
- Reduces the awkwardness of moving during meetings

Potential factors influencing motivation and sustained use

- Seamless integration with meeting software
- Long-term reward mechanisms

Anticipated use scenarios and additional benefits

- Online icebreaker activities
- Gamified decision-making

XSL•FO
RenderX

Perceived Value in Breaking Sedentary Behavior Patterns: "Now I Have a Reason to Move"

Participants affirmed that the designed bodily interactions provided a compelling reason to move during online meetings, offering innovative ways to disrupt sedentary routines. On the basis of their prior experiences, they could all relate to the physical strain and "discomfort" (P21) caused by the sedentary nature of online meetings. Introducing bodily interactions, as P22 noted, could effectively reduce sedentary behavior and encourage physical movement in online meeting scenarios by bringing "more physical activity in online meetings." A participant noted that when they were in an online meeting and wanted to move, they felt "embarrassed" (P15). Another participant shared as follows:

> *This game gives me a chance, now I have a reason to move.* [P22]

This view was reinforced by a participant:

> *You often remain static during online meetings, but adding such games will add more subjective movement.* [P23]

In addition, P12 and P14 indicated that after playing the games they felt more relaxed: "a little more relaxed than at the beginning for sure" (P12. Overall, participants expressed that gamified bodily interactions not only gave them a reason to engage in physical activity but also reduced the awkwardness associated with moving during online meetings.

Potential Factors Influencing Motivation and Sustained Use: "This Will Motivate Me to Stick With It"

Overview

Participants agreed that the games were effective in encouraging movement because they were novel, easy to learn, and facilitated highly interactive sessions that reduced social awkwardness. However, they acknowledged that the novelty and social benefits might diminish over time, particularly in routine online meetings where attendees are already familiar with the games. To address this, participants emphasized the need for strategies to sustain engagement over the long term. This supports our view that gamified bodily interactions should be integrated into meeting routines, rather than being transient or overly intense, promoting lasting behavioral change. The interviews revealed 2 key factors influencing user experience that are critical for achieving sustained success.

Seamless Integration With Meeting Software

A participant suggested adding prompts:

> *I think a pop-up prompt might remind me to use it more.* [P19]

Accordingly, it might be beneficial to incorporate reminder prompts on meeting platforms during breaks or between long sessions. However, the timing and delivery of these reminders need to be carefully considered. There is existing research indicating that notifications delivered at inappropriate times can distract users and potentially annoy them [19]. P21 expressed that seeing others engaging in antisedentary interactions could motivate them to participate as well. With seamless integration, users could be effectively prompted to initiate the bodily interactions.

Long-Term Reward Mechanisms

Extrinsic motivation arises from external factors, such as rankings and individual records [62]. Implementing these motivators (eg, points, badges, and leaderboards) in gamified bodily interactions can foster a competitive environment, thereby encouraging participants to remain engaged and enhance their performance. A participant noted that the leaderboard in Food Rain heightened competitiveness:

> *We played pretty aggressively because of the leaderboard.* [P13]

Another participant described the influence of the leaderboard as akin to "peer pressure" (P16). Furthermore, offering a variety of games can cater to diverse participant preferences and prevent monotony. Regular updates to existing games, including introducing "new game patterns" (P19) and "new visual and aural effects" (P13), are both convenient and effective strategies for sustaining player involvement. Indeed, this encapsulates the rationale behind our research use of gamification: to enhance extrinsic motivations for physical activity and reduce sedentary behavior.

Anticipated Use Scenarios and Additional Benefits: "It Could Be Used as an Icebreaker Activity"

When discussing the potential applications of gamified bodily interactions in online meetings, participants highlighted the possibility of integrating these games into various scenarios. Gamified bodily interactions can serve as icebreakers at the beginning of online meetings or during breaks, helping participants to become acquainted with each other and creating a comfortable atmosphere, as well as helping to break down communication barriers, especially in situations where participants have not met before or have had limited prior interactions; for instance, P13 expressed appreciation for the game Virus Hitter as a potential group icebreaker. P17 and P24 also highlighted the potential of such interactive games for serving as effective icebreakers in virtual settings:

> *It could be used as an icebreaker activity.* [P24]

This insight underscores the versatile applicability of gamified bodily interactions in online meetings, extending beyond mere physical engagement to fostering social connections.

In online meetings where decision-making is required, games can be used to facilitate discussions, promote active participation, and encourage team members to share their opinions. The interactive nature of these games can help maintain engagement during lengthy discussions and ensure that all voices are heard; for instance, a participant envisions the integration of these games as a creative tool for participant selection while simultaneously energizing the meeting atmosphere: "for example, to decide on a speaking order" (P13). In situations where a group of participants need to choose a member to speak or present but are unsure about whom to select, they can initiate the game and obtain a ranked list of candidates. Motivated by the ranking system and the prospect of being chosen to speak, participants are likely to engage actively in

XSL•FO
RenderX

the game and incorporate physical activity into their meeting experience. This innovative approach to participant selection not only encourages physical movement but also adds an element of excitement and competition, fostering a livelier and engaging online meeting environment.

Contextualizing the BIG-AOME Framework (RQ2)

Overview

In this section, we contextually examine and concretize the framework with the empirical data gathered from the evaluation sessions. As illustrated in Figure 7, participants were asked to position the 3 evaluated prototypes on each subdimension to verify initial design assumptions and gain an overview of how these design instances are distributed over this design space. Descriptive statistics are presented in Table 1. Using the gathered data, we analyze each design dimension and elaborate on relevant design choices to highlight the design space and considerations for future research.

Table 1. Descriptive statistics of user rating results across games and dimensions.

Dimensions and games	Ratings, mean (SD)	Ratings, median (IQR)
Exertion		
Food Rain	0.462 (0.251)	0.413 (0.306-0.528)
Virus Hitter	0.551 (0.156)	0.598 (0.466-0.649)
Frost	0.207 (0.109)	0.210 (0.158-0.263)
Stretch		
Food Rain	0.474 (0.276)	0.500 (0.278-0.636)
Virus Hitter	0.609 (0.210)	0.663 (0.453-0.750)
Frost	0.349 (0.144)	0.304 (0.250-0.418)
Body parts		
Food Rain	0.278 (0.172)	0.302 (0.157-0.430)
Virus Hitter	0.570 (0.212)	0.569 (0.419-0.750)
Frost	0.206 (0.220)	0.143 (0.090-0.236)
Attention		
Food Rain	0.781 (0.256)	0.781 (0.703-0.984)
Virus Hitter	0.523 (0.276)	0.577 (0.422-0.690)
Frost	0.228 (0.160)	0.235 (0.141-0.306)
Bodily interplay		
Food Rain	0.414 (0.270)	0.454 (0.220-0.594)
Virus Hitter	0.858 (0.176)	0.952 (0.799-1.000)
Frost	0.178 (0.273)	0.067 (0.000-0.248)
Duration		
Food Rain	0.478 (0.271)	0.485 (0.337-0.601)
Virus Hitter	0.491 (0.167)	0.500 (0.390-0.573)
Frost	0.503 (0.270)	0.523 (0.282-0.669)
Space type		
Food Rain	0.322 (0.310)	0.250 (0.080-0.395)
Virus Hitter	0.570 (0.311)	0.500 (0.330-0.784)
Frost	0.458 (0.340)	0.336 (0.225-0.722)

Bodily Engagement: How Does the Game Involve Bodies and Challenge Physically?

The first dimension of the framework focuses on the physical involvement and challenges presented by the game. This dimension encompasses 3 subdimensions: *exertion*, *stretch*, and *body parts*.

As shown in Figure 7, we can observe that the 3 prototypes follow a similar pattern in the 3 subdimensions: Virus Hitter involves the body more than Food Rain, which in turn involves the body more than Frost. This is in line with our initial design intentions. However, it is important to note that the ordering of the games in the 3 subdimensions is not always consistent; for instance, a game involving multiple body parts can still require

XSL•FO
RenderX

low exertion, such as matching the body's key points to a static shape—an idea proposed by the co-design participants.

Designs with lower bodily engagement levels appeal to a wider range of participants and involve less exertion, enabling players to engage more frequently or for longer periods. However, these designs may not effectively engage individuals who prefer high-energy activities or competitive environments, potentially leaving them feeling unchallenged or uninterested. By contrast, designs with higher bodily engagement levels offer more intense physical challenges. However, games requiring high bodily engagement may not be suitable for all participants due to physical limitations or personal preferences, leading some individuals to feel excluded or frustrated if they cannot fully engage in the activities; for example, after playing Virus Hitter, a participant said they were "too tired" (P19). Another participant stated as follows:

> *The last game is more obvious [a reference to the bodily engagement levels of Virus Hitter], the first two [Frost and Food Rain] are relatively easy.* [P12]

This highlights the varying levels of bodily engagement across the 3 games, confirming the noticeable differences in the physical challenges they offer.

Attention: What Level of Focus Is Required?

It can be observed from Figure 7 that Food Rain demands the highest level of attention, followed by Virus Hitter and Frost. The 3 games span almost the entire *attention* dimension, providing players with varying levels of cognitive engagement. This diverse range of attention levels can cater to the different preferences and scenarios in online meetings.

Bodily interactions that require low attention levels allow participants to engage without disrupting the meeting. They can maintain focus on the meeting content or personal tasks simultaneously, ensuring that the overall flow of the meeting remains uninterrupted while participating in peripheral bodily interactions. In addition, incorporating a design that requires only peripheral attention during a meeting break can be an effective way to encourage relaxation and mental rejuvenation. Participants can engage in the game while also performing other tasks, such as drinking water:

> *I can do something else [during a break while experiencing the gamified interaction].* [P19]

However, games that require only peripheral attention might not fully engage participants or provide a deeply immersive experience. P21 shared that Frost is "not like a game" and therefore they felt less enthusiasm to play it. Games that require high attention levels can create a more engaging and immersive experience. Moreover, such games can offer a complete shift from the meeting content, allowing participants to refresh their minds before returning to the meeting activities. However, playing games requiring high attention levels can lead to greater mental stress or cognitive load for participants.

Bodily Interplay: To What Extent Can Participants' Bodies Act Upon and React to Each Other in Gamified Bodily Interactions in Online Meetings?

As shown in Figure 7, Virus Hitter demonstrates a significantly higher level of bodily engagement than Food Rain and Frost. Collectively, the prototypes cover a wide range of bodily movements in this dimension.

Games requiring highly parallel bodily interplay provide flexibility in terms of participants' availability, allowing individuals to engage or disengage without impacting others. This feature is especially suitable for online meeting scenarios where participants may already be under mental strain. With minimal reliance on other players, individuals may feel less pressured to excel, fostering a more relaxed, enjoyable experience for those seeking less competitive interactions. However, such games may become repetitive and less captivating over time due to the limited participant interaction and challenge levels. Conversely, games requiring highly interdependent bodily interplay can foster collaboration and communication, potentially strengthening team dynamics. Such games, particularly when they incorporate competitive or cooperative elements, can enhance participants' engagement and motivation. Nonetheless, they also present challenges because the players' enjoyment and success depend heavily on their teammates' or opponents' actions, which can impact the overall experience.

Timeliness: When Should the Interaction Begin? How Long Should It Last?

As illustrated in Figure 7, regarding the preferred start time for each game, all participants suggested that both Food Rain and Virus Hitter are better suited for initiation during session breaks. Most participants felt that it would be more appropriate to initiate Frost in the middle of meeting activities. These insights correspond well with our design expectations, suggesting that our strategy regarding optimal game timings aligns with participant preferences.

Incorporating a game during a meeting break presents several benefits, including minimizing meeting disruption, facilitating social interaction among participants, and potentially enhancing team relationships. However, there are also drawbacks. Participants might prefer to use breaks for other tasks, such as attending to personal needs or engaging in work-related discussions. Furthermore, if some participants do not wish to participate in games during breaks, they might feel excluded or pressured to join. Conversely, initiating a game midway through a meeting offers a unique chance to invigorate participants. However, maintaining the right balance of attention is essential. Games incorporated into a meeting must be thoughtfully designed to avoid disrupting the flow or causing participants to lose focus on the main agenda.

Concerning the *duration* dimension, we noticed an intriguing pattern. All 3 games are positioned in the middle of the dimension, displaying minor differences. Our results indicate that user experiences across the 3 prototypes do not show considerable variations within the *duration* dimension. Short-duration games offer ease of integration into online

XSL•FO
RenderX

meetings, serving as refreshing breaks that do not demand much time. They serve as effective transitions between meeting segments or as energizing preludes to more serious discussions. Offering frequent opportunities for physical and mental breaks, these games counteract sedentary behavior effectively. Long-duration games promise a more immersive experience with complex mechanics, deeper narratives, and enhanced team-building opportunities. However, fitting them into the meeting agenda without disrupting the flow can be challenging, and extended play might lead to participant fatigue if the game is too demanding.

Virtual and Physical Environments: What Is the Optimum Videoconferencing Window Layout? Should Space Type Be Highly Private or Highly Public?

As shown in Figure 7, most participants believed that Virus Hitter should be experienced in an asymmetrical layout, while Food Rain and Frost should be experienced in a symmetrical layout. This aligns with our design considerations.

An asymmetrical layout, where a participant shares their screen, provides an expanded display space, potentially enhancing the game's visual impact. However, this might inadvertently shift focus toward the sharer's content, potentially reducing interaction with other attendees or causing discomfort for the sharer under scrutiny. A symmetrical layout, with equally sized videoconferencing windows for all participants, creates an equitable and harmonious environment, encouraging shared focus on the sharer and others. However, the smaller videoconferencing window size may impact game visibility and engagement, especially in larger meetings. Therefore, careful consideration and balance in the design process are essential to optimize user experience and participation.

Concerning the *space type* dimension, Food Rain was viewed as more suitable for private spaces, while Virus Hitter and Frost inhabit the middle ground. Playing in a private space offers comfort and security, allowing for uninhibited movement and gameplay without fear of social awkwardness or disturbing others. However, this may reduce social interaction and potentially lower motivation due to a lack of group dynamics. By contrast, playing in a public space enhances social interaction and participation, fostering a sense of community and group camaraderie. The visible engagement of others can encourage individual participation. However, playing in a public space can lead to feelings of self-consciousness during gameplay, potentially affecting enjoyment and immersion.

Discussion

Design Implications and Opportunities for Creating Antisedentary Online Meeting Environments

Overview

Our study aligns with the studies by Warburton et al [63] and Aldenaini et al [15], which have demonstrated the benefits of gamification and persuasive technology for antisedentary designs in general. In the context of online teaching, Shin et al [45] and Sachan and Peiris [46] conducted pilot studies to combine bodily interactions with online classroom scenarios. We further explored the vast design space and highlighted its underlying design implications. Furthermore, contributing to a design-oriented approach to this domain, we created prototypes for evaluation, enabling us to gather real-world experiences of participants within an online meeting environment. Here, we discuss 4 key design implications that emerged from our study.

Providing Proper and Playful Reasons for Moving the Body During Online Meetings

Sedentary behavior is regulated by personal habit strength [64] and is often socially and environmentally reinforced [13]. Many new technological interventions have been designed to reduce sedentary time, including games, reminders, wearable devices, and smart office equipment [42,65-67]. These interventions have typically adopted action planning as the key mechanism. However, an online meeting is a unique context where users are under additional stressors such as close-up, "face-to-face" communication and reduced mobility [47] or distractions at home and online [24]. Participants also reported that during online meetings they would feel highly embarrassed to stand up or move their body as they wanted even if they were feeling very tired. In such a situation, compared to a time-based prompt for a break, an incentive cue and a compelling reason for a break might appeal more to meeting participants.

Medical evidence shows that the quantified threshold for sedentary time varies by health condition, age, sex, and so on [13]. There are no rigorous recommendations about the optimum limit for sedentary time [68]. A study investigating the motivational processes underlying sedentary behaviors shows that action planning has a conditional effect on physical activity but no effect on limiting sedentary behavior [64]. This suggests that the focus should perhaps not be on setting strict activity goals or tracking exercise sessions. Instead, encouraging movement and breaking up sedentary patterns whenever possible may be more beneficial. This does not necessitate high-intensity training or strenuous workouts. Even light or moderate movements incorporated into online meetings can make a significant difference [69]. From a design research perspective, the challenge lies in creating socially engaging designs that provide meeting attendees with enjoyable, playful, and socially motivating reasons to engage in movement. Rather than strictly monitoring and regulating activity, the goal is to create an environment that naturally encourages movement and reduces sedentary behavior.

Designing Unobtrusive Bodily Movements as Secondary Tasks to Leverage Positive Multitasking in Online Meetings

Multitasking is a common behavior in online meetings, and it can have both positive and negative impacts [25]. Incorporating nondistracting, low-effort physical movements as secondary tasks during online meetings may serve as an effective design strategy for positive multitasking. Games with varying attention requirements offer diverse options for attention management, enabling participants to fluidly transition between relaxation and meeting focus; for instance, games that demand minimal attention can serve as a peripheral activity during the meeting. One such example is *Frost*, which features a minimalistic visual interface and is intentionally nondistracting, making it suitable

XSL•FO
RenderX

as a secondary task. The level of required attention is a crucial consideration in game design. By consciously defining the design goals and leveraging the benefits of positive multitasking, we can foster an antisedentary environment within online meeting routines.

Aiming for Fine-Grained Integration With Routines, Rather Than Intensity or Volume of Physical Activity

Our work presents a unique proposition that deviates from conventional exertion game design aimed at sports training or physical activity tracking. In addressing sedentary behaviors during online meetings, our emphasis is not on the intensity or volume of physical activity. Instead, we prioritize designing interactions that are mild in intensity, of moderate duration easily initiated, and accessible to most participants. These interactions allow for flexibility, enabling users to engage with or dismiss them as needed.

In terms of mitigating sedentary behaviors, as prior research argues, the goal is fulfilled not merely by increasing exertion intensity but also by incorporating light physical exercises [70] into daily working scenarios. It is important to note that sedentary behavior is completely different from a lack of physical activity. Most current exertion games are considered augmented sport [33], relying on game consoles and sensor equipment to create a game-like sports experience, thereby promoting physical activity [36]. Our work represents a different approach to exertion games: our games are designed specifically as playful antisedentary interventions. Instead of using these games solely for workouts or ad hoc entertainment, our objective is to seamlessly integrate them into online meeting routines. This integration aims to foster more lightweight, healthy behaviors among participants during their regular online engagements [42,71].

Expanding Design Variety and Integrating Games Into Existing Meeting Platforms to Enhance Real-World Impacts

In our exploration of the potential of gamified physical interactions within online meeting scenarios, we introduce the BIG-AOME framework. This framework serves as an initial foray into a 5D design space, revealing a broad spectrum of design opportunities that can guide the creation of a diverse array of experiences tailored to various user preferences. Our objective with the BIG-AOME framework is to equip both researchers and designers with a foundational tool to further investigate this innovative approach. By promoting the integration of exertion games [33] with persuasive interventions [14], we aim to inspire a collaborative effort in the development of engaging and health-promoting solutions tailored for online environments.

Recognizing the diversity in user preferences and motivational factors, it is vital to adopt a personalized approach to game-based health interventions. An individual's engagement with the game is influenced by factors such as motivation, interests, athletic abilities, social connections, and social status [72]. Hence, to inspire long-term positive behavior changes, it is crucial to craft inclusive and engaging experiences that resonate with the varied expectations of diverse users. During the evaluation phase, we noticed a diversity of user preferences for the different games, which signals a degree of success in catering to our audience's varied expectations. Nevertheless, we acknowledge that numerous areas across the various dimensions and subdimensions remain unexplored. This underscores the need for more diversified, design-driven research. By generating new design instances and extracting new design implications, designs in this domain could evolve toward more personalized, engaging, and effective game-based interventions.

Moreover, the importance of seamless integration with existing meeting platforms cannot be overstated. This echoes the principle proposed by Mandryk et al [42]: "providing an easy entry into play." Given the social nature of online meetings, users are often motivated to engage when they observe other users doing so on the platform. This affords new ways to increase user involvement [41]: with the growing ubiquity of online meetings, these platforms naturally present opportunities for an increased presence of gamified bodily interactions. By incorporating these interactions into existing meeting software and delivering appropriate signifiers about this new feature, users could become more conscious of their sedentary habits and are consequently more likely to participate in physical activities.

Limitations and Future Work

While our study contributes valuable insights into gamified physical activities within online meetings, it is essential to acknowledge certain limitations that necessitate further research. First, the evaluation period of our study was relatively brief, which may not adequately capture the long-term effects of these interventions. Future research should include longer evaluation periods to investigate more thoroughly the sustainability and long-term impact of such gamified interventions. Second, although we developed 3 high-fidelity prototypes that varied across different dimensions, we did not comprehensively explore some subdimensions. Consequently, while the data collected offers significant insights, these limitations may restrict the generalizability and robustness of our findings. Despite these constraints, we are optimistic about the extensive and promising potential of gamified bodily interactions in online meetings. Future studies could aim to experiment with more diverse and inclusive design features covering a wider range of physical activities. Expanding the scope of the interventions to include a broader variety of bodily movements and postures will enhance our understanding of the potential and effectiveness of integrating gamified physical activities into online meeting environments. In addition, investigating how gamified bodily interactions can stimulate users' intrinsic motivation to support sustained engagement over the long term is another valuable direction for future research.

Conclusions

We used a research through design methodology to craft gamified bodily interactions and explore their potential as antisedentary interventions within online meeting contexts. In collaboration with 11 users, we co-designed and iterated 3 prototypes, which led to the development of the BIG-AOME framework. Using these prototypes, we conducted user studies

XSL•FO
RenderX

with 3 groups totaling 15 participants. Empirical findings were gathered to understand user experiences with these prototypes and concretize the framework. The research findings indicate that antisedentary bodily interactions designed for online meetings have the potential to reduce sedentary behaviors while enhancing social connections. Furthermore, the proposed BIG-AOME framework explores the design space for antisedentary physical interactions in the context of online meetings.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Video demonstration.
[MP4 File (MP4 Video), 136133 KB - games_v13i1e62778_app1.mp4]

Multimedia Appendix 2
Gamified bodily interaction evaluation scale.
[PDF File (Adobe PDF File), 85 KB - games_v13i1e62778_app2.pdf]

Multimedia Appendix 3
Semistructured interview guide.
[PDF File (Adobe PDF File), 67 KB - games_v13i1e62778_app3.pdf]

References

1. Kaptelinin V, Danielsson K, Kaiser N, Kuenen C, Nordin M. Understanding the interpersonal space of online meetings: an exploratory study of “we-ness”. In: Proceedings of the Companion Publication of the 2021 Conference on Computer Supported Cooperative Work and Social Computing. 2021 Presented at: CSCW '21 Companion; October 23-27, 2021; Virtual Event. [doi: 10.1145/3462204.3481780]
2. Wu R, Yu Z. The influence of social isolation, technostress, and personality on the acceptance of online meeting platforms during the COVID-19 pandemic. Int J Hum Comput Interact 2022 Jul 19;39(17):3388-3405. [doi: 10.1080/10447318.2022.2097779]
3. Nesher Shoshan H, Wehrt W. Understanding “Zoom fatigue”: a mixed‐method approach. Appl Psychol 2021 Nov 14;71(3):827-852. [doi: 10.1111/apps.12360]
4. Peper E, Wilson V, Martin M, Rosegard E, Harvey R. Avoid Zoom fatigue, be present and learn. NeuroRegulation 2021 Mar 29;8(1):47-56. [doi: 10.15540/nr.8.1.47]
5. Riedl R. On the stress potential of videoconferencing: definition and root causes of Zoom fatigue. Electron Mark 2022;32(1):153-177 [FREE Full text] [doi: 10.1007/s12525-021-00501-3] [Medline: 35600914]
6. Mavrovouniotis F. Inactivity in childhood and adolescence: a modern lifestyle associated with adverse health consequences. Sport Sci Rev 2012 Aug;XXI(3-4):75-99. [doi: 10.2478/v10237-012-0011-9]
7. Ford ES, Caspersen CJ. Sedentary behaviour and cardiovascular disease: a review of prospective studies. Int J Epidemiol 2012 Oct;41(5):1338-1353 [FREE Full text] [doi: 10.1093/ije/dys078] [Medline: 22634869]
8. Liese AD, Ma X, Maahs DM, Trilk JL. Physical activity, sedentary behaviors, physical fitness, and their relation to health outcomes in youth with type 1 and type 2 diabetes: a review of the epidemiologic literature. J Sport Health Sci 2013 Mar;2(1):21-38. [doi: 10.1016/j.jshs.2012.10.005]
9. Tremblay MS, Aubert S, Barnes JD, Saunders TJ, Carson V, Latimer-Cheung AE, et al. Sedentary behavior research network (SBRN) - terminology consensus project process and outcome. Int J Behav Nutr Phys Act 2017 Jun 10;14(1):75 [FREE Full text] [doi: 10.1186/s12966-017-0525-8] [Medline: 28599680]
10. Luo Y, Lee B, Wohn DY, Rebar AL, Conroy DE, Choe E. Time for break: understanding information workers' sedentary behavior through a break prompting system. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018 Presented at: CHI '18; April 21-26, 2018; Montreal, QC. [doi: 10.1145/3173574.3173701]
11. Atkin AJ, Gorely T, Clemes SA, Yates T, Edwardson C, Brage S, et al. Methods of measurement in epidemiology: sedentary behaviour. Int J Epidemiol 2012 Oct;41(5):1460-1471 [FREE Full text] [doi: 10.1093/ije/dys118] [Medline: 23045206]
12. Young DR, Hivert MF, Alhassan S, Camhi SM, Ferguson JF, Katzmarzyk PT, et al. Sedentary behavior and cardiovascular morbidity and mortality: a science advisory from the American Heart Association. Circulation 2016 Sep 27;134(13):e262-e279. [doi: 10.1161/CIR.0000000000000440] [Medline: 27528691]
13. Dempsey PC, Biddle SJ, Buman MP, Chastin S, Ekelund U, Friedenreich CM, et al. New global guidelines on sedentary behaviour and health for adults: broadening the behavioural targets. Int J Behav Nutr Phys Act 2020 Nov 26;17(1):151 [FREE Full text] [doi: 10.1186/s12966-020-01044-0] [Medline: 33239026]

XSL•FO
RenderX

14. Wang Y, Wu L, Lange JP, Fadhil A, Reiterer H. Persuasive technology in reducing prolonged sedentary behavior at work: a systematic review. Smart Health 2018 Jun;7-8:19-30. [doi: 10.1016/j.smhl.2018.05.002]
15. Aldenaini N, Alqahtani F, Orji R, Sampalli S. Trends in persuasive technologies for physical activity and sedentary behavior: a systematic review. Front Artif Intell 2020 Apr 28;3:7 [FREE Full text] [doi: 10.3389/frai.2020.00007] [Medline: 33733127]
16. Wang Y, Reiterer H. The point-of-choice prompt or the always-on progress bar?: a pilot study of reminders for prolonged sedentary behavior change. In: Proceedings of the Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 Presented at: CHI EA '19; May 4-9, 2019; Glasgow, UK. [doi: 10.1145/3290607.3313050]
17. Morris D, Brush AJ, Meyers BR. SuperBreak: using interactivity to enhance ergonomic typing breaks. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2008 Presented at: CHI '08; April 5-10, 2008; Florence, Italy. [doi: 10.1145/1357054.1357337]
18. van Dantzig S, Geleijnse G, van Halteren AT. Toward a persuasive mobile application to reduce sedentary behavior. Pers Ubiquitous Comput 2012 Jul 12;17:1237-1246. [doi: 10.1007/s00779-012-0588-0]
19. Ojo SO, Bailey DP, Chater AM, Hewson DJ. Workplace intervention for reducing sitting time in sedentary workers: protocol for a pilot study using the behavior change wheel. Front Public Health 2022 Apr 12;10:832374 [FREE Full text] [doi: 10.3389/fpubh.2022.832374] [Medline: 35493386]
20. Ren X, Yu B, Lu Y, Chen Y, Pu P. HealthSit: designing posture-based interaction to promote exercise during fitness breaks. Int J Hum Comput Interact 2018 Aug 07;35(10):870-885. [doi: 10.1080/10447318.2018.1506641]
21. Cambo SA, Avrahami D, Lee ML. BreakSense: combining physiological and location sensing to promote mobility during work-breaks. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. 2017 Presented at: CHI '17; May 6-11, 2017; Denver, CO. [doi: 10.1145/3025453.3026021]
22. Bleakley A, Rough D, Edwards J, Doyle P, Dumbleton O, Clark L, et al. Bridging social distance during social distancing: exploring social talk and remote collegiality in video conferencing. Hum Comput Interact 2021 Nov 11;37(5):404-432 [FREE Full text] [doi: 10.1080/07370024.2021.1994859]
23. Warkentin ME, Sayeed L, Hightower R. Virtual teams versus face-to-face teams: an exploratory study of a web-based conference system. Decis Sci 2007 Jun 07;28(4):975-996. [doi: 10.1111/j.1540-5915.1997.tb01338.x]
24. Das M, Tang J, Ringland KE, Piper AM. Towards accessible remote work: understanding work-from-home practices of neurodivergent professionals. Proc ACM Hum Comput Interact 2021 Apr 22;5(CSCW1):1-30. [doi: 10.1145/3449282]
25. Cao H, Lee CJ, Iqbal S, Czerwinski M, Wong PN, Rintel S, et al. Large scale analysis of multitasking behavior during remote meetings. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. 2021 Presented at: CHI '21; May 8-13, 2021; Yokohama, Japan. [doi: 10.1145/3411764.3445243]
26. Deterding S, Dixon D, Khaled R, Nacke L. From game design elements to gamefulness: defining "gamification". In: Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments. 2011 Presented at: MindTrek '11; September 28-30, 2011; Tampere, Finland. [doi: 10.1145/2181037.2181040]
27. Lieder F, Chen PZ, Prentice M, Amo V, Tošić M. Gamification of behavior change: mathematical principle and proof-of-concept study. JMIR Serious Games 2024 Mar 22;12:e43078 [FREE Full text] [doi: 10.2196/43078] [Medline: 38517466]
28. Patel MS, Small DS, Harrison JD, Fortunato MP, Oon AL, Rareshide CA, et al. Effectiveness of behaviorally designed gamification interventions with social incentives for increasing physical activity among overweight and obese adults across the United States: the STEP UP randomized clinical trial. JAMA Intern Med 2019 Dec 01;179(12):1624-1632 [FREE Full text] [doi: 10.1001/jamainternmed.2019.3505] [Medline: 31498375]
29. Cugelman B. Gamification: what it is and why it matters to digital health behavior change developers. JMIR Serious Games 2013 Dec 12;1(1):e3 [FREE Full text] [doi: 10.2196/games.3139] [Medline: 25658754]
30. Schmidt-Kraepelin M, Warsinsky S, Thiebes S, Sunyaev A. The role of gamification in health behavior change: a review of theory-driven studies. Scholar Space. 2020 Jan 07. URL: http://hdl.handle.net/10125/63894 [accessed 2024-05-27]
31. Hamari J, Koivisto J, Sarsa H. Does gamification work? -- a literature review of empirical studies on gamification. In: Proceedings of the 47th Hawaii International Conference on System Sciences. 2014 Presented at: HICSS 2014; January 6-9, 2014; Waikoloa, HI. [doi: 10.1109/hicss.2014.377]
32. Mueller FF, Edge D, Vetere F, Gibbs MR, Agamanolis S, Bongers B, et al. Designing sports: a framework for exertion games. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2011 Presented at: CHI '11; May 7-12, 2011; Vancouver, BC. [doi: 10.1145/1978942.1979330]
33. Mueller F, Khot RA, Gerling K, Mandryk R. Exertion games. Found Trends Hum Comput Interact 2016;10(1):1-86. [doi: 10.1561/1100000041]
34. Berkovsky S, Freyne J, Coombe M. Physical activity motivating games: be active and get your own reward. ACM Trans Comput Hum Interact 2012 Dec 01;19(4):1-41. [doi: 10.1145/2395131.2395139]
35. Gao Y, Mandryk RL. GrabApple: the design of a casual exergame. In: Proceedings of the 10th International Conference on Entertainment Computing. 2011 Presented at: ICEC 2011; October 5-8, 2011; Vancouver, BC. [doi: 10.1007/978-3-642-24500-8]

XSL•FO
RenderX

36. Hernandez HA, Graham TC, Fehlings D, Switzer L, Ye Z, Bellay Q, et al. Design of an exergaming station for children with cerebral palsy. In: Proceedings of the CHI Conference on Human Factors in Computing Systems. 2012 Presented at: CHI '12; May 5-10, 2012; Austin, TX. [doi: 10.1145/2207676.2208652]
37. Lin SC, Lee JY, Yang Y, Fang CC, Fang HL, Hou TH. Exploring the design of upper limb strength training through high-intensity interval training combined with exergaming: usability study. JMIR Serious Games 2024 Apr 17;12:e51730 [FREE Full text] [doi: 10.2196/51730] [Medline: 38632713]
38. Manser P, Adcock-Omlin M, de Bruin ED. Design considerations for an exergame-based training intervention for older adults with mild neurocognitive disorder: qualitative study including focus groups with experts and health care professionals and individual semistructured in-depth patient interviews. JMIR Serious Games 2023 Jan 05;11:e37616 [FREE Full text] [doi: 10.2196/37616] [Medline: 36602851]
39. O'Loughlin EK, Sabiston CM, O'Rourke RH, Bélanger M, Sylvestre MP, O'Loughlin JL. The change in exergaming from before to during the COVID-19 pandemic among young adults: longitudinal study. JMIR Serious Games 2023 May 22;11:e41553 [FREE Full text] [doi: 10.2196/41553] [Medline: 36952329]
40. Mueller F, Agamanolis S, Picard R. Breakout for two: an example of an exertion interface for sports over a distance. MIT Media Lab. URL: https://web.media.mit.edu/~stefan/hc/publications/Mueller03BreakoutUBICOMP.pdf [accessed 2025-03-19]
41. Mueller FF, Gibbs MR, Vetere F, Edge D. Designing for bodily interplay in social exertion games. ACM Transact Comput Hum Interact 2017 May 10;24(3):1-41. [doi: 10.1145/3064938]
42. Mandryk RL, Gerling KM, Stanley KG. Designing games to discourage sedentary behaviour. In: Nijholt A, editor. Playful User Interfaces. Singapore, Singapore: Springer; Feb 18, 2014.
43. Windasari NA, Lin F. Why do people continue using fitness wearables? The effect of interactivity and gamification. Sage Open 2021 Nov 03;11(4):215824402110566. [doi: 10.1177/21582440211056606]
44. Škola F, Tinková S, Liarokapis F. Progressive training for motor imagery brain-computer interfaces using gamification and virtual reality embodiment. Front Hum Neurosci 2019 Sep 26;13:329 [FREE Full text] [doi: 10.3389/fnhum.2019.00329] [Medline: 31616269]
45. Shin S, Cho J, Kim SW. Jumple: interactive contents for the virtual physical education classroom in the pandemic era. In: Proceedings of the Augmented Humans International Conference 2021. 2021 Presented at: AHs '21; February 22-24, 2021; Rovaniemi, Finland. [doi: 10.1145/3458709.3458964]
46. Sachan MS, Peiris RL. Designing augmented reality based interventions to encourage physical activity during virtual classes. In: Proceedings of the CHI Conference on Human Factors in Computing Systems Extended Abstracts. 2022 Presented at: CHI EA '22; April 29-May 5, 2022; New Orleans, LA. [doi: 10.1145/3491101.3519749]
47. Bailenson JN. Nonverbal overload: a theoretical argument for the causes of Zoom fatigue. Technol Mind Behav 2021 Feb 23;2(1). [doi: 10.1037/tmb0000030]
48. Rudnicka A, Cook D, Cecchinato ME, Gould SJ, Newbold JW, Cox AL. The end of the active work break? Remote work, sedentariness and the role of technology in creating active break-taking norms. In: Proceedings of the 1st Annual Meeting of the Symposium on Human-Computer Interaction for Work. 2022 Presented at: CHIWORK '22; June 8-9, 2022; Durham, NH. [doi: 10.1145/3533406.3533409]
49. Khan M, Wantlin K, Patel Z, Glassman E, Maess P. Changing computer-usage behaviors: what users want, use, and experience. In: Proceedings of the Asian CHI Symposium 2021. 2021 Presented at: Asian CHI '21; May 8-13, 2021; Yokohama, Japan. [doi: 10.1145/3429360.3468180]
50. Epstein DA, Avrahami D, Biehl JT. Taking 5: work-breaks, productivity, and opportunities for personal informatics for knowledge workers. In: Proceedings of the CHI Conference on Human Factors in Computing Systems. 2016 Presented at: CHI '16; May 7-12, 2016; San Jose, CA. [doi: 10.1145/2858036.2858066]
51. Karl KA, Peluchette JV, Aghakhani N. Virtual work meetings during the COVID-19 pandemic: the good, bad, and ugly. Small Group Res 2022 Jun;53(3):343-365 [FREE Full text] [doi: 10.1177/10464964211015286] [Medline: 38603094]
52. Marlow J, van Everdingen E, Avrahami D. Taking notes or playing games?: understanding multitasking in video communication. In: Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing. 2016 Presented at: CSCW '16; February 27-March 2, 2016; San Francisco, CA. [doi: 10.1145/2818048.2819975]
53. Zimmerman J, Forlizzi J, Evenson S. Research through design as a method for interaction design research in HCI. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2007 Presented at: CHI '07; April 28-May 3, 2007; San Jose, CA. [doi: 10.1145/1240624.1240704]
54. Open broadcaster software homepage. Open Broadcaster Software. URL: https://obsproject.com/ [accessed 2024-05-22]
55. Welcome to MedlinePlus. MedlinePlus. URL: https://medlineplus.gov/ [accessed 2025-03-21]
56. Centers for Disease Control and Prevention homepage. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/index.html [accessed 2025-03-21]
57. World Health Organization homepage. World Health Organization. URL: https://www.who.int/ [accessed 2025-03-21]
58. Mueller FF, Byrne R, Andres J, Patibanda R. Experiencing the body as play. In: Proceedings of the CHI Conference on Human Factors in Computing Systems. 2018 Presented at: CHI '18; April 21-26, 2018; Montreal, QC. [doi: 10.1145/3173574.3173784]

XSL•FO
RenderX

59. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
60. Steinberg L. 6 yoga exercises you won't be embarrassed to do at your desk. Harvard Business Review. 2015 Dec 15. URL: https://hbr.org/2015/12/6-yoga-exercises-you-wont-be-embarrassed-to-do-at-your-desk [accessed 2024-05-22]
61. Chair workout - the great slim down. YouTube. URL: https://www.youtube.com/watch?v=HfrkguDu1c8 [accessed 2024-05-22]
62. Ryan RM, Deci EL. Intrinsic and extrinsic motivations: classic definitions and new directions. Contemp Educ Psychol 2000 Jan;25(1):54-67. [doi: 10.1006/ceps.1999.1020] [Medline: 10620381]
63. Warburton DE, Bredin SS, Horita LT, Zbogar D, Scott JM, Esch BT, et al. The health benefits of interactive video game exercise. Appl Physiol Nutr Metab 2007 Aug;32(4):655-663. [doi: 10.1139/H07-038] [Medline: 17622279]
64. Maher JP, Conroy DE. Habit strength moderates the effects of daily action planning prompts on physical activity but not sedentary behavior. J Sport Exerc Psychol 2015 Feb;37(1):97-107. [doi: 10.1123/jsep.2014-0258] [Medline: 25730895]
65. Torbeyns T, de Geus B, Bailey S, Decroix L, Meeusen R. The potential of bike desks to reduce sedentary time in the office: a mixed-method study. Public Health 2017 Mar;144:16-22. [doi: 10.1016/j.puhe.2016.11.006] [Medline: 28274379]
66. Ellingson LD, Meyer JD, Cook DB. Wearable technology reduces prolonged bouts of sedentary behavior. Transl J Am Coll Sport Med 2016 Apr 15;1(2):10-17. [doi: 10.1249/TJX.0000000000000001]
67. Thomsen T, Aadahl M, Beyer N, Hetland ML, Løppenthin K, Midtgaard J, et al. Motivational counselling and SMS-reminders for reduction of daily sitting time in patients with rheumatoid arthritis: a descriptive randomised controlled feasibility study. BMC Musculoskelet Disord 2016 Oct 18;17(1):434 [FREE Full text] [doi: 10.1186/s12891-016-1266-6] [Medline: 27756265]
68. Yoo S, Gough P, Kay J. Embedding a VR game studio in a sedentary workplace: use, experience and exercise benefits. In: Proceedings of the CHI Conference on Human Factors in Computing Systems. 2020 Presented at: CHI '20; April 25-30, 2020; Honolulu, HI. [doi: 10.1145/3313831.3376371]
69. Morton S, Fitzsimons C, Sivaramakrishnan D, Jepson R, Niven A. "Are we working (too) comfortably?": a focus group study to understand sedentary behaviour when working at home and identify intervention strategies. BMC Public Health 2024 Jun 06;24(1):1516 [FREE Full text] [doi: 10.1186/s12889-024-18892-1] [Medline: 38844904]
70. Levine JA, Eberhardt NL, Jensen MD. Role of nonexercise activity thermogenesis in resistance to fat gain in humans. Science 1999 Jan 08;283(5399):212-214. [doi: 10.1126/science.283.5399.212] [Medline: 9880251]
71. Ren X, An P, Bekker T, Chen Y, Ashok Khot R, ten Bhömer M, et al. Weaving healthy behaviors into new technology routines: designing in (and for) the COVID-19 work-from-home period. In: Proceedings of the Companion Publication of the 2020 ACM Designing Interactive Systems Conference. 2020 Presented at: DIS' 20 Companion; July 6-10, 2020; Eindhoven, The Netherlands. [doi: 10.1145/3393914.3395911]
72. Xu Y, Poole ES, Miller AD, Eiriksdottir E, Kestranek D, Catrambone R, et al. This is not a one-horse race: understanding player types in multiplayer pervasive health games for youth. In: Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work. 2012 Presented at: CSCW '12; February 11-15, 2012; Seattle, WA. [doi: 10.1145/2145204.2145330]

Abbreviations

BIG-AOME: Bodily Interaction Gamification towards Anti-sedentary Online Meeting Environments
RQ: research question

Edited by S Liu; submitted 31.05.24; peer-reviewed by A Faganel, EM Grech, AKK Priyo; comments to author 24.09.24; revised version received 19.11.24; accepted 27.01.25; published 18.04.25.

Please cite as:
Jiang J, Li S, Li X, Xu Y, Zhao J, An P
Playful Antisedentary Interactions for Online Meeting Scenarios: A Research Through Design Approach
JMIR Serious Games 2025;13:e62778
URL: https://games.jmir.org/2025/1/e62778
doi:10.2196/62778
PMID:

©Jiaqi Jiang, Shanghao Li, Xian Li, Yingxin Xu, Jian Zhao, Pengcheng An. Originally published in JMIR Serious Games (https://games.jmir.org), 18.04.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic

XSL•FO
RenderX

information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Experience of and Needs for Exergames in Older Adults With Mild Cognitive Impairment: Qualitative Interview Study

Xi Chen[1*], PhD; Dian Jiang[1*], PhD; Hongting Ning[2], PhD; Lina Wu[3], PhD; Yifei Chen[4], MD; Chi Zhang[1], PhD; Ruotong Peng[1], MD; Yishu Zhu[1], MD; Hui Feng[1], PhD

[1]Xiangya School of Nursing, Central South University, Changsha, China

[2]School of Nursing, Hunan University of Chinese Medicine, Changsha, China

[3]Sun Yat-sen University School of Nursing, Guangzhou, China

[4]Shanxi Technology and Business University, Taiyuan, China

[*]these authors contributed equally

Corresponding Author:
Hui Feng, PhD
Xiangya School of Nursing, Central South University
No. 172, Tongzipo Road
Yuelu District
Changsha, 410013
China
Phone: 86 15173121969
Email: feng.hui@csu.edu.cn

Abstract

Background: As a novel intervention method that combines exercise and games, exergames have demonstrated a positive impact on enhancing the cognitive and physical functions of older adults with mild cognitive impairment (MCI). However, there remains a dearth of knowledge and evidence regarding the experiences and needs of the older adult population in China with MCI about exergames.

Objective: This qualitative study aimed to investigate the experience of and needs for exergames among older adults with MCI.

Methods: We adopted a phenomenological methodology for this study, and conducted it at a community and nursing home in Changsha, Hunan Province, from June to August 2023. We used the purpose sampling method to conduct semistructured interviews with 21 older people with MCI. Older people with MCI were allowed to experience exergames using our preselected exergame device, the Nintendo Switch, and they were interviewed to understand their experience and needs for exergames. The interviews were recorded and transcribed verbatim, and the data were uploaded to NVivo 12 software for encoding. The corresponding text was then reviewed for data analysis. Data analysis was guided by the methodology proposed by Giorgi and was carried out simultaneously with data collection. This study's trustworthiness was evaluated according to credibility, dependability, confirmability, and transferability criteria.

Results: Overall, 21 participants (mean age 70.2, SD 7.6 y; n=17, 81% women; mean Montreal Cognitive Assessment score 18.8, SD 3.6) were interviewed. Moreover, 21 interviews were conducted. By the 18th interview, the data were saturated, and to make sure no new topics came up, we conducted 3 more interviews. The experience of older people with MCI with exergames includes five parts: their attitudes toward exergames vary, they are both entertaining and interesting, they promote physical activity and exercise, they pass the time and relieve loneliness, and their conditions of use are not restricted. The needs of older people with MCI for exergames include the desire to design older people–friendly exergames, ensure scientific validity and safety in the process of sports, provide a good gaming experience, exercise physical and cognitive function, and provide support and training.

Conclusions: This study provides an interpretative understanding of the experiences and needs associated with exergames in older people with MCI, which could inform exergame development appropriate for this population and guide the implementation of exergame interventions in this population. Most older people with MCI expressed a positive attitude toward exergames, but not all were interested in them. Older people with MCI viewed exergames as both entertaining and fun, promoting physical activity and exercise, passing the time, relieving loneliness, and the conditions of use were not restricted. Exergames for older people with MCI should be older people–friendly, scientific, safe, provide a good play experience, exercise physical and cognitive

XSL•FO
RenderX

function, and provide training and support. In the future, exergames should be tailored to meet the unique needs of older people with MCI, which is critical to improving their well-being.

(JMIR Serious Games 2025;13:e53631) doi:10.2196/53631

KEYWORDS

exergaming; older adults; mild cognitive impairment; experience; qualitative study

Introduction

Background

Dementia is recognized as a priority global public health issue by the World Health Organization [1]. A new case of dementia is diagnosed every 3 seconds worldwide [2]. It is estimated that the global prevalence of dementia is doubling roughly every 20 years and will reach 131.5 million by 2050 [2]. Dementia costs up to US $81 billion a year in health care costs and lost income and is estimated to increase to US $2 trillion by 2030 [2]. China currently accounts for approximately one-quarter of all patients with dementia globally, making it the country with the largest number of patients with dementia [3], which needs more attention.

There is currently no successful cure for dementia. Therefore, many researchers are focusing on mild cognitive impairment (MCI), the intermediate stage between the cognitive changes caused by normal aging and the onset of dementia. This may be the best opportunity to prevent or delay the onset of dementia [4]. The prevalence of MCI is estimated to exceed one-quarter in people aged >60 years, with an increased risk associated with increasing age [5]. The conversion rate to dementia within 3 years among patients with MCI is approximately 14 times higher than that among individuals without MCI over the same period [6]. Therefore, timely and effective intervention measures should be taken for older adults with MCI.

Exergames are a combination of exercise and games, also known as active video games, which incorporate physical activity into a video game environment through body-controlled movements [7]. Exergames provide both physical and cognitive training, and the combination of these interventions has demonstrated synergistic effects [8]. Previous systematic reviews indicated that exergames are one of the most effective exercise interventions for improving health outcomes, including cognition and physical function, in older adults [9,10]. Furthermore, promising findings from studies involving patients with MCI suggest that exergames have the potential to enhance both physical and cognitive function in this population [11,12].

However, the practical implementation of most exergames for older adults is not ideal [13]. Most exergames, especially commercially available exergames, such as the Nintendo Wii or Xbox Kinect, were not specifically developed for older adults and do not consider their characteristics and preferences [14-16]. Therefore, a user-centered design is needed to develop appropriate exergames for older adults with MCI. In addition, our team recently published a systematic review and qualitative synthesis [17] of all the qualitative studies on the experience of older adults using exergames worldwide. However, it is important to note that none of these studies focused specifically on older adults in China. Given that older adults are a susceptible group that may engage with or use exergames, it is crucial to investigate their experience and needs for exergames. Therefore, to develop user-friendly, adaptive, and personalized exergames suitable for older adults with MCI in China, we conducted this qualitative study to explore their experience and needs with regard to exergames.

Objectives

This study aimed to investigate the experience and needs of older adults with MCI in China about exergames. The findings from this study hold substantial importance in informing the design and development of exergames, as well as guiding the implementation of exergame interventions for older adults, particularly those with MCI.

Methods

Qualitative Approach

To achieve the purpose of the study, we used the semistructured interview method. This method of in-depth interview is an effective approach for explanatory questioning, because it allows for a deep exploration and examination of an individual's views [18]. At the same time, to better understand the experiences and actual needs of older adults with MCI, we used the phenomenological methodology [19]. It focuses on the individual and helps the participants clarify the basic elements of personal experience [20,21]. At the same time, our approach was guided by the approach taken in the study by Giorgi [20,22], which is a systematic approach to data collection, analysis, and synthesis suitable for scientific analysis, and emphasizes the importance of researcher engagement and reflection in the aforementioned process. We used the COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist [23] to ensure that this study met the recommended standards for reporting qualitative data.

Study Setting and Patient Recruitment

From June 2023 to July 2023, we used the purpose sampling method to recruit participants in a community and nursing home in Changsha City, Hunan Province, by pasting posters, distributing leaflets, or recommending potential participants by the nursing home staff. First, members of our research team contacted interested older adults with MCI by phone to provide detailed information about the study procedures. Those who expressed interest in participating were then asked for their consent and subsequently screened for eligibility. The inclusion criteria were as follows: (1) aged ≥60 years and (2) pass the diagnostic criteria for MCI by Petersen [24] (self-reported or insider reported memory decline; scores based on the Montreal Cognitive Assessment [25] were ≤13 points for the illiterate

XSL•FO
RenderX

group, ≤19 points for the primary school group, and ≤24 points for the junior high school group and higher; the score of the Activity of Daily Living Scale was ≤23 points; no clinical diagnosis of dementia). The exclusion criteria were as follows: (1) patients with serious heart, lung, kidney, and other organ diseases; (2) patients with severe limb motor dysfunction, visual impairment, and hearing impairment; and (3) patients with severe mental illness. Participants were continually recruited until data were saturated and no new topics emerged from the final interviews. Participants who were willing to participate in the semistructured interview and met the inclusion criteria were fully informed of the study objectives and procedures, and all participants provided written informed consent.

Data Collection

All interviews were conducted by 2 trained researchers (XC and DJ) in the form of face-to-face interviews with the participants during July to August 2023. Each interview lasted 40 to 60 minutes, and the 2 researchers transcribed the recordings verbatim into text files within 24 hours of the interview. In addition, the researchers also made field notes for subsequent analysis. The researchers continued to conduct interviews until the data were saturated. Demographic questionnaires were also conducted after each interview. The interview was divided into 2 parts: the first part involved allowing older adults with MCI to experience exergames using our preselected exergame device Nintendo Switch. The second part was to interview older adults with MCI about their experience of and needs for exergames.

The first part was an exergaming experience for older adults with MCI using our preselected, existing commercial exergame device Nintendo Switch. Given that older adults are a susceptible group in contact with exergames and most of them are not familiar with exergames, this study allowed older adults with MCI to play exergames for 10 to 20 minutes under the guidance of the rigorously trained research team members before the interview. The research team members promptly answered any questions from the older adults with MCI during this period to ensure familiarity and mastery of exergames. The Nintendo Switch was chosen because it has been shown to be safe, feasible, and acceptable for use in older patients [26], while also improving their cognitive and physical function and overall quality of life [27]. We invited older patients with MCI to experience a badminton game called *Nintendo Switch Sports* on the device Nintendo Switch (Figure 1). In this game, they manipulated the Joy-Con handles to execute a range of actions associated with playing badminton. These actions directly controlled the hand movements of animated characters displayed on the screen, thereby facilitating interactive engagement between older adults with MCI and the animated characters. In the process of older adults with MCI experiencing the exergame, we adjusted the settings for each participant's motor function. In general, the games were implemented in a standing position. However, if the participants had limited motor function, we also allowed them to play exergames while sitting in a stool or wheelchair. In addition, the single-player mode is the basis for using the 2-player mode. To let the participants fully understand and experience the exergame in a limited time, we only let them experience the basic mode, that is, the single-player mode.

XSL•FO
RenderX

Figure 1. Nintendo Switch Sports.



The second part was to interview older adults with MCI regarding their experience and needs for exergames. It consisted of five broad, open-ended questions: (1) How did you feel after playing exergames? (2) Did you have any worries when playing exergames? (3) What do you hope to achieve by playing exergames? (4) What do you think the existing exergames need to improve? (5) How do you think exergames suitable for older adults should be designed?

Data Analysis

Data analysis was guided by the methodology proposed by Giorgi [20], which represents a systematic approach to data collection, analysis, and synthesis suitable for scientific analysis. It emphasizes the importance of researcher engagement and reflection in these processes. The method developed by Giorgi [20] encompasses 4 steps. First, the transcripts were read in full to achieve immersion in the experiences and needs of the participants. Second, the transcripts were reread, and statements that conveyed the experiences and needs of the participants were recorded as units of meaning (codes). Third, codes related to the research objectives were grouped into themes. Finally, the units of meaning were investigated according to the research question, their basic structure was revealed, and these transformed units of meaning were synthesized into descriptive statements.

The data analysis was carried out in parallel with data collection. Two researchers (XC and DJ) listened to the audio recordings at the end of each interview to assess data saturation. Two researchers (XC and DJ) transcribed the audio recordings into text and a third researcher (HN) checked the two transcripts for consistency, and, if there were discrepancies, the recordings were replayed for verification and correction. To facilitate data management and analysis, transcripts were uploaded to the qualitative data management software NVivo 12. Two researchers (XC and DJ) simultaneously extracted, analyzed, and validated the information from the transcripts. In case of disputes, the aforementioned researchers discussed them, and qualitative research experts were consulted. The initial coding process for all collected data was independently carried out by the principal investigator (XC). The research team met weekly to discuss emerging themes; resolve differences in judgment; and review data analysis, coding, and results to facilitate interpretation of results as well as to ensure the credibility of the analytical process. The full dataset was reviewed by a member of this research team (HN) to validate the completed analyses. The research team concluded that the data reached saturation before recruitment ceased.

The trustworthiness of this study was evaluated according to four criteria: (1) credibility, (2) dependability, (3) confirmability, and (4) transferability. A variety of methods were used to ensure the trustworthiness of the study. These included taking reflective notes, triangulating data from multiple perspectives, conducting research team meetings, reporting on the study in detail, using purposive sampling, creating an audit trail, and utilizing verbatim quotes to support themes. In addition, the process of analyzing the data adhered to the standards required by the

XSL•FO
RenderX

study's design, and the final results were sent to all participants for validation and approval.

Ethical Considerations

This study was approved by the ethics review committee of Nursing and Behavioral Medicine Research, School of Nursing, Central South University (E202397). All participants provided their written informed consent before enrollment in the study and retained the right to withdraw at any point in time. The data provided have been deidentified. No compensation was offered to the participants.

Results

Overview

A total of 21 interviews were conducted for this study. By the 18th interview, the data had become saturated, and we conducted 3 more interviews to ensure no new topics emerged. The average age of the study participants was 70.2 (SD 7.6) years, ranging from 60 to 85 years. Of the 21 participants, 17 (81%) were women participants. Regarding educational background, 24% (5/21) of participants had completed primary education, 33% (7/21) had completed junior high school education, 33% (7/21) had completed senior high school education, and 10% (2/21) had obtained a college education. All interviews were conducted in a face-to-face manner, with 76% (16/21) carried out in community settings and the remaining 24% (5/21) conducted in nursing homes. Complete demographic information is presented in Table 1.

Table 1. Participants' characteristics (N=21).

Characteristics	Values
Age (y), mean (SD; range)	70.2 (7.6; 60-85)
Woman participants, n (%)	17 (81)
Education level completed, n (%)	
Primary school	5 (24)
Junior high school	7 (33)
Senior high school	7 (33)
College level	2 (10)
Geographical location, n (%)	
Community	16 (76)
Nursing home	5 (24)
MoCA[a], mean (SD; range)	18.8 (3.6; 11-24)

[a]MoCA: Montreal Cognitive Assessment.

The Experience of Older Adults With MCI Participating in Exergames

Attitudes Toward Exergames Vary

Most older adults with MCI (18/21, 86%) exhibited a positive attitude toward exergames, while a small number of them (3/21, 14%) were not interested in exergames.

Specifically, older adults with MCI found that exergames were easier to do than real exercise, less strenuous, easier to use, and provided an immersive experience that made them feel like they were doing real exercise:

> *Playing badminton with exergames is less strenuous than playing badminton in real life. We only need to use the remote control on our hands, which is not so tiring.* [P10]

> *This exergames is a good one to play at home. It is also a sport, better than outdoor sports, and it is convenient and good. It's easy to use, easy to learn, and a little easier than the real thing. In reality, I may not catch the ball when playing badminton, but in exergames, as long as my hand moves, I usually hit the ball, which is very good for sports.* [P13]

> *In real life, I don't know how to play badminton. But in exergames, I can play badminton, because I can control the ball well as long as I wave my hand at will.* [P21]

> *In general, it's good, especially for me. Just when I was playing exergames, I wanted to go to the ground to pick up the ball, because I thought I was playing real badminton.* [P4]

A minority of older adults with MCI (3/21, 14%) were not interested in exergames due to several factors. First, they perceived exergames as primarily designed for younger individuals or children, thereby rendering them unsuitable for older adults. Second, they felt that the amount of exercise in exergames was excessive. Finally, some might have lacked interest in sports or have limited energy to engage in exergames:

> *Exergames are good, if you have time, you can still play this exergame. But I don't want to play it anymore, because it's for my son and grandson, not*

XSL•FO
RenderX

for us old people, and I don't think it's suitable for old people. [P12]

It's a little bit of exercise, which we certainly don't normally do. We usually take a walk, we don't run at this age, this sport is not suitable for us. [P8]

I'm not very interested in these games, especially the ball game. I'm not really interested. I don't have much energy for this. [P3]

Both Entertaining and Interesting

Older adults with MCI found exergames enjoyable; they experienced a sense of accomplishment after successfully catching the ball and felt rejuvenated by interacting with the animated characters within the game. In addition, they felt that their mood would become happier and their spirit would become better after using exergames. In addition, they thought exergames were fun, so they would not be bored when they exercised:

I was happy when I caught the ball. I had a sense of accomplishment. [P12]

We're old people. It's no fun playing with old people. After playing with the young people in this game, I feel that my mentality is much better, and I can feel younger in every way. [P16]

After each activity, I will become happy and energetic, and I will not nod off. If I were sitting at home right now, I'd be dozing off. But now I feel refreshed after doing this exercise. [P17]

I thought it was funny. I had a great time, ha-ha. Especially when I win. [P21]

It's fun. I feel comfortable when I play. [P1]

This game is very interesting. I like it very much. It's good for me, I love it. I think it's interesting, it doesn't get boring. Exercise can't be boring, otherwise, you always feel that you are just there, no fun. Sports should be as fresh as this, or you can't stick to it. In addition, there's an opponent in the game to compete with me, which I think is fine. [P4]

Promote Physical Activity and Exercise

In this interview, older adults with MCI suggested that exergames could encourage older adults to engage in physical activity and recognize its benefits over prolonged periods of sitting. In addition, they believed that using exergames could exercise various body systems, including training the brain, limbs, eyes, reaction speed, reaction ability, and coordination of various parts of the body:

If this is placed in the community activity center, it will certainly make older adults who have been sitting and playing cards increase their physical activity. [P15]

Exergames are not only fun, but also can exercise the body. I usually sit at home on the sofa watching TV, not doing any physical activity, if there are exergames I can move my body. [P1]

Exergames can exercise the body, especially the flexibility of the brain and hands, reaction speed and reaction ability. [P16]

This game is good for our brains. While playing games, older adults can not only exercise their bodies, but also exercise the coordination and cooperation of the limbs and brain. [P20]

This game is very good, it can exercise the body, especially for the strength of the arms. Old people just need to be active. In addition, exergames are good for the brain, eyesight and reaction capacity. [P2]

Playing this exergame is beneficial for older people. Many older adults sit for a long time, which is not good for their health. After engaging in exercise, they can at least make their limbs more coordinated. In addition, we can also walk while playing badminton, which can play a role in exercising the whole body. Finally, by engaging in exergames, our muscles in the limbs are less likely to atrophy and arms become more flexible. [P6]

Pass Time and Relieve Loneliness

Older adults with MCI found exergames as a way to pass the time and relieve loneliness. They thought that exergames can be used as entertainment tools to occupy their free time. Moreover, some older people would feel lonely because they lived alone, and playing exergames made their house lively:

If you live alone like me, when you are lonely or bored, you can do this exercise, which is very feasible. [P15]

I like entertainment myself, not only to move the body, but also to spend time. I live alone, I can only watch TV when I rest, it is very quiet. I really want to go outside and have fun, not so lonely. The livelier, the better. [P16]

Compared with real badminton, this is a pure entertainment, to pass the time. It suits us seniors who are retiring because we have nothing to do but pass the time. [P9]

Conditions of Use Are Not Restricted

Older adults with MCI mentioned that the use of exergames was not restricted by factors such as location, number of participants, time, and weather conditions. First, exergames can be used both in the community and at home, but when used in the community, some older people may feel shy or embarrassed by being observed. Second, exergames can be used for both single training and double training, and they mentioned that they were willing to play exergames even if they were alone. Moreover, the use of exergames is not limited by time, older adults with MCI can use exergames at any time according to personal preference, and they can play alone without cooperating with others' schedules. Finally, the use of exergames will not be limited by the weather conditions and eliminates the need to worry about not being able to participate in sports because of wind and rain:

XSL•FO
RenderX

Exergames can also be played at home. But it is not suitable for the community, because I'm not good at games, and there are a lot of people in the community, I feel uncomfortable and a little embarrassed by being observed. [P12]

I can play badminton at home with my family through exergames, especially with my husband. [P14]

I am willing to play this exergame by myself. [P1]

This will allow us to exercise at home. Because the outdoor wind is not conducive to playing badminton, the stadium is far from home, and it is not convenient to go out when it rains. Exergames allow me to exercise at home and even do full-body exercises. [P8]

This can be exercised indoors, especially at home and at any time. Even if there is no one to play with, I can play alone, and do not have to cooperate with other people's time. [P7]

I wouldn't have to go out and play if I had an exergame. I just can't walk because my knee is out of shape and I can't move easily. And there are opponents in exergames, which can make me more interested in playing. [P2]

The Need for Exergames in Older Adults With MCI

Design Older Adult–Friendly Exergames

Older adult–friendly design should be centered around older adults, consider their needs for easy-to-understand products, and integrate them into the product conception and design [28,29]. Some older adults with MCI (6/21, 29%) thought that the pace and speed of the game were too fast and the operation was too complicated. A small number of older adults with MCI (3/21, 14%) preferred not to use game pads because they were complicated to use and easy to lose. Furthermore, some older adults with MCI expressed a desire for exergames to be designed to accommodate both seated and standing positions, to cater to the needs of older adults with physical limitations and to prevent prolonged sitting. Finally, older adults with MCI also hoped to design exergames that conform to Chinese culture, to facilitate their familiarity with and the use of exergames:

I'm old. This is a young man's game. I can't keep up. If the game is designed for older people, the speed and pace of the game should be slower. This is too complicated for me, we old people usually watch movies and TV for entertainment. [P12]

This exergame is a bit complicated and difficult to use. I hope that as soon as I open it, I can enter the formal game matchmaking interface, and there will be no redundant selection interface, so that it will be faster for older people to use. [P13]

I still don't know how to use this gamepad. It would be much easier if we did not have to use a gamepad. In addition, it's easy to forget where you put it, and you have to look for it everywhere before using exergames. [P20]

It's easier if I don't have to use this gamepad and hold it all the time, because it's complicated to use. [P21]

Older people react more slowly, so the game should be designed to be slower and simpler so that we can react in time. [P9]

This game can be played both sitting on a stool and standing up. But sitting all the time is not good for health, standing all the time is also not good, it is best to sit for a while and then stand for a while. [P1]

Exergames should be designed to be able to sit on a stool to play the game, many older people's legs and feet are not convenient, and it can also be convenient for disabled people with lower limbs to play this. [P5]

I am not very good at using this gamepad. Not only are there many buttons on it, but all the words displayed on the gamepad are in English, which I can't understand. It should be changed to Chinese. In addition, the whole exergame is very tedious to use, and there are many unnecessary steps, you should design a very simple and convenient game for older adults, so that we older people will take the initiative to play this game. [P4]

Scientific Validity and Safety

Some older individuals with MCI expressed concern about the possibility of sustaining injuries while engaging in exergames. On the one hand, they were worried that they did not receive scientific guidance during the process of exercise and could cause injuries, such as muscle strains and sprained feet. On the other hand, many older people experienced various medical conditions (eg, high blood pressure, heart disease, knee problems, or vision impairment) and were worried about secondary injuries caused by these exercises. Consequently, all of these individuals hoped to ensure the scientific validity and safety of sports in the process of using exergames:

I'm afraid of getting hurt during sports now. Because I bought a grip strengthener last time, maybe the exercise was too intense, it hurt my cervical spine, and I still have pain in my cervical spine, which has never hurt before. I don't know how to use the grip strengthener, I just grab it by myself. Although I thought I was doing well, using the grip 20 times a day, a few days later I injured my neck. We should not exert too much force in the process of exercise. Therefore, I now think about whether I will get injured when I exercise. [P4]

For us old people, we are afraid that we are too old to do sports. Because some old people have high blood pressure and heart disease, they are afraid that he is too excited in the process of exercise, resulting in a sudden heart attack. [P7]

We dare not play badminton now, because we are afraid that we will hurt our feet easily without protection. In addition, our age is not suitable for strenuous exercise, we can only gently do exercise at home. [P8]

XSL•FO
RenderX

> *I used to do square dancing, but I haven't danced for three to four years now because I have coronary heart disease and can't be too tired. Besides, my knee joint is not good. It often hurts. This exergame is good, but I'm just afraid of hurting my joints or falling down. In addition, my heart can't keep up, and I dare not do strenuous exercise, at most do some gentle exercise, such as Taijiquan and so on. Finally, I am afraid that we old people in the process of playing the game fell to the ground, resulting in a sudden death.* [P10]

> *I dare not play this exergame too much. Because I used to play video games when I was young, I even stayed up late playing games, which made my eyes not so visible. Thus, I don't play video games now, for fear of hurting my eyes.* [P18]

Provide a Good Gaming Experience

Some older adults with MCI (3/21, 14%) were hopeful that future exergame developments would provide them with a good gaming experience. For example, older adults with MCI expected the actions of the game's animated characters to align with their movements, expected to experience victories over defeats, and expected exergames to offer a diverse range of games for them to select and switch at any point:

> *Anyway, when my body moves, the game character inside the screen has to move, but if I don't move, it shouldn't move, its movement has to be consistent with mine. However, the movements of the animated characters in this game are not consistent with mine. I don't like this routine of the game. It is not completely according to your actions, and there is no sense of reality.* [P19]

> *I'm playing ball with robots now. I always feel that I can't win, so there is a sense of psychological loss.* [P14]

> *I hope you develop as many games as possible so that I can play different games. For example, I can play this game and then play that game, or play one game today and play another game tomorrow, and I will be happier.* [P4]

Exercise Physical and Cognitive Function

Older adults with MCI also expressed a desire for the future development of exergames to cater to their physical and cognitive needs. Regarding physical function, they aimed to exercise their bodies, improve the agility of their hands and eyes, prevent shoulder periarthritis and cervical spondylitis, improve the coordination of various parts of the body, help them keep healthy and fit, and strengthen their physique. Regarding cognitive function, they hoped to exercise their brains more during exergames, thereby enhancing memory function and reaction speed and mitigating the risk of Alzheimer disease:

> *I hope I can exercise my body through exergames, and at the same time I can use my brain more. If I don't use my brain, my memory will get worse and worse.* [P10]

> *I hope to improve my reaction ability through exergames, as people's reflexes in all aspects will slow down when they get old.* [P11]

> *I want to do this for fun, to pass the time and to exercise. Plus, hopefully this will help us lose weight and keep us in shape. Finally, more exercise can prevent Alzheimer's disease.* [P9]

> *I hope to improve my hand-eye dexterity and reaction speed through this exergame. In addition, it would be better if I could exercise my legs through exergames.* [P5]

> *This exergame is good for body function. If I insist on playing it for a long time, it may help me prevent shoulder periarthritis and cervical spondylitis, and this activity is very good for the body.* [P6]

> *I hope to improve the coordination of all parts of my body through exergames, especially the coordination between the hands, eyes and brain. Besides, I can keep fit and strengthen my body through this.* [P7]

> *I mainly hope to exercise my knee through this. My knees hurt when I walk, but they feel better when I sit.* [P3]

Provide Support and Training

There was also a need to provide older adults with MCI with the necessary support and training in the use of exergames. Older adults with MCI may encounter various problems and difficulties in the process of playing exergames. At such moments, incentive mechanisms and game tips are needed to help them enhance their confidence and overcome difficulties. In addition, most older adults with MCI lacked experience and knowledge of exergames, so they wanted to get the necessary training before using exergames:

> *This game can add some encouragement so that it will be more interesting to play, and I will be more willing to play. For example, add some language communication, which will make people happier. The more you play, the more fun you have.* [P16]

> *It would be better to provide game tips, for example, telling me when to serve and receive the ball during badminton.* [P1]

> *I think voice interaction could be added to this exergame. For example, using voice to ask the user whether to play alone or in pairs, older adults can directly answer orally, which will be very convenient. In addition, the game should be very clear to tell me how to play the game, and what to play. Finally, it needs to provide as many game tips as possible, such as if I don't know how to use the gamepad, the game automatically tells me which key to press at each step. Although I can understand it by reading the instructions, it will take a lot of time.* [P4]

> *To design suitable exergames for MCI seniors, it is necessary to ensure that the pace of the game is slow. At the same time, it is necessary to familiarize older adults with exergames and give relevant training to them before formal use.* [P12]

XSL•FO
RenderX

I think this exergame is great, but I would have been more receptive and enjoyed it if someone had shown me how to use it before the game started. [P6]

Discussion

Principal Findings

To the best of our knowledge, this is the first qualitative study to have explored the experience and needs of older adults with MCI in China regarding exergames. This study included the perspectives of 21 older individuals with MCI and has contributed new knowledge to the exergaming domain through 2 themes: experiences of older adults with MCI regarding exergames and their needs for exergames. Given that exergaming is an innovative, fun, and relatively safe form of exercise [30], successful exergaming may be a promising approach to promote health in older adults with MCI. Therefore, this qualitative study is both timely and meaningful.

We begin to discuss our findings by reflecting on our methods. Our results were combined with our previous knowledge, experience, ideas, and opinions about our research topic, while we acknowledge that these subjective views could influence the results and conclusions of the research. Therefore, in order to maximize the credibility and confirmability of the findings, we thoroughly informed participants and obtained informed consent, established a non-judgmental and open environment during interviews, adopted active listening and open question techniques to encourage honest and comprehensive answers, and engaged in continuous reflection during data collection and analysis. We also conducted regular research team discussions in the process of data analysis. All differences and issues were thoroughly discussed and resolved through collaborative efforts within the research team, with all authors confirming the findings in this study. Therefore, we identified our themes in the process of reflection. It is worth noting that the principal investigator (XC) is a doctoral student with a background in geriatric nursing and digital health research, who received training in systematic qualitative research methods and completed a six-month internship in the Sports Rehabilitation and Medical Robotics team at the Ningbo Institute of Materials Technology and Engineering. As a result, her knowledge and insights in these areas increase the richness of the data, improve the reliability of data collection, and deepen the interpretation of the findings. At the same time, the authors (HN and HF) are experts in geriatric nursing, digital health, qualitative research, and nursing informatics research. The work and research experience of the other coauthors in this field also enhanced the analytical process by introducing new perspectives.

In addition, this study used the COREQ checklist to improve reliability. To enhance confirmability, researchers actively engaged in thorough self-reflection, maintaining a high level of honesty and openness. To support transferability, we provided a detailed description of the study background, participant inclusion criteria, participant demographics, data collection and analysis procedures, and findings to enable readers to assess the applicability of our findings to other settings. Reflective note-taking, interview scheduling, and minutes of supervisory research group meetings provided audit clues for the analysis. The paper also provided many quotes from different participants and used verbatim quotes to support the themes to provide authenticity.

The Experience of Older Adults With MCI Participating in Exergames

Most older adults with MCI expressed a positive attitude toward exergames, citing them as a fun, enjoyable, and novel experience. Previous qualitative studies have also demonstrated pleasurable experiences when participants engage in exergames [31,32]. These previous findings, along with our own, suggest that exergames may be a promising tool for creating fun during physical exercise. However, it should be noted that not everyone is interested in exergames, and not all exergames are suitable for older people. Our team previously published a systematic review and qualitative meta-analysis of older adults' experiences with exercise gaming programs [17]. The reasons we identified for older people's lack of interest in exergames are as follows. First, lack of ability to use exergames, due to age and health-related factors (vision, hearing, motor skills, or cognitive impairment). Second, lack of experience in playing exergames, leading them to worry that they cannot understand and play such games correctly. Third, older people in East Asian countries (such as China, Japan, and South Korea) may feel embarrassed when using exergames due to potential observation or judgment by others. Finally, most existing exergames require a high degree of flexibility and adaptability, which may not be entirely suitable for older people. Nonetheless, most barriers can be addressed by designing exergames for different target groups and giving older people enough time to train and familiarize themselves with exergames.

Our findings found that the vast majority of older adults with MCI believe that exergames are both entertaining and interesting. The possible reason is that exergames are a combination of sports and games, which can not only allow older adults to enjoy immersive game experiences but also encourage them to participate in sports through games, thereby increase the entertainment and fun of their sports [33-35]. Many previous studies have also concluded that exergames are an entertaining, interesting, and attractive form of physical exercise [36-38]. In addition, a recent systematic review and meta-synthesis of 11 studies [39] found that older adults view exergames as a pleasurable experience and suggested that health care professionals and aged care centers may consider using exergames to encourage older adults to participate in and adhere to regular exercise [39].

Physical activity is an effective, cost-effective, and nonpharmacological intervention to reduce the risk of dementia [40,41]. Relevant studies have shown that physical activity can substantially improve the cognitive function, physical function, and quality of life of patients with MCI [42,43]. However, older adults with MCI have low levels of physical activity and are sedentary [44,45]. A cross-sectional survey of global aging and adult health conducted by the World Health Organization in 6 low- and middle-income countries, including China, showed that people with MCI were 1.28 times more likely than people without MCI to not meet internationally recommended physical activity standards [46].

XSL•FO
RenderX

Our findings indicated that older adults with MCI believe that exergames can encourage them to engage in physical activity and recognize its benefits compared to prolonged sitting. This result may help address low levels of physical activity and sedentary behavior among older adults with MCI. Moreover, our results are consistent with previous studies, which also suggest that exergames can encourage and motivate older adults to participate in physical activity [8,15,47].

In addition, exergames are expected to be an important way to improve the mental health of older adults. Our results suggest that some older adults with MCI believe that exergames can alleviate their loneliness. The results are consistent with previous studies. An intervention study involving 20 healthy older adults indicated that exergames may help address social isolation and loneliness among this population [48]. In addition to loneliness, other studies have suggested that exergames can improve other aspects of mental health, including apathy, anxiety, and depression [49,50]. A systematic review of 10 randomized controlled trials showed that exergames had a positive effect on mental health in older adults, reducing the incidence of apathy, anxiety, and depression among older adults [50]. A systematic review of 18 randomized controlled trials with a sample size of 1023 individuals aged >60 years revealed that the longer duration of the exergame intervention, the greater improvement in depression among older adults [49]. Furthermore, a randomized controlled experiment [26] involving 55 older patients with cognitive impairment showed that exergames could substantially decrease depression scores and improve depressive symptoms in these individuals. Finally, our findings indicated that the utilization of exergames is not restricted by factors such as location, number of participants, time, and weather conditions. This is consistent with previous findings. A systematic review and meta-analysis of 18 randomized controlled trials [49] has proposed that the use of exergames are not limited by sports facilities, time, place, space, meteorological environment, or the number of people. Multiple studies have also demonstrated that exergames can be used in different settings, such as home environments, hospitals, communities, and nursing homes [15,49,51]. Moreover, one study suggested that exergames can not only provide individual training but also can be performed in small groups [14].

It is worth noting that we encouraged the participants to express themselves freely and fully respect the opinions they express. However, in addition to aspects related to the use of exergames for older adults with MCI, our study found that participants were less likely to report other multidimensional health aspects, such as social interactions and quality of life. Considering the overall health of older adults, especially in the context of MCI, it is crucial to explore how exergames can contribute to these other dimensions of health beyond cognitive and physical exercise alone. Therefore, it is suggested that future studies should increase the investigation on this aspect, which can provide more comprehensive guidance for the design of exergames. In addition, it is recommended to pay more attention to the integration of exergames with other multidimensional health aspects in the design and development of exergames, as well as the implementation of exergame intervention and promotion, which will help improve the overall health of older people with MCI.

The Need for Exergames in Older Adults With MCI

First, age-friendly exergames should be designed, which is particularly important to improve older people's exergame experience and participation enthusiasm [15,52]. The results of our research found that some individuals with MCI felt that the pace and speed of the game were too fast, and that the controls in the game and the use of the gamepad were too complex. The possible reason for this is that the target users of existing exergames are teenagers and young people, who have too high requirements for game skills, and the game design is not only too complicated and difficult but also lacks attraction, entertainment, and interaction [53,54]. In addition, the experience of older adults with MCI with exergames also influenced their experience. One study suggested that older adults lacked experience in using exergames, which makes it difficult for them to get familiar with them, results in their insufficient acceptance and lack of motivation to use exergames [55]. As a result, existing exergames (such as the Wii and Kinect) may not have been designed to be used by older people [17]. Therefore, it is imperative to design and develop age-appropriate exergames that cater to the specific needs and characteristics of older adults with MCI to enhance their acceptance and engagement in these activities.

Second, it is essential to guarantee the secure and scientific use of exergames by older adults with MCI. Most of the existing exergames are designed for entertainment, without considering the principles of sports [15], and lack scientific training standards and sports programs, which limits older adults from obtaining the maximum health benefits from exergames [14]. According to a previous study [56], one of the most important reasons for the poor rehabilitation effect of exergames on patients with musculoskeletal back and neck pain is that existing exergames are designed for entertainment. Therefore, future studies should target older adults with MCI, take health benefits as the purpose, and add rigorous exercise programs with scientific validity to exergames to better meet the needs of older adults with MCI. In addition, exergames may cause some harm to older adults with MCI. For example, older adults with MCI may be prone to visual and hearing impairment due to prolonged exposure to high-brightness screens and noisy game sounds [57]. There are also many older individuals with various diseases, particularly high blood pressure and heart disease, and at times, even multiple diseases coexist. This may result in potential accidents in the process of playing exergames. Therefore, it is necessary to ensure the safety of older adults with MCI when engaging in exergames to prevent any potential injuries.

Third, it is necessary to ensure that older people with MCI have a good gaming experience during the use of exergames. Older adults with MCI may have difficulty understanding and using exergames due to learning and cognitive decline. The complex project function and level design in exergames increases the learning and cognitive challenges of older adults with MCI, often causing them to have doubts when using exergames, which can easily make them feel frustrated and withdraw, thus reducing

XSL•FO
RenderX

their enthusiasm to participate in exergames [28]. However, the mode can be simplified to make older adults directly understand the operation logic and rules of exergames and reduce the possibility of their wrong operation, to improve the usability of exergames and optimize the experience of exergames for older individuals [28]. Exergames also do not always accurately track and record the movement of older adults, which makes gameplay frustrating [48]. Exergames that incorporate speed elements may further exacerbate the frustration of physically restricted users, reducing their self-efficacy, motivation, and enjoyment of exergames [48]. Therefore, this aspect should be taken into account when designing exergames for older adults with MCI.

Fourth, exergames need to exercise the physical and cognitive functions of older adults with MCI. MCI is characterized by a decline in cognitive function that exceeds the expected level for an individual's age and education level, but with no substantial disruption to activities of daily living [58]. In addition, increasing age is associated with declines in cognitive and physical function [59]. Previous studies have also shown that physical functional parameters, such as gait speed and grip strength, correlate with cognitive function in older adults [60-62]. Cognitive function and physical function are major determinants of quality of life in older adults, and any reduction in these 2 factors may affect the ability of older adults to perform activities of daily living. Previous systematic reviews have indicated that exergames are one of the most effective exercise interventions to improve health outcomes (such as cognition and function) in older adults [9,10]. Therefore, to avert the decline in cognitive and physical functions in older adults with MCI, the design of exergames targeted at this population should focus on the exercise of physical and cognitive functions.

Finally, older adults with MCI need to be provided with the necessary support and training in the use of exergames. Older adults with MCI may easily feel frustrated and disappointed when they encounter various problems or difficulties during the gaming process. Therefore, timely support and guidance should be provided to enhance the self-confidence of older adults, overcome the "digital divide," and ensure their continued participation in exergames [28]. Two studies [32,63] have proposed that frequent and positive feedback from people around them can increase the enthusiasm of older adults who are just beginning to participate in exercise, effectively improve their confidence, create a sense of pride and achievement, and improve their acceptance of exergames and participation motivation. Furthermore, most older individuals lack the experience and knowledge concerning exergames [64], and necessary training should be provided for older adults with MCI before using exergames, including theoretical instruction, a hands-on demonstration of operation, and practical experience for the participants.

Limitations

There are some limitations in this study. First, this study only focused on the experiences and needs of older adults for exergames, and future research should also explore how other stakeholders view exergames, such as family members, caregivers, and health care professionals. Second, our sample size was relatively small, but our interview data reached saturation, indicating that adding new participants will not provide any additional or valuable insights. At the same time, Patton [65] also emphasized that sample size is not as important as other methods in qualitative research. Third, most older adults lack access to new technologies, and only an initial gaming experience was provided for older adults with MCI in this study, which may have limited their comprehension of exergames. However, initial saturation results have been achieved. Fourth, the uneven gender distribution was also a limitation of our study. Out of the total participants, 17 were woman participants while only 4 were man participants, with a substantial overrepresentation of woman participants. This may be because woman individuals are more prone to Alzheimer disease [66]. Fifth, the location of participants was not uniform. This is because the severity of cognitive impairment in older adults in nursing institutions is generally high, while our study only needs older adults with MCI, making it challenging to recruit suitable participants in nursing institutions. Sixth, the selection of exergames in this study is limited to Nintendo Switch Sports on Nintendo Switch, and there should be a wider range of games and platforms to choose from, so that the participants can better understand the nature of games. Finally, this study was limited to a community and nursing home in Changsha, Hunan Province, China, so the findings may have limited generalizability to other settings and countries.

Conclusions

This study offers an interpretative understanding of the experience and needs of older adults with MCI in relation to exergames. The experience of older adults with MCI with exergames included 5 parts: attitudes toward exergames vary, both entertaining and interesting, promote physical activity and exercise, pass time and relieve loneliness, and conditions of use are not restricted. The needs of older adults with MCI for exergames include the desire to design older adult–friendly exergames, ensure science and safety in the process of sports, provide a good gaming experience, exercise physical and cognitive functions, and provide support and training. This study can provide qualitative evidence to guide the development of exergames tailored for older adults with MCI. Future studies are needed to extend the effectiveness of our findings to a larger population of older adults.

Acknowledgments

This study was funded by the National Natural Science Foundation of China (grant 72174212) and the Natural Science Foundation Project of Hunan Province (grant 2025JJ60537).

XSL•FO
RenderX

Data Availability

The datasets generated or analyzed in the course of this study may be provided by the corresponding author on reasonable request.

Authors' Contributions

XC contributed to the investigation, methodology, conceptualization, data curation, formal analysis, and writing of the original draft. DJ contributed to the investigation, methodology, conceptualization, data curation, formal analysis, and writing of the original draft. XC and DJ contributed equally as shared first authors. HN contributed to the investigation, conceptualization, writing the original draft, and visualization. LNW and YFC contributed to the methodology, data curation, review, and editing of the manuscript. CZ, RTP, and YSZ contributed to the methodology, software, review, and editing of the manuscript. HF contributed to the funding acquisition, project administration, resources, and supervision. HN is the co-corresponding author of this paper and can be reached at Hongting4960@hnucm.edu.cn. All authors revised the manuscript and approved the version submitted for publication and agree to be accountable for all aspects of the work.

Conflicts of Interest

None declared.

References

1. Global action plan on the public health response to dementia 2017 - 2025. World Health Organization. 2017. URL: https://www.who.int/publications/i/item/global-action-plan-on-the-public-health-response-to-dementia-2017---2025 [accessed 2025-05-29]
2. Prince M, Wimo A, Guerchet M, Ali G, Wu Y, Prina M. World Alzheimer report 2015: the global impact of dementia: an analysis of prevalence, incidence, cost and trends. Alzheimer's Disease International. 2015. URL: https://www.alzint.org/resource/world-alzheimer-report-2015/ [accessed 2025-05-29]
3. Jia L, Quan M, Fu Y, Zhao T, Li Y, Wei C, Group for the Project of Dementia Situation in China. Dementia in China: epidemiology, clinical management, and research advances. Lancet Neurol 2020 Jan;19(1):81-92. [doi: 10.1016/S1474-4422(19)30290-X] [Medline: 31494009]
4. Hu M, Hu H, Shao Z, Gao Y, Zeng X, Shu X, et al. Effectiveness and acceptability of non-pharmacological interventions in people with mild cognitive impairment: overview of systematic reviews and network meta-analysis. J Affect Disord 2022 Aug 15;311:383-390. [doi: 10.1016/j.jad.2022.05.043] [Medline: 35597472]
5. GBD 2016 Dementia Collaborators. Global, regional, and national burden of Alzheimer's disease and other dementias, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet Neurol 2019 Jan;18(1):88-106 [FREE Full text] [doi: 10.1016/S1474-4422(18)30403-4] [Medline: 30497964]
6. Tschanz JT, Welsh-Bohmer KA, Lyketsos CG, Corcoran C, Green RC, Hayden K, Cache County Investigators. Conversion to dementia from mild cognitive disorder: the Cache County Study. Neurology 2006 Jul 25;67(2):229-234. [doi: 10.1212/01.wnl.0000224748.48011.84] [Medline: 16864813]
7. Mat Rosly M, Mat Rosly H, Davis Oam GM, Husain R, Hasnan N. Exergaming for individuals with neurological disability: a systematic review. Disabil Rehabil 2017 Apr;39(8):727-735. [doi: 10.3109/09638288.2016.1161086] [Medline: 27108475]
8. Stojan R, Voelcker-Rehage C. A systematic review on the cognitive benefits and neurophysiological correlates of exergaming in healthy older adults. J Clin Med 2019 May 23;8(5):3390 [FREE Full text] [doi: 10.3390/jcm8050734] [Medline: 31126052]
9. Fragala MS, Cadore EL, Dorgo S, Izquierdo M, Kraemer WJ, Peterson MD, et al. Resistance training for older adults: position statement from the national strength and conditioning association. J Strength Cond Res 2019 Aug;33(8):2019-2052. [doi: 10.1519/JSC.0000000000003230] [Medline: 31343601]
10. Di Lorito C, Long A, Byrne A, Harwood RH, Gladman JR, Schneider S, et al. Exercise interventions for older adults: a systematic review of meta-analyses. J Sport Health Sci 2021 Jan;10(1):29-47 [FREE Full text] [doi: 10.1016/j.jshs.2020.06.003] [Medline: 32525097]
11. Cai Z, Ma Y, Li L, Lu GZ. Effects of exergaming in older individuals with mild cognitive impairment and dementia: a systematic review and meta-analysis. Geriatr Nurs 2023;51:351-359. [doi: 10.1016/j.gerinurse.2023.03.028] [Medline: 37099867]
12. Zhao Y, Feng H, Wu X, Du Y, Yang X, Hu M, et al. Effectiveness of exergaming in improving cognitive and physical function in people with mild cognitive impairment or dementia: systematic review. JMIR Serious Games 2020 Jun 30;8(2):e16841 [FREE Full text] [doi: 10.2196/16841] [Medline: 32602841]
13. Gerling K, Mandryk R. Custom-designed motion-based games for older adults: a review of literature in human-computer interaction. Gerontechnology 2014 May 30;12(2):68-80. [doi: 10.4017/gt.2013.12.2.001.00]
14. Choi SD, Guo L, Kang D, Xiong S. Exergame technology and interactive interventions for elderly fall prevention: a systematic literature review. Appl Ergon 2017 Nov;65:570-581. [doi: 10.1016/j.apergo.2016.10.013] [Medline: 27825723]
15. Skjæret N, Nawaz A, Morat T, Schoene D, Helbostad JL, Vereijken B. Exercise and rehabilitation delivered through exergames in older adults: an integrative review of technologies, safety and efficacy. Int J Med Inform 2016 Jan;85(1):1-16 [FREE Full text] [doi: 10.1016/j.ijmedinf.2015.10.008] [Medline: 26559887]

XSL•FO
RenderX

16. Subramanian S, Dahl Y, Skjæret Maroni N, Vereijken B, Svanæs D. Assessing motivational differences between young and older adults when playing an exergame. Games Health J 2020 Feb;9(1):24-30. [doi: 10.1089/g4h.2019.0082] [Medline: 31532235]
17. Ning H, Jiang D, Du Y, Li X, Zhang H, Wu L, et al. Older adults' experiences of implementing exergaming programs: a systematic review and qualitative meta-synthesis. Age Ageing 2022 Dec 05;51(12):118. [doi: 10.1093/ageing/afac251] [Medline: 36571772]
18. Denzin NK, Lincoln YS. The SAGE Handbook of Qualitative Research. 5th edition. Thousand Oaks, CA: Sage Publication; 1994.
19. Creswell JW, Poth CN. Qualitative Inquiry and Research Design: Choosing Among Five Approaches. Thousand Oaks, CA: Sage Publication; 2016.
20. Giorgi A. An application of phenomenological method in psychology. Duq Stud Phen Psychol 1975;2:82-103. [doi: 10.5840/dspp197529]
21. Finlay L. Applying phenomenology in research: problems, principles and practice. Br J Occup Ther 1999 Jul 01;62(7):299-306. [doi: 10.1177/030802269906200705]
22. Giorgi A. Concerning a serious misunderstanding of the essence of the phenomenological method in psychology. J Phenomenol Psychol 2008;39(1):33-58. [doi: 10.1163/156916208x311610]
23. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
24. Petersen RC. Mild cognitive impairment as a diagnostic entity. J Intern Med 2004 Sep;256(3):183-194 [FREE Full text] [doi: 10.1111/j.1365-2796.2004.01388.x] [Medline: 15324362]
25. Lu J, Li D, Li F, Zhou A, Wang F, Zuo X, et al. Montreal cognitive assessment in detecting cognitive impairment in Chinese elderly individuals: a population-based study. J Geriatr Psychiatry Neurol 2011 Dec;24(4):184-190. [doi: 10.1177/0891988711422528] [Medline: 22228824]
26. Takei K, Morita S, Watanabe Y, Suganami M, Inao M. Safety, feasibility, and acceptability of physiotherapy combined with strength training using active video games for older patients with musculoskeletal conditions. Disabil Rehabil Assist Technol 2024 Apr;19(3):641-647. [doi: 10.1080/17483107.2022.2115562] [Medline: 36001528]
27. Tuan SH, Chang LH, Sun SF, Lin KL, Tsai YJ. Using exergame-based exercise to prevent and postpone the loss of muscle mass, muscle strength, cognition, and functional performance among elders in rural long-term care facilities: a protocol for a randomized controlled trial. Front Med (Lausanne) 2022;9:1071409 [FREE Full text] [doi: 10.3389/fmed.2022.1071409] [Medline: 36582297]
28. Manser P, Adcock-Omlin M, de Bruin ED. Design considerations for an exergame-based training intervention for older adults with mild neurocognitive disorder: qualitative study including focus groups with experts and health care professionals and individual semistructured in-depth patient interviews. JMIR Serious Games 2023 Jan 05;11:e37616 [FREE Full text] [doi: 10.2196/37616] [Medline: 36602851]
29. Nawaz A, Skjæret N, Helbostad JL, Vereijken B, Boulton E, Svanaes D. Usability and acceptability of balance exergames in older adults: a scoping review. Health Informatics J 2016 Dec;22(4):911-931 [FREE Full text] [doi: 10.1177/1460458215598638] [Medline: 26303810]
30. Ozdogar AT, Baba C, Kahraman T, Sagici O, Dastan S, Ertekin O, et al. Effects and safety of exergaming in persons with multiple sclerosis during corticosteroid treatment: a pilot study. Mult Scler Relat Disord 2022 Jul;63:103823. [doi: 10.1016/j.msard.2022.103823] [Medline: 35523062]
31. Rogers C, Shamley D, Amosun S. Older adults' experience of an exergaming intervention to improve balance and prevent falls: a nested explanatory qualitative study. Appl Sci 2021 Dec 09;11(24):11678. [doi: 10.3390/app112411678]
32. Meekes W, Stanmore EK. Motivational determinants of exergame participation for older people in assisted living facilities: mixed-methods study. J Med Internet Res 2017 Jul 06;19(7):e238 [FREE Full text] [doi: 10.2196/jmir.6841] [Medline: 28684385]
33. Bakker J, Donath L, Rein R. Balance training monitoring and individual response during unstable vs. stable balance Exergaming in elderly adults: Findings from a randomized controlled trial. Exp Gerontol 2020 Oct 01;139:111037. [doi: 10.1016/j.exger.2020.111037] [Medline: 32730797]
34. Ismail NA, Hashim HA, Ahmad Yusof H. Physical activity and exergames among older adults: a scoping review. Games Health J 2022 Feb;11(1):1-17. [doi: 10.1089/g4h.2021.0104] [Medline: 34851732]
35. Stanmore E, Stubbs B, Vancampfort D, de Bruin ED, Firth J. The effect of active video games on cognitive functioning in clinical and non-clinical populations: a meta-analysis of randomized controlled trials. Neurosci Biobehav Rev 2017 Jul;78:34-43 [FREE Full text] [doi: 10.1016/j.neubiorev.2017.04.011] [Medline: 28442405]
36. Altorfer P, Adcock M, de Bruin ED, Graf F, Giannouli E. Feasibility of cognitive-motor exergames in geriatric inpatient rehabilitation: a pilot randomized controlled study. Front Aging Neurosci 2021;13:739948 [FREE Full text] [doi: 10.3389/fnagi.2021.739948] [Medline: 34912206]

XSL•FO
RenderX

37. Valenzuela T, Okubo Y, Woodbury A, Lord SR, Delbaere K. Adherence to technology-based exercise programs in older adults: a systematic review. J Geriatr Phys Ther 2018;41(1):49-61. [doi: 10.1519/JPT.0000000000000095] [Medline: 27362526]
38. National Academies of Sciences, Engineering, and Medicine, Health and Medicine Division, Board on Health Sciences Policy, Committee on Preventing Dementia and Cognitive Impairment. Preventing Cognitive Decline and Dementia: A Way Forward. Washington, DC: National Academies Press; 2017.
39. Risk reduction of cognitive decline and dementia: WHO guidelines. World Health Organization. 2019. URL: https://www.who.int/publications/i/item/risk-reduction-of-cognitive-decline-and-dementia [accessed 2025-05-29]
40. Demurtas J, Schoene D, Torbahn G, Marengoni A, Grande G, Zou L, European Society of Geriatric Medicine Special Interest Group in Systematic Reviews and Meta-Analyses, Frailty, Sarcopenia, Dementia. Physical activity and exercise in mild cognitive impairment and dementia: an umbrella review of intervention and observational studies. J Am Med Dir Assoc 2020 Oct;21(10):1415-22.e6 [FREE Full text] [doi: 10.1016/j.jamda.2020.08.031] [Medline: 32981668]
41. Roheger M, Hennersdorf XS, Riemann S, Flöel A, Meinzer M. A systematic review and network meta-analysis of interventions for subjective cognitive decline. Alzheimers Dement (N Y) 2021;7(1):e12180 [FREE Full text] [doi: 10.1002/trc2.12180] [Medline: 34268448]
42. Falck RS, Landry GJ, Best JR, Davis JC, Chiu BK, Liu-Ambrose T. Cross-sectional relationships of physical activity and sedentary behavior with cognitive function in older adults with probable mild cognitive impairment. Phys Ther 2017 Oct 01;97(10):975-984 [FREE Full text] [doi: 10.1093/ptj/pzx074] [Medline: 29029554]
43. Varma VR, Watts A. Daily physical activity patterns during the early stage of Alzheimer's disease. J Alzheimers Dis 2017;55(2):659-667 [FREE Full text] [doi: 10.3233/JAD-160582] [Medline: 27716669]
44. Vancampfort D, Stubbs B, Lara E, Vandenbulcke M, Swinnen N, Koyanagi A. Mild cognitive impairment and physical activity in the general population: findings from six low- and middle-income countries. Exp Gerontol 2017 Dec 15;100:100-105. [doi: 10.1016/j.exger.2017.10.028] [Medline: 29104092]
45. Di Lorito C, Bosco A, Booth V, Goldberg S, Harwood RH, Van der Wardt V. Adherence to exercise interventions in older people with mild cognitive impairment and dementia: a systematic review and meta-analysis. Prev Med Rep 2020 Sep;19:101139 [FREE Full text] [doi: 10.1016/j.pmedr.2020.101139] [Medline: 32793408]
46. Gallou-Guyot M, Mandigout S, Bherer L, Perrochon A. Effects of exergames and cognitive-motor dual-task training on cognitive, physical and dual-task functions in cognitively healthy older adults: an overview. Ageing Res Rev 2020 Nov;63:101135. [doi: 10.1016/j.arr.2020.101135] [Medline: 32768594]
47. Uzor S, Baillie L, Skelton D. Senior designers: empowering seniors to design enjoyable falls rehabilitation tools. In: Proceedings of the 2012 SIGCHI Conference on Human Factors in Computing Systems. 2012 Presented at: CHI '12; May 5-10, 2012; Austin, TX p. 1179-1188 URL: https://dl.acm.org/doi/10.1145/2207676.2208568 [doi: 10.1145/2207676.2208568]
48. Freed SA, Sprague BN, Stephan AT, Doyle CE, Tian J, Phillips CB, et al. Feasibility and enjoyment of exercise video games in older adults. Front Public Health 2021;9:751289 [FREE Full text] [doi: 10.3389/fpubh.2021.751289] [Medline: 34805074]
49. Yen HY, Chiu HL. Virtual reality exergames for improving older adults' cognition and depression: a systematic review and meta-analysis of randomized control trials. J Am Med Dir Assoc 2021 May;22(5):995-1002. [doi: 10.1016/j.jamda.2021.03.009] [Medline: 33812843]
50. Fernandes CS, Magalhães B, Lima A, Nóbrega P, Silva M, Santos C. Impact of exergames on the mental health of older adults: a systematic review and GRADE evidence synthesis. Games Health J (Forthcoming) 2022 Jul 11. [doi: 10.1089/g4h.2021.0229] [Medline: 35819413]
51. Li J, Theng YL, Foo S. Effect of exergames on depression: a systematic review and meta-analysis. Cyberpsychol Behav Soc Netw 2016 Jan;19(1):34-42. [doi: 10.1089/cyber.2015.0366] [Medline: 26716638]
52. Corregidor-Sánchez AI, Segura-Fragoso A, Rodríguez-Hernández M, Jiménez-Rojas C, Polonio-López B, Criado-Álvarez JJ. Effectiveness of virtual reality technology on functional mobility of older adults: systematic review and meta-analysis. Age Ageing 2021 Feb 26;50(2):370-379. [doi: 10.1093/ageing/afaa197] [Medline: 33068106]
53. Eun SJ, Kim EJ, Kim JY. Development and evaluation of an artificial intelligence-based cognitive exercise game: a pilot study. J Environ Public Health 2022;2022:4403976 [FREE Full text] [doi: 10.1155/2022/4403976] [Medline: 36203500]
54. Torre MM, Temprado JJ. Effects of exergames on brain and cognition in older adults: a review based on a new categorization of combined training intervention. Front Aging Neurosci 2022;14:859715 [FREE Full text] [doi: 10.3389/fnagi.2022.859715] [Medline: 35431905]
55. Schaie KN, Charness N. Impact of Technology on Successful Aging. Cham, Switzerland: Springer; 2003.
56. Jansen-Kosterink SM, Huis In 't Veld RM, Schönauer C, Kaufmann H, Hermens HJ, Vollenbroek-Hutten MM. A serious exergame for patients suffering from chronic musculoskeletal back and neck pain: a pilot study. Games Health J 2013 Oct;2(5):299-307 [FREE Full text] [doi: 10.1089/g4h.2013.0043] [Medline: 24761327]
57. Wang Y, Huang Y, Xu J, Bao D. Interaction preference differences between elderly and younger exergame users. Int J Environ Res Public Health 2021 Nov 29;18(23):37 [FREE Full text] [doi: 10.3390/ijerph182312583] [Medline: 34886309]
58. Gauthier S, Reisberg B, Zaudig M, Petersen RC, Ritchie K, Broich K, et al. Mild cognitive impairment. Lancet 2006 Apr 15;367(9518):1262-1270. [doi: 10.1016/S0140-6736(06)68542-5] [Medline: 16631882]

XSL•FO
RenderX

59. Ma L, Chan P. Understanding the physiological links between physical frailty and cognitive decline. Aging Dis 2020 Apr;11(2):405-418 [FREE Full text] [doi: 10.14336/AD.2019.0521] [Medline: 32257550]
60. Buracchio T, Dodge HH, Howieson D, Wasserman D, Kaye J. The trajectory of gait speed preceding mild cognitive impairment. Arch Neurol 2010 Aug;67(8):980-986 [FREE Full text] [doi: 10.1001/archneurol.2010.159] [Medline: 20697049]
61. Demnitz N, Esser P, Dawes H, Valkanova V, Johansen-Berg H, Ebmeier KP, et al. A systematic review and meta-analysis of cross-sectional studies examining the relationship between mobility and cognition in healthy older adults. Gait Posture 2016 Oct;50:164-174 [FREE Full text] [doi: 10.1016/j.gaitpost.2016.08.028] [Medline: 27621086]
62. Kobayashi-Cuya KE, Sakurai R, Suzuki H, Ogawa S, Takebayashi T, Fujiwara Y. Observational evidence of the association between handgrip strength, hand dexterity, and cognitive performance in community-dwelling older adults: a systematic review. J Epidemiol 2018 Sep 05;28(9):373-381 [FREE Full text] [doi: 10.2188/jea.JE20170041] [Medline: 29526916]
63. Planinc R, Nake I, Kampel. Exergame design guidelines for enhancing elderly's physical and social activities. In: Proceedings of the 2013 International Conference on Ambient Computing, Applications, Services and Technologies. 2013 Presented at: AMBIENT '13; September 29-October 3, 2013; Porto, Portugal p. 58-63 URL: https://www.researchgate.net/publication/304098838_Exergame_Design_Guidelines_for_Enhancing_Elderly's_Physical_and_Social_Activities#:~:text=%5B17%5D%20Induced%20guidelines%20for%20user%20interface%3B%207
64. Gerling KM, Schild J, Masuch M. Exergame design for elderly users: the case study of SilverBalance. In: Proceedings of the 7th International Conference on Advances in Computer Entertainment Technology. 2010 Presented at: ACE '10; November 17-19, 2010; Taipei, Taiwan p. 66-69 URL: https://dl.acm.org/doi/10.1145/1971630.1971650 [doi: 10.1145/1971630.1971650]
65. Patton MQ. Qualitative Research & Evaluation Methods. Thousand Oaks, CA: Sage Publication; 2002.
66. Podcasy JL, Epperson CN. Considering sex and gender in Alzheimer disease and other dementias. Dialogues Clin Neurosci 2016 Dec;18(4):437-446 [FREE Full text] [doi: 10.31887/DCNS.2016.18.4/cepperson] [Medline: 28179815]

Abbreviations

COREQ: Consolidated Criteria for Reporting Qualitative Research
MCI: mild cognitive impairment

Edited by A Coristine; submitted 13.10.23; peer-reviewed by H Jin, K Takei; comments to author 29.02.24; revised version received 23.05.24; accepted 13.05.25; published 17.06.25.

Please cite as:
Chen X, Jiang D, Ning H, Wu L, Chen Y, Zhang C, Peng R, Zhu Y, Feng H
The Experience of and Needs for Exergames in Older Adults With Mild Cognitive Impairment: Qualitative Interview Study
JMIR Serious Games 2025;13:e53631
URL: https://games.jmir.org/2025/1/e53631
doi:10.2196/53631
PMID:

©Xi Chen, Dian Jiang, Hongting Ning, Lina Wu, Yifei Chen, Chi Zhang, Ruotong Peng, Yishu Zhu, Hui Feng. Originally published in JMIR Serious Games (https://games.jmir.org), 17.06.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Preliminary Efficacy of a Gamified Mobile App for Promoting Self-Health Management Among Nurses in the Post-COVID Era: Single-Group Pre-Post Study

Shao Huan Hsu[1], MSc; Li Jung Lu[1], PhD; Pei Chin Chou[2], PhD

[1]College of Management, Institute of Healthcare Information Management, National Chung Cheng University, Chiayi County, Taiwan

[2]Department of Respiratory Care, Chang Gung University of Science and Technology, Chiayi County, Taiwan

Corresponding Author:
Li Jung Lu, PhD
College of Management
Institute of Healthcare Information Management
National Chung Cheng University
No. 168, Section 1
Daxue Road, Minxiong Township
Chiayi County, 621301
Taiwan
Phone: 886 +886 972091305
Email: e4810.y0058@msa.hinet.net

Abstract

Background: The COVID-19 pandemic has significantly affected health care professionals, especially nurses, who have experienced elevated levels of stress, burnout, and physical health challenges. In the postpandemic era, supporting their well-being is crucial. Gamification, which is the application of game design elements in nongame contexts, has emerged as a promising strategy to promote engagement in health behaviors. This study explores the use of a gamified mobile app to support self-health management among nurses recovering from the COVID-19 experience.

Objective: This study aimed to evaluate the preliminary efficacy of a gamified mobile app for promoting self-health management among nurses who experienced the COVID-19 pandemic. The study examined whether gamification could enhance engagement, improve physical health outcomes, and encourage sustainable behavior change.

Methods: A single-arm pre-post intervention study was conducted using a user-centered design. The app was developed based on the Octalysis framework and goal-setting theory, incorporating personalized exercise prescriptions and health monitoring features. Nurses from a regional hospital in Hsinchu, Taiwan, participated in the 8-week intervention. Data were collected through interviews, pre- and postintervention surveys, and app usage analytics. Key outcomes included changes in step counts, BMI, and user engagement.

Results: After the intervention, BMI classification improved significantly. The proportion of participants classified as obese decreased from 38.5% (90/234) to 13.7% (32/234), and the proportion of those classified as overweight increased from 24.8% (58/234) to 34.6% (81/234). Overall, the combined proportion of overweight or obese participants declined from 63.2% (148/234) to 48.3% (113/234), and that of participants with normal BMI increased from 18.4% (43/234) to 33.8% (79/234) (χ^2_4=29.98; P<.001). Octalysis tool results showed strong motivational engagement, with the highest scores in *development and accomplishment* (mean 7.29), *epic meaning and calling* (mean 7.05), and *empowerment of creativity and feedback* (mean 6.55).

Conclusions: The gamified mobile app demonstrated promising efficacy in enhancing self-health management among nurses in the post-COVID era by increasing physical activity and improving BMI. Gamification elements, such as achievement, purpose, and feedback, were effective in sustaining engagement. Further studies are recommended to assess long-term outcomes and broader applicability.

(JMIR Serious Games 2025;13:e66262) doi:10.2196/66262

XSL•FO
RenderX

KEYWORDS

gamification; mobile health; nurse wellness; self-health management; post-COVID; health care technology; behavioral intervention; digital health; health promotion; exercise adherence

Introduction

Background and Research Motivation

Since the COVID-19 outbreak in late 2019, health care systems have faced unprecedented stress. Nurses, particularly those on the frontlines, have experienced increased exposure to infection, mortality, psychological distress, and burnout [1-6]. Female nurses have been especially vulnerable [7-10]. Reports, such as *Who Cares Wins* by the United Nations and *ESG White Paper* by Taiwan's Ministry of Health and Welfare, highlight the importance of workforce well-being within environmental, social, and governance (ESG) frameworks [11]. Despite these initiatives, nurse health continues to be threatened by excessive workload and chronic stress, compromising both individual well-being and the sustainability of health care systems.

To support the physical and mental health of health care workers, both self-management and external interventions are essential. Organizations like the World Health Organization (WHO) and the National Center for Chronic Disease Prevention and Health Promotion (NCCDPHP) recommend that adults engage in at least 150 minutes of moderate exercise weekly [12], but nurses often find it difficult due to shift work and fatigue [13]. As such, digital health platforms and mobile apps have gained popularity for providing accessible health management tools [14]. These include psychological screenings, mental health apps like BetterHelp and Calm [15], dietary tracking tools, and weight management apps that support goal-setting and outcomes [16-22].

Although digital health platforms offer structural support for health management, sustaining user engagement over time remains a major challenge. A key reason is the lack of intrinsic motivation and enjoyment, often resulting in user attrition [14,23]. This gap highlights the urgent need for more engaging and user-centered interventions.

Introduction to the Literature Review

The American College of Sports Medicine advocates the FITT (frequency, intensity, time, and type) principle, a widely accepted method for exercise prescription. However, Burnet et al [24] argued that the FITT model lacks the element of "fun" and introduced the FITTF (frequency, intensity, time, type, and fun) model to improve user adherence. Yet, many health apps still fail to incorporate an enjoyable design, limiting user compliance. Gamification, which is the application of game elements to nongame contexts, has emerged as a promising strategy to enhance health behavior engagement. Initial gamified apps often lacked theoretical rigor, but more recent efforts have used motivational theories such as goal-setting, behavior change, and social influence. Alahäivälä and Oinas-Kukkonen [25] integrated gamification into health apps to boost self-management behaviors. Yang et al. [14] found that incorporating goal-based tasks and social interaction improved users' health outcomes [26,27]. Competitive features like leaderboards can also enhance motivation, as shown by Landers et al [28,29], though some studies caution about the potential negative impacts of competition.

Chou [30] developed the Octalysis framework, combining game design, motivational psychology, and behavioral economics into 8 core drives. Although it is widely used in education and marketing [31-34], its application in health care remains limited, representing an area for further exploration. A systematic review by Xu et al [35] analyzed 50 studies on gamified health apps and reported mixed findings in terms of engagement and effectiveness. External motivators like rewards and rankings are common [36,37], but lasting behavior change often depends on internal motivation, such as personalized goal-setting. Landers et al [29] emphasized that well-designed gamification strategies can transform intrinsic desires into actionable goals [29]. While many apps employ external incentives, personal goals more effectively reflect users' health needs and sustain engagement. Therefore, understanding how to balance achievement incentives and personal goal-setting is key to designing effective health management apps.

Despite growing interest in gamified health tools, few studies have examined their use among health care workers, particularly nurses, in the postpandemic context. Furthermore, the integration of the Octalysis framework with the FITTF model remains underexplored. There is a clear research gap in addressing how gamification can be tailored to health care professionals, whose unique work stressors and responsibilities differ from those of the general population.

This gap sets the stage for our study, which seeks to explore how gamification strategies, rooted in the Octalysis framework and guided by the FITTF principle, can be tailored for the postpandemic health management of nurses.

Rationale for Selecting the Research Setting

This study was conducted at a branch of a medical center in Hsinchu City, Taiwan, which was selected for 3 main reasons. First, the hospital actively promotes employee health and participates in the Health Promoting Hospital initiative, providing a supportive environment for health interventions. Second, it has a stable and diverse nursing workforce, including various shift types and specialties, ensuring the inclusion of participants with different health needs. Third, the hospital is open to technological innovation and expressed a willingness to collaborate in the co-design and testing of the gamified app. These factors ensure practical feasibility, institutional support, and contextual relevance for evaluating the effectiveness of the intervention.

Research Questions and Objectives

To address the identified gap, this study aims to design and evaluate a gamified mobile health management app specifically tailored for nurses. The app will incorporate the Octalysis framework and FITTF exercise prescription principles and will be iteratively refined through user feedback. The key research

XSL•FO
RenderX

objectives are as follows: (1) to identify essential gamification features using the Octalysis framework, which effectively support nurses' health needs after the COVID-19 pandemic, and (2) to develop and validate a FITTF-based exercise prescription that enhances nurses' physical health and health management outcomes.

By focusing on health care workers, a frontline group severely affected by the pandemic, this study contributes to theoretical advancements in gamified health interventions and the practical development of scalable and sustainable self-management tools for health care professionals.

Based on previous literature, we hypothesized that gamification interventions would improve physical activity and BMI among health care workers in the postpandemic era and promote sustained engagement through motivational gamification features.

Methods

Overview

In light of the current literature, there has been limited research focused on developing gamified health management apps specifically designed for nurses in the postpandemic era. This study aims to address this gap by integrating the Octalysis gamification framework with the goal-setting theory to develop a gamified health management app tailored to nurses' needs. The app's design will draw on exercise prescriptions recommended by leading global health organizations and will be refined through in-depth interviews conducted with staff at a selected case hospital to ensure it meets user needs effectively.

Study Design

This study adopted a single-group pre-post intervention design and was conducted over an 8-week period to evaluate the preliminary efficacy of a gamified mobile app in promoting self-health management among nurses in the post-COVID context.

Participants and Recruitment

This study employed a purposive sampling strategy to recruit eligible participants. Anonymous online questionnaires were distributed via a secure web-based platform, and a total of 234 valid responses were included in the final analysis.

The inclusion criteria were as follows: (1) being a full-time employee of the participating hospital, (2) willingness to participate in the 8-week intervention program, and (3) ability to operate the mobile app independently.

The exclusion criteria were as follows: (1) medical conditions that limited or prohibited physical activity, (2) pregnancy, and (3) concurrent participation in other structured health or wellness programs during the study period.

System Development Methodology and Design Philosophy

Two primary development methodologies were considered for this study: system development life cycle (SDLC) and prototyping.

SDLC Methodology

SDLC is a well-defined system development methodology consisting of several distinct phases, including system design, construction, testing, and delivery, which are all aimed at producing a high-quality system that meets or exceeds customer expectations [38].

Prototyping

Prototyping, on the other hand, is a system development methodology that emphasizes user feedback and rapid iteration. It involves creating a simplified initial version (prototype) of the system, which allows users to visualize and interact with the system's design and features early on. The development team then refines the system based on user feedback, ultimately achieving a product that meets user needs. This approach is particularly suitable for projects with unclear or frequently changing requirements, as it shortens development cycles and enhances efficiency.

While SDLC offers a structured and comprehensive approach, it is time-consuming and costly, and often struggles to accommodate evolving requirements. Conversely, prototyping provides a quicker, more cost-effective solution that allows for iterative improvements based on user input. Given these advantages, this study opted for the prototyping method.

This study adopted a prototyping approach to system development, which emphasizes user-centered design, iterative refinement, and feedback integration. Prototyping allows users to interact with early-stage versions of the system, enabling developers to incorporate their feedback in real time and create a product that closely aligns with users' actual needs. Compared to the more rigid and time-consuming SDLC methodology, prototyping is more flexible and efficient, making it suitable for projects with evolving requirements.

From a design perspective, this approach reflects a clear empowering orientation. By involving end users (nurses) in the design process through interviews and iterative feedback loops, the system fosters a sense of ownership and autonomy. This aligns with the principles of empowering technologies, which aim to enhance users' self-efficacy and enable independent health management.

In addition to its empowering foundations, the system also incorporated hedonic elements in its gamification design. Drawing from the Octalysis framework, the final product included features, such as goal progression, achievement rewards, and friendly competition, which introduced playfulness, unpredictability, and social influence. These elements were designed to enhance enjoyment and intrinsic motivation, thereby supporting sustained user engagement in health behavior change.

Together, these complementary design philosophies, empowering users through co-design and engaging them through gamified, hedonic features, form the theoretical and practical backbone of the mobile app developed in this study. Moreover, although the system was not formally structured based on a cognitive behavioral therapy (CBT) framework, several features were aligned with CBT behavioral activation strategies. These included personalized goal-setting, self-monitoring dashboards,

XSL•FO
RenderX

and timely performance feedback, which have been shown to support users in recognizing and modifying maladaptive behavior patterns. These strategies promote sustained engagement in physical activity and reflect core techniques commonly adopted in CBT-informed digital health interventions, such as those described in the work of Tudor-Sfetea et al [39]

System Interviews

In-Depth Interviewing

In-depth interviewing, also known as qualitative interviewing, was categorized by Patton [40] into 3 forms: structured interview, unstandardized interview, and guided interview.

Structured interview is a quantitative research method commonly used in survey research. This approach standardizes interview questions and uses precise language to ensure that each respondent experiences the same questions and sequence, thus ensuring the reliability of responses.

Unstandardized interview is an open-ended interview method without standardized questions. The interviewer engages the respondent in a free-form discussion around a central theme, collecting data flexibly without the respondent's awareness.

Guided interview is a semistructured interview method where the interviewer prepares a general outline of topics to be discussed. This approach allows for guided yet flexible conversations, enabling respondents to express their thoughts in a relaxed environment while ensuring that the discussion remains focused.

Given these advantages, this study employed guided interviews. The researcher prepared an interview outline and key questions in advance and made adjustments based on respondent feedback during the interviews. The complete interview guide is available in Multimedia Appendix 1. This approach provided rich, diverse data and deep insights into users' real needs and challenges, informing the design of the gamified health management app.

Additionally, while the study did not implement a formal CBT protocol, several interview-informed design elements aligned with CBT principles. Specifically, the interviews helped identify behavioral barriers and health beliefs, which were subsequently addressed through in-app features such as goal-setting, self-monitoring, and activity prompts, which are key components associated with behavioral activation strategies in CBT. These elements encouraged users to recognize and alter unhelpful behavior patterns, supporting sustainable health management behavior.

The hospital's Health Promotion Committee, led by the Deputy Director, has implemented several measures to enhance the health of employees, patients, and their families, creating a healthy hospital environment. The key measures include promoting mental support activities (providing psychological counseling and support to enhance employee mental health), implementing health checks and tracking (regular health checks with follow-up management of the results), conducting health promotion training (offering health knowledge and skills training to increase employee health literacy), enhancing health awareness (raising awareness and understanding of health management among employees through various forms of education and communication), and encouraging participation in health promotion activities (organizing and encouraging participation in health activities to foster a sense of engagement and motivation).

Qualitative Interview Procedures

The initial interview outline was designed to address the challenges and needs nurses face in their daily work and health management, without being strictly based on any specific theoretical framework. Through open-ended interviews, the study aimed to uncover users' real needs and difficulties. The interview results were used to establish the necessary functions for the app.

To inform the design of the gamified mobile health management app, a qualitative phase was conducted prior to system implementation. This phase involved semistructured interviews with frontline nurses using a guided interview protocol that explored challenges in self-health management and preferences for gamification features.

Participants were recruited through purposive sampling from among full-time nursing staff at the study hospital. All interviews were conducted in a private setting, audio recorded with participant consent, and transcribed verbatim. Thematic analysis was applied to the transcribed data using open coding techniques. Codes were iteratively grouped into categories, and overarching themes were identified through constant comparison, following established qualitative data analysis protocols.

Thematic analysis was conducted following the 6-phase framework by Braun and Clarke [41], including data familiarization, initial code generation, theme searching, theme review, definition, and reporting. Two researchers independently coded the interview transcripts manually and met regularly to compare interpretations, resolve discrepancies, and consolidate themes. Coding matrices and summary tables were used to organize and refine emerging categories. To enhance trustworthiness, peer debriefing was conducted and representative quotes were selected to support each theme.

System Architecture Workflow

The gamified health management app developed in this study is designed to operate across various platforms. The app provides personalized exercise prescriptions based on user health data and allows users to store and query their health records. Nurses begin by registering a personal account and entering basic health data, such as age, height, weight, BMI, and fitness scores, via mobile devices. The system then evaluates and analyzes the data to generate a tailored exercise prescription, ensuring the appropriateness and effectiveness of the health plan.

Demographic Data Collection

Prior to system implementation, participants completed a baseline self-administered questionnaire that collected demographic and health-related information. Items included age, gender, education level, occupation, regular exercise habits, and prior gaming behavior. The instrument consisted of structured multiple-choice and Likert-scale questions. Data were

XSL•FO
RenderX

anonymized, and only fully completed responses were used for analysis.

Gamification Design Features

Participants can engage in gamified challenges by completing the prescribed exercises, with the system offering rewards for successful completion to encourage ongoing engagement. The app was developed using the Octalysis framework, which informed the implementation of several core motivational drivers. Although visual screenshots were unavailable due to poststudy system access restrictions, the following features were integrated and aligned with key Octalysis dimensions: development and accomplishment (users tracked their daily steps via a visible progress bar and earned digital badges when achieving specific milestones), social influence and relatedness (a real-time leaderboard displayed peer rankings to foster a sense of friendly competition), empowerment of creativity and feedback (the app allowed avatar customization and delivered immediate feedback on user performance).

These features were designed to enhance intrinsic motivation and promote sustained behavior change in accordance with the Octalysis gamification model. In addition to aligning with core motivational drivers, these components also represent key hedonic (eg, badges and leaderboards) and empowering (eg, self-monitoring and avatar customization) gamification elements, consistent with the framework outlined by Lin et al [42].

Due to poststudy access restrictions, the system interface was not available for screenshot capture. Therefore, visual illustrations of the app's design cannot be provided. However, detailed descriptions of each core component have been included to convey its structure and function clearly.

Development Environment and Tools

The development environment and tools used in this study are outlined in Table 1.

Table 1. Development environment and tools.

Category	Environment/tool
Platform or system	Full platform (web version)
Development tools	Visual Studio Code and Firebase
Programming languages	HTML, CSS, JavaScript, Node.js, and SQL

System Evaluation

This study developed a gamified health management app based on the Octalysis framework, which was used as the foundation for the app's design. The efficacy of the system was assessed using the Octalysis tool, a specialized evaluation instrument that measures the app's performance across 8 core drives of gamification. Additionally, a survey method was employed to gather user feedback and further evaluate the app's efficacy.

Due to the anonymized and aggregate nature of the Octalysis tool output, individual participant-level responses were not stored. Therefore, SDs could not be computed for the core drive scores.

Evaluation Methods

Octalysis Tool

The Octalysis tool was used for a comprehensive evaluation of the system. This professional tool assesses the efficacy of the gamified design in 8 core drives, including epic meaning and calling, development and accomplishment, and social influence. These core drives play a critical role in evaluating the system's ability to motivate and engage users.

Survey Method

A survey was conducted using a Likert scale. The 10-point scale (ranging from "1" to "10") was designed to collect user feedback on various aspects of the system's functionality and user experience. The survey covered all core drives to ensure a balanced and comprehensive evaluation of the system's overall efficacy.

Evaluation Process

Survey Design and Distribution

The survey, structured around the 8 core drives of the Octalysis framework, was distributed to system users. The survey included multiple questions aimed at assessing user satisfaction and experience across different dimensions of the system.

Data Collection and Analysis

After collecting the survey responses, the data were analyzed using the Octalysis tool. The analysis involved calculating scores for each core drive, and an Octalysis diagram was generated to visually represent the system's performance across these drives.

Result Interpretation and Improvement Recommendations

Based on the evaluation results, the system's strengths and areas for improvement were identified. Specific recommendations were made to optimize the gamification design, enhance user experience, and improve the overall efficacy of health management within the system.

Statistical Analysis Clarification

All statistical comparisons were conducted using SPSS version 25 (IBM Corp). To assess the significance of changes in BMI classifications before and after the 8-week intervention, a chi-square test was applied. The results indicated a statistically significant improvement in BMI distribution (χ^2_4=29.98; P<.001).

For Octalysis tool data, only aggregate average scores were available due to the tool's design. Individual-level responses and SDs were not accessible, which prevented the use of

XSL•FO
RenderX

inferential statistical tests, such as *t* tests or ANOVA, for gamification-related outcomes. As a result, these findings were reported descriptively.

Daily step data from the top 10 users on the leaderboard were also analyzed descriptively due to the limited subgroup sample size. Participants were categorized post hoc into "stable" and "unstable" engagement groups based on their activity patterns, which were further interpreted qualitatively through follow-up interviews.

Descriptive findings were transparently reported for all datasets where inferential testing was not feasible. Future studies are encouraged to incorporate larger samples and adopt analytical tools and designs, such as time-series models or regression-based approaches, that allow for more robust inferential analysis and causal inference.

Ethical Considerations

This study was reviewed and approved by the Institutional Review Board of MacKay Memorial Hospital, Hsinchu Branch (approval number: 23MMHISO16e). Data were collected via an anonymous, web-based questionnaire embedded in a gamified mobile app. The survey did not collect any personal identifiers or sensitive health information. Participation was entirely voluntary. Before accessing the questionnaire, participants were presented with an electronic informed consent statement that explained the study's purpose, procedures, data handling practices, and confidentiality assurances. Proceeding to complete the questionnaire was considered to indicate implied consent, in accordance with institutional ethical guidelines. All responses were anonymized, securely stored, and analyzed in aggregate. Only group-level results were reported to ensure the privacy and confidentiality of participants.

Results

Findings From In-Depth Interviews

Table 2 summarizes the key insights from the 3 interview sessions, which informed the design of the gamified mobile health management app.

During the interviews, users expressed needs that aligned with 4 core drives in the Octalysis framework: development and accomplishment, social influence and relatedness, unpredictability and curiosity, and loss and avoidance. These findings reflect key motivational drivers that informed the design of the intervention. These insights supported the selection of Octalysis as the theoretical foundation for the platform's gamified feature design, ensuring both motivational alignment and sustained user engagement. This ensures that the app's features align with nurses' needs while leveraging gamification to enhance overall system efficacy.

Table 2. Interview results.

Session	Duration	Participants	Conclusion	Function requirements
1	2 hours	4	Main goal of the app: Weight control is a primary goal. Participants agreed that the app should focus on helping nurses maintain or achieve an ideal weight due to the nature of their busy and stressful work.	Pedometer, body data tracking, virtual island tour, leaderboard, and task and goal setting
2	2 hours	4	The app should provide personalized exercise prescriptions based on the user's health status and must follow the FITT[a] principle.	Exercise prescriptions, video tutorial links, and time-limited activities
3	2 hours	4	To enhance ongoing use, the app should include an achievement system where users can earn badges for reaching exercise goals.	Achievement system

[a]FITT: frequency, intensity, time, and type.

Data Analysis

This study conducted an in-depth data analysis based on feedback from the questionnaire. To ensure the accuracy and completeness of the data, invalid responses were excluded, and valid responses were carefully categorized and summarized. Considering the specificity of this study, the Octalysis tool, which is designed to evaluate the system's alignment with the Octalysis framework, was employed. The Octalysis tool, grounded in gamification theory, provides a comprehensive evaluation of the system's performance in areas such as user experience, engagement, and motivational incentives. Quantitative analysis using the Octalysis tool was conducted to gain insights into the system's strengths and areas for improvement.

Questionnaire Response Rate and Demographic Analysis

This study used a purposive sampling strategy to recruit eligible participants. An anonymous online questionnaire was distributed via a secure web-based platform, resulting in 234 valid responses included in the final analysis. The analysis revealed that the majority of respondents were female (145/234, 62.0%), were aged between 31 and 50 years (123/234, 52.6%), held a university degree (137/234, 58.5%), and were engaged in nursing (143/234, 61.1%). Among the 234 respondents, 106 (45.3%) reported having a regular exercise habit, with most exercising less than 2 hours per week (79/106, 74.5%), and 128 (54.7%) did not have a regular exercise habit. Furthermore, of the 234 respondents, 17 (7.3%) reported playing sports-related games, with most of them playing for less than 5 hours per week (13/17, 76.5%), and the remaining 217 (92.7%) did not engage in such games. The data suggest that there is considerable room for improving exercise habits among the participants. The detailed findings of the demographic analysis are presented in Table 3.

XSL•FO
RenderX

Table 3. Participant demographics.

Variable category	Value (N=234), n (%)
Gender	
Male	89 (38.0)
Female	145 (62.0)
Age (years)	
≤30	78 (33.3)
31-50	123 (52.6)
≥51	33 (14.1)
Education level	
High school/vocational	38 (16.2)
Associate degree	46 (19.7)
University	137 (58.5)
Graduate school	13 (5.6)
Occupation	
Health care related	143 (61.1)
Non–health care related	91 (38.9)
Exercise habit	
Yes	106 (45.3)
No	128 (54.7)
Sports-related game	
Yes	17 (7.3)
No	217 (92.7)

Octalysis Tool Analysis

Participants rated *development and accomplishment* and *epic meaning and calling* as the most engaging motivational dimensions, as shown in Table 4. These high scores likely reflect strong user interaction with gamification elements, such as the step tracking system, achievement badges, and progress bars, which were directly linked to these Octalysis core drives.

It is worth noting that a total of 198 valid responses were collected for the Octalysis survey, which was less than the responses for the demographic questionnaire. This discrepancy can be attributed to the different nature of the 2 questionnaires. The demographic survey was completed by participants before engaging with the gamified app, collecting general information such as background and occupation. Since it was administered early, participation was relatively high, leading to a larger number of responses. Conversely, the Octalysis assessment survey was completed after participants had engaged with the walking game, which may have led to lower response rates due to factors such as participant fatigue or decreased willingness to complete the survey after the intervention.

Table 4. Octalysis tool analysis.

Core drives	Average score[a]
Epic meaning and calling	7.05
Development and accomplishment	7.29
Empowerment of creativity and feedback	6.55
Ownership and possession	5.92
Social influence and relatedness	5.97
Scarcity and impatience	6.53
Curiosity and unpredictability	6.30
Loss and avoidance	6.30

[a]SDs could not be reported as the Octalysis tool provides only aggregate mean scores and does not retain individual-level response data.

XSL•FO
RenderX

Epic Meaning and Calling

Participants scored an average of 7.05 points in this dimension, indicating a relatively high level of engagement with the sense of mission and purpose within the walking game. This suggests that participants perceived the game as more than just a physical activity, attributing additional meaning and value to it, which enhanced their motivation and involvement.

Development and Accomplishment

In this dimension, participants scored an average of 7.29 points, reflecting positive feedback on the sense of achievement and personal progress within the game. This result suggests that the walking game successfully created a sense of accomplishment among participants, fostering a desire for personal growth and encouraging continued participation.

Empowerment of Creativity and Feedback

Participants scored an average of 6.55 points in this dimension, indicating a moderate response regarding exercise creativity and receiving immediate feedback within the game. This result implies that there is room for improvement in the game's ability to stimulate creative thinking and provide timely and effective feedback. Future iterations of the gamified health management system should focus on enhancing these aspects to better meet participant expectations and needs.

Ownership and Possession

This dimension received the lowest average score of 5.92 points, indicating that participants felt a relatively low sense of ownership and possession within the game. This suggests that there is significant room for improvement in how the game fosters a sense of ownership and engagement with in-game items and related elements.

Social Influence and Relatedness

Participants scored an average of 5.97 points in this dimension, reflecting a relatively low need for social interaction and relatedness within the game. This finding suggests that the integration of social interaction elements in gamified tools should be approached with caution, ensuring that these elements align with the target audience's preferences.

Scarcity and Impatience

In this dimension, participants scored an average of 6.53 points, indicating a moderate need for experiences of scarcity and urgency within the game. This suggests that incorporating elements that introduce challenge and a sense of time pressure could enhance participant engagement.

Curiosity and Unpredictability

Participants scored an average of 6.30 points in this dimension, indicating a moderate expectation for curiosity and unpredictability within the game. This highlights the importance of maintaining participants' curiosity and excitement in future gamified tool designs.

Loss and Avoidance

Participants scored an average of 6.30 points in this dimension, indicating a moderate need to avoid losses within the game. This suggests that participants may have a desire to avoid failure and loss, providing a potential area for improvement in future iterations of the system.

Overall, the Octalysis tool analysis revealed the motivational structure of health care–related participants when engaging with the walking game. Participants responded most positively to the "epic meaning and calling" and "development and accomplishment" dimensions, while "ownership and possession" and "social influence and relatedness" received lower scores. These insights not only guide future improvements to gamified tools but also highlight the challenges of fostering social connections among health care professionals. These findings contribute to the precise adjustment of gamified elements to more effectively motivate and support health care workers in health management activities. Table 4 presents the average scores across the 8 core drives. Due to the design of the Octalysis tool, only aggregate mean scores were available, and individual-level response data could not be stored or accessed. As such, SDs could not be calculated for this dataset.

We agree that reporting SD is important for interpreting data variability and maintaining statistical rigor. In future studies, we recommend using engagement assessment tools that allow for the export of individual-level responses, which would enable more robust statistical analyses, including the calculation of SDs and inferential comparisons.

Daily Step Data of the Top 10 Users on the Leaderboard

A key aspect of this study was to analyze the daily step data of the top 10 users on the leaderboard. According to the recommendations of the WHO and the American Heart Association (AHA), a daily step count of at least 10,000 steps is suggested to maintain basic health. This study used step count as the primary metric due to its ease of measurement, comprehensibility, and efficacy in reflecting daily activity levels [40]. Step count data are crucial indicators for evaluating individual physical activity levels, especially in research on health behaviors and lifestyle. The top 10 users on the leaderboard represented the most active participants in this study in terms of walking activity. Their step count trends revealed some interesting patterns that warrant further exploration. By conducting a detailed analysis of these users' step data, we aimed to gain deeper insights into factors that may influence walking activity and provide valuable recommendations for encouraging healthier lifestyles and activity levels. In this section, we will present a detailed analysis and discussion of the daily step data of the top 10 users on the leaderboard, aiming to uncover trends and potential correlations.

Based on the variation in user step counts, these 10 users were categorized into 2 groups: stable and unstable. A more detailed analysis was conducted for each group to understand the reasons behind their step count changes and potential influencing factors.

Analysis of the Stable Group

User K825 consistently had a high step count from the beginning. On further interviews, it was revealed that this user had a pre-existing focus on health and exercise, along with a regular exercise routine.

XSL•FO
RenderX

Although there were some fluctuations in the step count of user L171, this user's overall step count showed a gradual increase. This trend was attributed to the user's growing awareness of the importance of exercise for health, leading to a deliberate effort to increase daily activity.

Analysis of the Unstable Group

The step count of user L821 was more volatile but showed a gradual increase, particularly in the final days, where there was a noticeable uptick. This change might be due to an increased health awareness or changes in the user's lifestyle.

Analysis of Other Unstable Group Users

The step counts of these users were more erratic, with no clear trends. These users may require more support and encouragement to maintain a stable exercise routine.

As the activity period approached its conclusion, it appears that the users became aware that their rankings were unlikely to change, leading to a decrease in their motivation to continue increasing their step counts. In such scenarios, users may feel that further competition is unnecessary or may believe that additional steps will not significantly affect their rankings.

During the result analysis, interview findings revealed that users were motivated to exercise more when they saw the leading user putting in effort. This behavior indicates that users' behavior changes were primarily driven by social influence rather than the Hawthorne effect. Social influence refers to the impact of others' actions on an individual's behavior, particularly when those individuals are seen as role models or leaders.

When users observed the leading user actively exercising, they felt encouraged to increase their own activity levels. This imitation behavior suggests that in group activities, individual behavior is often influenced by other group members. This underscores the importance of leveraging social influence when designing health management programs to motivate behavioral change.

This social influence can create a ripple effect within a group, where one person's actions inspire others to follow suit, leading to an overall increase in group activity levels. Therefore, future research and interventions should capitalize on this by designing health behavior change programs that incorporate incentivizing social interactions and competitive elements.

Pre- and Postintervention BMI Comparisons for All Users

To assess the changes in BMI among users before and after the fitness training program, this study conducted pre- and posttest comparisons of users' BMI data, as summarized in Table 5, followed by further analysis.

Table 5 presents the distribution of participants' BMI classifications before and after the 8-week intervention. The table includes both the number and percentage of participants in each BMI category (underweight, slightly underweight, normal, overweight, and obese), with notable shifts observed following the intervention. Specifically, the proportion of participants classified as obese significantly decreased from 38.5% (90/234) to 13.7% (32/234), while the proportion of those classified as overweight increased from 24.8% (58/234) to 34.6% (81/234). The combined percentage of overweight or obese participants declined from 63.2% (148/234) to 48.3% (113/234), accompanied by an increase in participants with normal BMI from 18.4% (43/234) to 33.8% (79/234). These changes were statistically significant, as confirmed by a chi-square test (χ^2_4=29.98; P<.001), indicating a meaningful improvement in participants' weight status following the intervention.

Further analysis revealed that many individuals who were previously classified as obese successfully transitioned to the overweight or normal BMI category. Additionally, small changes were observed in the underweight category, with 1 participant shifting into the underweight category and 1 into the normal BMI category, possibly reflecting improved body awareness or behavior change. These findings support the efficacy of the gamified mobile app in promoting positive BMI outcomes among nurses in the post-COVID era.

Table 5. BMI classification before and after the intervention.

BMI category	Pretest[a] (N=234), n (%)	Posttest[a] (N=234), n (%)
Underweight	23 (9.8)	24 (10.3)
Slightly underweight	20 (8.5)	18 (7.7)
Normal	43 (18.4)	79 (33.8)
Overweight	58 (24.8)	81 (34.6)
Obese	90 (38.5)	32 (13.7)

[a]A chi-square test was conducted to compare pre- and postintervention distributions (χ^2_4=29.98; P<.001).

Impact of the Gamified Health App on BMI Category Transitions

The data in Table 5 reveal significant shifts in BMI distribution following the fitness training intervention.

The number of users in the underweight category increased by 1, indicating that 1 user moved from the slightly underweight category to the underweight category. This may suggest that the user's weight did not increase as expected after the intervention, possibly due to the absence of a dietary management feature in the study's app.

XSL•FO
RenderX

The number of slightly underweight users decreased by 2, with 1 user moving to the underweight category and another successfully entering the normal BMI range, indicating a positive impact of the training on weight control for these users.

There was a significant increase in the number of users within the normal BMI range, rising from 43 in the pretest period to 79 in the posttest period. This increase was largely due to users from the overweight and obese categories moving toward a healthier range, suggesting that the majority of users who were previously overweight or slightly overweight achieved better weight control, reaching a normal BMI after the intervention.

Both the overweight and obese categories saw notable reductions in user numbers. The number of overweight users increased slightly from 58 to 81, primarily due to users from the obese category successfully reducing their weight and moving into the overweight range. However, the number of obese users decreased significantly from 90 to 32, indicating that a large portion of obese users were able to reduce their BMI to healthier levels through the training program.

Discussion

Principal Findings

This study evaluated the preliminary efficacy of a gamified mobile app designed to promote self-health management among nurses in the post-COVID context. The principal findings revealed that participants showed a significant improvement in BMI classification following the 8-week intervention, with a notable reduction in the proportion of individuals classified as obese. The app's gamified features, developed using the Octalysis framework, successfully enhanced user engagement, with the highest motivational scores observed in the domains of *development and accomplishment* and *epic meaning and calling*. Additionally, qualitative interviews confirmed that nurses valued the app's goal-setting, achievement rewards, and personalization features, which supported sustained participation. These results suggest that a well-designed gamified mobile health intervention can effectively promote physical activity and self-management behaviors in health care professionals.

User Engagement Insights and Implications for Sustainable Health Interventions

This study's participants primarily consisted of young, female health care professionals with university-level education, reflecting a demographic that makes up a significant portion of the nursing workforce. However, this concentration may limit the generalizability of findings to populations with different age distributions, gender ratios, or educational backgrounds. Age distribution showed that 33.3% (78/234) of participants were younger than 30 years, 52.6% (123/234) were between 31 and 50 years, and 14.1% (33/234) were older than 51 years. Interviews suggested that younger participants were more inclined to adopt new technologies and gamified apps, leading to higher engagement with the health management game. In contrast, older participants did not exhibit common motivations, indicating the need for tailored strategies across age groups. In terms of gender, female participants made up 62.0% (145/234) of the population and male participants made up 38.0% (89/234). Regarding education, 58.5% (137/234) had a university degree, 16.2% (38/234) had completed high school or vocational school, and 19.7% (46/234) held an associate degree. Only 5.6% (13/234) had a graduate degree. These findings suggest that women and individuals with higher educational attainment are more likely to engage in health management activities, which is crucial for future intervention design. Regarding exercise habits, 45.3% (106/234) of participants reported a regular exercise routine, with most of these exercising less than 2 hours per week. Only 7.3% (17/234) played exercise-related games, and 76.5% of these did so for fewer than 5 hours per week. This low participation in physical activity and gamified health apps underscores the need for enhanced strategies to increase engagement, especially among health care workers with limited leisure time due to heavy workloads. Activity tracking data revealed that step counts generally increased over time but declined near the end of the study. This pattern suggests that behavior was influenced by the perception of being monitored. Some users increased their activity due to perceived observation but relaxed their efforts once accountability diminished. This highlights the motivational role of structured timelines and perceived oversight in behavior change. Qualitative interviews emphasized the influence of peer leadership. For example, user K825 consistently demonstrated high activity levels and was seen as a motivational figure. This kind of role modeling aligns with the organizational health promotion theory. To succeed, institutions should foster a health-oriented culture by empowering peer leaders and providing the necessary support. Lastly, the long-term sustainability of health interventions depends on more than individual motivation. Continuous evaluation, user feedback, and supportive educational structures are needed. While identifying highly engaged individuals is a strong starting point, systemic support is essential for maintaining engagement and realizing sustained improvements in employee wellness.

Comparison With Prior Work

This study contributes to the growing body of research on gamified health management apps, particularly in the context of health care workers, by addressing gaps and building on previous studies. Previous research has demonstrated the efficacy of gamification in various health-related interventions, such as promoting physical activity, enhancing adherence to healthy behaviors, and improving overall well-being [26,27]. However, much of the existing literature has focused on the general population or specific patient groups rather than health care workers, who face unique stressors and challenges in maintaining their health due to the demanding nature of their work [2,13].

Unlike earlier studies that predominantly examined the impact of gamification on a broader audience, this research specifically targeted health care professionals, particularly nurses in a post–COVID-19 environment. This focus allowed for a more nuanced understanding of how gamified interventions can be tailored to meet the specific needs of health care workers, addressing both physical and mental health challenges exacerbated by the pandemic. While previous studies have highlighted the potential of gamification to increase engagement

XSL•FO
RenderX

and motivation [14,23], this study goes further by integrating the Octalysis framework and the FITTF principle, which adds a structured approach to designing gamified health interventions that are both engaging and effective.

Moreover, this study's use of in-depth interviews and feedback loops to iteratively refine the app distinguishes it from prior research that often employed more static or one-size-fits-all gamification strategies. The focus on iterative design based on user feedback ensures that the app remains relevant and user-friendly, which has been identified as a critical factor in the success of health-related interventions [25].

In comparison to earlier studies, the findings of this study suggest that health care workers respond positively to gamified health management tools when these tools are specifically designed to align with their work environment and health needs. This contrasts with the more mixed results seen in broader population studies, where the efficacy of gamification can be less consistent due to the diversity of user backgrounds and motivations [35].

Overall, this research not only confirms the potential benefits of gamification in health management but also highlights the importance of context-specific design in enhancing the efficacy of these interventions for health care workers. It builds on prior work by offering a more targeted approach, contributing to a deeper understanding of how gamification can be effectively applied in high-stress, high-demand professional environments.

Limitations

This section addresses the limitations encountered in this study and proposes potential avenues for improvement. Understanding these limitations is crucial for future researchers to consider the study's findings more comprehensively and to guide further exploration in this field.

First, the participants in this study were selected from a specific health care institution, with a primary focus on health care workers within that institution. This focus allowed for an in-depth understanding of the particular needs and challenges within this specific work environment, enabling the provision of practical and tailored recommendations for the staff in this setting. Given that health care workers' job characteristics, stressors, and health needs are significantly different from those of other professional groups, this targeted approach ensures the relevance and efficacy of the study's conclusions.

However, this focus also limits the generalizability of the findings. In other types of health care institutions, such as large general hospitals, small clinics, or primary care facilities in remote areas, the work environment and staff needs may vary considerably. As a result, the study's findings may not be fully applicable in these different contexts. Additionally, health care personnel in other functional areas, such as administrative management, logistical support, or technical roles, may experience different stressors and health needs compared to the clinical workers primarily considered in this study. Therefore, applying the study's results to these other contexts should be done with caution. To address this limitation, future research should consider including participants from multiple institutions, settings, and functional roles to enhance the external validity and applicability of the findings.

Second, another limitation is the self-selection of participants, which introduces the potential for selection bias and may affect the representativeness of the study's findings. Volunteers may be more motivated, health conscious, or interested in gamified interventions, which could lead to more favorable attitudes and outcomes during the study period. As such, the results may primarily reflect the characteristics of this specific subgroup rather than those of the broader population of health care workers. Furthermore, the sample was drawn from a single hospital, which may limit the generalizability of the findings to other clinical environments or to health care personnel in nonclinical roles. These factors should be carefully considered when interpreting the study outcomes.

Third, although 271 participants completed the baseline questionnaire, only 198 provided valid responses to the postintervention Octalysis assessment. This drop in the response rate may be attributed to participant fatigue, reduced motivation following the intervention period, or competing clinical duties. Such attrition may have introduced response bias and reduced the completeness and representativeness of the postintervention data. To mitigate this limitation in future studies, researchers should consider using shorter or more engaging follow-up instruments, sending scheduled reminders, or providing appropriate incentives to improve response rates and ensure data integrity.

Fourth, one of the most important limitations of this study is the use of a single-group pre-post design without a control or comparison group. While this approach allowed for the observation of changes in physical activity and BMI among the same participants, it inherently limited the ability to draw causal conclusions. Without a comparison group, it is not possible to rule out the influence of confounding factors or natural changes over time on the observed outcomes. External factors, such as seasonal variations, organizational initiatives, peer influence, or participants' increased self-awareness, may have contributed to the improvements.

The absence of a control group was primarily due to practical and logistical constraints. The intervention was implemented in a real-world hospital setting where participants were also full-time clinical staff. Limited personnel availability, high clinical workload, and institutional restrictions on randomization made it infeasible to assign participants to a control or waitlist group without disrupting operations.

In addition, the sample consisted of voluntary participants recruited from a single institution. Individuals more interested in personal health or gamified technologies may have been more likely to participate, potentially leading to selection bias and limiting the generalizability of findings to the wider health care workforce.

Future studies should adopt randomized controlled trial (RCT) designs where possible. When full randomization is not feasible, alternative quasiexperimental designs, such as waitlist-control or matched cohort approaches, could provide a practical yet

XSL•FO
RenderX

rigorous alternative to strengthen internal validity and causal inference while accommodating real-world constraints.

To address the limitations identified in this study, future research should consider adopting RCT designs to better establish causal relationships and minimize potential sources of bias. RCTs can provide more robust evidence of efficacy by controlling for external influences, accounting for natural changes over time, and improving the internal validity of the findings. Incorporating a control or comparison group would allow for a more accurate assessment of the intervention's effects, and the use of stratified or random sampling methods would enhance the generalizability of the results across diverse populations and institutional settings. Although this study yielded valuable insights into the impact of a gamified intervention among health care workers in a specific institutional context, these findings should be applied with caution when extending to broader settings. Future studies are encouraged to explore long-term outcomes and assess intervention sustainability over time using more rigorous experimental designs.

Recommendations

Based on the findings, hospitals should consider systematically incorporating gamified apps into employee wellness programs to improve nurses' health behaviors. Management support, training programs, and continuous user feedback mechanisms should be established to sustain engagement. Future studies should adopt RCTs and broaden participant diversity to confirm the efficacy and generalizability of gamified health interventions in health care contexts.

Future Directions and Follow-Up

In future research, we plan to adopt more rigorous experimental designs to strengthen the internal validity of the findings. Where feasible, an RCT design will be considered. Alternatively, quasiexperimental approaches, such as a waitlist-control group or matched cohort design, could offer a practical yet methodologically sound solution in clinical environments where full randomization is not feasible. These approaches would allow for clearer attribution of effects to the intervention while balancing real-world implementation constraints. Although this study focused primarily on physical health indicators, such as daily step counts and BMI, psychological dimensions, such as stress, anxiety, and emotional exhaustion, are equally important in understanding the holistic impact of health interventions. Future iterations of the app evaluation should incorporate validated tools, such as the Perceived Stress Scale (PSS) and Generalized Anxiety Disorder-7 (GAD-7), to assess mental well-being outcomes. While this study demonstrated short-term improvements in physical activity and BMI following the 8-week gamified intervention, long-term efficacy remains an important consideration. To address this, a follow-up study is planned to assess the sustainability of outcomes at 3 and 6 months after the intervention. This follow-up will involve reassessing participants' daily step counts and BMI to determine whether behavior changes are maintained over time. Additional qualitative interviews will explore user experiences and identify factors influencing sustained engagement. These data will help inform refinements to the app and guide broader implementation strategies in future clinical settings.

Conclusion

This study aimed to evaluate the preliminary efficacy of a gamified mobile app in promoting self-health management among nurses in the post-COVID era. The findings demonstrated statistically significant improvements in participants' BMI classifications and strong motivational engagement driven by key gamification features designed using the Octalysis framework. The app successfully supported physical activity and goal-oriented behaviors through achievement systems, feedback mechanisms, and social influence. These results suggest that gamified digital health tools can serve as effective self-care interventions for health care professionals in high-stress settings. Further research is needed to assess the long-term effects and scalability of such interventions.

Acknowledgments

We thank the participants who contributed their time and insights during the interviews and surveys. We also acknowledge the staff and administrators of the medical institutions for their cooperation and assistance in facilitating the data collection process. Finally, we are grateful to our friends and family for their understanding, encouragement, and continued support throughout the research process. The gamified health management app was developed and implemented independently, and no financial or personal relationships exist that could have influenced the findings.

Data Availability

The data sets generated and analyzed during this study are not publicly available due to institutional data protection policies and participant privacy considerations. However, deidentified data may be made available by the corresponding author upon reasonable request.

Authors' Contributions

LJL and SHH conceptualized the study. SHH and LJL designed the methodology, performed the investigations, supervised the study, and wrote the original manuscript draft. LJL validated the findings, acquired funding, and reviewed and edited the manuscript. SHH curated the study and performed the formal analysis. PCC synthesized feedback and contributed to writing–review and editing.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Draft interview outline.
[DOCX File , 28 KB - games_v13i1e66262_app1.docx]

References

1. Nguyen H, Huang H, Wong M, Yang Y, Huang T, Teng C. Moderator roles of optimism and weight control on the impact of playing exergames on happiness: the perspective of social cognitive theory using a randomized controlled trial. Games Health J 2018 Aug;7(4):246-252 [FREE Full text] [doi: 10.1089/g4h.2017.0165] [Medline: 29708775]
2. Chutiyami M, Bello U, Salihu D, Ndwiga D, Kolo M, Maharaj R, et al. COVID-19 pandemic-related mortality, infection, symptoms, complications, comorbidities, and other aspects of physical health among healthcare workers globally: An umbrella review. Int J Nurs Stud 2022 May;129:104211 [FREE Full text] [doi: 10.1016/j.ijnurstu.2022.104211] [Medline: 35278750]
3. David HMSL, Rafael RMR, de Mello Alves MG, Breda KL, de Araújo Faria MG, Neto M, et al. Infection and mortality of nursing personnel in Brazil from COVID-19: A cross-sectional study. International Journal of Nursing Studies 2021;124:104089 [FREE Full text] [doi: 10.1016/j.ijnurstu.2021.104089]
4. Saragih I, Tonapa S, Saragih I, Advani S, Batubara S, Suarilah I, et al. Global prevalence of mental health problems among healthcare workers during the Covid-19 pandemic: A systematic review and meta-analysis. Int J Nurs Stud 2021 Sep;121:104002 [FREE Full text] [doi: 10.1016/j.ijnurstu.2021.104002] [Medline: 34271460]
5. Yifan T, Ying L, Chunhong G, Jing S, Rong W, Zhenyu L, et al. Symptom cluster of ICU nurses treating COVID-19 pneumonia patients in Wuhan, China. J Pain Symptom Manage 2020 Jul;60(1):e48-e53 [FREE Full text] [doi: 10.1016/j.jpainsymman.2020.03.039] [Medline: 32276095]
6. Yunitri N, Chu H, Kang X, Jen H, Pien L, Tsai H, et al. Global prevalence and associated risk factors of posttraumatic stress disorder during COVID-19 pandemic: A meta-analysis. Int J Nurs Stud 2022 Feb;126:104136 [FREE Full text] [doi: 10.1016/j.ijnurstu.2021.104136] [Medline: 34856503]
7. Dai Y, Hu G, Xiong H, Qiu H, Yuan X. Psychological impact of the coronavirus disease 2019 (COVID-19) outbreak on healthcare workers in China. medrxiv. 2020. URL: https://www.medrxiv.org/content/10.1101/2020.03.03.20030874v1 [accessed 2025-06-19]
8. Lai J, Ma S, Wang Y, Cai Z, Hu J, Wei N, et al. Factors associated with mental health outcomes among health care workers exposed to coronavirus disease 2019. JAMA Netw Open 2020 Mar 02;3(3):e203976-e203977 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.3976] [Medline: 32202646]
9. Pappa S, Ntella V, Giannakas T, Giannakoulis VG, Papoutsi E, Katsaounou P. Prevalence of depression, anxiety, and insomnia among healthcare workers during the COVID-19 pandemic: A systematic review and meta-analysis. Brain Behav Immun 2020 Aug;88:901-907 [FREE Full text] [doi: 10.1016/j.bbi.2020.05.026] [Medline: 32437915]
10. Shaukat N, Ali D, Razzak J. Physical and mental health impacts of COVID-19 on healthcare workers: a scoping review. Int J Emerg Med 2020 Jul 20;13(1):40 [FREE Full text] [doi: 10.1186/s12245-020-00299-5] [Medline: 32689925]
11. Chen WH, Chen YZ. ESG sustainable development: A sustainability guide for businesses, understanding what needs to be done for ESG. Businessweekly. 2024. URL: https://www.businessweekly.com.tw/carbon-reduction/blog/3010720 [accessed 2024-09-01]
12. Piercy KL, Troiano RP, Ballard RM, Carlson SA, Fulton JE, Galuska DA, et al. The physical activity guidelines for Americans. JAMA 2018 Nov 20;320(19):2020-2028 [FREE Full text] [doi: 10.1001/jama.2018.14854] [Medline: 30418471]
13. Melnyk BM, Hsieh AP, Tan A, Gawlik KS, Hacker ED, Ferrell D, et al. The state of mental health and healthy lifestyle behaviors in nursing, medicine and health sciences faculty and students at Big 10 Universities with implications for action. J Prof Nurs 2021;37(6):1167-1174 [FREE Full text] [doi: 10.1016/j.profnurs.2021.10.007] [Medline: 34887036]
14. Yang H, Du HS, Wang L, Wu T. The influence of social support networks on health conditions via user engagement: Gender as a moderator. Journal of Electronic Commerce Research 2019;20(1):35-54 [FREE Full text]
15. Ross J. The mental health impact of COVID on nurses: a patient care concern. J Perianesth Nurs 2022 Dec;37(6):973-974 [FREE Full text] [doi: 10.1016/j.jopan.2022.09.010] [Medline: 36273968]
16. Aguilar-Martínez A, Solé-Sedeño J, Mancebo-Moreno G, Medina F, Carreras-Collado R, Saigí-Rubió F. Use of mobile phones as a tool for weight loss: a systematic review. J Telemed Telecare 2014 May 29;20(6):339-349 [FREE Full text] [doi: 10.1177/1357633x14537777]
17. Bardus M, Smith J, Samaha L, Abraham C. Mobile phone and web 2.0 technologies for weight management: a systematic scoping review. J Med Internet Res 2015 Nov 16;17(11):e259 [FREE Full text] [doi: 10.2196/jmir.5129] [Medline: 26573984]
18. Quelly S, Norris A, DiPietro J. Impact of mobile apps to combat obesity in children and adolescents: A systematic literature review. J Spec Pediatr Nurs 2016 Jan;21(1):5-17 [FREE Full text] [doi: 10.1111/jspn.12134] [Medline: 26494019]

XSL•FO
RenderX

19. Riaz S, Sykes C. Are smartphone health applications effective in modifying obesity and smoking behaviours? A systematic review. Health Technol 2015 Mar 25;5(2):73-81 [FREE Full text] [doi: 10.1007/s12553-015-0104-4]
20. Flores Mateo G, Granado-Font E, Ferré-Grau C, Montaña-Carreras X. Mobile phone apps to promote weight loss and increase physical activity: a systematic review and meta-analysis. J Med Internet Res 2015 Nov 10;17(11):e253 [FREE Full text] [doi: 10.2196/jmir.4836] [Medline: 26554314]
21. Bardus M, Smith J, Samaha L, Abraham C. Mobile and Web 2.0 interventions for weight management: an overview of review evidence and its methodological quality. Eur J Public Health 2016 Aug;26(4):602-610 [FREE Full text] [doi: 10.1093/eurpub/ckw090] [Medline: 27335330]
22. Schippers M, Adam P, Smolenski D, Wong H, de Wit JBF. A meta-analysis of overall effects of weight loss interventions delivered via mobile phones and effect size differences according to delivery mode, personal contact, and intervention intensity and duration. Obes Rev 2017 Apr;18(4):450-459 [FREE Full text] [doi: 10.1111/obr.12492] [Medline: 28187246]
23. Tortorella G, Fogliatto F, Espôsto K, Vergara A, Vassolo R, Mendoza D, et al. Effects of contingencies on healthcare 4.0 technologies adoption and barriers in emerging economies. Technological Forecasting and Social Change 2020 Jul;156:120048 [FREE Full text] [doi: 10.1016/j.techfore.2020.120048]
24. Burnet K, Kelsch E, Zieff G, Moore J, Stoner L. How fitting is F.I.T.T.?: A perspective on a transition from the sole use of frequency, intensity, time, and type in exercise prescription. Physiol Behav 2019 Feb 01;199:33-34 [FREE Full text] [doi: 10.1016/j.physbeh.2018.11.007] [Medline: 30414400]
25. Alahäivälä T, Oinas-Kukkonen H. Understanding persuasion contexts in health gamification: A systematic analysis of gamified health behavior change support systems literature. Int J Med Inform 2016 Dec;96:62-70 [FREE Full text] [doi: 10.1016/j.ijmedinf.2016.02.006] [Medline: 26944611]
26. Johnson D, Horton E, Mulcahy R, Foth M. Gamification and serious games within the domain of domestic energy consumption: A systematic review. Renewable and Sustainable Energy Reviews 2017 Jun;73:249-264 [FREE Full text] [doi: 10.1016/j.rser.2017.01.134]
27. Sardi L, Idri A, Fernández-Alemán JL. A systematic review of gamification in e-Health. J Biomed Inform 2017 Jul;71:31-48 [FREE Full text] [doi: 10.1016/j.jbi.2017.05.011] [Medline: 28536062]
28. Hunicke R, LeBlanc M, Zubek R. MDA: A Formal Approach to Game Design and Game Research. AAAI. 2004. URL: https://cdn.aaai.org/Workshops/2004/WS-04-04/WS04-04-001.pdf [accessed 2025-06-19]
29. Landers R, Bauer K, Callan R. Gamification of task performance with leaderboards: A goal setting experiment. Computers in Human Behavior 2017 Jun;71:508-515 [FREE Full text] [doi: 10.1016/j.chb.2015.08.008]
30. Chou YK. Actionable Gamification: Beyond Points, Badges, and Leaderboards. Fremont, CA: Octalysis Media; 2017.
31. Marisa F, Sakinah S, Izzah Z, L A, David R, Aris A. Evaluation of student core drives on e-learning during the Covid-19 with octalysis gamification framework. IJACSA 2020;11(11):104-116 [FREE Full text] [doi: 10.14569/ijacsa.2020.0111114]
32. Oliveira S, Cruz M. The gamification octalysis framework within the primary english teaching process: the quest for a transformative classroom. Revista Lusófona de Educação 2018;41(41):63-82 [FREE Full text] [doi: 10.24140/issn.1645-7250.rle41.04]
33. Sulispera T, Recard M. Octalysis gamification framework for enhancing students' engagement in language learning. DIALEKTIKA Jurnal Pendidikan Bahasa Inggris 2020;8(2):103-128 [FREE Full text] [doi: 10.58436/jdpbi.v8i2.670]
34. Mustikasari AA, Yusuf M, Rejekiningsih T. Storyboard visualization for gamification design for deaf children's education using octalysis approach. Journal of International Conference Proceedings 2022;5(6):51-60 [FREE Full text] [doi: 10.32535/ijcp.v5i1.1443]
35. Xu L, Shi H, Shen M, Ni Y, Zhang X, Pang Y, et al. The effects of mHealth-based gamification interventions on participation in physical activity: systematic review. JMIR Mhealth Uhealth 2022 Feb 03;10(2):e27794 [FREE Full text] [doi: 10.2196/27794] [Medline: 35113034]
36. Fried Y, Slowik L. Enriching goal-setting theory with time: an integrated approach. Academy of Management Review 2004;29(3):404-422 [FREE Full text] [doi: 10.2307/20159051]
37. Locke E, Latham G. Building a practically useful theory of goal setting and task motivation. A 35-year odyssey. Am Psychol 2002 Sep;57(9):705-717 [FREE Full text] [doi: 10.1037//0003-066x.57.9.705] [Medline: 12237980]
38. Daniels A, Yeates D. Basic Training in Systems Analysis. Lanham, MD: Pitman; 1969.
39. Tudor-Sfetea C, Rabee R, Najim M, Amin N, Chadha M, Jain M, et al. Evaluation of two mobile health apps in the context of smoking cessation: qualitative study of cognitive behavioral therapy (CBT) versus non-CBT-based digital solutions. JMIR Mhealth Uhealth 2018 Apr 18;6(4):e98 [FREE Full text] [doi: 10.2196/mhealth.9405] [Medline: 29669708]
40. Patton MQ. Qualitative Research & Evaluation Methods: Integrating Theory and Practice. Thousand Oaks, CA: SAGE Publications, Inc; 1990.
41. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
42. Lin Y, Tudor-Sfetea C, Siddiqui S, Sherwani Y, Ahmed M, Eisingerich AB. Effective behavioral changes through a digital mHealth app: exploring the impact of hedonic well-being, psychological empowerment and inspiration. JMIR Mhealth Uhealth 2018 Jun 15;6(6):e10024 [FREE Full text] [doi: 10.2196/10024] [Medline: 29907557]

XSL•FO
RenderX

Abbreviations

CBT: cognitive behavioral therapy
FITT: frequency, intensity, time, and type
FITTF: frequency, intensity, time, type, and fun
RCT: randomized controlled trial
SDLC: system development life cycle
WHO: World Health Organization

Edited by A Coristine; submitted 08.09.24; peer-reviewed by M Theodoratou, G La Torre, D Patel, A Eisingerich; comments to author 01.03.25; revised version received 21.05.25; accepted 11.06.25; published 03.07.25.

Please cite as:
Hsu SH, Lu LJ, Chou PC
Preliminary Efficacy of a Gamified Mobile App for Promoting Self-Health Management Among Nurses in the Post-COVID Era: Single-Group Pre-Post Study
JMIR Serious Games 2025;13:e66262
URL: https://games.jmir.org/2025/1/e66262
doi:10.2196/66262
PMID:

©Shao Huan Hsu, Li Jung Lu, Pei Chin Chou. Originally published in JMIR Serious Games (https://games.jmir.org), 03.07.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Active Video Games to Improve Behavioral Intentions and Cognitive Function in Patients With Schizophrenia: Randomized Controlled Trial

Huan-Hwa Chen[1], PhD; Ching-Ching Lin[2], BS; Man-Ling Yu[2], MSD; Hsiu-Lan Wu[2], BS; Hui-Chu Shen[2], BS; Hsiu-Fen Hsieh[3,4], PhD

[1]Department of Nursing, Chung Hwa University of Medical Technology, Tainan, Taiwan

[2]Department of Nursing, Taipei Veterans General Hospital, Yuanshan Branch, Yilan, Taiwan

[3]School of Nursing, Kaohsiung Medical University, Kaohsiung, Taiwan

[4]Department of Medical Research, Kaohsiung Medical University Hospital, Kaohsiung Medical University, Kaohsiung, Taiwan

Corresponding Author:
Hsiu-Fen Hsieh, PhD
School of Nursing, Kaohsiung Medical University
100 Shih-Chuan 1st Rd., Sanmin District
Kaohsiung, 80708
Taiwan
Phone: 886 073121101 ext 2626
Fax: 886 07218364
Email: hsiufen96@gmail.com

Abstract

Background: Schizophrenia is a severe mental illness that affects the cognitive, social, and daily functions of patients. Physical activity has been found to be important for maintaining these functions in patients with schizophrenia, but many lack the motivation to participate in physical activities.

Objective: This study aimed to explore the efficacy of active video games (AVGs) on the behavioral intention and cognitive function of patients with schizophrenia.

Methods: In this experimental study, 103 participants were recruited from 2 medical centers. All participants were randomly assigned to the experimental or control group, and 82 participants (n=41, 50% in the experimental group and n=41, 50% in the control group) completed all the processes of our protocol. The experimental group was provided with AVGs for 30 minutes twice per week for 6 weeks. The Mini Mental State Examination and a behavioral intention questionnaire were administered before and after playing the AVGs. Data were collected between April 2021 and January 2022. Generalized estimating equations and 2-tailed paired *t* tests were used for data analysis.

Results: The experimental group showed significant improvements in behavioral intention to participate in AVGs compared with the control group at both T1 (β=4.88; *P*=.009) and T2 (β=4.24; *P*=.04). In addition, the experimental group experienced significant improvements in orientation (T2: β=0.66; *P*=.04) and language (T2: β=0.28; *P*=.03) among cognitive functions compared to the control group. In contrast, there was no significant change in these variables in the control group.

Conclusions: Playing AVGs can effectively enhance the behavioral intention of patients with schizophrenia to participate in physical activity and exercise and significantly improve their orientation and language. AVGs are inexpensive and easily operated tools for people with mental or physical disabilities.

Trial Registration: ClinicalTrials.gov NCT05933356; https://clinicaltrials.gov/study/NCT05933356

(JMIR Serious Games 2025;13:e69116) doi:10.2196/69116

KEYWORDS

schizophrenia; behavioral intention; cognitive function; active video games

XSL•FO
RenderX

Introduction

Background

Schizophrenia can lead to high functional disability and is associated with a heavy economic burden worldwide. The World Health Organization lists it as one of the top 10 conditions in terms of the global burden of disease [1,2]. Schizophrenia is a disease characterized by impairments in cognitive functions. Positive symptoms of this disorder, such as delusions and hallucinations, respond relatively well to antipsychotic medication, whereas the management of negative symptoms remains a clinical challenge [3,4]. Negative symptoms are associated with severe deterioration in cognition and global functioning [5,6]. Patients with negative symptoms often exhibit weak motivation for activities, which accelerates the degradation of cognitive functions. Schizophrenia is a condition that inherently leads to impaired cognitive function, compounded by patients' low motivation to engage in activities, which further accelerates cognitive decline. A study conducted in Singapore evaluated the utility of the Mini Mental State Examination (MMSE) among long-term hospitalized patients diagnosed with schizophrenia or schizoaffective disorder. The findings indicated that MMSE scores were significantly influenced by sociodemographic and clinical variables, including age, educational attainment, and duration of illness [7]. A study conducted in the United States assessed partially remitted outpatients attending a depot antipsychotic clinic or a clozapine clinic (N=272) using the MMSE. The findings indicated that the cognitive gap did not appear to widen significantly with increasing age. The most commonly observed deficits among patients with poorer performance were in memory, attention, and constructional tasks. The MMSE was found to be easy to administer and applicable for use among outpatient populations with schizophrenia [8].

Exercise not only improves physical function but also serves as a crucial factor in enhancing cognitive function. Multiple factors hinder the efficacy of treating patients with schizophrenia, including limited insight, poor adherence, treatment resistance [9], antipsychotic-related side effects [10], and cultural factors [11]. Therefore, new treatment options are required to promote recovery and optimize rehabilitation [12]. Many studies have indicated that exercise can be an additional option for the prevention and treatment of schizophrenia [4,13-15] and it can improve patients' positive and negative symptoms, quality of life, and cognitive function [13,16]. In addition, exercise enhances physical well-being and decreases potential somatic comorbidities such as metabolic syndrome [3] and cardiovascular diseases in patients with schizophrenia [3]. A systematic review indicated that exercise significantly improved patients' working memory, social cognition, and attention and vigilance [17].

A previous systematic review identified exercise interventions as one of the most effective strategies for improving mental health. These interventions typically lasted for 1 to 3 months, with a frequency of 2 to 3 sessions per week and most sessions lasting between 30 and 60 minutes [18]. Exercise not only increases cerebral blood flow—providing the brain with more oxygen and nutrients—but also reduces neuronal apoptosis, promotes neurogenesis and brain plasticity, and enhances cognitive reserve. A previous study systematically reviewed the relationship among exercise, brain-derived neurotrophic factor, neural plasticity, and cognitive function. The results indicated that exercise can elevate brain-derived neurotrophic factor levels in the brain, thereby promoting neural plasticity and enhancing cognitive function [19].

Most patients with schizophrenia lack motivation for physical activities such as exercise and social interaction owing to various factors, including the effects of negative symptoms [3,4,18,19]. However, in recent decades, they have been recognized as interested in modern technology-assisted activities [20], and such activities are a possible method of rehabilitation. For example, video games, active video games (AVGs) [21-24], and virtual reality [25] have been documented as novel interventions for cognitive training and for alleviating psychotic symptoms in patients with schizophrenia [24,26,27]. A systematic review also found a positive relationship between video games and cognitive functions such as memory [28-30], social cognition [31], executive function [32], attention [33], and reaction time [34]. It has been suggested that playing video games can have cognitive, motivational, emotional, and social benefits [23] with a lower cost than those of other cognitive interventions [35].

With the rapidly changing technologies in today's world, AVGs not only are fun but can also serve as exercise tools, making it more convenient for people to exercise. AVGs refer to exercises that use video games to make exercise more fun [36]. A previous systematic review focused on interactive motion-sensing games that required players to engage in physical movements such as waving, jumping, kicking, or full-body motion to control internet-based characters or complete in-game tasks. These games integrate entertainment with physical activity, breaking away from the traditional sedentary gaming model [37]. AVGs are considered capable of enhancing an individual's behavioral intention to engage in these activities by fostering imagination, control, challenge, and curiosity. Engagement in exercise can increase either through consistently finding satisfaction in familiar activities or by trying new ones to sustain interest and motivation. [38]. One study reported that >90% of the participants were satisfied with playing AVGs, and most participants (76.9%) agreed that this intervention promotes healthier lifestyles and is an acceptable alternative way to perform physical activity [39]. AVGs significantly improve cognitive domains, including processing speed, attention and vigilance, working memory, verbal learning, visual learning, reasoning and problem-solving [40], social cognition, social functioning, and motivation [24] and also reduce cognitive impairments and psychological stress [41]. A systematic review indicated that AVGs were globally beneficial for patients with schizophrenia, particularly with regard to their cognitive function [42].

In summary, AVGs are accessible and easy to operate for patients with disabilities [43,44], and many studies indicate that interventions integrating gameplay with physical exercise can enhance motivation for device-assisted activities; alleviate residual symptoms; and improve cognitive, social, and

XSL•FO
RenderX

occupational functioning, thereby contributing to an improved quality of life among patients with schizophrenia. Nevertheless, despite the growing global emphasis on digital health interventions, studies specifically investigating the impact of AVGs on this population in Taiwan remain scarce. This highlights a critical need to develop innovative, accessible, and engaging strategies to facilitate recovery and rehabilitation in patients living with schizophrenia. Furthermore, addressing cognitive deficits—a core and profoundly disabling characteristic of the disorder—is imperative due to their significant influence on daily functioning and overall well-being [19].

Objectives

This study aimed to explore the role of AVGs in influencing the behavioral intention of patients with schizophrenia to participate in novel device-assisted activities and examine their efficacy in improving cognitive function.

We hypothesized that the experimental group would show greater improvement in cognitive function than the control group at each assessment point (3 and 6 weeks) during the intervention. In addition, we expected that playing AVGs would increase the experimental group's behavioral intention to continue engaging with AVGs after the intervention.

Methods

Study Design

The protocol was registered with ClinicalTrials.gov (NCT05933356) at the time of submission. We used an experimental design with random sampling, treating each day care center as a unit. Patients within these centers were assigned to either the experimental or control group using a random number generator (Microsoft Excel). This study was single blinded, with the research assistants being the blinded party. All participants received usual care, whereas those in the experimental group additionally engaged with AVGs over the 6-week study period.

Participants and Setting

This was an experimental study with participants recruited from day care centers from 2 hospitals in southern Taiwan. The MMSE and a behavioral intention questionnaire were administered before and after playing the AVGs.

The sample size was calculated using G*Power (version 3.1.1) for the F_1 test and its repeated measures with an effect size of 0.25, a significance level of .05, a power (1–β) of 0.80, and an estimated minimum sample size of 78 [45]. A total of 103 patients with schizophrenia participated in our study, and 82 (79.6%) completed all intervention processes and questionnaires. Before data collection, the principal investigator explained the purpose and process to the nurses in each ward and invited them to participate in the study.

The inclusion criteria for patients were as follows: (1) diagnosis of schizophrenia according to the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* [46]; (2) age of 20 to 65 years; (3) ability to read and understand traditional Chinese; and (4) current admission to a psychiatric day care center. The exclusion criteria were as follows: (1) patients with unstable mental conditions, such as delusions or hallucinations; and (2) patients with an intellectual disability or severe cognitive function impairment (MMSE score of <14) [47].

Study Protocol

Control Group

The control group only received scheduled activities in the day care center as usual without any intervention.

Experimental Group

The experimental group received routine care and participated in AVGs from Nintendo Ltd played for 30 minutes twice a week over 6 weeks. At the start, participants were instructed to play games designed to enhance attention, spatial intelligence, and enjoyment in a safe, spacious environment. Each patient acted as an independent player, engaging in fair competition that fostered intrinsic motivation and enhanced the ability to observe others' behaviors.

Participants were guided in playing the AVGs by the research team. Each session lasted approximately 30 minutes. Measurements were recorded at baseline, week 3, and week 6, designated as T0, T1, and T2, respectively.

Data Collection

Each participant was required to complete a demographic characteristic and behavioral intention questionnaire, and their cognitive functions were assessed using the MMSE. The principal investigator initially collected participants' demographic information through structured interviews. To ensure the accuracy and completeness of the data, the collected information was subsequently verified in collaboration with nurses at the day care center. Data were collected between April 2021 and September 2022.

Instruments

The Mario Party Series (Nintendo Ltd)

Mario Party is a party video game series featuring characters from the Mario franchise where up to 4 local players or computer-controlled characters compete in a board game interspersed with mini games. The games are currently developed and published by Nintendo. The *Mario Party* games feature at least 80 mini games across various types. Four-player games are typically free-for-all, where players compete individually. In other mini games, players form teams to cooperate and win. We used the Joy-Cons of the Nintendo Switch to control the game. Joy-Cons consist of 2 individual units, each containing an analog stick and a set of buttons. They can be used while attached to the main Nintendo Switch console unit or detached for wireless use. When detached, a pair of Joy-Cons can be used by a single player or split between 2 players as individual controllers. Each Joy-Con includes an accelerometer and gyroscope for motion tracking (Table 1).

XSL•FO
RenderX

Table 1. Name and description of each minigame.

Minigame	Image	Goals	Content
Sizzling Stakes		Enhance the patient's orientation	Shake the controller to control the pan. All 6 sides of the beef must be pan-seared to secure a win. Gameplay: 4 individuals play the game at the same time, independently but concurrently.
Snack Attack		Enhance the patient's orientation	Catch popcorn in a cup by controlling the remote button; 1point for each caught piece of popcorn. Gameplay: 4 individuals play the game at the same time, independently but concurrently.
Candy Shakedown		Enhance the patient's pace and orientation	Pour out all the candies from the jar by shaking the controller. The player with the fastest time is the winner. Gameplay: 4 individuals play the game at the same time, independently but concurrently.
Precision Gardening		Enhance the patient's hand strength and reasoning	Shake both controllers in a downward direction to water the flowers; the flower blooms as fast as the speed of shaking. The fastest player wins the game. Gameplay: 4 individuals play the game at the same time, independently but concurrently.
Pie Hard		Enhance the patient's communication, attention and reasoning	Throw pies to the opposite side by clicking the button; a point is earned when a pie hits a competitor. Gameplay: 4 individuals play the game at the same time, independently but concurrently.
Tall Order		Enhance the patient's orientation	Use the controller to catch as ice in the bowl as possible. Gameplay: 4 individuals play the game at the same time, independently but concurrently.
Rowboat Uprising		Enhance the patient's communication, attention and reasoning	The faster the controller is shaken, the less time is needed to sail to the goal. Players play the game in teams of two.
Net Worth		Enhance the patient's communication, attention and reasoning	Two players are in one team. The arms are lifted up together to catch fish while they swim through the net. Players play the game in teams of two.
Strike It Rich		Enhance the patient's attention and reasoning	The hands are lifted up when the character encounters a brick in the race. One point is won when the character hit a brick successfully. Gameplay: 4 individuals play the game at the same time, independently but concurrently.
Take a Stab		Enhance the patient's attention and reasoning	The arms are stretched forward when the fruit is close to the sword. Player can use beats shown on screen for timing. Gameplay: 4 individuals play the game at the same time, independently but concurrently.
Wiped Out		Enhance the patient's executive function and reasoning	Wipe soap from windows by waving the hands in a circular motion. A point is earned each time a soap bubble is wiped off. Gameplay: 4 individuals play the game at the same time, independently but concurrently.
Clearing the Table		Enhance the patient's attention and reasoning	Pull the tablecloth quickly when enough cups are placed on the table by shaking the hands. One success leads to 1 point earned. Gameplay: 4 players play together simultaneously, each controlling their own character.

Behavioral Intention Questionnaire Instrument

We measured participants' behavioral intention questionnaire to play AVGs using a self-developed questionnaire that was formulated with input from 4 experts and consisted of 4 items. Each item was rated on a 5-point Likert scale to assess the level of agreement with each statement, ranging from 0 (*strongly disagree*) to 4 (*strongly agree*). The questionnaire contained the following questions: "In your opinion, will you play and try different AVGs in the future? Do you think that AVGs can make you relaxed or vigorous?" "In your opinion, will you be interested in trying different types of AVGs?" "Do you think that AVGs can make you vigorous?" The content validity index of the questionnaire was 1.00, indicating excellent content validity. The total score for the 4 items ranged from 0 to 16. Higher scores indicated a stronger intention to play AVGs. The test-retest (2 weeks) intraclass correlation coefficient was 0.88. The Cronbach α for behavioral intention was 0.93 in this study.

MMSE Instrument

Cognitive tests were conducted using the MMSE [47], which is one of the best-known and most widely used instruments for

XSL•FO
RenderX

assessing cognitive function in clinical practice worldwide. The MMSE comprises 11 questions covering 6 cognitive domains: orientation, attention and calculation, memory, language, reasoning and problem-solving, and executive function [48]. The total score ranges from 0 to 30, with higher scores indicating better cognitive function. The MMSE sensitivity was good (0.83, 95% CI 0.62-0.95), and the specificity was very good (0.99, 95% CI 0.92-0.99). The Cronbach α was 0.78 [49]. The behavioral intention questionnaire and MMSE questionnaires are available in Multimedia Appendix 1.

Data Analysis

Descriptive analyses using chi-square and independent-sample *t* tests were conducted using the SPSS software (version 21.0; IBM Corp). Descriptive analysis was based on participants' demographic characteristics, such as age, gender, educational level, marital status, and religious beliefs. After controlling for the confounding factor of age, a generalized estimating equation was used to assess changes in behavioral intention and cognitive functions (including orientation, memory, attention and calculation, language, reasoning and problem-solving, and executive function) from baseline to the end of the intervention in the 2 groups. All tests were 2-tailed, with a significance level set at $P<.05$.

Ethical Considerations

This study was approved by the institutional review boards of Kaohsiung Medical University Chung-Ho Memorial Hospital (KMUHIRB-SV (I)–20200105) and Taipei Veterans General Hospital (2023-07-002A). Written informed consent was obtained from all participants, and they were assured of data confidentiality and anonymity and that they could withdraw from the study at any time without providing a reason. The principal investigator visited each psychiatric day care center, explained the study's purpose and process to all participants, and inquired regarding their willingness to take part. Each participant was offered TWD $400 (US $13-$14) as compensation for their taking part.

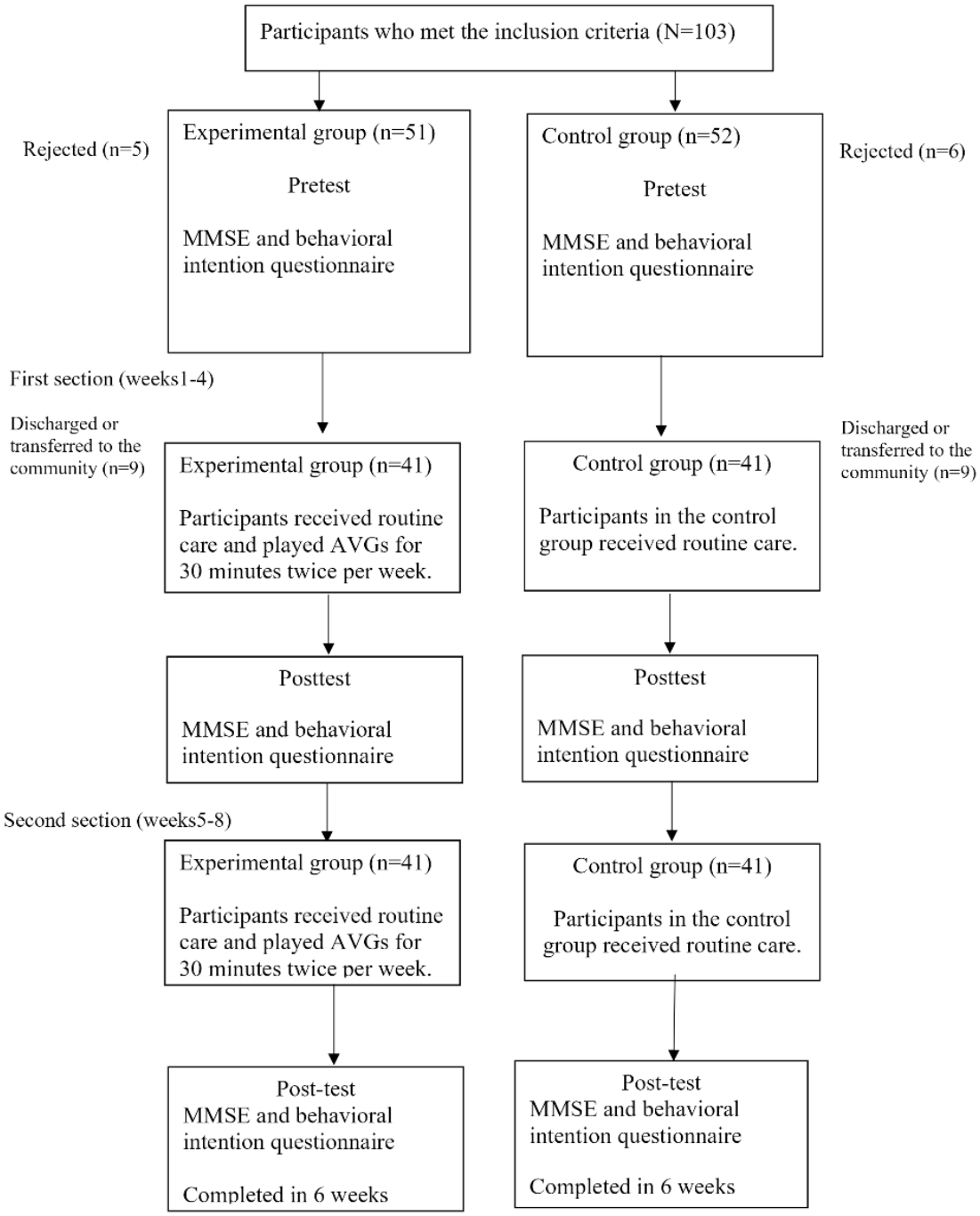
Results

Baseline Data

A total of 82 participants met the inclusion criteria and completed all the processes of our protocol. The participants were divided into the experimental and control groups (Figure 1), and we compared demographic data between these groups. The independent-sample 2-tailed *t* test and chi-square test results showed no statistically significant differences between the 2 groups (Table 2).

XSL•FO
RenderX

Figure 1. Participant flowchart. AVG: active video game; BIQ: behavioral intention questionnaire; MMSE: Mini Mental State Examination.



XSL•FO
RenderX

Table 2. Demographic characteristics and baseline outcomes.

Characteristic	Experimental group (n=41)	Control group (n=41)	Chi-square (*df*)	*t* test (*df*)	*P* value
Age (y), mean (SD)	43.61 (12.2)	47.95 (10.6)	—[a]	1.72 (80)	.09
Sex, n (%)			1.00 (1)	—	.59
Male	19 (46)	19 (46)			
Female	22 (54)	22 (54)			
Hypertension, n (%)			0.746 (1)	—	.50
No	35 (85)	36 (88)			
Yes	6 (15)	5 (12)			
Diabetes mellitus, n (%)			0.312 (1)	—	.16
No	34 (83)	38 (93)			
Yes	7 (17)	3 (7)			
Level of education, n (%)			3.57 (5)	—	.61
Junior high school	6 (15)	6 (15)			
Senior high school	25 (61)	22 (54)			
College	10 (24)	11 (27)			
Higher than college	0 (0)	2 (5)			
Marital status, n (%)			3.20 (3)	—	.36
Single	34 (83)	32 (78)			
Divorced	3 (7)	6 (15)			
Married	4 (10)	3 (7)			
Religious beliefs, n (%)			3.13 (1)	—	.08
Yes	31 (76)	25 (61)			
No	10 (24)	16 (39)			
Onset age (y), mean (SD)	27.40 (12.3)	24.74 (8.6)	—	−1.09 (72)	.28
BMI (kg/m^2), mean (SD)	26.40 (5.4)	25.58 (6.8)	—	−0.59 (16)	.55
Living status, n (%)			6.44 (1)	—	.17
Living alone	6 (15)	9 (22)			
Living with their family or friends	35 (85)	32 (78)			
MMSE[b] score (0-30), mean (SD)	26.76 (3.2)	25.34 (4.2)	—	−1.720 (81)	.09
Behavioral intention score (0-16), mean (SD)	13.75 (3.8)	13.66 (3.3)	—	−0.124 (81)	.90

[a]Not applicable.

[b]MMSE: Mini Mental State Examination.

Changes in Cognitive Function and Behavioral Intention From Baseline to the End of the Intervention

We controlled the age and educational level as the MMSE scores were related to both of these factors. We used generalized estimating equations to analyze changes in participants' cognitive functions and found significant improvements in the cognitive domains of orientation (T2: β=0.66; *P*=.04) and language (T2: β=0.28; *P*=.03), as well as in behavioral intention (T1: β=4.88 and *P*=.009; T2: β=4.24; *P*=.04), following the intervention with AVGs compared to baseline (T0) in the experimental group (Table 3).

XSL•FO
RenderX

Table 3. Changes in cognitive function from baseline to the end of the intervention.

Parameter	EG[a] vs CG[b]		T2[c] vs T0[d]		T1[e] vs T0		EG × T2 vs CG × T2		EG × T1 vs CG × T1	
	β (SE)	*P* value	β (SE)	*P* value	β (SE)	*P* value	β (SE)	*P* value	β (SE)	*P* value
MMSE[f]	0.29 (2.90)	.92	−0.42 (1.53)	.79	0.08 (1.44)	.95	−0.13 (0.67)	.85	0.36 (0.60)	.55
Orientation	0.44 (1.43)	.56	0.57 (0.65)	.35	0.66 (0.70)	.40	1.90 (0.33)	.04[g]	1.76 (0.32)	.08
AC[h]	1.42 (1.27)	.26	−0.79 (0.88)	.37	0.90 (0.70)	.20	−0.15 (0.45)	.73	−0.45 (0.34)	.18
Language	1.09 (0.39)	.89	1.05 (0.30)	.11	0.98 (0.28)	.29	1.24 (0.13)	.03[g]	1.22 (0.13)	.09
RPS[i]	−0.18 (0.39)	.65	−0.22 (0.38)	.56	−0.52 (0.32)	.10	0.07 (0.16)	.66	0.21 (0.14)	.13
Memory	−0.34 (0.51)	.50	0.06 (0.49)	.90	0.44 (0.40)	.28	−0.01 (0.20)	.94	0.19 (0.20)	.35
EF[j]	−0.04 (0.24)	.87	−0.26 (0.27)	.33	−0.38 (0.25)	.12	0.17 (0.13)	.20	0.09 (0.11)	.43
Behavioral intention	69.88 (2.61)	.10	0.82 (0.46)	.67	1.16 (0.38)	.70	4.24 (0.69)	.04[g]	4.88 (0.60)	.009[k]

[a]EG: experimental group.

[b]CG: control group.

[c]T2: 6-week time point.

[d]T0: baseline.

[e]T1: 3-week time point.

[f]MMSE: Mini Mental State Examination.

[g]*P*<.05.

[h]AC: attention and calculation.

[i]RPS: reasoning and problem-solving.

[j]EF: executive function.

[k]*P*<.01.

Discussion

Principal Findings

The experimental group had significantly higher behavioral intention to participate in AVGs and showed significant improvements in the cognitive domains of orientation and language compared to the control group. These findings indicate that the intervention not only enhanced motivation among patients with schizophrenia to participate in AVG-based activities but also led to measurable cognitive gains. Taken together, the results suggest that AVG interventions are both feasible and potentially effective in simultaneously addressing cognitive deficits and motivational challenges in this population.

Comparison to Previous Work

Behavioral Intention

Our results indicated a significant enhancement in the behavioral intention of patients with schizophrenia to participate in AVGs after the intervention. This result is similar to the results of previous studies [50] that have indicated that interventions incorporating multisensory stimulation can enhance the motivation of patients with schizophrenia to engage in device-assisted activities [51,52]. A plausible explanation for the observed improvements is that AVGs simultaneously activate multiple sensory modalities, including visual, auditory, and tactile systems. This multisensory stimulation fosters a more immersive and engaging experience, which may contribute to increased participation. Such experiences can help stimulate cognitive processing and enhance attention, memory, and motor skills. In addition, the dynamic and interactive nature of AVGs can create an environment that promotes relaxation, reduces stress, and fosters a sense of accomplishment.

The combination of enjoyable gameplay and positive emotional responses such as satisfaction and pleasure could further contribute to improved cognitive function by increasing motivation and engagement during the intervention. Due to these features of novel device-assisted activities, it is feasible to use AVGs as a tool to change the behavioral intention for physical activity or exercise in patients with schizophrenia.

Cognitive Function: Significant Improvements in the Language and Orientation Domains

We also found that language improvement was consistent with that found in previous research [40], which reports enhanced verbal abilities following AVG engagement. One plausible explanation for this outcome is that AVGs activate the brain's reward system, which may, in turn, facilitate verbal learning and memory consolidation in patients with schizophrenia [53]. These gains were reflected in both expressive and receptive capacities. Specifically, participants demonstrated enhanced abilities in verbal communication, including word retrieval and fluency, as well as improved comprehension of spoken language. Such improvements are particularly meaningful for patients with schizophrenia as language dysfunction is closely associated with impaired social interaction, reduced functional capacity, and poor treatment adherence. Thus, the observed enhancements underscore the potential of structured, technology-assisted

XSL•FO
RenderX

interventions to promote not only cognitive processing but also communicative competence, which is an essential element in real-world rehabilitation and recovery.

In addition to gains in language, our study revealed significant enhancement in orientation. Interestingly, such outcomes have been less commonly reported in previous AVG intervention studies, suggesting a novel contribution of our findings. Most previous research has focused on improvements in general cognitive functioning, executive function, or memory, with limited emphasis on orientation as a distinct outcome [54]. The observed improvement in orientation may be attributed to the specific design characteristics of the games used—namely, the *Mario Party* series. These games incorporate spatial and directional challenges that require players to navigate complex virtual environments, follow sequential directions, and make decisions based on their spatial understanding. Such tasks likely stimulate cognitive processes related to spatial awareness, planning, and environmental integration, all of which are integral to orientation. Therefore, our findings expand the scope of known cognitive benefits associated with AVG interventions by highlighting their potential to enhance orientation through immersive spatial gameplay.

Cognitive Function: No Significant Changes in Attention and Calculation, Reasoning and Problem-Solving, Memory, and Executive Function

One plausible explanation for this finding pertains to the selection of the cognitive assessment instrument. In this study, cognitive function was measured using the MMSE, a widely used screening tool in clinical settings. While the MMSE offers the advantage of brevity and ease of administration, it is well recognized for its limited sensitivity in detecting deficits across several higher-order cognitive domains, including attention and calculation, reasoning and problem-solving, memory, and executive function [47]. These limitations may have constrained the detection of subtle but meaningful changes resulting from the intervention and may account for the absence of significant findings within these domains in our sample. Given that these cognitive functions are critical to daily functioning and are often impaired in neuropsychiatric populations such as patients with schizophrenia, reliance on the MMSE alone may underestimate the true extent of cognitive deficits [55]. Consequently, we recommend that future research incorporate more sensitive and domain-specific assessment instruments such as the Montreal Cognitive Assessment, which has demonstrated superior efficacy in detecting subtle cognitive impairments, particularly in patients with schizophrenia [56]. Using such tools will likely yield a more comprehensive and accurate profile of cognitive dysfunction, thereby informing targeted interventions and improving clinical outcomes.

Previous studies have implemented AVG interventions using gaming consoles such as the Nintendo Wii and Xbox Kinect primarily targeting physical activity or cognitive stimulation in diverse clinical populations. A systematic review reported that 44% of AVG-based interventions used the Nintendo Wii console and 36% used the Xbox Kinect console, reflecting the widespread adoption of these platforms in therapeutic settings [57]. Diener et al [53] further noted that, although these interventions often shared similar durations and frequencies, critical procedural details such as session content, gameplay structure, and supervision level were frequently underreported. This lack of detail limits both replicability and generalizability [53]. In contrast, this study adopted a clearly defined and systematically implemented 6-week AVG intervention protocol. However, it is possible that the intervention duration and exposure frequency were insufficient to elicit more robust improvements in higher-order cognitive domains. Complex cognitive abilities such as memory and executive function typically require more intensive and sustained engagement to show measurable gains [58].

Implications for Rehabilitation and Clinical Practice

In patients with schizophrenia, cognitive impairment and motivational deficits are well-documented barriers to rehabilitation [59], and our results suggest that AVGs may help address both. Compared to traditional rehabilitation methods, which often require extensive professional oversight and may lack engaging elements, AVGs offer an interactive, motivating, and low-barrier alternative [60,61]. Traditional therapeutic activities in psychiatric day care settings, such as arts and crafts, group discussions, or paper-based cognitive tasks, while beneficial, are often passive, repetitive, or fail to sufficiently motivate patients with negative symptoms or attentional difficulties [62]. In contrast, AVGs offer a dynamic and patient-centered alternative with several distinct advantages. AVGs are interactive, provide real-time feedback, and present task-oriented challenges within immersive environments. These features are particularly effective in sustaining attention and promoting active engagement. Importantly, AVGs can be administered with minimal supervision and infrastructural demands, making them well suited for integration into community-based programs and other low-resource contexts [58-62].

Our findings support the potential of AVGs to address both cognitive and motivational deficits. Participants in the intervention group not only showed significant improvements in the cognitive domains of orientation and language but also demonstrated increased behavioral intention to engage with AVGs. The structured yet enjoyable gameplay likely fostered a sense of mastery and self-efficacy, which are key factors in enhancing intrinsic motivation. This is particularly relevant for patients with schizophrenia, who often struggle with motivation and social withdrawal.

While overall cognitive function did not show significant improvement, we observed notable gains in the domains of orientation and language. These domains are often relatively preserved and more responsive to interventions in patients with mild cognitive impairment or certain psychiatric conditions [63,64]. In addition, AVGs can significantly impact behavioral intention by motivating patients to engage in physical activity. For patients with schizophrenia, who often experience difficulties with motivation and social engagement, the interactive nature of AVGs provides a compelling, enjoyable way to encourage participation in exercise. The games' structured yet fun activities help foster a sense of achievement, which can translate into more consistent engagement in physical

XSL•FO
RenderX

activity. Taken together, these findings highlight the feasibility and clinical utility of AVGs as a scalable, low-barrier intervention for improving select cognitive domains and enhancing motivation among patients with schizophrenia. Their affordability, accessibility, and adaptability to various patient needs position them as promising adjuncts to traditional rehabilitation strategies in resource-limited or community-based psychiatric programs.

Future Directions

First, it is unclear whether the cognitive and behavioral gains associated with AVGs persist over time or depend on continued engagement. Second, future studies should explore how AVGs can be integrated into comprehensive, multidisciplinary care models. Personalizing AVG content according to patients' cognitive status, physical capabilities, or symptom severity may further improve intervention efficacy. Third, conducting subgroup analyses across age groups, illness stages, and care settings could inform more personalized and scalable implementation strategies.

Limitations

Our study has several limitations. First, we provided a relatively lower intensity and shorter duration of the intervention for patients with schizophrenia, which may limit the benefits of AVGs observed in our participants. Second, our sample size was sufficient but relatively small, which may limit the generalizability of the findings to the broader population. Further studies with a larger sample size, higher intervention intensity, and longer duration, along with long-term follow-up, are needed to overcome these limitations. Finally, we used the MMSE to assess cognitive function. However, the MMSE has limited sensitivity to executive function and working memory, which may explain the lack of significant findings in these domains. For future research, we recommend using the Montreal Cognitive Assessment as it is more effective in detecting cognitive impairments, particularly in patients with schizophrenia or mild cognitive impairment.

Conclusions

This study contributes to the growing body of evidence supporting AVGs as effective and engaging interventions for improving orientation, language, and behavioral motivation in patients with schizophrenia. Our structured 6-week program yielded meaningful improvements in these domains, which are often difficult to impact through traditional approaches. Although preliminary, the results highlight the potential of well-designed AVG interventions as scalable, low-barrier tools for psychiatric rehabilitation. Unlike conventional therapies that frequently rely on specialized resources or highly trained professionals, AVGs offer a flexible alternative suitable for diverse settings, including psychiatric wards, outpatient clinics, and community centers.

Acknowledgments

The authors thank all participants for their generous involvement and also thank the Ministry of Science and Technology of Taiwan for their support (110-2314-B-037-028). The article processing charge is supported by the National Science and Technology Council, Taiwan (114-2314-B-037-075), and Kaohsiung Medical University (M113007).

Data Availability

The datasets generated or analyzed during this study are not publicly available due to the inclusion of sensitive and confidential information but are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The questionnaire on behavioral intention and the Mini-Mental State Examination in both the Chinese (original) and English versions.
[PDF File (Adobe PDF File), 762 KB - games_v13i1e69116_app1.pdf]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1065 KB - games_v13i1e69116_app2.pdf]

References

1. Schizophrenia. World Health Organization. 2022 Jan 10. URL: https://www.who.int/news-room/fact-sheets/detail/schizophrenia [accessed 2025-08-30]
2. Chong HY, Teoh SL, Wu DB, Kotirum S, Chiou CF, Chaiyakunapruk N. Global economic burden of schizophrenia: a systematic review. Neuropsychiatr Dis Treat 2016;12:357-373 [FREE Full text] [doi: 10.2147/NDT.S96649] [Medline: 26937191]

XSL•FO
RenderX

3. Vancampfort D, Stubbs B, Mitchell AJ, De Hert M, Wampers M, Ward PB, et al. Risk of metabolic syndrome and its components in people with schizophrenia and related psychotic disorders, bipolar disorder and major depressive disorder: a systematic review and meta-analysis. World Psychiatry 2015 Oct 25;14(3):339-347 [FREE Full text] [doi: 10.1002/wps.20252] [Medline: 26407790]
4. Maurus I, Hasan A, Schmitt A, Roeh A, Keeser D, Malchow B, et al. Aerobic endurance training to improve cognition and enhance recovery in schizophrenia: design and methodology of a multicenter randomized controlled trial. Eur Arch Psychiatry Clin Neurosci 2021 Mar 03;271(2):315-324 [FREE Full text] [doi: 10.1007/s00406-020-01175-2] [Medline: 32748261]
5. Knapp M, Mangalore R, Simon J. The global costs of schizophrenia. Schizophr Bull 2004;30(2):279-293. [doi: 10.1093/oxfordjournals.schbul.a007078] [Medline: 15279046]
6. Heinrichs RW, Zakzanis KK. Neurocognitive deficit in schizophrenia: a quantitative review of the evidence. Neuropsychology 1998 Jul;12(3):426-445. [doi: 10.1037//0894-4105.12.3.426] [Medline: 9673998]
7. Ong HL, Subramaniam M, Abdin E, Wang P, Vaingankar JA, Lee SP, et al. Performance of Mini-Mental State Examination (MMSE) in long-stay patients with schizophrenia or schizoaffective disorders in a psychiatric institute. Psychiatry Res 2016 Jul 30;241:256-262. [doi: 10.1016/j.psychres.2016.04.116] [Medline: 27203153]
8. Ganguli R, Brar JS, Vemulapalli H, Jafar H, Ahuja R, Sharma S, et al. Mini-Mental State Examination (MMSE) performance of partially remitted community-dwelling patients with schizophrenia. Schizophr Res 1998 Sep 07;33(1-2):45-52. [doi: 10.1016/s0920-9964(98)00044-9] [Medline: 9783343]
9. Sanford N, Veckenstedt R, Moritz S, Balzan RP, Woodward TS. Impaired integration of disambiguating evidence in delusional schizophrenia patients. Psychol Med 2014 Oct;44(13):2729-2738. [doi: 10.1017/S0033291714000397] [Medline: 25065271]
10. Stroup TS, Gray N. Management of common adverse effects of antipsychotic medications. World Psychiatry 2018 Oct;17(3):341-356 [FREE Full text] [doi: 10.1002/wps.20567] [Medline: 30192094]
11. Mucci A, Kawohl W, Maria C, Wooller A. Treating schizophrenia: open conversations and stronger relationships through psychoeducation and shared decision-making. Front Psychiatry 2020;11:761 [FREE Full text] [doi: 10.3389/fpsyt.2020.00761] [Medline: 32903708]
12. Suenderhauf C, Walter A, Lenz C, Lang UE, Borgwardt S. Counter striking psychosis: commercial video games as potential treatment in schizophrenia? A systematic review of neuroimaging studies. Neurosci Biobehav Rev 2016 Sep;68:20-36. [doi: 10.1016/j.neubiorev.2016.03.018] [Medline: 27090742]
13. Bueno-Antequera J, Munguía-Izquierdo D. Exercise and schizophrenia. Adv Exp Med Biol 2020;1228:317-332. [doi: 10.1007/978-981-15-1792-1_21] [Medline: 32342467]
14. Dauwan M, Begemann MJ, Heringa SM, Sommer IE. Exercise improves clinical symptoms, quality of life, global functioning, and depression in schizophrenia: a systematic review and meta-analysis. Schizophr Bull 2016 May;42(3):588-599 [FREE Full text] [doi: 10.1093/schbul/sbv164] [Medline: 26547223]
15. Dauwan M, Begemann MJ, Slot MI, Lee EH, Scheltens P, Sommer IE. Physical exercise improves quality of life, depressive symptoms, and cognition across chronic brain disorders: a transdiagnostic systematic review and meta-analysis of randomized controlled trials. J Neurol 2021 Apr 14;268(4):1222-1246 [FREE Full text] [doi: 10.1007/s00415-019-09493-9] [Medline: 31414194]
16. Girdler SJ, Confino JE, Woesner ME. Exercise as a treatment for schizophrenia: a review. Psychopharmacol Bull 2019 Feb 15;49(1):56-69 [FREE Full text] [Medline: 30858639]
17. Firth J, Stubbs B, Rosenbaum S, Vancampfort D, Malchow B, Schuch F, et al. Aerobic exercise improves cognitive functioning in people with schizophrenia: a systematic review and meta-analysis. Schizophr Bull 2017 May 01;43(3):546-556 [FREE Full text] [doi: 10.1093/schbul/sbw115] [Medline: 27521348]
18. Rocamora-Montenegro M, Compañ-Gabucio LM, Garcia de la Hera M. Occupational therapy interventions for adults with severe mental illness: a scoping review. BMJ Open 2021 Oct 29;11(10):e047467 [FREE Full text] [doi: 10.1136/bmjopen-2020-047467] [Medline: 34716157]
19. Oyovwi MO, Ogenma UT, Onyenweny A. Exploring the impact of exercise-induced BDNF on neuroplasticity in neurodegenerative and neuropsychiatric conditions. Mol Biol Rep 2025 Jan 20;52(1):140. [doi: 10.1007/s11033-025-10248-1] [Medline: 39832087]
20. Correll CU, Schooler NR. Negative symptoms in schizophrenia: a review and clinical guide for recognition, assessment, and treatment. Neuropsychiatr Dis Treat 2020;16:519-534 [FREE Full text] [doi: 10.2147/NDT.S225643] [Medline: 32110026]
21. Stubbs B, Firth J, Berry A, Schuch FB, Rosenbaum S, Gaughran F, et al. How much physical activity do people with schizophrenia engage in? A systematic review, comparative meta-analysis and meta-regression. Schizophr Res 2016 Oct;176(2-3):431-440. [doi: 10.1016/j.schres.2016.05.017] [Medline: 27261419]
22. Schrank B, Sibitz I, Unger A, Amering M. How patients with schizophrenia use the internet: qualitative study. J Med Internet Res 2010 Dec 19;12(5):e70 [FREE Full text] [doi: 10.2196/jmir.1550] [Medline: 21169176]
23. Hilgard J, Engelhardt CR, Bartholow BD. Individual differences in motives, preferences, and pathology in video games: the gaming attitudes, motives, and experiences scales (GAMES). Front Psychol 2013;4:608 [FREE Full text] [doi: 10.3389/fpsyg.2013.00608] [Medline: 24058355]

XSL•FO
RenderX

24. Campos C, Mesquita F, Marques A, Trigueiro MJ, Orvalho V, Rocha NB. Feasibility and acceptability of an exergame intervention for schizophrenia. Psychol Sport Exerc 2015 Jul;19:50-58 [FREE Full text] [doi: 10.1016/j.psychsport.2015.02.005]
25. Granic I, Lobel A, Engels RC. The benefits of playing video games. Am Psychol 2014 Jan;69(1):66-78. [doi: 10.1037/a0034857] [Medline: 24295515]
26. Fleming TM, Bavin L, Stasiak K, Hermansson-Webb E, Merry SN, Cheek C, et al. Serious games and gamification for mental health: current status and promising directions. Front Psychiatry 2016;7:215 [FREE Full text] [doi: 10.3389/fpsyt.2016.00215] [Medline: 28119636]
27. Chan CL, Ngai EK, Leung PK, Wong S. Effect of the adapted virtual reality cognitive training program among Chinese older adults with chronic schizophrenia: a pilot study. Int J Geriatr Psychiatry 2010 Jun;25(6):643-649. [doi: 10.1002/gps.2403] [Medline: 19806599]
28. Amado I, Brénugat-Herné L, Orriols E, Desombre C, Dos Santos M, Prost Z, et al. A serious game to improve cognitive functions in schizophrenia: a pilot study. Front Psychiatry 2016;7:64 [FREE Full text] [doi: 10.3389/fpsyt.2016.00064] [Medline: 27148093]
29. Lau HM, Smit JH, Fleming TM, Riper H. Serious games for mental health: are they accessible, feasible, and effective? A systematic review and meta-analysis. Front Psychiatry 2016;7:209 [FREE Full text] [doi: 10.3389/fpsyt.2016.00209] [Medline: 28149281]
30. Waris O, Jaeggi SM, Seitz AR, Lehtonen M, Soveri A, Lukasik KM, et al. Video gaming and working memory: a large-scale cross-sectional correlative study. Comput Human Behav 2019 Aug;97:94-103. [doi: 10.1016/j.chb.2019.03.005] [Medline: 31447496]
31. Palva S, Kulashekhar S, Hämäläinen M, Palva JM. Localization of cortical phase and amplitude dynamics during visual working memory encoding and retention. J Neurosci 2011 Mar 30;31(13):5013-5025 [FREE Full text] [doi: 10.1523/JNEUROSCI.5592-10.2011] [Medline: 21451039]
32. Pallavicini F, Ferrari A, Mantovani F. Video games for well-being: a systematic review on the application of computer games for cognitive and emotional training in the adult population. Front Psychol 2018 Nov 7;9:2127 [FREE Full text] [doi: 10.3389/fpsyg.2018.02127] [Medline: 30464753]
33. Rose A, Vinogradov S, Fisher M, Green MF, Ventura J, Hooker C, et al. Randomized controlled trial of computer-based treatment of social cognition in schizophrenia: the TRuSST trial protocol. BMC Psychiatry 2015 Jul 03;15:142 [FREE Full text] [doi: 10.1186/s12888-015-0510-1] [Medline: 26138715]
34. Wang P, Zhu XT, Qi Z, Huang S, Li HJ. Neural basis of enhanced executive function in older video game players: an fMRI study. Front Aging Neurosci 2017 Nov 21;9:382 [FREE Full text] [doi: 10.3389/fnagi.2017.00382] [Medline: 29209202]
35. Irons JL, Remington RW, Mclean JP. Not so fast: rethinking the effects of action video games on attentional capacity. Australian Journal of Psychology 2020 Nov 20;63(4):224-231. [doi: 10.1111/j.1742-9536.2011.00001.x]
36. Deleuze J, Christiaens M, Nuyens F, Billieux J. Shoot at first sight! First person shooter players display reduced reaction time and compromised inhibitory control in comparison to other video game players. Comput Hum Behav 2017 Jul;72:570-576 [FREE Full text] [doi: 10.1016/j.chb.2017.02.027]
37. Zhao Y, Soh KG, Saad HB, Rong W, Liu C, Wang X. Effects of active video games on mental health among college students: a systematic review. BMC Public Health 2024 Dec 18;24(1):3482 [FREE Full text] [doi: 10.1186/s12889-024-21011-9] [Medline: 39695572]
38. Roberts MT, Lloyd J, Välimäki M, Ho GW, Freemantle M, Békefi AZ. Video games for people with schizophrenia. Cochrane Database Syst Rev 2021 Mar 04;2(2):CD012844 [FREE Full text] [doi: 10.1002/14651858.CD012844.pub2] [Medline: 33539561]
39. Benzing V, Schmidt M. Exergaming for children and adolescents: strengths, weaknesses, opportunities and threats. J Clin Med 2018 Nov 08;7(11):422. [doi: 10.3390/jcm7110422] [Medline: 30413016]
40. Ahmer Z, Tanzil S. Internet addiction among social networking sites users: emerging mental health concern among medical undergraduates of Karachi. Pak J Med Sci 2018;34(6):1473-1477 [FREE Full text] [doi: 10.12669/pjms.346.15809] [Medline: 30559806]
41. López-Martín O, Segura Fragoso A, Rodríguez Hernández M, Dimbwadyo Terrer I, Polonio-López B. Effectiveness of a programme based on a virtual reality game for cognitive enhancement in schizophrenia. Gac Sanit 2016;30(2):133-136 [FREE Full text] [doi: 10.1016/j.gaceta.2015.10.004] [Medline: 26611240]
42. Nahum M, Fisher M, Loewy R, Poelke G, Ventura J, Nuechterlein KH, et al. A novel, online social cognitive training program for young adults with schizophrenia: a pilot study. Schizophr Res Cogn 2014 Mar 01;1(1):e11-e19 [FREE Full text] [doi: 10.1016/j.scog.2014.01.003] [Medline: 25267937]
43. Reddy LF, Irwin MR, Breen EC, Reavis EA, Green MF. Social exclusion in schizophrenia: psychological and cognitive consequences. J Psychiatr Res 2019 Jul;114:120-125. [doi: 10.1016/j.jpsychires.2019.04.010] [Medline: 31059992]
44. Brun G, Verdoux H, Couhet G, Quiles C. Computer-assisted therapy and video games in psychosocial rehabilitation for schizophrenia patients. Encephale 2018 Sep;44(4):363-371. [doi: 10.1016/j.encep.2017.12.009] [Medline: 29501256]
45. Faul F, Erdfelder E, Buchner A, Lang AG. Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. Behav Res Methods 2009 Nov;41(4):1149-1160. [doi: 10.3758/BRM.41.4.1149] [Medline: 19897823]

XSL•FO
RenderX

46. Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition. Washington, DC: American Psychiatric Publishing; 2013.
47. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975 Nov;12(3):189-198. [doi: 10.1016/0022-3956(75)90026-6] [Medline: 1202204]
48. Kabátová O, Puteková S, Martinková J, Súkenníková M. Analysis of psychometric features of the Mini-Mental State Examination and the Montreal Cognitive Assessment methods. Clin Soc Work Health Interv 2016;7(2):62-69 [FREE Full text] [doi: 10.22359/cswhi_7_2_08]
49. Ministry OH. Dementia Clinical Practice Manual. Taipei: Ministry of Health and Welfare; 2017. URL: https://www.ntshb.gov.tw/df_ufiles/226/衛生福利部失智症診療手冊1060223.pdf [accessed 2025-07-19]
50. Moller AC, Sousa CV, Lee KJ, Alon D, Lu AS. Active video game interventions targeting physical activity behaviors: systematic review and meta-analysis. J Med Internet Res 2023 May 16;25:e45243. [doi: 10.2196/45243] [Medline: 37191992]
51. Mura G, Carta MG, Sancassiani F, Machado S, Prosperini L. Active exergames to improve cognitive functioning in neurological disabilities: a systematic review and meta-analysis. Eur J Phys Rehabil Med 2018 Jun;54(3):450-462. [doi: 10.23736/s1973-9087.17.04680-9]
52. Comeras-Chueca C, Marin-Puyalto J, Matute-Llorente A, Vicente-Rodriguez G, Casajus JA, Gonzalez-Aguero A. Effects of active video games on health-related physical fitness and motor competence in children and adolescents with overweight or obesity: systematic review and meta-analysis. JMIR Serious Games 2021 Oct 18;9(4):e29981 [FREE Full text] [doi: 10.2196/29981] [Medline: 34661549]
53. Diener J, Rayling S, Bezold J, Krell-Roesch J, Woll A, Wunsch K. Effectiveness and acceptability of e- and m-health interventions to promote physical activity and prevent falls in nursing homes-a systematic review. Front Physiol 2022;13:894397 [FREE Full text] [doi: 10.3389/fphys.2022.894397] [Medline: 35669573]
54. Wang W, Ye M, Zhong Y, Zhang M, Wang Y, Chen C, et al. Smartphone video games for cognitive functions and negative symptoms in male with chronic schizophrenia: a randomized controlled trial. Heliyon 2025 Mar 15;11(3):e42049 [FREE Full text] [doi: 10.1016/j.heliyon.2025.e42049] [Medline: 39991209]
55. Arevalo-Rodriguez I, Smailagic N, Roqué I Figuls M, Ciapponi A, Sanchez-Perez E, Giannakou A, et al. Mini-Mental State Examination (MMSE) for the detection of Alzheimer's disease and other dementias in people with mild cognitive impairment (MCI). Cochrane Database Syst Rev 2015 Mar 05;2015(3):CD010783 [FREE Full text] [doi: 10.1002/14651858.CD010783.pub2] [Medline: 25740785]
56. Rosca EC, Cornea A, Simu M. Montreal Cognitive Assessment for evaluating the cognitive impairment in patients with schizophrenia: a systematic review. Gen Hosp Psychiatry 2020;65:64-73. [doi: 10.1016/j.genhosppsych.2020.05.011] [Medline: 32502879]
57. Green CS, Bavelier D. Learning, attentional control, and action video games. Curr Biol 2012 Mar 20;22(6):R197-R206 [FREE Full text] [doi: 10.1016/j.cub.2012.02.012] [Medline: 22440805]
58. Yu Y, Wang J, Xu J. Optimal dose and type of exercise to improve cognitive function in patients with mild cognitive impairment: a systematic review and network meta-analysis of RCTs. Front Psychiatry 2024;15:1436499. [doi: 10.3389/fpsyt.2024.1436499] [Medline: 39328348]
59. Lohse K, Shirzad N, Verster A, Hodges N, Van der Loos HF. Video games and rehabilitation: using design principles to enhance engagement in physical therapy. J Neurol Phys Ther 2013 Dec;37(4):166-175. [doi: 10.1097/NPT.0000000000000017] [Medline: 24232363]
60. Pope Z, Zeng N, Gao Z. The effects of active video games on patients' rehabilitative outcomes: a meta-analysis. Prev Med 2017 Mar;95:38-46. [doi: 10.1016/j.ypmed.2016.12.003] [Medline: 27939260]
61. Choi J, Medalia A. Intrinsic motivation and learning in a schizophrenia spectrum sample. Schizophr Res 2010 May;118(1-3):12-19 [FREE Full text] [doi: 10.1016/j.schres.2009.08.001] [Medline: 19716270]
62. Kueider AM, Parisi JM, Gross AL, Rebok GW. Computerized cognitive training with older adults: a systematic review. PLoS One 2012;7(7):e40588 [FREE Full text] [doi: 10.1371/journal.pone.0040588] [Medline: 22792378]
63. Kemik K, Ada E, Çavuşoğlu B, Aykaç C, Savaş DD, Yener G. Detecting language network alterations in mild cognitive impairment using task-based fMRI and resting-state fMRI: a comparative study. Brain Behav 2024 May;14(5):e3518 [FREE Full text] [doi: 10.1002/brb3.3518] [Medline: 38698619]
64. Stanmore E, Stubbs B, Vancampfort D, de Bruin ED, Firth J. The effect of active video games on cognitive functioning in clinical and non-clinical populations: a meta-analysis of randomized controlled trials. Neurosci Biobehav Rev 2017 Jul;78:34-43 [FREE Full text] [doi: 10.1016/j.neubiorev.2017.04.011] [Medline: 28442405]

Abbreviations

AVG: active video game
MMSE: Mini Mental State Examination

XSL•FO
RenderX

Edited by A Coristine; submitted 22.11.24; peer-reviewed by HC Li, LY Liao; comments to author 07.05.25; revised version received 06.06.25; accepted 26.08.25; published 01.10.25.

Please cite as:
Chen HH, Lin CC, Yu ML, Wu HL, Shen HC, Hsieh HF
Active Video Games to Improve Behavioral Intentions and Cognitive Function in Patients With Schizophrenia: Randomized Controlled Trial
JMIR Serious Games 2025;13:e69116
URL: https://games.jmir.org/2025/1/e69116
doi:10.2196/69116
PMID:

©Huan-Hwa Chen, Ching-Ching Lin, Man-Ling Yu, Hsiu-Lan Wu, Hui-Chu Shen, Hsiu-Fen Hsieh. Originally published in JMIR Serious Games (https://games.jmir.org), 01.10.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Exploring Mixed-Reality Exergames for Sports Rehabilitation: Design Insights and Evaluation Findings

Michelle C Haas[1], MSc; Larissa Wild[2], MA; Leander Schneeberger[2], MA; Eveline S Graf[1*], PhD; Anna Lisa Martin-Niedecken[2,3*], PhD

[1]Institute of Physiotherapy, School of Health Sciences, ZHAW Zurich University of Applied Sciences, Winterthur, Switzerland

[2]Institute for Design Research, Department Design, Zurich University of the Arts, Zurich, Switzerland

[3]Digital Health Design Living Lab, Zurich University of the Arts, Zurich, Switzerland

[*]these authors contributed equally

Corresponding Author:
Anna Lisa Martin-Niedecken, PhD
Institute for Design Research
Department Design
Zurich University of the Arts
Pfingstweidstrasse 96
Zurich, 8031
Switzerland
Phone: 41 434463246
Email: anna.martin@zhdk.ch

Abstract

Background: Exergaming, involving physically active play, provides a means for motivating and functional training. It may also offer an innovative solution for rehabilitation after sports injuries such as anterior cruciate ligament (ACL) injury. Rehabilitation is lengthy and only partially prepares athletes for the demands of sports. As a result, only 65% of athletes return to the same performance level as before the injury. Exergames specific to ACL rehabilitation are missing. Their unique combination of physical and cognitive challenge may contribute to preparing patients for their return to sports. Collaboration of sport scientists, game designers, and physiotherapists enables comprehensive and user-centered development of an innovative training concept incorporating the needs of both patients and therapists along with scientific evidence.

Objective: This project aimed to develop a specific exergame scenario for sports rehabilitation after ACL injuries. A research-driven, user-centered, iterative approach was followed. The project was structured into four phases: (1) assessment of motor performance during a fitness exergame, (2) investigation and establishment of user requirements, (3) development of a new exergame, and (4) validation of the new exergame.

Methods: For assessment of motor performance, lower extremity kinematics during a fitness exergame scenario in the ExerCube were assessed in 24 athletes (6 after ACL injury) using marker-based motion capture. Regarding user requirements, focus groups with physiotherapists and patients were conducted. Development of a new exergame was from the kinematic analysis results and user requirements guided the iterative, interdisciplinary design of a new exergame scenario and movement concept for the ExerCube. For validation, developed exergame scenarios were recurrently evaluated with patient and therapist focus groups and transformed into a final exergame scenario.

Results: For assessment of motor performance, there was a main effect of exercise in maximal knee valgus, knee internal rotation, and hip flexion with $P<.001$. Squats showed the lowest knee valgus (4.23°, 95% CI 4.12-4.51) and knee internal rotation angle (3.68°, 95% CI 3.33-4.03). Focus groups revealed that patients want to return to sports as quickly as possible but have concerns about sensory overload during training. Physiotherapists desire a device that allows additional independent training with new therapeutic stimuli. Concerning the development of a new exergame, the movement concept for the new exergame included strength and balance exercises, and a mini-exergame for endurance, reaction, and skill training. Three difficulty levels allow for varying complexity and speed. For validation, focus groups on the newly developed exergame scenarios highlighted their motivational potential, suitability for autonomous use, and the need for therapist-controlled adaptation and integration with conventional tools.

XSL•FO
RenderX

Conclusions: The interdisciplinary, evidence-based approach facilitated a systematic development of requirements for an exergame for rehabilitation, ensuring user-centered implementation. A final evaluation confirmed its motivational potential and applicability, leading to its deployment at all ExerCube locations.

(JMIR Serious Games 2025;13:e68431) doi:10.2196/68431

KEYWORDS

interdisciplinary; exergames; sports rehabilitation; physiotherapy; research and development; user-centered design; evaluation; anterior cruciate ligament

Introduction

Rehabilitation After Sports Injuries

Rupture of the anterior cruciate ligament (ACL) is a commonly known sports injury resulting in at least 6 months of time loss until a full return to sport is possible [1-3]. Between 1988-1989 and 2003-2004, the incidence rate (IR) of ACL injuries increased by 1.3% per year among collegiate athletes in the United States [4]. As reporting of sports injuries is not standardized, differences in IRs exist. While Gornitzky et al (2016) [5] report an IR of 0.062 per 1000 athletic exposures among different types of sport, Moses et al (2012) [6] report an IR of 0.05 per average person per year and an IR of 0.15%-3.7% for professional athletes. In female floorball, the risk of an ACL injury equals 3.6 per 1000 game hours [7]. Therefore, injury rates differ depending on factors such as the type or level of sport.

This difference is also present when assessing injury risk for a second ACL rupture, either on the ipsilateral or contralateral side. The latest systematic review and meta-analysis reports a reinjury rate of 21.9% while another meta-analysis reported 15% [8,9]. An even lower reinjury rate, 9.6%, was reported in a prospective study following 1082 patients after ACL reconstruction [10]. The high rates of reinjury suggest that current rehabilitation protocols still have to be improved. Evidence-based rehabilitation protocols are currently mainly available for patients who undergo reconstructive surgery of the ACL. Immediate mobilization, weight-bearing, strength training, and neuromuscular training are recommended after surgery [1-3,11-14]. In contrast, no benefits of bracing could be shown, and it is therefore not recommended [2,11]. Rehabilitation protocols without surgery include proprioceptive training, strength training, and counseling to avoid high-risk movements [12]. To guide progression throughout the rehabilitation process, the concept of “control to chaos” can be followed. This concept provides a structured framework stating that in early stages, exercises with high control of body movement should be performed. As rehabilitation advances, the level of complexity and uncertainty is gradually increased, introducing different stimuli as elements of “chaos” to better reflect real-life scenarios and prepare for a return to sports [15,16].

Only 65% return to their preinjury level of sports after an ACL reconstruction and consecutive rehabilitation [17]. Multiple reasons could explain this low rate: poorly performed surgery or rehabilitation, lax criteria for rehabilitation progression or return to sport, or the lack of addressing psychological factors during rehabilitation [17]. Additionally, patients still exhibit altered movement patterns, which are associated with a higher risk for ACL injury, and diminished explosive neuromuscular performance after ACL rehabilitation [18-21]. According to Buckthorpe (2019) [18], a more holistic and interdisciplinary approach to rehabilitation protocols is needed in order to overcome these deficits. Potentially, exergaming could assist in overcoming these deficits.

Exergaming

Oh and Yang (2010) [22] define exergaming as “a combination of exertion and video games including strength, balance, and flexibility activities.”

Exergames, which require physical effort and are controlled by (full-body) movements, have been proposed as effective training and therapy alternatives by interdisciplinary research and development communities. These games are designed to be attractive and effective, which makes them suitable for application areas where additional motivation and physical-cognitive training load, as well as a safe environment, are needed [23].

Studies in various target audiences have shown that exergaming has the potential to increase training adherence [24], long-term motivation [25], and engagement [26]. Exergaming has also been found to have positive effects on mental health, including social interaction, self-esteem, self-concept, motivation, and mood [27-32].

Furthermore, exergames can have positive effects on cognitive functions, such as executive functions, attention, and visual-spatial skills [32-37] as well as physical activity levels, energy expenditure, and heart rate. By immersing players in an audiovisual, narrative-based game scenario, exergames shift their cognitive and emotional focus to the playful experience, facilitating engagement in physically challenging training [38,39].

Compared to traditional single-task training or therapy approaches, exergames combine physical and cognitive training and thus provide an incorporated dual task training approach [39], which is beneficial for different target groups [36,40-42].

Dual Task and Motor Performance

When dual-tasking in a jump landing task by counting backwards, healthy participants (CONs) have shown altered movement patterns such as decreased knee flexion (KF) angles at initial contact, decreased jump height, or increased peak posterior and vertical ground reaction forces in the initial phase of landing [43]. Therefore, landing mechanics associated with

XSL•FO
RenderX

increased ACL loading occur when a simultaneous cognitive challenge is added to jump-landings.

In people after musculoskeletal injury, such as an ACL injury, motor performance is also altered when dual-tasking [44]. For gait and postural stability, dual task effects in people after ACL injury have already been investigated and compared to CONs [45,46]. Compared to a single task, patients after ACL reconstruction showed smaller peak KF angles and peak knee extension moments during the loading response phase of walking. Additionally, the difference between legs in peak hip adduction angle (middle stance and terminal stance) and peak hip abduction moment (loading response phase) was decreased in the dual-task condition [45]. Anteropostural stability, as well as overall stability during a standing task, was lower with the additional auditory Stroop task in patients after ACL reconstruction [46]. Overall, compared to CONs, patients after ACL reconstruction sacrifice performance to compensate for the additional cognitive load of a dual task, which could result in increased injury risk.

Exergaming in Rehabilitation

Extended reality is a continuum that encompasses virtual reality (VR), augmented reality, and mixed reality. While in VR, the user is visually and aurally occluded from the physical environment, in augmented reality, the user interacts with virtual objects in a physical environment. Mixed reality blends the experience of both VR and augmented reality [47]. In rehabilitation, the use of extended reality and serious games is increasing in popularity [48]. In comparison to traditional rehabilitation, using extended reality in rehabilitation provides more excitement as well as physical and cognitive fidelity [48]. Moreover, there is an external focus of attention which leads to a better movement execution [49]. Extended reality devices such as head-mounted displays or monitors can represent immersive systems which, together with physical presence, involvement, and social- or self-presence, lead to an immersive experience [47]. The immersive nature of a digital environment also has positive effects, such as improved movement patterns in patients after ACL injury [50]. Exergames, physically active games performed with extended reality, are known for their promising potential as complementary or alternative rehabilitation tools, but not every exergame is equally suitable. For example, the implementation of an adequate movement concept and cognitive challenges is just as important as the implementation of a motivating and immersive, multisensory game experience. Furthermore, matching the exergame to the needs of the primary (athletes) and secondary (therapists) users is crucial for the successful long-term use of an exergame in sports rehabilitation. While exergames have been used in a variety of clinical settings, such as neurorehabilitation or geriatrics [51,52], their application for sports rehabilitation is very limited [53].

ExerCube: Physically Immersive Fitness Game Environment

One exergame setting featuring various game and training experiences, which was iteratively designed by involving different target audiences and interdisciplinary research and development teams, is the ExerCube [54-57].

The ExerCube consists of a physical immersive training hardware (platform) and several exergame experiences with different physical-cognitive training foci. The Sphery Racer (Figure 1), which was the first exergame experience that could be played on the cube, takes the player on a digital science-fiction underwater race.

XSL•FO
RenderX

Figure 1. A person playing the exergaming scenario “Sphery Racer” in the ExerCube (source: Sphery Ltd).



The digital game scenario is projected onto the 3 cube walls surrounding the player. To control the game, the player performs functional full-body workout exercises, which are tracked by 4 HTC Vive (HTC Corporation) trackers that the player wears on both wrists and ankles. The game starts with low-to-moderate intensity exercises, which increase to high-intensity exercises over time.

Additionally, the player’s physical (via movement execution) and cognitive performances (via reaction time) are tracked and used to automatically adjust the game’s difficulty and complexity in real time. Generally, the workout intensity in the Sphery Racer is scalable from moderate to high intensity.

Among other things, the Sphery Racer experience was proven to be an attractive and effective alternative to functional high-intensity interval training with a personal trainer [39,58], to provide high exercise intensities (86% of maximum heart rate) [59], and to yield significant positive effects on cognitive (motor) skills (faster reaction times), especially on concentration, cognitive flexibility and divided attention in young game sports athletes after a training intervention [60].

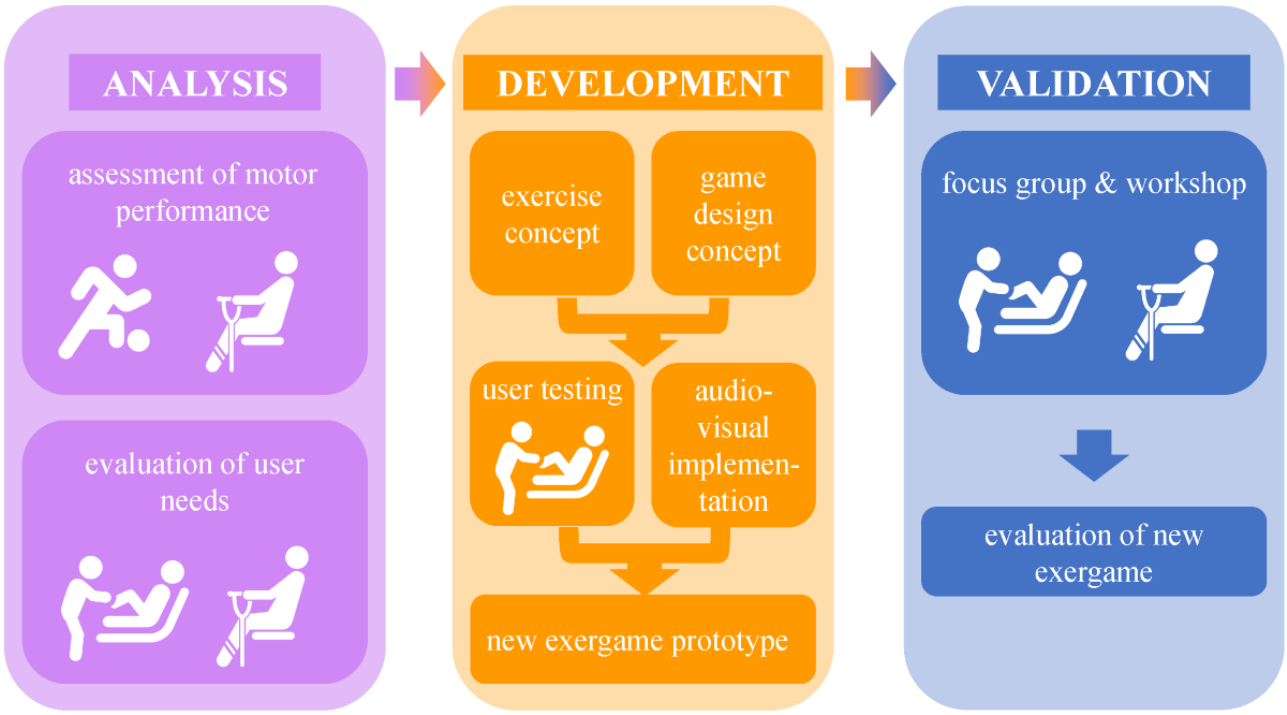
Goal of This Project

The overarching aim of this project was to design and evaluate a mixed reality exergame specifically for sports rehabilitation. The iterative research and development stages of the interdisciplinary project are presented as follows.

The project consisted of several project phases (Figure 2). As this figure shows, the results of the various phases formed the basis for the subsequent phases. Before the design work could start, motor performance during an existing fitness exergame scenario and user needs for a rehabilitation-specific exergame had to be evaluated. This knowledge was then incorporated into the design work, which included the definition of an exercise concept with the subsequent creation of a game design concept. An iterative cycle of internal testing, feedback from therapists, and implementation of audiovisual cues led to the final prototype of the new exergame. Focus groups and workshops with patients and physiotherapists (primary and secondary end users) were performed to evaluate the prototype. Each section of this paper is divided into subsections that correspond to the 4 different phases.

XSL•FO
RenderX

Figure 2. Illustration of the different project phases.



Methods

Motor Performance During ExerCube Training

Objective

The goal of this part of the project was to analyze the movement patterns during exergaming to determine during which exercises the participants' lower extremities are in a more injury-prone position.

Participants

Participants were recruited between November 2021 and July 2022 in Winterthur, Switzerland. They were either recruited at the university campus via mail or at local sports clubs via flyers. Before enrolment, participants were screened based on preselected inclusion and exclusion criteria. All participants had to be aged between 18 and 40 years, participate in at least 3 hours per week in a volleyball, handball, soccer, or floorball sports club, and understand verbal and written instructions in German. Exclusion criteria for all participants were an acute or chronic musculoskeletal, neurological, or cardiopulmonary disease, severe pain during exercise execution, and pregnancy. Additionally, CONs was excluded if there was a prior injury to the ACL or prior knee surgery. Patients were only included if they were still in physical therapy but were cleared to fully participate in athletic activities.

In total, 18 CONs (9 females) and 6 patients (2 females) were included. The type of sports was volleyball (n=3), handball (n=5, two patients), soccer (n=9, two patients), and floorball (n=7, two patients). Characteristics of all participants are displayed in Table 1.

Table 1. Mean and SD of participant characteristics based on athletes who participated in the biomechanical analysis during exergaming.

Characteristic	Patients (n=6)		CONs[a] (n=18)		Difference	
	Mean	SD	Mean	SD	Welch *t* test[b] (*df*)	*P* value
Age (years)	25.00	5.90	25.22	3.26	0.09 (1,6)	.93
Height (m)	1.76	0.07	1.71	0.08	–1.31 (1,10)	.22
Weight (kg)	77.22	2.84	69.07	9.01	–3.37 (1,22)	<.01
BMI ($kg*m^{-2}$)	25.00	1.33	23.42	1.45	–2.47 (1,9)	<.05
Sport-specific training (hours per week)	7.17	3.8	7.58	3.41	0.24 (1,8)	.82

[a]CON: healthy participant.

[b]2-tailed.

Data Collection and Analysis

Data was collected in the movement laboratory of Zurich University of Applied Sciences. Both data collection and analysis are described in detail elsewhere [61,62]. Participants performed a 25-minute high-intensity exergame (Sphery Racer, ExerCube) including different exercises. Neutral (side-unspecific) exercises included burpees, jumps, squats, and

XSL•FO
RenderX

tripples (skipping). Side-specific exercises were punches, middle touches, high touches, low touches, and lunges. As the game changes speed depending on the participant's performance, the movement speed and intensity vary depending on the speed of the game. There were 5 levels for speed, and participants needed to perform exercises continuously correctly for 20 seconds to increase the speed level. After each timing or exercise mistake, the participant dropped 1 level. Consequently, the level of speed varied between participants as well as within 1 data collection. The speed of the game was the only variable to change. The order of the exercises was constant for all participants. Critical angles refer to the peak angles of the corresponding joint direction during 10°-30° KF.

Statistical Analysis

Data were excluded for the left leg of 1 CONs due to the slipping of marker clusters. Therefore, only data from the right leg was statistically analyzed and will be presented here. First, using data of CONs, it was analyzed during which of the exercises the biomechanical risk factors are more pronounced than in others. Then, data of the 2 exercises most at risk and the exercise least at risk were compared between CONs and patients.

For CONs, a 1-way repeated measures ANOVA with concurrent pairwise *t* test was applied using the package *rstatix* (version 0.7.1) in R (version 4.2.2; R Foundation) [63,64]. For result interpretation, marginal means were estimated and graphically displayed with the package *emmeans* [65]. Maximal angle of either knee valgus (KV), knee internal rotation (KIR), or minimal angle of hip flexion (HF) was the outcome of interest. Y_i, exercise was the within-subject covariate ($exercise_k$), β_0 represented the intercept, β_k the effect of the k-covariate, and ε_i the independent and normally distributed errors $\varepsilon_i \sim N(0,\sigma^2)$. The variable exercise had 14 levels, 1 for each combination of side (if applicable) and exercise.

$$Y_i=\beta_0+\beta_1 exercise+\varepsilon_i$$

No deviations from homoscedasticity or normality were found with residual analysis through visual observation. Significance level was set to $P<.05$.

Comparison of patients and CONs was performed by means of an independent Welch *t* test in the neutral exercise squat and the side-specific exercises, high touch and punch. Significance level was set to $P<.05$.

User Requirement for Rehabilitative ExerCube Use

Objective

The goal of this part of the project was to gain an in-depth understanding of the people directly involved in the rehabilitation after ACL injury (patients and physical therapists), extract their needs and desires during rehabilitation, and make this knowledge accessible to the game designer.

Participants

Four physical therapists specialized in sports rehabilitation, as well as 1 athlete who underwent rehabilitation after ACL injury, were purposely chosen to participate in interviews. Additionally, a sports physical therapist was observed during a workday to observe their interaction with various patients with an ACL injury.

Data Collection and Analysis

Data Collection Methods

Different data were collected to assess user requirements: (1) guideline-based observation in a clinical setting; and (2) guideline-based, semistructured interviews and focus groups.

Observation in Clinical Setting

The interaction between patients and physical therapists during ACL rehabilitation was observed by the game designers and movement scientists. The focus of the observation was on the therapeutic content specific to different phases of the rehabilitation and the interaction between the patient and therapist. Subsequently, the observed therapeutic content was matched to phases of a rehabilitative pathway based on functional outcome measures [3,66]. This matching was done by a movement scientist together with the observed physical therapist.

Interviews and Focus Groups

Physical therapists (n=4) and 1 patient were interviewed based on a semistructured interview guideline. The physical therapists were interviewed in 2 separate focus groups, while the patient was interviewed alone. The interviews included questions about attitude and experience toward gaming or exergaming (eg, "Have you had experience with exergames?"), therapeutic concepts (eg, "How would you perceive your role as therapist while your patient plays an exergame in the therapy session?"), motivation during rehabilitation (eg, "Which exercises motivate you and feel particularly good, strong, like "achievement" or "powerful"?"), feedback methods (eg, "What kind of feedback (feedback on current performance or over time) do you like in rehab?"), and design preferences (eg, "Would you prefer a realistic or an abstract scenario in the exergame?").

Three authors with varied disciplinary expertise (human-computer interaction, game research and physiotherapy, and movement sciences) undertook the qualitative evaluation of interviews. Using an iterative thematic coding method inspired by the thematic analysis by Braun and Clark (2021) [67], emerging results were thoroughly discussed until consensus was attained. The results were then transferred into a user persona for both therapists as well as patients who might potentially use the intended prototype.

Rehabilitation-Specific Motivating Workout Scenarios for the ExerCube

Objective

The goal of this part of the project was to develop a movement concept and exergame design concept based on the information from previous project parts to consecutively design a motivating workout scenario specifically for rehabilitation purposes.

Iterative Design Process

Following an iterative, interdisciplinary, and holistic design process, a new exergame scenario for the "back to sports" phase of sports rehabilitation after knee injury was developed.

XSL•FO
RenderX

General components of the technological setup of the original setting (cube hardware and tracking with HTC Vive) were kept for the purpose of this project. The movement concept, as well as the software, was iteratively developed. Throughout the iterative concept and prototyping phases, various guided and sequential walkthrough [68] and bodystorming [69] sessions with participants from different backgrounds (physiotherapy, movement science, and game design) were performed. During the sessions, participants were asked to think aloud [70] regarding their experience with the respective concept or prototype. Additionally, participants were observed by the interdisciplinary team and asked questions during the sessions. The team took notes during each session, which were analyzed and discussed afterward from the different disciplinary perspectives and transferred into concrete design decisions.

This iterative procedure enabled the implementation of a new exergame concept consisting of a knee rehabilitation-specific, functional movement concept including motor-cognitive components, as well as a matching, user-centered, engaging audiovisual and narrative game design to be applied in the ExerCube setting, which is described in more detail as follows.

Validation of the ExerUp Game

Objective

The goal of this project part was to obtain feedback as well as first impressions on the newly developed workout scenario by secondary end users.

Participants

Physical therapists specialized in sports rehabilitation (N=3) and a personal trainer (N=1) were purposely chosen to participate in a focus group where the newly developed exergame scenario was presented, and their impressions and feedback were recorded.

Focus Group

As part of the focus group, all participants were able to (1) observe a person performing a session with the newly developed *ExerUp Game* and (2) try out the game themselves and provide feedback (think aloud) during the try-out, as well as while observing others.

Additionally, a guideline-based semistructured interview was performed focusing on relevant aspects of the digital game scenario, the tracking technology, the movement concept, the overall experience, as well as potential application and implementation cases.

The interviews and the feedback were assessed by 3 authors with different disciplinary backgrounds (human-computer interaction or game research and physiotherapy or movement sciences) following an iterative thematic coding approach based on thematic analysis by Braun and Clark (2021) [67]. Emerging results were discussed until agreement was reached.

Ethical Considerations

Objective

The present study was conducted according to the local legal requirements as well as the guidelines of the Declaration of Helsinki [71] and the principles and procedures for integrity in scientific research involving human beings.

Motor Performance During ExerCube Training

This phase of the project was approved by the ethics committee of the canton of Zurich (Basec-No.: 2021-01700). All participants read and signed an informed written consent form before being included in this study. All study data were recorded without identifiable information to ensure privacy and confidentiality. A participant identification list was stored separately from any study data. Participants did not receive any compensation for study participation, though related travel expenses were remunerated.

User Requirement for Rehabilitative ExerCube Use, Rehabilitation-Specific Motivating Workout Scenarios for the ExerCube, or Validation of the ExerUp Game

These phases did not require ethical approval as no health-related data were captured. Nevertheless, all study data were recorded without identifiable information (eg, during interviews) to ensure privacy and confidentiality. Participants did not receive any compensation for study participation.

Results

Motor Performance During ExerCube Training

About CONs

There was a main effect of exercise in all maximal angles of the investigated variables: KV ($F_{13,10958}$=133.65, P<.001), KIR ($F_{13,10958}$=136.50, P<.001), and HF ($F_{13,10952}$=169.68, P<.001).

Pairwise comparison showed significantly lower KV (P<.001) and KIR (P<.001) during the squat compared to all other exercises. Moreover, the tripples and jump showed a lower KV (P<.001) compared to 11, respectively 9, other exercises. In contrast, the high-touch (both sides), the middle touch (to the right), and the punch (to the right) had a higher KV compared to 8 other exercises (P<.001). For maximal KIR, the punch to the right side displayed higher maximal angles than all other exercises (P<.001). Additionally, the high touch (to the right) and the middle touch (to the right) displayed higher KIR than 12, respectively 10, other exercises (P<.001). A lower HF was present during the left lunge (compared to 11 other exercises), during the left punch and burpee (compared to 10 other exercises), as well as during the tripples (compared to 9 other exercises). A higher HF compared to 11 other exercises was present in the low-touch (to the left), while a higher HF than 10 other exercises was present in the right punch and higher than in 9 other exercises in the squat. An overview of the estimated marginal means, including confidence level for each variable, can be found in Figures 3-5 and in Tables S2-S4 in Multimedia Appendix 1.

XSL•FO
RenderX

Figure 3. Estimated marginal means for maximal knee valgus of the right leg during 10°-30° knee flexion based on the biomechanical analysis during exergaming of athletes without any prior knee injury. BR: burpee; JU: jump; l_HT: left high touch; l_LT: left low touch; l_LU: left lunge; l_MT: left middle touch; l_PU: left punch; r_HT: right high touch; r_LT: right low touch; r_LU: right lunge; r_MT: right middle touch; r_PU: right punch; SQ: squat; TR: tripples.

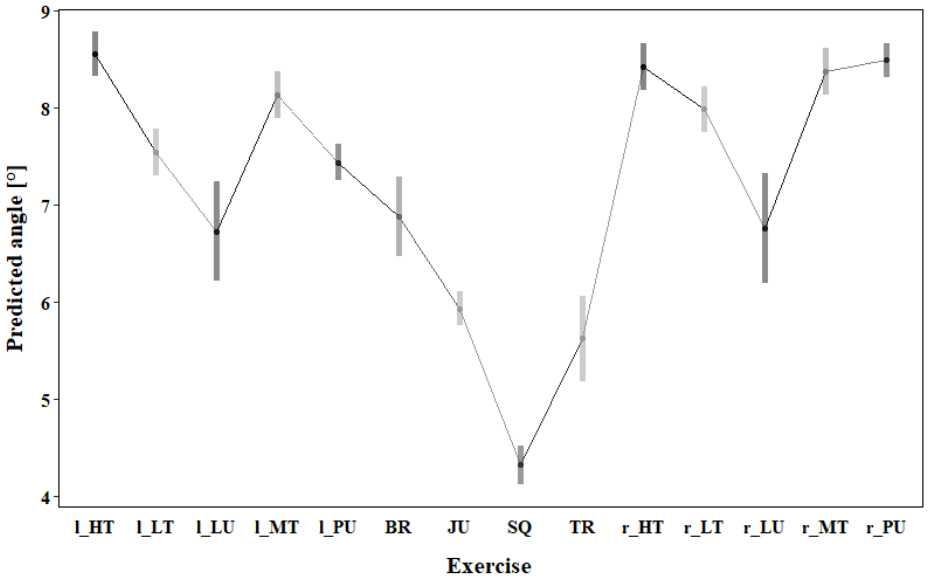
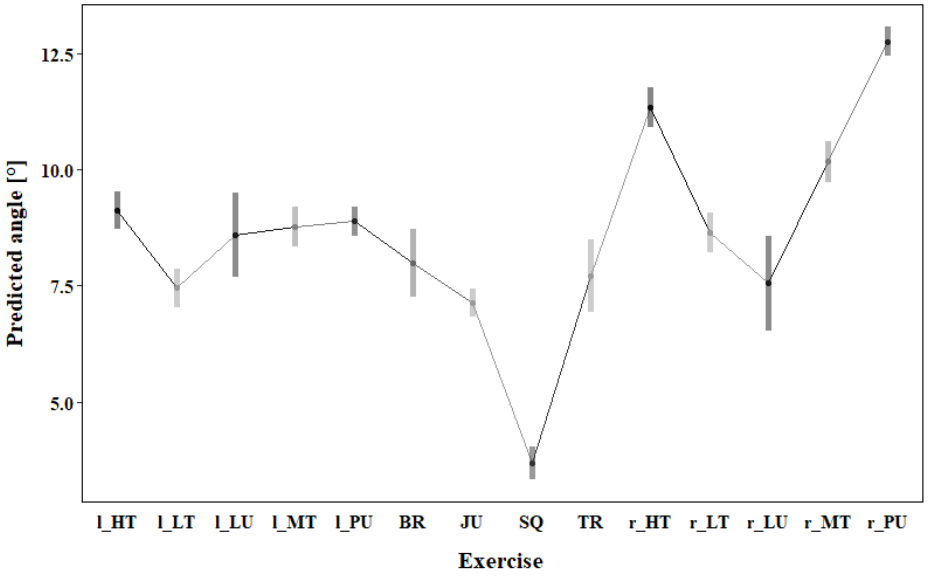
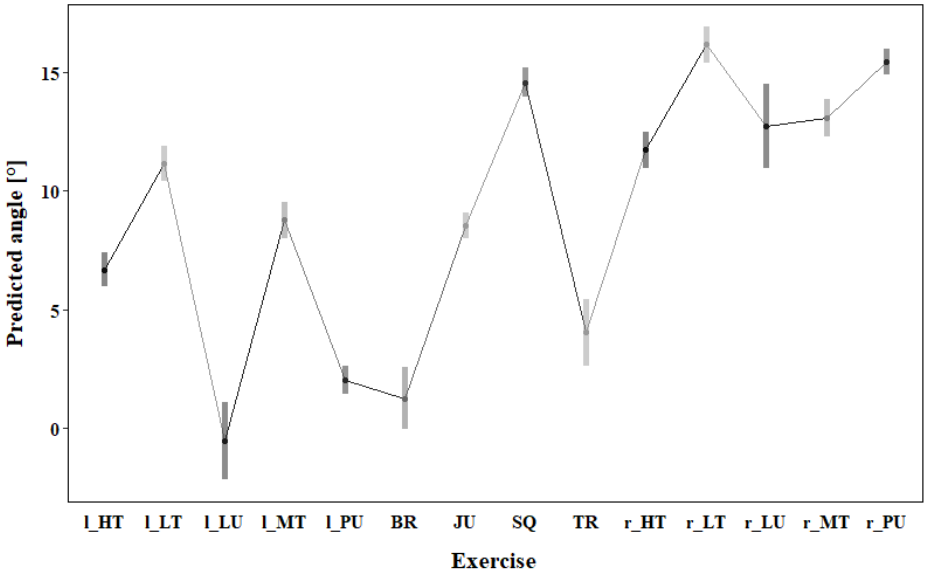


Figure 4. Estimated marginal means for maximal knee internal rotation of the right leg during 10°-30° knee flexion based on the biomechanical analysis during exergaming of athletes without any prior knee injury. BR: burpee; JU: jump; l_HT: left high touch; l_LT: left low touch; l_LU: left lunge; l_MT: left middle touch; l_PU: left punch; r_HT: right high touch; r_LT: right low touch; r_LU: right lunge; r_MT: right middle touch; r_PU: right punch; SQ: squat; TR: tripples.



XSL•FO
RenderX

Figure 5. Estimated marginal means for minimal hip flexion of the right side during 10°-30° knee flexion based on the biomechanical analysis during exergaming of athletes without any prior knee injuries. BR: burpee; JU: jump; l_HT: left high touch; l_LT: left low touch; l_LU: left lunge; l_MT: left middle touch; l_PU: left punch; r_HT: right high touch; r_LT: right low touch; r_LU: right lunge; r_MT: right middle touch; r_PU: right punch; SQ: squat; TR: tripples.



Patients

In comparison to CONs, patients showed significant differences in maximal KV and KIR during 10°-30° KF of the right leg during all investigated exercises, except for the KV of the right high touch. However, in minimal HF, significant differences between CONs and patients were only found in the squat and left punch (Table 2).

Table 2. Comparison of the biomechanical results of different exercises between CONs[a] and ACL[b]-injured participants.

	CONs (n=18)		Patients (n=6)		Difference	
Location and exercise (side)	Mean	SD	Mean	SD	Welch *t* test[c] (*df*)	*P* value
Knee valgus [°]						
Squat	4.32	3.48	4.75	3.15	–2.29 (1,745)	.02
High touch (left)	8.55	3.54	8.39	2.34	0.86 (1,719)	.39
High touch (right)	8.42	3.28	9.38	2.89	–4.53 (1,504)	<.001
Punch (left)	7.44	3.71	8.26	3.25	–4.40 (1,808)	<.001
Punch (right)	8.49	3.32	9.58	3.29	–6.18 (1,780)	<.001
Knee internal rotation [°]						
Squat	4.56	6.39	3.21	4.91	6.12 (1,1778)	<.001
High touch (left)	6.75	4.59	9.11	5.90	7.00 (1,605)	<.001
High touch (right)	11.35	5.78	8.38	5.16	7.85 (1,500)	<.001
Punch (left)	8.89	6.80	7.09	5.31	5.68 (1,912)	<.001
Punch (right)	12.77	6.96	9.45	7.36	8.51 (1,741)	<.001
Hip flexion [°]						
Squat	14.32	10.64	16.29	9.54	–4.83 (1,1501)	<.001
High touch (left)	6.67	10.43	5.89	9.15	1.20 (1,533)	.23
High touch (right)	11.72	10.01	11.87	10.11	–0.22 (1,441)	.83
Punch (left)	2.04	11.81	–0.46	10.59	–4.08 (1,758)	<.001
Punch (right)	15.42	10.54	14.68	10.02	1.37 (1,797)	.17

[a]CON: healthy participant.

[b]ACL: anterior cruciate ligament.

[c]2-tailed.

XSL•FO
RenderX

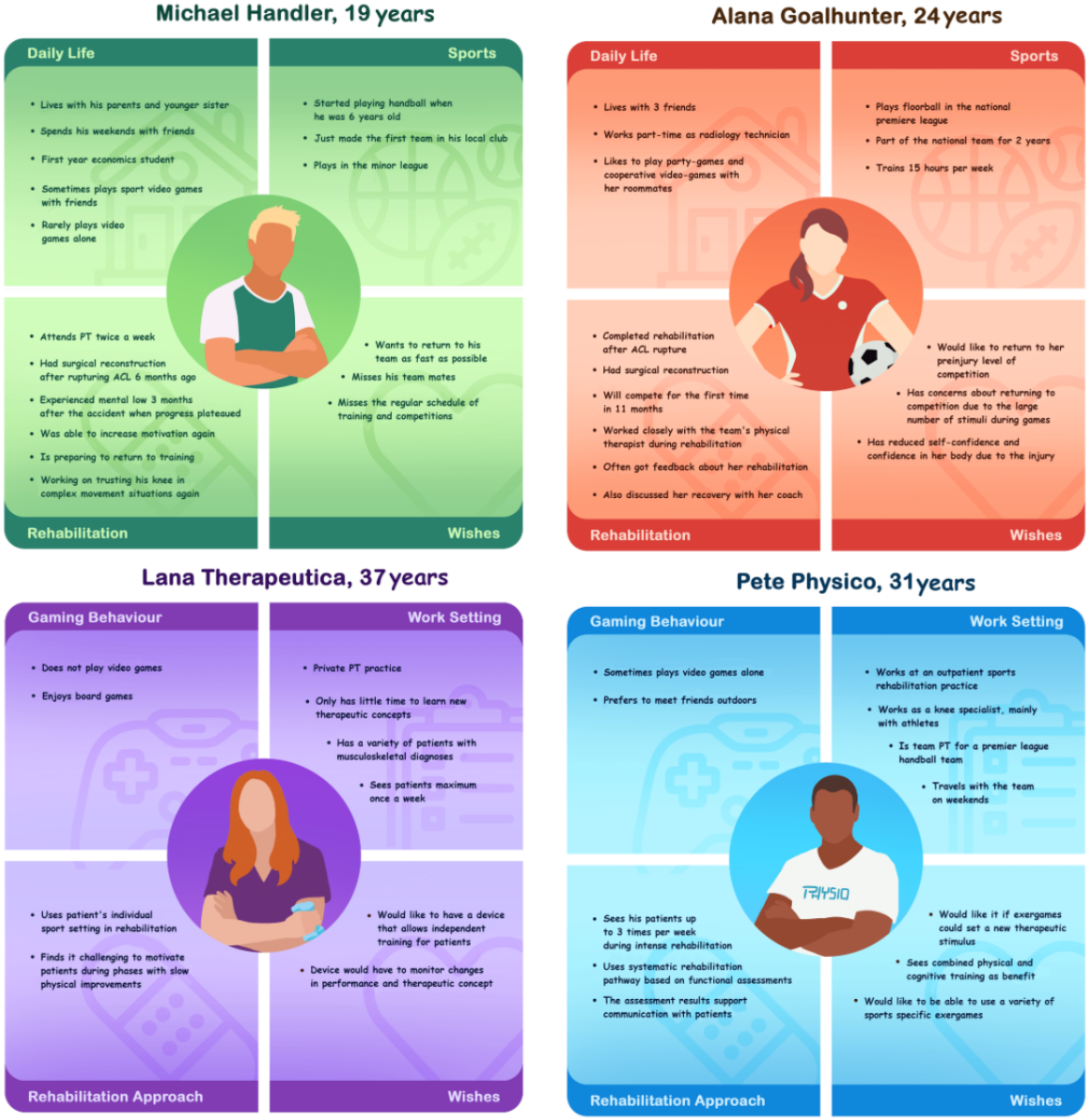
User Requirement for Rehabilitative ExerCube Use Evaluation Outcomes

The observation in the clinical setting, as well as the focus groups, revealed that there is a need for a motivating enhancement of conventional physiotherapy. Conventional physiotherapy can provide effective interventions, but it is challenging to keep patients motivated over the long-term period of rehabilitation after an ACL injury. This is especially true after about 6 months of rehabilitation, when progress may slow down and more complex tasks need to be trained, for example, with agility drills, jumping and hopping exercises, and full strength and balance need to be obtained. In this phase, it would be important to introduce sport-specific, motor-cognitive tasks; however, it is often not possible to allocate sufficient time for such exercises during the limited time for patient-therapist interaction. Further, as patients have different levels of activity preinjury and very different goals for rehabilitation, a digital intervention needs to be adjustable for different levels of performance.

The specific requirements for an exergaming solution for ACL rehabilitation based on patient and therapists' feedback are that the patient should be independent in its use, while the therapist should be able to manage and monitor the activity. Both physical and cognitive tasks should be challenging with increasing complexity. The exergame should provide feedback to the patient and provide a form of social interaction, which could be competitive or cooperative.

Based on the results of the observations, interviews, and walkthroughs, personas for patients as well as therapists were developed (Figure 6). Two athlete personas and 2 personas describing therapists were defined, representing a spectrum of individual personal characteristics but also rehabilitation paths and expectations.

Figure 6. Personas for athletes or patients undergoing rehabilitation and physical therapists were developed based on observations, interviews, and walkthroughs with patients and therapists. ACL: anterior cruciate ligament; PT: physiotherapy.



XSL•FO
RenderX

Rehabilitation-Specific Motivating Workout Scenarios for the ExerCube

Movement Concept

Based on the findings from the biomechanical analysis of the Sphery Racer [72-77], different exercises were excluded as they are not suitable for sports rehabilitation after knee injury, whereas the findings from the focus groups with patients (primary users) and physios (secondary users) as well as the consultation of experts from the field of movement science and physiotherapy allowed us to derive and connect different functional exercises which are well known for being used in traditional sports rehabilitation after knee injury, and which were then combined in a sports rehabilitation-specific functional movement concept featuring the following foci and movements:

Focus 1 included (1) balance and stability, which used side steps and touches, single-leg deadlift, skater jumps, and lunges; and (2) strength and reaction, which used tuck jumps, burpees, and lateral hops.

Focus 2 included coordination, agility, reaction or speed, and motor-cognitive performance, which used free, dynamic movements.

Exergame Experience Design

Furthermore, the focus groups revealed target group-specific preferences and requirements for the look and feel, for motivational aspects, as well as for meaningful game mechanics and gameplay.

Therefore, two complementary exergame concepts were derived: (1) concept 1: clean and sterile environment + little number of stimuli (Figure 7), and (2) concept 2: natural rather overloaded environment + many stimuli (Figure 8).

Figure 7. The Tube-Game of the newly developed exergame scenario being played in the ExerCube (source: ZHAW Zurich University of Applied Sciences).

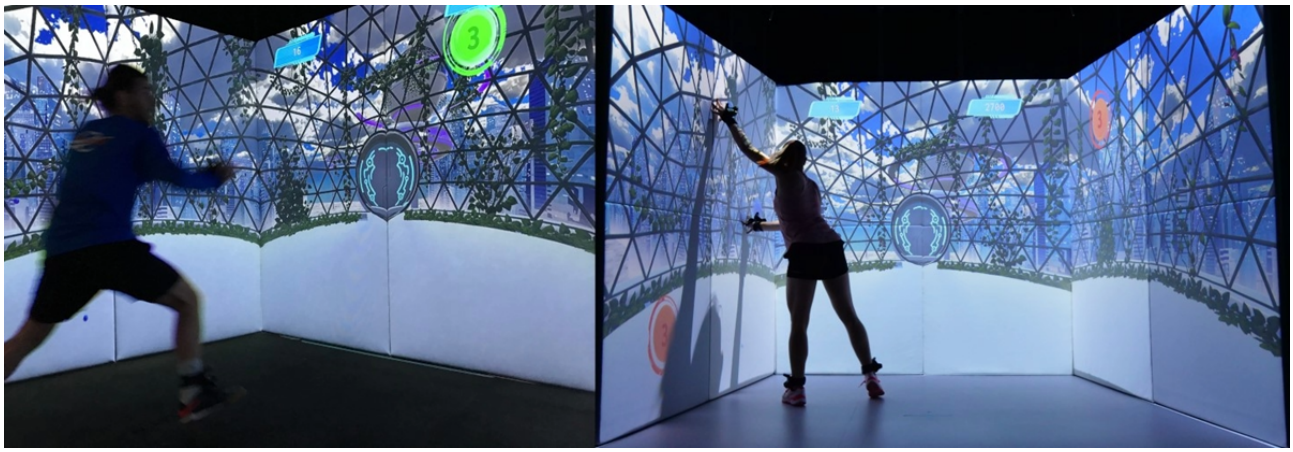


Exergame Description

Following the movement concept as well as the implications for the gameplay, game mechanics, and the audiovisual exergame design were translated into 2 meaningful exergame experiences with slightly different foci. Both experiences provide athletes with an immersive stimulus that allows them to shift their often very internal focus on the body part that was previously injured and set and keep it in their sports environment, such as with their team members, opponents, or a sports device. This is crucial for athletes to be ready to perform safely and well when returning to their sports.

The new exergame scenario (ExerUp-Game) consists of different parts—a main game (Tube-Game, Figure 7) and a mini game (Dome-Game, Figure 8). Both exergame parts focus on different training aspects:

Figure 8. The Dome-Game of the newly developed exergame scenario being played in the ExerCube (source: Zurich University of the Arts and ZHAW Zurich University of Applied Sciences).



The Tube-Game (control) mainly focuses on jumps, balance, stability, and coordination, whereas the Dome-Game (chaos)

XSL•FO
RenderX

focuses on agility, reaction, cognition, and inhibition. Depending on the respective focus of the therapy, the games can either be used sequentially and in combination or in a stand-alone version.

The ExerUp-Game features different presets of intensity and complexity (easy, medium, or hard), which allow patients and therapists to individually choose the matching intensity.

To match the transition from a secure and controlled environment in sports rehabilitation to the specific physical, cognitive, and mental requirements on the field (eg, during a match), the ExerUp-Game represents the principle of "control to chaos" [15].

When playing the ExerUp-Game, the player is elevated onto a platform in a nature-inspired science-fiction world. The player will then be performing the differently colored and challenging exercises to control the avatar on the digitalracing track (Tube-Game). After a while, the player will then enter the Dome-Game, which takes the player into a digital sphere, providing differently colored digital targets (green, yellow, and red) on all 3 cube walls surrounding the player, which need to be reached in a short time (includes a countdown). Red targets need to be avoided.

The game plays in a nonsports-specific scenario because it should be suitable for different athletes from different sports disciplines. A video snippet of the ExerUp-Game is available in Multimedia Appendix 2.

Validation of the ExerUp-Game

Objective

The participant's feedback was organized into different categories, depending on the design component or topic it was related to.

Attitudes Toward Gaming and Exergames

In addition to the feedback directly related to design and implementation, participants shared more general attitudes toward gaming and exergames. Some linked the use of games to nostalgic memories from their childhood, suggesting a personal familiarity that could enhance engagement. At the same time, others expressed that they do not intuitively associate gaming with physiotherapy, highlighting a perceptual disconnect between playful digital environments and structured therapeutic contexts. Nevertheless, the use of exergames was seen as a potentially motivating intervention. Particularly for individuals who are typically less engaged in conventional physical activity.

Game Content or Training Concept

The 2 parts of the game, Tube-Game and Dome-Game, were rated as being suitable for 2 different time points during rehabilitation, with the Tube-Game being appropriate earlier during rehabilitation than the Dome-Game. Additionally, the Dome-Game was deemed to be complementary to conventional physical therapy as it allows for a training regimen that is not possible in the conventional setting.

Further, it was highlighted that the games are not sport-specific. While this was considered a nonissue, as the games may be complementary to sport-specific training, it remains unclear whether the potential learnings from the games can be transferred to the sport-specific scenarios that the athletes will face after the rehabilitation.

Additional comments indicated that the focus of patients during gameplay could shift away from movement quality and toward game success, raising questions about whether the games sufficiently support therapeutic movement goals. Participants also questioned whether and how the system's technological structure influences movement execution, and how this aligns with physiotherapeutic best practices.

Implementation in Rehabilitation

The therapeutic and training experts deem the games to be playable by the patients themselves, which would free timely, cognitive, and physical resources of the therapists or trainers. The games could be played before or after conventional therapy. For the games to be supplementary to the conventional therapy, they should be adjustable by the therapists so they can define which exercises and which intensity a patient needs to train at. Further, the therapists should receive a report outlining which exercises are performed badly or well in order for the therapists to track the progress of an individual patient. However, the participants reported that they do not see a benefit in integrating the game into the conventional therapy session.

Additionally, participants expressed uncertainty regarding the role of the therapist within the exergame context, in particular, how much influence they could exert on the gameplay and exercise selection, and how visible or trackable therapeutic outcomes would be for them. These ambiguities point to a need for clearer therapist interfaces and configurable training modules that ensure transparency and flexibility in exercise design.

Use of Traditional Therapeutic Devices

Overall, the participants see potential to include traditional, therapeutic devices such as balls, weights, or agility ladders in the exergame and expressed the benefit if examples of such combinations of devices are provided by the manufacturer. The simplicity of the use of the ExerCube, together with the potential of individualizing the setting, for example, in combination with traditional devices, was mentioned as a strong benefit of the exergame.

This point was further emphasized through calls for more guidance and best-practice examples from the manufacturer. The integration of analog equipment was seen as a promising way to bridge the gap between familiar physical therapy routines and the novel experience of digital, game-based movement training.

Discussion

Principal Findings

The aim of this study was the development of an exergame for knee rehabilitation in sports based on the needs of users, such as patients and therapists, as well as biomechanical principles. Given the motivating environment provided by exergames, there is tremendous potential to disrupt the lengthy and repetitive rehabilitation process that can lead to low motivation [78] and consequently improved adherence [48], which is crucial for

XSL•FO
RenderX

good rehabilitation outcomes [79,80]. Additionally, exergames allow for motor-cognitive training that has positive effects on executive functions [60], which are impaired in patients after ACL injury [81].

Motor Performance During ExerCube Training

This project part aimed to analyze which of the exercises performed during exergaming in CONs elicit the most pronounced biomechanical risk factors associated with ACL injury. Moreover, data of the 2 exercises most at risk and the exercise least at risk were compared between CONs and patients.

Movement strategies and, therefore, maximal KV, KIR, and minimal HF differ between the 14 different exercises. Squat, jump, and tripples demonstrated the lowest peak angles of biomechanical risk factors for ACL injury. In contrast, punch, middle touch, and high touch showed higher KV and KIR angles than the other exercises, indicating a greater potential for ACL strain. Therefore, these exercises present more pronounced biomechanical risk factors for ACL injury and should be performed at a later stage of rehabilitation. If several different levels of the new exergame are to be designed, punches and touches should only be incorporated in the level intended after strength and neuromuscular control are sufficiently restored. Patients in rehabilitation after an ACL injury move differently than CONs during the exercises squat, punch, and high-touch. Patients present with more KV and less HF than CONs. In contrast, patients showed less KIR, suggesting a complex interplay of compensatory strategies. Overall, in 2 of the 3 investigated biomechanical risk factors for an ACL injury, patients showed more extreme movement patterns associated with injury risk. Of the 6 patients, only 2 were female, which could possibly explain the differences to CONs. An analysis of sex differences using data of CONs showed that females move with more KIR than males during touches and punches [62,75]. However, differences between males and females were dependent on the exercise and side, which supports the main effect of exercise in this analysis. Hence, the difference might result from the imbalance of the sexes in the patient group, whereas the CONs group had the same number of male and female participants. Additionally, high interindividual variability in movement patterns, as indicated by the high SDs, could have contributed to the difference [61]. As the analysis was limited to the right leg, a bias related to leg dominance might have been introduced. Incorporating biomechanical analysis into exergame design can help ensure that patients are challenged appropriately at each stage of recovery.

User Requirement for Rehabilitative ExerCube Use

This qualitative evaluation aimed to develop an understanding of the people directly involved in the rehabilitation after ACL injury, namely the athletes and the physical therapist. Further, key aspects of the rehabilitation after ACL injury were defined to inform the game designers.

In the early phases of rehabilitation, controlled movements are being performed. With the progression of the rehabilitation, more complex and chaotic tasks can be performed with the focus shifting from stability to stop-and-go and 3D, powerful movements as suggested by literature [15,82]. To ensure that an exergame is appropriate for the different levels of performance between athletes, as well as for the changing capabilities of a specific athlete, the game should include different levels of physical and cognitive challenge. A physical therapist should be able to choose which level of challenge to assign to a specific patient at a specific time point during the rehabilitation. Remaining with the concept of "control to chaos," both more controlled movements with limited cognitive challenge, as well as complex or chaotic movement patterns in combination with high cognitive demand, should be possible with the exergame. In a further step, adaptive difficulty of the motor and cognitive demand would be beneficial to specifically tailor the game to the needs of an individual athlete.

Maintaining motivation and, therefore, adherence to exercising appears to be one of the key challenges during the rehabilitation process after ACL injury [79]. With exergames being highly motivating [48], there could be tremendous potential, especially when the exergame targets a rehabilitation phase where motivation is generally low. The results indicate that elements such as feedback, as well as interaction with other athletes, for example, through leaderboards, could be ways to maintain the motivation of the athletes.

Rehabilitation-Specific Motivating Workout Scenarios for the ExerCube

The framework guiding the development of the exergaming scenario was chosen as "control to chaos," which aligns with the suggested progression from controlled movements while building strength and neuromuscular control to more complex and reactive movements later in rehabilitation [82]. Focusing on balance and stability incorporated with the exercises side steps and touches, single leg deadlift, skater jumps, and lunges is especially important at earlier stages of rehabilitation, while later on the focus shifts to strength and reaction, which is incorporated by the exercises tuck jumps, burpees, and lateral hops. Incorporation of the focus on agility, reactivity, and motor-cognitive performance as a second focus corresponds well to the needs of patients and therapists. To create a motivational environment, the look and feel was derived from the requirements of patients and therapists while incorporating an immersive stimulus, shifting the focus externally. Literature suggests that an external focus of attention is beneficial to elicit safer movement patterns when jumping [49]. Application of the workout scenario was intended to be broad; hence, no sport-specific scenarios were incorporated, thus limiting the transfer to sport-specific tasks.

Validation of the ExerUp-Game

Generally, the newly developed ExerUp-Game was well received. Participants were questioning whether exergames could be a beneficial alternative or addition to traditional physical therapy. The following three aspects were mentioned as crucial: (1) does the ExerUp-Game allow physical therapists to modify the content, (2) does the exergame technology ensure adequate movement quality and safety, and (3) do patients have an internal or external focus of attention (eg, focus on the game or their movement execution) during the exergame session as this determines the time point of the application of the

XSL•FO
RenderX

ExerUp-Game or the individual Dome-Game or Tube-Game during the rehabilitation.

Overall

Combining the detailed analysis of the kinematics during exergaming with user requirements established the need for different levels of physical and cognitive challenge during an exergame for sports rehabilitation. Different movement tasks performed during the exergame were associated with different knee angles. The assessment of qualitative user-specific needs indicated the need for different levels of challenge, starting from controlled situations and ranging to chaotic movement patterns. This information was translated into a movement concept and a design concept that were the baseline for the development of the new, target group-specific ExerUp-Game consisting of 2 individual, but combinable games (Dome-Game and Tube-Game).

To subsequently implement the exergame into the daily practice of the target group, re-evaluation of whether target-specific requirements were met is needed. Only based on the results of an evaluation can the exergame be further improved and find its application in practice.

Limitations

Each phase of the project had specific limitations, which are outlined below in the order in which the phases are presented in this paper. First, data on biomechanical movement patterns were collected in a realistic setting. Therefore, exercises and their execution and performance were neither normalized nor standardized. As a result, high variability of biomechanical variables was observed. Second, due to the adaptive speed levels of the Sphery Racer according to exercise performance, participants performed a different number of exercises. However, the exercise sequence was fixed, which is why comparability between exercises is still feasible. Third, the peak angle within 10°-30° KF does not necessarily correspond to the peak angle during the entire trial. According to literature, the injury mechanism of ACL ruptures is likely multiplanar [83], justifying the assessment of multiple risk factors simultaneously. Fourth, the sample size was small and unbalanced between CONs and patients, which limits the generalizability of the results of this first project part. Future studies should aim for larger and more balanced samples to strengthen representativeness.

Fifth, in the second part of the project, there were only a few physical therapists and only 1 athlete after an ACL injury was involved. The reason for this was to find experienced people with availability to participate in this project phase. Consequently, individual opinions may have largely influenced the results of interviews and focus groups.

Sixth, there is a lack of scientifically valid and universally accepted rehabilitation pathways after ACL injury. Consequently, the work in the third project phase, the development of rehabilitation-specific workout scenarios for the ExerCube, could not be based on a standardized rehabilitation program. This limits the evidence base used for this work.

Seventh, with the evaluation, it was not established how effective the new rehabilitation game is, what its long-term benefits are, and how it could be implemented in sports rehabilitation. While initial feedback from focus groups was positive, these statements have limited generalizability. These aspects will be covered in future research.

Recommendations

The development of an exergame for sports rehabilitation requires an interdisciplinary team. As team members often come from diverse professional backgrounds, this can lead to specific challenges. For example, establishing a common terminology often demands extensive discussions and frequent exchanges to avoid misunderstandings. Allocating sufficient resources to address these challenges should therefore be an integral part of planning and implementing similar projects in the future.

Additionally, a user-centered design approach is essential. A thorough understanding of the needs of both primary and secondary end users is crucial to creating a meaningful exergaming experience that is truly tailored to the target population.

Conclusions

An interdisciplinary, evidence-based approach combining quantitative and qualitative methods led to the development of a tailored exergame scenario for sports rehabilitation following ACL injuries. Subsequent evaluations with end users indicated that the scenario has strong potential to be both motivating and suitable for independent use. However, further adaptations under the supervision of therapists and closer integration with conventional rehabilitation tools are still necessary.

Acknowledgments

The authors kindly thank Sphery Ltd. for providing the infrastructure and technical support throughout this study's duration, which ensured a smooth implementation of this study. This study was funded by the Digital Innovation Program of the Canton of Zurich (DIZH; 17).

Data Availability

Data used for biomechanical analysis is available on the web [84].

XSL•FO
RenderX

Authors' Contributions

AMN and EG acquired the funding for the project. AMN, EG, and MH contributed equally to the conception, design, and drafting of the manuscript. The studies were designed and conducted by AMN, EG, and MH. MH coordinated and executed the first study, under the supervision of EG in collaboration with AMN. LW and LS contributed to the focus groups and co-developed the new exergame concepts under the supervision of AMN and in collaboration with EG, MH, and the Sphery Ltd. team. AMN, EG, and MH contributed to the interpretation of the findings and critically revised the manuscript for important intellectual content. All authors approved the final version of the manuscript.

Conflicts of Interest

Besides her academic position at Zurich University of the Arts, ALM-N is also cofounder of Sphery Ltd, which developed the ExerCube based on the results of her research and development work. No revenue was paid (or promised to be paid) directly to ALM-N, to Sphery, or the research institutions. The other authors declare no conflicts of interest.

Multimedia Appendix 1
Detailed information on the performed exercises and biomechanical outcomes.
[DOCX File , 285 KB - games_v13i1e68431_app1.docx]

Multimedia Appendix 2
Video snippet of the newly developed exergame for sports rehabilitation.
[MP4 File (MP4 Video), 116150 KB - games_v13i1e68431_app2.mp4]

References

1. Cavanaugh JT, Powers M. ACL rehabilitation progression: where are we now? Curr Rev Musculoskelet Med 2017;10(3):289-296 [FREE Full text] [doi: 10.1007/s12178-017-9426-3] [Medline: 28791612]
2. Kruse L, Gray B, Wright R. Rehabilitation after anterior cruciate ligament reconstruction: a systematic review. J Bone Jt Surg Am 2012;94(19):1737-1748 [FREE Full text] [doi: 10.2106/JBJS.K.01246] [Medline: 23032584]
3. van Melick N, van Cingel REH, Brooijmans F, Neeter C, van Tienen T, Hullegie W, et al. Evidence-based clinical practice update: practice guidelines for anterior cruciate ligament rehabilitation based on a systematic review and multidisciplinary consensus. Br J Sports Med 2016;50(24):1506-1515. [doi: 10.1136/bjsports-2015-095898] [Medline: 27539507]
4. Hootman JM, Dick R, Agel J. Epidemiology of collegiate injuries for 15 sports: summary and recommendations for injury prevention initiatives. J Athl Train 2007;42(2):311-319 [FREE Full text] [Medline: 17710181]
5. Gornitzky AL, Lott A, Yellin JL, Fabricant PD, Lawrence JT, Ganley TJ. Sport-specific yearly risk and incidence of anterior cruciate ligament tears in high school athletes: a systematic review and meta-analysis. Am J Sports Med 2016;44(10):2716-2723. [doi: 10.1177/0363546515617742] [Medline: 26657853]
6. Moses B, Orchard J, Orchard J. Systematic review: annual incidence of ACL injury and surgery in various populations. Res Sports Med 2012 Jul;20(3-4):157-179. [doi: 10.1080/15438627.2012.680633] [Medline: 22742074]
7. Pasanen K, Parkkari J, Kannus P, Rossi L, Palvanen M, Natri A, et al. Injury risk in female floorball: a prospective one-season follow-up. Scand J Med Sci Sports 2008;18(1):49-54. [doi: 10.1111/j.1600-0838.2007.00640.x] [Medline: 17490461]
8. Patel AD, Bullock GS, Wrigley J, Paterno MV, Sell TC, Losciale JM. Does sex affect second ACL injury risk? A systematic review with meta-analysis. Br J Sports Med 2021;55(15):873-882. [doi: 10.1136/bjsports-2020-103408] [Medline: 34001504]
9. Wiggins AJ, Grandhi RK, Schneider DK, Stanfield D, Webster KE, Myer GD. Risk of secondary injury in younger athletes after anterior cruciate ligament reconstruction: a systematic review and meta-analysis. Am J Sports Med 2016;44(7):1861-1876 [FREE Full text] [doi: 10.1177/0363546515621554] [Medline: 26772611]
10. Shelbourne KD, Gray T, Haro M. Incidence of subsequent injury to either knee within 5 years after anterior cruciate ligament reconstruction with patellar tendon autograft. Am J Sports Med 2009;37(2):246-251. [doi: 10.1177/0363546508325665] [Medline: 19109531]
11. Andrade R, Pereira R, van Cingel R, Staal JB, Espregueira-Mendes J. How should clinicians rehabilitate patients after ACL reconstruction? A systematic review of clinical practice guidelines (CPGs) with a focus on quality appraisal (AGREE II). Br J Sports Med 2020;54(9):512-519. [doi: 10.1136/bjsports-2018-100310] [Medline: 31175108]
12. Cheatham SA, Johnson DL. Current concepts in ACL Injuries. Phys Sportsmed 2010;38(3):61-68. [doi: 10.3810/psm.2010.10.1809] [Medline: 20959697]
13. Wilk K, Arrigo C. Rehabilitation principles of the anterior cruciate ligament reconstructed knee: twelve steps for successful progression and return to play. Clin Sports Med 2017;36(1):189-232. [doi: 10.1016/j.csm.2016.08.012] [Medline: 27871658]
14. Yabroudi M, Irrgang J. Rehabilitation and return to play after anatomic anterior cruciate ligament reconstruction. Clin Sports Med 2013;32(1):165-175. [doi: 10.1016/j.csm.2012.08.016] [Medline: 23177470]
15. Taberner M, Allen T, Cohen DD. Progressing rehabilitation after injury: consider the 'control-chaos continuum'. Br J Sports Med 2019;53(18):1132-1136 [FREE Full text] [doi: 10.1136/bjsports-2018-100157] [Medline: 30737202]

XSL•FO
RenderX

16. Gokeler A, McKeon PO, Hoch MC. Shaping the functional task environment in sports injury rehabilitation: a framework to integrate perceptual-cognitive training in rehabilitation. Athletic Train Sports Health Care 2020;12(6):283-292. [doi: 10.3928/19425864-20201016-01]
17. Ardern CL, Taylor NF, Feller JA, Webster KE. Fifty-five per cent return to competitive sport following anterior cruciate ligament reconstruction surgery: an updated systematic review and meta-analysis including aspects of physical functioning and contextual factors. Br J Sports Med 2014;48(21):1543-1552. [doi: 10.1136/bjsports-2013-093398] [Medline: 25157180]
18. Buckthorpe M. Optimising the late-stage rehabilitation and return-to-sport training and testing process after ACL reconstruction. Sports Med 2019;49(7):1043-1058. [doi: 10.1007/s40279-019-01102-z] [Medline: 31004279]
19. Stearns KM, Pollard CD. Abnormal frontal plane knee mechanics during sidestep cutting in female soccer athletes after anterior cruciate ligament reconstruction and return to sport. Am J Sports Med 2013;41(4):918-923. [doi: 10.1177/0363546513476853] [Medline: 23425687]
20. Goerger BM, Marshall SW, Beutler AI, Blackburn JT, Wilckens JH, Padua DA. Anterior cruciate ligament injury alters preinjury lower extremity biomechanics in the injured and uninjured leg: the JUMP-ACL study. Br J Sports Med 2015;49(3):188-195. [doi: 10.1136/bjsports-2013-092982] [Medline: 24563391]
21. Lee S, Chow J, Tillman M. Persons with reconstructed ACL exhibit altered knee mechanics during high-speed maneuvers. Int J Sports Med 2014;35(6):528-533. [doi: 10.1055/s-0033-1358466] [Medline: 24408765]
22. Oh Y, Yang S. 2010 Presented at: Meaningful Play 2010; October 21-23, 2010; Michigan p. 21.
23. Pirovano M, Surer E, Mainetti R, Lanzi PL, Alberto Borghese N. Exergaming and rehabilitation: a methodology for the design of effective and safe therapeutic exergames. Entertainment Comput 2016;14:55-65. [doi: 10.1016/j.entcom.2015.10.002]
24. Valenzuela T, Okubo Y, Woodbury A, Lord SR, Delbaere K. Adherence to technology-based exercise programs in older adults: a systematic review. J Geriatr Phys Ther 2018;41(1):49-61. [doi: 10.1519/JPT.0000000000000095] [Medline: 27362526]
25. Macvean A, Robertson J. Understanding exergame users' physical activity, motivation and behavior over time. : ACM; 2013 Presented at: CHI '13: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; April 27, 2013; Paris, France p. 1251-1260. [doi: 10.1145/2470654.2466163]
26. Lyons EJ. Cultivating engagement and enjoyment in exergames using feedback, challenge, and rewards. Games Health J 2015;4(1):12-18. [doi: 10.1089/g4h.2014.0072] [Medline: 26181675]
27. Byrne AM, Kim M. The exergame as a tool for mental health treatment. J Creativity Ment Health 2019;14(4):465-477. [doi: 10.1080/15401383.2019.1627263]
28. Joronen K, Aikasalo A, Suvitie A. Nonphysical effects of exergames on child and adolescent well-being: a comprehensive systematic review. Scand J Caring Sci 2017;31(3):449-461. [doi: 10.1111/scs.12393] [Medline: 27862145]
29. Lee S, Kim W, Park T, Peng W. The psychological effects of playing exergames: a systematic review. Cyberpsychol Behav Soc Netw 2017;20(9):513-532. [doi: 10.1089/cyber.2017.0183] [Medline: 28837353]
30. Li J, Theng Y, Foo S. Effect of exergames on depression: a systematic review and meta-analysis. Cyberpsychol Behav Soc Netw 2016;19(1):34-42. [doi: 10.1089/cyber.2015.0366] [Medline: 26716638]
31. Röglin L, Stoll O, Ketelhut S, Martin-Niedecken A, Ketelhut K. A 3-month school-based exergaming intervention increases students' physical self-concept: a randomized controlled trial. Ger J Exerc Sport Res 2024;54(4):598-608. [doi: 10.1007/s12662-024-00954-6]
32. Staiano A, Calvert S. Exergames for physical education courses: physical, social, and cognitive benefits. Child Dev Perspect 2011;5(2):93-98 [FREE Full text] [doi: 10.1111/j.1750-8606.2011.00162.x] [Medline: 22563349]
33. Benzing V, Heinks T, Eggenberger N, Schmidt M. Acute cognitively engaging exergame-based physical activity enhances executive functions in adolescents. PLOS One 2016;11(12):e0167501. [doi: 10.1371/journal.pone.0167501] [Medline: 28030542]
34. Best JR, Nagamatsu LS, Liu-Ambrose T. Improvements to executive function during exercise training predict maintenance of physical activity over the following year. Front Hum Neurosci 2014;8:353 [FREE Full text] [doi: 10.3389/fnhum.2014.00353] [Medline: 24904387]
35. Mura G, Carta MG, Sancassiani F, Machado S, Prosperini L. Active exergames to improve cognitive functioning in neurological disabilities: a systematic review and meta-analysis. Eur J Phys Rehabil Med 2018;54(3):450-462 [FREE Full text] [doi: 10.23736/S1973-9087.17.04680-9] [Medline: 29072042]
36. Stojan R, Voelcker-Rehage C. A systematic review on the cognitive benefits and neurophysiological correlates of exergaming in healthy older adults. J Clin Med 2019;8(5):734 [FREE Full text] [doi: 10.3390/jcm8050734] [Medline: 31126052]
37. Xiong S, Zhang P, Gao Z. Effects of exergaming on preschoolers' executive functions and perceived competence: a pilot randomized trial. J Clin Med 2019;8(4):469 [FREE Full text] [doi: 10.3390/jcm8040469] [Medline: 30959860]
38. Ermi L, Mäyrä F. Fundamental components of the gameplay experience: analysing immersion. 2005 Presented at: Digital Games Research Conference 2005, Changing Views: Worlds in Play; June 16-20, 2005; Vancouver, British Columbia, Canada. [doi: 10.26503/dl.v2005i1.119]

XSL•FO
RenderX

39. Martin-Niedecken A, Mahrer A, Rogers K, de Bruin E, Schättin A. “HIIT” the ExerCube: comparing the effectiveness of functional high-intensity interval training in conventional vs. exergame-based training. Front Comput Sci 2020;2:33 [FREE Full text] [doi: 10.3389/fcomp.2020.00033]
40. Ballesteros S, Voelcker-Rehage C, Bherer L. Editorial: cognitive and brain plasticity induced by physical exercise, cognitive training, video games, and combined interventions. Front Hum Neurosci 2018;12:169 [FREE Full text] [doi: 10.3389/fnhum.2018.00169] [Medline: 29867403]
41. Egger F, Benzing V, Conzelmann A, Schmidt M. Boost your brain, while having a break! The effects of long-term cognitively engaging physical activity breaks on children's executive functions and academic achievement. PLOS One 2019;14(3):e0212482 [FREE Full text] [doi: 10.1371/journal.pone.0212482] [Medline: 30840640]
42. Schättin A, Arner R, Gennaro F, de Bruin ED. Adaptations of prefrontal brain activity, executive functions, and gait in healthy elderly following exergame and balance training: a randomized-controlled study. Front Aging Neurosci 2016;8:278 [FREE Full text] [doi: 10.3389/fnagi.2016.00278] [Medline: 27932975]
43. Dai B, Cook RF, Meyer EA, Sciascia Y, Hinshaw TJ, Wang C, et al. The effect of a secondary cognitive task on landing mechanics and jump performance. Sports Biomech 2018;17(2):192-205. [doi: 10.1080/14763141.2016.1265579] [Medline: 28632053]
44. Burcal CJ, Needle AR, Custer L, Rosen AB. The effects of cognitive loading on motor behavior in injured individuals: a systematic review. Sports Med 2019;49(8):1233-1253. [doi: 10.1007/s40279-019-01116-7] [Medline: 31066022]
45. Shi H, Huang H, Yu Y, Liang Z, Zhang S, Yu B, et al. Effect of dual task on gait asymmetry in patients after anterior cruciate ligament reconstruction. Sci Rep 2018;8(1):12057 [FREE Full text] [doi: 10.1038/s41598-018-30459-w] [Medline: 30104568]
46. Mohammadi-Rad S, Salavati M, Ebrahimi-Takamjani I, Akhbari B, Sherafat S, Negahban H, et al. Dual-tasking effects on dynamic postural stability in athletes with and without anterior cruciate ligament reconstruction. J Sport Rehabil 2016;25(4):324-329. [doi: 10.1123/jsr.2015-0012] [Medline: 27632858]
47. Lee HA. A unified conceptual model of immersive experience in extended reality. Comput Hum Behav Rep 2025;18:100663 [FREE Full text] [doi: 10.31234/osf.io/sefkh]
48. Howard MC. A meta-analysis and systematic literature review of virtual reality rehabilitation programs. Comput Hum Behav 2017;70:317-327. [doi: 10.1016/j.chb.2017.01.013]
49. Gokeler A, Benjaminse A, Welling W, Alferink M, Eppinga P, Otten B. The effects of attentional focus on jump performance and knee joint kinematics in patients after ACL reconstruction. Phys Ther Sport 2015;16(2):114-120. [doi: 10.1016/j.ptsp.2014.06.002] [Medline: 25443228]
50. Gokeler A, Bisschop M, Myer GD, Benjaminse A, Dijkstra PU, van Keeken HG, et al. Immersive virtual reality improves movement patterns in patients after ACL reconstruction: implications for enhanced criteria-based return-to-sport rehabilitation. Knee Surg Sports Traumatol Arthrosc 2016;24(7):2280-2286. [doi: 10.1007/s00167-014-3374-x] [Medline: 25311052]
51. Vieira C, Ferreira da Silva Pais-Vieira C, Novais J, Perrotta A. Serious game design and clinical improvement in physical rehabilitation: systematic review. JMIR Serious Games 2021;9(3):e20066 [FREE Full text] [doi: 10.2196/20066] [Medline: 34554102]
52. Bonnechère B, Jansen B, Omelina L, Van Sint Jan S. The use of commercial video games in rehabilitation: a systematic review. Int J Rehabil Res 2016;39(4):277-290. [doi: 10.1097/MRR.0000000000000190] [Medline: 27508968]
53. Fernandes CS, Magalhães B, Gomes JA, Santos C. Exergames to improve rehabilitation after anterior cruciate ligament injury: systematic review and GRADE evidence synthesis. Int J Orthop Trauma Nurs 2022;44:100917. [doi: 10.1016/j.ijotn.2021.100917] [Medline: 35033981]
54. Martin-Niedecken AL. Towards balancing fun and exertion in exergames: exploring the impact of movement-based controller devices, exercise concepts, game adaptivity and player modes on player experience and training intensity in different exergame settings. CORE (Connecting Repositories). 2021. URL: https://core.ac.uk/download/478632112.pdf [accessed 2025-09-19]
55. Martin-Niedecken AL, Mekler ED. The ExerCube: Participatory Design of an Immersive Fitness Game Environment. Serious Games Cham: Springer International Publishing; 2018:263-275.
56. Schättin A, Pickles J, Flagmeier D, Schärer B, Riederer Y, Niedecken S, et al. Development of a novel home-based exergame with on-body feedback: usability study. JMIR Serious Games 2022;10(4):e38703 [FREE Full text] [doi: 10.2196/38703] [Medline: 36472900]
57. Ringgenberg N, Mildner S, Hapig M, Hermann S, Kruszewski K, Martin-Niedecken AL, et al. ExerG: adapting an exergame training solution to the needs of older adults using focus group and expert interviews. J Neuroeng Rehabil 2022;19(1):89 [FREE Full text] [doi: 10.1186/s12984-022-01063-x] [Medline: 35974409]
58. Martin-Niedecken AL, Rogers K, Turmo VL, Mekler ED, Márquez SE. ExerCube vs. personal trainer: evaluating a holistic, immersive, and adaptive fitness game setup. USA: Association for Computing Machinery; 2019 Presented at: CHI '19: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems; May 4-9, 2019; Glasgow Scotland UK p. 1-15. [doi: 10.1145/3290605.3300318]

XSL•FO
RenderX

59. Ketelhut S, Ketelhut RG, Kircher E, Röglin L, Hottenrott K, Martin-Niedecken AL, et al. Gaming instead of training? Exergaming induces high-intensity exercise stimulus and reduces cardiovascular reactivity to cold pressor test. Front Cardiovasc Med 2022;9:798149 [FREE Full text] [doi: 10.3389/fcvm.2022.798149] [Medline: 35155627]
60. Martin-Niedecken AL, Bucher V, Adcock M, de Bruin ED, Schättin A. Impact of an exergame intervention on cognitive-motor functions and training experience in young team sports athletes: a non-randomized controlled trial. Front Sports Act Living 2023;5:1170783 [FREE Full text] [doi: 10.3389/fspor.2023.1170783] [Medline: 38077282]
61. Haas MC, Martin-Niedecken AL, Wild L, Schneeberger L, Graf ES. Intra- and inter-subject variability in a realistic, non-standardized exergaming scenario. ISBS Proc Arch 2024;42(1):48 [FREE Full text]
62. Haas MC, Martin-Niedecken AL, Wild L, Schneeberger L, Graf ES. Biomechanical risk factors for ACL injury during a high-intensity exergame differ between the sexes based on exercise type. PLOS One 2025;20(5):e0324702 [FREE Full text] [doi: 10.1371/journal.pone.0324702] [Medline: 40397966]
63. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. 2021. URL: https://www.R-project.org/ [accessed 2025-09-19]
64. Kassambara A. rstatix: pipe-friendly framework for basic statistical tests. The Comprehensive R Archive Network. 2022. URL: https://CRAN.R-project.org/package=rstatix [accessed 2025-09-19]
65. Lenth RV. emmeans: estimated marginal means, aka least-squares means. The Comprehensive R Archive Network. 2021. URL: https://CRAN.R-project.org/package=emmeans [accessed 2025-09-19]
66. van Grinsven S, van Cingel REH, Holla CJM, van Loon CJM. Evidence-based rehabilitation following anterior cruciate ligament reconstruction. Knee Surg Sports Traumatol Arthrosc 2010;18(8):1128-1144. [doi: 10.1007/s00167-009-1027-2] [Medline: 20069277]
67. Braun V, Clarke V. Thematic Analysis: A Practical Guide. Los Angeles London New Delhi Singapore Washington DC Melbourne: SAGE; 2021.
68. Cavagnuolo M, Capozza V, Matrella A. The walkthrough method: state of the art, innovative aspects, and application fields. In: Punziano G, Delli Paoli A, editors. Handbook of Research on Advanced Research Methodologies for a Digital Society. Hershey, Pennsylvania: IGI Global Scientific Publishing; 2022:461-486.
69. Schleicher D, Jones P, Kachur O. Bodystorming as embodied designing. interactions 2010;17(6):47-51. [doi: 10.1145/1865245.1865256]
70. Eccles DW, Arsal G. The think aloud method: what is it and how do I use it? Qual Res Sport Exerc Health 2017;9(4):514-531. [doi: 10.1080/2159676x.2017.1331501]
71. World Medical Association. World medical association declaration of Helsinki: ethical principles for medical research involving human participants. JAMA 2025;333(1):71-74. [doi: 10.1001/jama.2024.21972] [Medline: 39425955]
72. Haas M, Martin-Niedecken A, Wild L, Schneeberger L, Graf E. Biomechanische analyse von risikofaktoren für knieverletzungen während eines exergame-trainings. ZHAW digitalcollection. 2023. URL: https://digitalcollection.zhaw.ch/handle/11475/29787 [accessed 2025-09-19]
73. Haas MC, Martin-Niedecken AL, Wild L, Schneeberger L, Graf ES. Comparison of biomechanical risk factors for ACL injury between patients and healthy subjects during exergaming. CISS 2024;9(2):045. [doi: 10.36950/2024.2ciss045]
74. Haas M, Martin-Niedecken A, Wild L, Schneeberger L, Graf E. Biomechanical analysis of the knee during an immersive, high-intensity exergame training in healthy athletes. ZHAW Zürcher Hochsch Angew Wiss 2023:1. [doi: 10.3218/3725-8]
75. Haas MC, Martin-Niedecken AL, Wild L, Schneeberger L, Graf ES. Biomechanical risk factors for ACL injury differ between the sexes when performing a high-intensive exergame. CISS 2024;9(2):044. [doi: 10.36950/2024.2ciss044]
76. Martin-Niedecken AL, Haas M, Sascha K, Graf E. Exergames – Innovative Trainingsansätze für Prävention, Rehabilitation und Sport!? ZHAW digitalcollection. 2023. URL: https://digitalcollection.zhaw.ch/handle/11475/29810 [accessed 2025-09-19]
77. Graf E, Haas M, Wild L, Schneeberger L, Martin-Niedecken A. Forschungsbasiertes exergame-design für die sportrehabilitation. ZHAW digitalcollection. 2023. URL: https://digitalcollection.zhaw.ch/handle/11475/29811 [accessed 2025-09-19]
78. Tannous H, Istrate D, Benlarbi-Delai A, Sarrazin J, Gamet D, Ho Ba Tho M, et al. A new multi-sensor fusion scheme to improve the accuracy of knee flexion kinematics for functional rehabilitation movements. Sensors (Basel) 2016;16(11):1914 [FREE Full text] [doi: 10.3390/s16111914] [Medline: 27854288]
79. Pizzari T, McBurney H, Taylor N, Feller J. Adherence to anterior cruciate ligament rehabilitation: a qualitative analysis. J Sport Rehabil 2022;11(2):90-102. [doi: 10.1123/jsr.11.2.90]
80. Brewer B, Cornelius A, Van Raalte J, Brickner J, Sklar J, Corsetti J, et al. Rehabilitation adherence and anterior cruciate ligament reconstruction outcome. Psychol, Health Med 2004;9(2):163-175. [doi: 10.1080/13548500410001670690]
81. Piskin D, Benjaminse A, Dimitrakis P, Gokeler A. Neurocognitive and neurophysiological functions related to ACL injury: a framework for neurocognitive approaches in rehabilitation and return-to-sports tests. Sports Health 2022;14(4):549-555 [FREE Full text] [doi: 10.1177/19417381211029265] [Medline: 34236003]
82. Jenkins SM, Guzman A, Gardner BB, Bryant SA, Del Sol SR, McGahan P, et al. Rehabilitation after anterior cruciate ligament injury: review of current literature and recommendations. Curr Rev Musculoskelet Med 2022;15(3):170-179 [FREE Full text] [doi: 10.1007/s12178-022-09752-9] [Medline: 35381974]

XSL•FO
RenderX

83. Quatman CE, Quatman-Yates CC, Hewett TE. A 'plane' explanation of anterior cruciate ligament injury mechanisms: a systematic review. Sports Med 2010;40(9):729-746. [doi: 10.2165/11534950-000000000-00000] [Medline: 20726620]
84. Haas M, Graf ES, Martin-Niedecken AL. ExerUP! Kinematics during exergaming. Harvard Dataverse. URL: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/XBJXC4 [accessed 2025-09-19]

Abbreviations

ACL: anterior cruciate ligament
CON: healthy participant
HF: hip flexion angle
IR: incidence rate
KF: knee flexion
KIR: knee internal rotation angle
KV: knee valgus angle
VR: virtual reality

Edited by A Coristine; submitted 14.11.24; peer-reviewed by S Göbel, M Gasmi ; comments to author 05.05.25; revised version received 30.06.25; accepted 05.08.25; published 31.10.25.

Please cite as:
Haas MC, Wild L, Schneeberger L, Graf ES, Martin-Niedecken AL
Exploring Mixed-Reality Exergames for Sports Rehabilitation: Design Insights and Evaluation Findings
JMIR Serious Games 2025;13:e68431
URL: https://games.jmir.org/2025/1/e68431
doi:10.2196/68431
PMID:

©Michelle C Haas, Larissa Wild, Leander Schneeberger, Eveline S Graf, Anna Lisa Martin-Niedecken. Originally published in JMIR Serious Games (https://games.jmir.org), 31.10.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Achieving Gameplay Independence in Virtual Reality Exergames for Individuals With Mild Intellectual Disabilities: Pilot Study

Julia Ciążyńska[1], PhD; Janusz Maciaszek[1], Prof Dr

Department of Physical Activity and Health Promotion Science, Akademia Wychowania Fizycznego im. Eugeniusza Piaseckiego w Poznaniu, Poznan, Greater Poland Voivodeship, Poland

Corresponding Author:
Julia Ciążyńska, PhD
Department of Physical Activity and Health Promotion Science
Akademia Wychowania Fizycznego im. Eugeniusza Piaseckiego w Poznaniu
Królowej Jadwigi 27/39
Poznan, Greater Poland Voivodeship, 61-883
Poland
Phone: 48 (61) 835 52 10
Email: ciazynska@awf.poznan.pl

Abstract

Background: Individuals with mild intellectual disabilities (ID) often face cognitive and functional challenges, which can lead to low physical activity (PA) and a higher risk of obesity. While virtual reality (VR) exergames show promise for promoting PA in typically developing children, a key barrier for individuals with ID is the lack of a structured teaching methodology. This study argues that a tailored approach is essential to help children with mild ID gain independence in gameplay. By learning specific patterns, they can achieve greater autonomy, which not only facilitates increased PA but also improves motor competence, physical fitness, functional abilities, and overall well-being.

Objective: This study aims to evaluate the effectiveness of the WISH (Warm-up, Imitation, Settings, Half-hour exergame session) and WON (Warm-up, Objective evaluation, No problem!) training protocols in improving participant independence, exergame performance, and overall gameplay experience in VR.

Methods: We used a multisession, single-group research design involving 16 training sessions in this pilot study. The 16 sessions were conducted during scheduled physical education classes at a special school in Poland from October 2023 to May 2024. The intervention comprised two main protocols: the WISH protocol (sessions 1-4), an introductory phase focused on familiarization with VR technology and gameplay mechanics, and the WON protocol (sessions 5-16), designed for gradual reduction of trainer assistance to promote gameplay independence.

Results: The statistical analysis confirmed the effectiveness of both the WISH and WON protocols. A Wilcoxon signed-rank test on the WISH protocol revealed a statistically significant improvement in understanding instructions for the warm-up (r=0.87; P=.009), the projector imitation (r=0.91; P=.007), and participant exergame performance (r=0.90; P=.03). Within the WON protocol, the Wilcoxon test also showed a significant increase in participant exergame performance (r=0.89; P=.008). Further analysis using Spearman rank-order correlation indicated a very strong association between increased independence and better exergame performance (ρ=0.91; P=.002) and overall gameplay experience (ρ=0.63; P<.05).

Conclusions: This pilot study suggests that the structured WISH and WON training protocols may have the potential to enhance functional autonomy, exergame performance, and overall gameplay experience in individuals with mild ID. The observed improvements indicate that such structured pedagogical approaches could be beneficial for this population. These preliminary findings warrant further investigation through larger-scale, controlled studies to confirm efficacy and explore the transferability of these benefits to broader contexts and other VR exergames.

(JMIR Serious Games 2025;13:e71823) doi:10.2196/71823

KEYWORDS

active video game; exergame; gameplay experience; health promotion; independence; intellectual disabilities; physical activity; physical education; usability; virtual reality

XSL•FO
RenderX

Introduction

Intellectual disabilities (ID) comprise a spectrum of severity levels: mild (F70), moderate (F71), severe (F72), and profound disability (F73), as per the International Classification of Diseases, 10th Revision, by the World Health Organization [1]. Mild severity is characterized by an intelligence quotient (IQ) between 50 and 70, moderate by an IQ between 35 and 49, severe by an IQ between 20 and 34, and profound by an IQ of less than 20 [1]. Individuals with mild ID have difficulties in many cognitive and functional abilities [2]. They may struggle with expressing their thoughts, retaining information, interpreting sensations [3], and engaging in both fine and gross motor movements and motor competence [4]. Despite these obstacles, most people with mild ID can lead productive lives with some assistance. This category accounts for about 85% of those with ID and may need help with some areas of life but remain at least partly independent [5]. People with mild ID tend to be less physically active than their peers without disabilities [6]. This reduced physical activity (PA) can contribute to poor physical health, including low cardiovascular fitness, weak muscles, and an increased risk of overweight and obesity. Given these challenges, developing effective strategies to increase PA participation among individuals with ID is crucial [7]. By promoting regular exercise, we can help improve their overall health and well-being.

Recent initiatives to promote PA among typically developing children have involved incorporating exergames into school physical education (PE) classes [8-10]. These exergames use innovative technology to create immersive environments that require physical gestures and movements to control the gameplay. Additionally, fully immersive virtual reality (VR) exergames transport users into the game world, often through head-mounted displays, eliminating distractions and focusing attention on the game. Exergames incorporating interactive technology and immersive VR experiences have shown promise in promoting PA for people with ID [11-13]. However, before the potential gains of PA exergames that people with ID may derive from exergaming are examined, a teaching methodology should be developed that will combine manual and digital techniques for teaching correct movement patterns [14,15]. Knowing these patterns and practicing them can help improve general physical fitness and functional ability and prevent injury [14]. Consequently, an appropriate step-by-step guide [16] should be developed for introducing people with ID to the VR exergame to achieve independence in gameplay [17]. Recent research suggests that technology systems designed to provide step-by-step guide instructions can be effective tools for teaching and supporting individuals with ID to independently complete tasks and develop motor competence [17]. Building independence in the sphere of exergaming with occasional help from a PE teacher can translate into a sense of independence in taking actions in everyday life for people with ID. Independence, like self-reliance, is a term that involves the ability to take actions to manage one's affairs and provide for oneself [18]. This entails relying on one's own efforts, resources, judgment, and abilities without requiring support from others. PE teachers can aim to guide people with ID toward the most optimal level of independence that is attainable for an individual.

Beyond the specific context of ID, the creation of standardized methods for teaching VR goggle use is a critical and unmet need in the broader fields of PE and rehabilitation. While the potential of VR to enhance engagement in PA and therapeutic exercises is widely recognized [19], its effective application is often limited by the lack of structured instructional protocols. Other studies, such as that by Villada Castillo et al [20], have outlined protocols for introducing exergames for stroke rehabilitation, focusing on a brief introduction, system setup, and free play to gather feedback. Similarly, a methodology proposed by Pirovano et al [21] emphasizes a 3-step design process for therapeutic exergames: identifying exercise requirements, transforming exercises into a virtual environment, and adding game elements for entertainment.

This study builds upon this work by providing a comprehensive teaching methodology that is uniquely structured to facilitate independence. Unlike existing protocols that are often brief or focused on system design and feedback gathering, our approach integrates with a standard PE lesson structure. This division into a warm-up, main part, and cool-down over 45 minutes addresses the practical needs of a school setting, ensuring that the protocols are not only effective but also highly adaptable. By providing a successful protocol for teaching the use of VR technology, our research aims to address this gap. Creating tailored, step-by-step methods is essential for making VR an accessible and effective tool, ultimately fostering greater engagement in activities crucial for well-being across a wide spectrum of users.

Despite evidence that VR exergaming can increase PA, to the best of the authors' knowledge, no previous research has reported on a methodical training protocol for introducing VR gameplay into PE lessons, specifically designed for individuals with mild ID.

The aim of this study was to evaluate the effectiveness of the WISH (Warm-up, Imitation, Settings, Half-hour exergame session) and WON (Warm-up, Objective evaluation, No problem!) training protocols in improving participants' independence, exergame performance, and overall experience playing a VR exergame.

This pilot study aimed to explore whether (1) a structured teaching methodology could be beneficial for implementing VR technology in individuals with mild ID, (2) the proposed protocols could lead to improvements in participants' functional independence and sense of motor competence, and (3) the proposed protocols show potential to serve as a transferable approach for teaching other VR games based on similar movement mechanisms, warranting further investigation.

Methods

Ethical Considerations

The institutional review board at the Karol Marcinkowski Medical University in Poznań approved this study (no 684/23). All procedures were conducted in accordance with the ethical

XSL•FO
RenderX

standards of the institutional and national research committees and with the 1964 Helsinki Declaration, and its later amendments or comparable ethical standards.

All participants were pupils in the special school complex in Poland. Prior to their involvement, comprehensive written informed consent was obtained from the parents or legal guardians of all participants. Additionally, each participant provided verbal and written assent, ensuring their voluntary agreement to take part in the study. Participants did not receive any monetary compensation for their involvement.

To ensure the privacy and confidentiality of participants, all collected data were anonymized and deidentified immediately upon collection. Unique numerical codes were assigned to each participant, and all personally identifying information was stored separately in a secure, password-protected database accessible only to the principal investigator. No personally identifiable data were included in the research database or in any reports or publications.

Individuals visible in the table-of-contents picture are either the authors or pupils who participated in the study. Explicit written consent for the publication of their identifiable images was obtained from these individuals or their legal guardians. These consent forms have been uploaded as supplementary material alongside the manuscript for verification.

Study Design

The WISH protocol (Textbox 1) and the WON protocol (Textbox 2) were developed as a coherent training methodology to gradually increase independence in children with mild ID when using VR exergames. Their names are acronyms that reflect the structure and purpose of each stage. The WISH and WON protocols were initially tested on a sample of healthy adults to identify foundational inconsistencies and refine the core procedures. Subsequently, the protocols were evaluated in a cohort of healthy children aged 15 years to detect and address specific challenges related to a younger population, ensuring the methodology was appropriately adapted for the target group. Finally, before being used, the protocols were evaluated by a school psychologist, a PE teacher from a special school, and an impartial researcher, who gave positive opinions based on their knowledge and the potential usefulness of the protocols.

The WISH (Warm-up, Imitation, Settings, Half-hour exergame session) protocol is a 4-session introductory step. It was created to familiarize participants with the basics of VR exergames, acclimating them to the technology and gameplay mechanics with support. The WON (Warm-up, Objective evaluation, No problem!) protocol is the main one, running from session 5 to 16. It focuses on gradually reducing trainer assistance to promote full independence in gameplay. The name “No problem!” refers to the questionnaire used to measure the level of independence.

Both protocols were developed based on existing physical therapy and special education methodologies, adapted for the specifics of VR, and were validated through the continuous evaluation of participant progress across successive sessions.

Each of the 16 WISH and WON sessions was conducted once every 2 weeks during PE classes at a special school, following a standard lesson structure of a warm-up, a main part, and a cool-down. The duration of a single exergame session (the exergame “OhShape!” was used) was approximately 30 minutes, with each protocol having a defined time structure for its respective stages. Overall graphical information for both protocols is available in Figure 1.

In OhShape!, users engage in a rhythmic fitness game with musical elements, in which the goal is to match their bodies to the shapes that appear on walls. Players stand before colorful walls with silhouettes cut into them, and their task is to position themselves correctly to pass through. The game also features additional elements, such as special walls that must be avoided by squatting or broken with the hands, and coins that are collected by performing movement sequences. The game engages the entire body, requiring players to perform squats, dodges, and dynamic poses to the rhythm of the music. This effectively combines fun with PA, offering a compelling fitness experience. A video of the “I” step with OhShape! exergame gameplay is provided in Multimedia Appendix 1.

XSL•FO
RenderX

Textbox 1. WISH (Warm-up, Imitation, Settings, Half-hour exergame session): a 4-step protocol (sessions 1-4).

Step 1. "W" (10 minutes)

- The warm-up included movements inspired directly by in-game elements, designed to prepare participants for the main part of the training. The exercises performed were:
  - arm swings (in-game: catching coins)
  - hand movements imitating boxing punches (in-game: smashing walls)
  - squats (in-game: avoiding walls that appear above the head)
  - step-and-reach (in-game: avoiding walls)
  - a "mirror" game, consisting of imitating movements performed by the trainer

Step 2. "I" (3 minutes)

- Imitating the exergame by using a projector and video recorded from the perspective of a participant in virtual reality (VR) goggles:
  - the movements learned during the warm-up are used
  - all participants position themselves in front of the image displayed by the projector and "play" without using VR goggles
  - the trainer verifies the physical abilities of the participants and their understanding of the instructions. He reminds them aloud what to do with the appropriate element of the exergame (eg, avoid the transparent wall—do a squat; break the red wall—use boxing with your hands).

Step 3. "S" (2 minutes)

- Setting up the VR goggles and the exergame zone:
  - marking the goggles with the participant symbol (A-H) and adjusting the goggles to the participants (setting the distance between the lenses, fastening straps)
  - setting a safe zone for the participant in which they could move freely
  - connecting the view from the goggles to the app on the trainer's phone, which allowed for a real-time view of the exergame

Step 4. "H" (30 minutes)

- Half-hour exergame session:
  - launching the game remotely for participants (the "A" button on the controller allowed voice launch of game elements through a voice message, eg, "launch game," "again," and "menu")
  - the game was launched at the lowest level of difficulty ("try out")
  - game control by viewing the progress of participants in the gym and from their perspective in the goggles through an app on the phone
  - overall scores were recorded by the trainer (stars)

XSL•FO
RenderX

Textbox 2. WON (Warm-up, Objective evaluation, No problem!): a 3-step protocol (sessions 5-16).

Step 1. "W" (10 minutes)

- The warm-up included more advanced movements directly inspired by in-game elements, designed to prepare participants for the main part of the training. The exercises performed were:
  - arm swings (in-game: catching coins)
  - both hands movements imitating boxing punches (in-game: smashing walls)
  - squats (in-game: avoiding walls that appear above the head)
  - step-and-reach (in-game: avoiding walls)
  - combo movement combining squats and step-and-reach exercises
  - a "mirror" game, consisting of imitating movements performed by the trainer

Step 2. "O" (30 minutes)

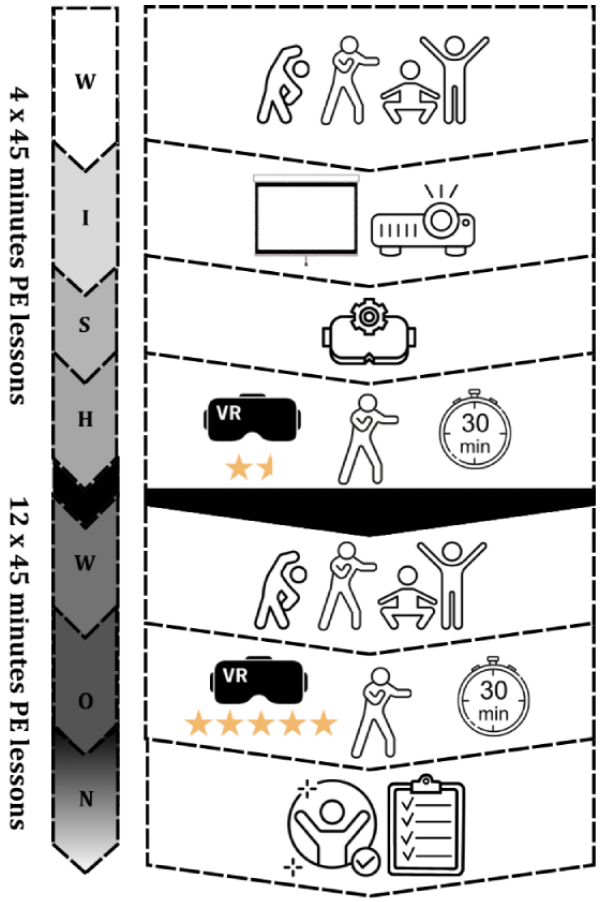
- Objective evaluation of exergame performance:
  - based on the overall score (stars) obtained in the earlier session, the level of difficulty of the game on the day of the session was selected for each participant separately
  - participants were not helped in the game unless they asked for it
  - the exergame was not launched for the participant, they launched it themselves

Step 3. "N" (5 minutes)

- No problem! questionnaire:
  - each participant completed the questionnaire and interview sessions independently

XSL•FO
RenderX

Figure 1. Overall design of the WISH (Warm-up, Imitation, Settings, Half-hour exergame session) and WON (Warm-up, Objective evaluation, No problem!) structured teaching protocols, illustrating their sequential intervention delivery over 16 biweekly sessions in a pilot study involving 8 adolescents with mild intellectual disabilities. The intervention was conducted at a special school in Poland from October 2023 to May 2024, spanning an 8-month period within the 2023-2024 academic year. PE: physical education; VR: virtual reality.



Study Sample

All participants (N=8) were approached on the basis that they were pupils of a special school and had a mild intellectual disability (based on data provided by the diagnosis carried out in psychological and pedagogical counselling centers, aimed at recognizing the disability, determining the degree of ID, and issuing a decision on the need for special education). This was a convenience sample, comprising all eligible students from 2 distinct classes at the specific special school who provided parental/guardian informed consent and their own assent to participate. These 2 classes (totaling 8 pupils) represented the entire available cohort of students within the specified age range and disability level who typically participate together in PE lessons at this institution. Due to the pilot nature of the study and the specificity of the population, formal sample size calculations were not performed. The PE teacher and psychologist of their school assessed the level of physical fitness and mental health as good, enabling participation in the study. All participants exercise regularly in PE lessons in their special school.

A total of 8 participants (male: n=3; female: n=5) aged between 15 and 20 years (mean age 17.63 years, SD 1.77) took part in the study.

Inclusion criteria were as follows: no previous VR experience, mild ID, current pupils of a special school, teenagers or young adults aged 15-25 years, wearing or not wearing corrective glasses (a special overlay for glasses was applied), right- or left-handed, no use of psychotropic medicines, absence of long-term or recent injuries, and good physical fitness and mental health to participate in PE lessons (confirmed by a statement from a school psychologist and the current PE teacher of the special school).

Exclusion criteria were as follows: previous VR experience, no mild ID, former pupils of a special school, age greater than 25 years, history of eye surgery, use of psychotropic medicines, long-term or recent injuries, and below-average physical fitness or mental health to participate in PE lessons (as confirmed by a statement from a psychologist and the current PE teacher of the special school).

Measures

Technology

A VR Meta Quest 2 128 GB system (Meta Platforms Technologies Ireland Limited, 2023) was used as the VR headset in this study. This system includes a wireless headset for viewing and exergaming, as well as 2 hand controllers. To ensure optimal performance, a minimum room size of 2.0 × 2.0 meters is recommended. The VR headset features a 5.7-inch OLED display with a resolution of 1832 × 1920 pixels per eye and a refresh rate of 60-90 Hz. The exergame used in this study, OhShape!, was downloaded from the Meta Quest Store app.

XSL•FO
RenderX

Procedure

The study was conducted over 16 sessions held as part of PE lessons at a special school. The entire procedure was based on two protocols, WISH and WON, which were designed to gradually increase participants' independence in using a VR exergame.

The WISH protocol comprised 4 steps aimed at developing correct movement technique and ensuring safety and comprehension of the game's rules. In the initial "W" (Warm-up) and "I" (Imitation), the trainer assessed whether participants understood the instructions and performed the movements correctly, recording the results on a task sheet (task sheet with 3-point scale: "YES," 2 points, indicated complete understanding and correct execution; "PARTIALLY," 1 point, indicated partial understanding or execution requiring minor cues; and "NO," 0 points, indicated no understanding or incorrect execution). These scores were based on predefined, detailed criteria, ensuring consistent assessment across sessions and participants. During the "S" (Settings), the trainer ensured participant safety by setting up the VR goggles and play zones, documenting the settings to maintain consistency between sessions. The final stage, "H" (Half-hour exergame session), involved a 30-minute exergame gameplay segment during which the trainer remotely launched the game and monitored participant progress, recording their score (stars; exergame performance). The WISH protocol lasted 4 sessions, with the objective for each participant to achieve full understanding of the instructions ("W" and "I" stages) and earn at least 1.5 stars in the game.

The WON protocol was implemented following the 4 WISH sessions, with the primary objective of developing full independence in gameplay. The first step, "W" (Warm-up), continued to support correct movement technique. In the subsequent "O" (Objective evaluation) stage, participants transitioned to independent play, where they were responsible for launching the exergame themselves. The trainer adjusted the difficulty level individually for each participant based on their exergame performance (stars) from previous sessions and provided assistance only when requested. Participant progress or regression was monitored by recording their overall exergame performance score on a task sheet. The final step, "N" (No problem! questionnaires), involved a mixed methods assessment using the No problem! independence questionnaire, which contained 7 functional levels, and a separate gameplay experience questionnaire with a 3-point Likert scale.

Independence and Gameplay Experience Questionnaires

The independence questionnaire, titled No problem!, was combined with interviews and used an adapted 7-level functional independence scale, directly derived from the FIM levels of the Functional Independence Measure (FIM) [22,23]. This adapted scale ranged from total assistance to total independence. Data on independence level were collected from training sessions 5 to 16. For clarity and transparency, the specific definitions of these adapted 7 functional levels are provided in Multimedia Appendix 2.

In this study, the definition of the 7 functional levels and corresponding 7-point scale was described as follows. A score of 7 points indicated total independence, where all tasks were completed without the need for assistance from third parties. In the interview, participants asked about problems during the exergame played used the phrase, "I had no problem."

A score of 6 points reflected incomplete independence. The activity required longer (than expected) time to complete the gameplay, but the rest of the gameplay was correctly followed. In the interview, participants asked about problems during the exergame played used, for example, the phrase, "I had some problems, but I didn't need supervision."

A score of 5 points referred to supervision. The participant required no more assistance than standing next to a third person, including assignment, guidance, or persuasion to perform the activity, but no physical contact. In the interview, participants asked about problems during the exergame played used, for example, the phrase, "I had a few problems, and I needed a voice support."

A score of 4 points indicated minimal assistance. Regarding physical contact, the participants required no more assistance than touching and performing most (over 75%) of the activity. In the interview, participants asked about problems during the exergame played used, for example, the phrase, "I had a few problems, and I needed minimal physical support."

A score of 3 points represented moderate assistance, where the participant required more assistance or performed half or more of the activity/effort (50%-74%). In the interview, participants asked about problems during the exergame played used, for example, the phrase, "I had a few problems, and I needed moderate physical support."

A score of 2 points reflected maximum assistance. The participant performed half or less of the specified activity (25%-49%), but at least 25%. In the interview, participants asked about problems during the exergame played used, for example, the phrase, "I had a lot of problems, and I needed a moderate physical support."

A score of 1 point indicated total assistance. The participant could perform little, if any, of the planned activity/effort (less than 25%). In the interview, participants asked about problems during the exergame played used, for example, the phrase, "I had a lot of problems, and I needed total assistance."

Self-reported answers on the gameplay experience questionnaire [24] were assessed using a 3-point Likert scale. Respondents were asked to indicate the extent to which they agreed with each statement by selecting one of the following options: "A Lot," "A Little," or "No." Because adults and children with ID often struggle with text comprehension, simplifying assessment scales is a frequent practice [25-27]. To improve understanding of the Likert-type items in this study, 2 strategies were used: each of the 3 response options was paired with a corresponding emoticon, and the researcher also presented the questions orally. The emoticon card is shown in Figure 2. For question number 5, different emoticons were used due to the reverse structure of the question. To assess usability in the context of VR, the focus

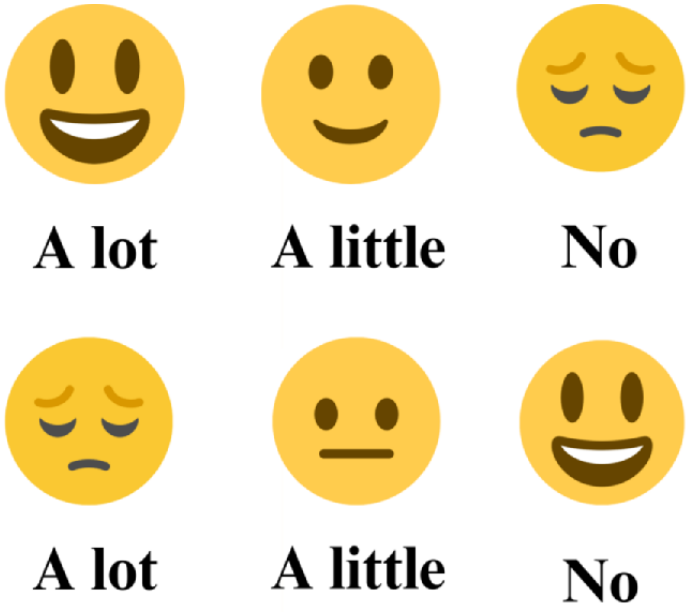
XSL•FO
RenderX

should be on areas such as ease to use, feeling safe, and cumbersome handling [28,29].

The gameplay experience questionnaire included the following seven items: question 1, “Did you get used to the game quickly?”; question 2, “Were the controls easy to use?”; question 3, “Did the things you saw look real?”; question 4, “Were the VR goggles comfortable?”; question 5, “Were you worried about putting on the VR goggles?”; question 6, “Did it feel like you were in control?”; and question 7, “Did the way you moved look real?”

The gameplay experience questionnaire included seven items: getting used to the game (question 1), ease of controls (question 2), visual realism (question 3), comfort of VR goggles (question 4), concerns of VR goggles (question 5), sense of control (question 6) and realism of movements (question 7). To derive an overall gameplay experience score, the numerical ratings (1=”No,” 2=”A Little,” and 3=”A Lot”) from the 7 items were summed, with the score for question number 5 (“Were you worried about putting on the VR goggles?”) being reverse-coded prior to summation (ie, “A Lot” mapped to 1, “A Little” to 2, and “No” to 3), ensuring that a higher total score consistently reflected a more positive overall experience.

Figure 2. Emoticon card used as a visual aid to simplifying Likert-type assessment scales in a pilot study involving 8 adolescents with mild intellectual disabilities (October 2023 to May 2024). Each of the 3 response options was paired with a corresponding emoticon to enhance comprehension, with a modified set of emoticons used for the reversely structured question number 5.



Statistical Analysis

Calculations were performed using Statistica 13.0 software (TIBCO Software Inc). To assess whether the WISH and WON protocols induced statistically significant improvements: Wilcoxon signed-rank tests were used to analyze changes in independence levels (from the No problem! questionnaire) and exergame performance (number of stars earned); McNemar tests were used for analyzing changes in categorical data from the gameplay experience questionnaire; chi-square tests were used for other categorical comparisons, where appropriate; and Spearman rank-order correlations were performed to investigate relationships between participants’ independence levels (data from the No problem! questionnaire) and their exergame performance (number of stars earned) and overall gameplay experience (the final composite score from the gameplay experience questionnaire).

Results

WISH

Understanding the task in” W” step was assessed by the trainer after the warm-up. The data for each session in WISH protocol are presented in Table 1. During session 1, the results were as follows: median 1.0 (IQR 0.25; 0.75–1.00) point. This means that there were participants (6/8, 75%) who partially understood the instructions (and therefore received 1.00 point), but there were also participants (2/8, 25%) who did not understand the instructions at all (0.00 point). The score increased from session 1 to session 4. In session 4, all participants scored 2.00 points (8/8, 100%). A Wilcoxon signed-rank test was performed to compare the results of the participants in the WISH protocol between session 1 and session 4. The *P* value was .009 and the effect size was 0.87 for warm-up instruction understanding.

Understanding the task in “I” step was assessed by the trainer after completing the session in front of the projector. The data for each session in the WISH protocol are presented in Table 2. During session 1, the results were as follows: median 0.50 (IQR 1.00; 0.00–1.00), range (minimum-maximum) 0.00-1.00

XSL•FO
RenderX

point. This means that there were participants (4/8, 50%) who partially understood the instructions (and therefore received 1.00 point), but there were also participants (4/8, 50%) who did not understand the instructions at all (0.00 point). The mean score increased from session 1 to session 4. In session 4, all participants scored 2.00 points.

The star achievement in the "H" step was assessed by the trainer after VR exergame gameplay. The data for each session in the WISH protocol are presented in Table 3. During session 1, the results were as follows: median 0.75 (IQR 0.50; 0.50–1.00) points. This means that there were participants (4/8, 50%) who scored 1 full star (1.00 point), but there were also participants (4/8, 50%) who scored only half a star (0.50 point). The stars provided information about the result achieved in the VR exergame gameplay—the higher the score, the more stars summarizing the gameplay experience. The median score increased from session 1 to session 4. In session 4, the results were as follows:median 1.00 (IQR 0.25; 1.00–1.25) points, range (minimum-maximum) 0.50-2.00 points. One participant scored 2 points, one participant scored 1.5 points, 5 participants scored 1 point, and only one participant who completed the protocol scored 0.5 points.

Table 1. Trainer's assessment of participants' understanding of warm-up instructions during the "W" (Warm-up) step of the WISH (Warm-up, Imitation, Settings, Half-hour exergame session) protocol across sessions 1-4 in a pilot study involving adolescents with mild intellectual disabilities (N=8). Data collected from a special school in Poland (October 2023 to May 2024) are presented as median (IQR; Q1–Q3), mean (SD), and range for each session. The Wilcoxon signed-rank test performed to compare session 1 and session 4 revealed a statistically significant difference (*P*=.009) with a corresponding large effect size (r=0.87).

Variable	Session 1	Session 2	Session 3	Session 4
Understanding of warm-up instructions, points[a]				
Median (IQR; Q1–Q3)	1.00 (0.25; 0.75–1.00)	1.00 (0.00; 1.00–1.00)	2.00 (0.25; 1.75–2.00)	2.00 (0.00; 2.00–2.00)
Mean (SD)	0.75 (0.46)	1.13 (0.35)	1.75 (0.46)	2.00 (0.00)
Range	0.00-1.00	1.00-2.00	1.00-2.00	2.00-2.00
Number of participants achieving score, n (%)				
0.00 points (No)	2 (25)	0 (0)	0 (0)	0 (0)
1.00 points (Partially)	6 (75)	7 (87)	2 (25)	0 (0)
2.00 points (Yes)	0 (0)	1 (13)	6 (75)	8 (100)

[a]It indicates that the trainer recorded on the task sheet whether each participant understood the instructions and performed the warm-up correctly (yes=2 points, partially=1 point, and no=0 points).

Table 2. Trainer's assessment of participants' understanding of imitation instructions during the "I" (Imitation) step of the WISH (Warm-up, Imitation, Settings, Half-hour exergame session) protocol across sessions 1-4 in a pilot study involving adolescents with mild intellectual disabilities (N=8). Data collected from a special school in Poland (October 2023 to May 2024) are presented as median (IQR; Q1–Q3), mean (SD), and range for each session. The Wilcoxon signed-rank test performed to compare session 1 and session 4 revealed a statistically significant difference *P*=.007 with a corresponding large effect size (r=0.91).

Variable	Session 1	Session 2	Session 3	Session 4
Understanding of imitation instructions, points[a]				
Median (IQR; Q1–Q3)	0.50 (1.00; 0.00–1.00)	1.00 (0.50; 1.00–1.50)	2.00 (0.00; 2.00–2.00)	2.00 (0.00; 2.00–2.00)
Mean (SD)	0.50 (0.53)	1.25 (0.46)	1.86 (0.35)	2.00 (0.00)
Range	0.00-1.00	1.00-2.00	1.00-2.00	2.00-2.00
Number of participants achieving score, n (%)				
0.00 points (No)	4 (50)	0 (0)	0 (0)	0 (0)
1.00 points (Partially)	4 (50)	6 (75)	1 (13)	0 (0)
2.00 points (Yes)	0 (0)	2 (25)	7 (87)	8 (100)

[a]It indicates that the trainer recorded on the task sheet whether each participant understood the instructions and performed the warm-up correctly (yes=2 points, partially=1 point, and no=0 points).

XSL•FO
RenderX

Table 3. Trainer's assessment of participants' exergame performance (star achievement) during the "H" (Half-hour exergame session) step of the WISH (Warm-up, Imitation, Settings, Half-hour exergame session) protocol across sessions 1-4 in a pilot study involving adolescents with mild intellectual disabilities (N=8). Data collected from a special school in Poland (October 2023 to May 2024) are presented as median (IQR; Q1–Q3), mean (SD), and range for each session. The Wilcoxon signed-rank test performed to compare session 1 and session 4 revealed a statistically significant difference (*P*=.03) with a corresponding large effect size (r=0.90).

Variable	Session 1	Session 2	Session 3	Session 4
Exergame performance (stars), points[a]				
Median (IQR; Q1–Q3)	0.75 (0.50; 0.50–1.00)	1.00 (0.50; 0.75–1.25)	1.00 (0.25; 0.75–1.00)	1.00 (0.25; 1.00–1.25)
Mean (SD)	0.75 (0.27)	1.06 (0.52)	1.00 (0.53)	1.13 (0.44)
Range	0.50-1.00	0.50-2.00	0.50-2.00	0.50-2.00
Number of participants achieving score, n (%)				
0.00 points	0 (0)	0 (0)	0 (0)	0 (0)
0.50 points	4 (50)	2 (25)	2 (25)	1 (13)
1.00 points	4 (50)	4 (50)	5 (63)	5 (63)
1.50 points	0 (0)	1 (13)	0 (0)	1 (13)
2.00 points	0 (0)	1 (13)	1 (13)	1 (13)

[a]It indicates that the stars achievement system ranged from 0 to 5 (including half points), reflecting the result achieved in the VR exergame gameplay (higher scores indicate more stars and better performance).

WON

Understanding the task in the "W" step was assessed by the trainer after the warm-up. The data are presented in Table 4. During session 5, the results were as follows: median 2.00 (IQR 1.00; 1.00–2.00) points, range (minimum-maximum) 1.00-2.00 points. This means that there were still participants (3/8, 37.50%) who partially understood the instructions (and therefore received 1.00 point), but there were no participants (n=0) who did not understand the instructions at all. The median score increased from session 5 to session 9. There were no changes from session 9 to session 16. In session 9, all participants scored 2.00 points (8/8, 100%).

The star achievement in the "O" step was assessed by the trainer after the VR exergame gameplay. The data for each session in WON protocol are presented in Table 5. During session 5, the results were as follows: median 1.00 (IQR 0.50; 1.00–1.50) points, range (minimum-maximum) 0.50-2.00 points. This means that there were participants (2/8, 25%) who scored 2 full stars (2.00 points), but there was also one participant who scored only half a star (0.50 points). The stars provided information about the result achieved in the VR exergame gameplay (the higher the score, the more stars that summarized the gameplay experience). The mean score increased from session 5 to session 16. In session 16, the results were as follows: median 4.00 (IQR 0.75; 3.75–4.50) points, range (minimum-maximum) 3.50-5.00 points. This means that there was one participant who scored 5 full stars (5.00 points), which was the maximum that could be obtained. Two participants received 4.5 points, 3 participants received 4 stars, and 2 participants received 3.5 stars. This was the score with which they concluded the study.

The gameplay experience questionnaire shows the fluctuation of participants' VR exergame gameplay experience between the fourth and last sessions. The questionnaire was divided into 7 sections: getting used to the game (question 1), ease of controls (question 2), visual realism (question 3), comfort of VR goggles (question 4), concerns of VR goggles (question 5), sense of control (question 6), and realism of movements (question 7). The data for each item by session are presented in Table 6.

A Wilcoxon signed-rank test was performed to compare participants' scores on understanding instructions within the WON protocol between session 5 and session 16. The analysis revealed that the observed improvement was statistically significant (*P*=.01), with a very large effect size (*r*=0.87). This confirms that the protocols effectively enhanced participants' independence and ability to navigate the virtual environment. To assess the influence of the protocols on gameplay experience, a McNemar test was performed on questionnaire data, comparing responses from session 5 and session 16. The analysis showed a statistically significant improvement in several aspects of perceived game usability. Participants adapted to the game more quickly (*P*=.01), found the controls easier to use (*P*=.01), and felt more in control of their virtual avatar (*P*=.01). Additionally, there was a significant reduction in the level of worry associated with using VR goggles (*P*=.02) and a significant improvement in their perceived comfort (*P*=.01). Conversely, changes in perceived visual realism (*P*=.16) and movement realism (*P*=.32) were not statistically significant. These findings provide crucial quantitative evidence supporting the positive impact of the protocols on user experience.

The No problem! questionnaire was an independence gameplay questionnaire combined with interviews and included 7 functional levels of independence, ranging from total assistance to total independence. Data on independence in gameplay level were collected from sessions 5 to 16 (the WON protocol, "N" step) and are presented in Table 7.

A Wilcoxon signed-rank test was conducted to assess the impact of the WON protocol on participants' level of independence in gameplay by comparing data from session 5 and session 16.

XSL•FO
RenderX

The analysis revealed that the observed improvement in independence levels was statistically significant (*P*=.008). This value is well below the standard .05 significance threshold, confirming that the increase in independence was not due to random chance. Furthermore, the effect size (*r*=0.89) indicates a very large effect of the intervention, providing strong quantitative evidence of its significant impact on developing participants' autonomy.

To investigate the relationship between the improvement in participant independence and their gameplay experience, a Spearman rank-order correlation analysis was conducted. The analysis compared the final scores from the No problem! independence questionnaire with the final scores from the gameplay experience questionnaire. The results revealed a very strong and statistically significant correlation between the two variables.

As shown in Table 8, the high Spearman rank-order correlation coefficients for most items indicate that as participants' independence increased, so did their positive assessment of the VR exergame. This provides strong evidence that the protocol not only enhanced autonomy but also significantly improved the overall user experience. Furthermore, a Spearman rank-order correlation between the participants' independence levels and the aggregated overall gameplay experience score revealed a moderately strong and statistically significant positive association (ρ=0.63; *P*<.05).

A Spearman rank-order correlation was conducted to investigate the relationship between participants' independence levels (data from the No problem! questionnaire) and their exergame performance (number of stars earned). This analysis used data from the final session (session 16) for all 8 participants. The analysis revealed a statistically significant positive correlation (*P*=.002). The high correlation coefficient (ρ=0.91) indicates that as participants became more independent during the exergaming session, their exergame achievements also consistently improved.

Table 4. Trainer's assessment of participants' understanding of warm-up instructions during the "W" (Warm-up) step of the WON (Warm-up, Objective evaluation, No problem!) protocol across sessions 5-16 in a pilot study involving adolescents with mild intellectual disabilities (N=8). Data collected from a special school in Poland (October 2023 to May 2024) are presented as median (IQR; Q1–Q3), mean (SD), and range for each session, along with the distribution of participants achieving specific points. The Wilcoxon signed-rank test performed to compare session 5 and session 16 revealed a statistically significant difference (*P*=.06) with a corresponding large effect size (r=0.87).

Variable	Session 5	Sessions 9-16
Understanding of warm-up instructions, points[a]		
Median (IQR; Q1–Q3)	2.00 (1.00; 1.00–2.00)	2.00 (0.00; 2.00–2.00)
Mean (SD)	1.63 (0.48)	2.00 (0.00)
Range (minimum-maximum)	1.00-2.00	2.00-2.00
Number of participants achieving points, n (%)		
0.00 points (No)	0 (0)	0 (0)
1.00 points (Partially)	3 (37.5)	0 (0)
2.00 points (Yes)	5 (62.5)	8 (100)

[a]It indicates that the trainer recorded on the task sheet whether each participant understood the instructions and performed the warm-up correctly (yes=2 points, partially=1 point, and no= 0 points).

XSL•FO
RenderX

Table 5. Trainer's assessment of participants' exergame performance during the "O" (Objective evaluation) step of the WON (Warm-up, Objective evaluation, No problem!) protocol across sessions 5-16 in a pilot study involving adolescents with mild intellectual disabilities (N=8). Data collected from a special school in Poland (October 2023 to May 2024) are presented as median (IQR; Q1–Q3), mean (SD), and range for each session, along with the distribution of participants achieving specific points. The Wilcoxon signed-rank test performed to compare session 5 and session 16 revealed a statistically significant difference (*P*=.008) with a corresponding large effect size (r=0.89).

Variable	Session 5	Session 16
Exergame performance (stars), points[a]		
Median (IQR; Q1–Q3)	1.00 (0.50; 1.00–1.50)	4.00 (0.75; 3.75–4.50)
Mean (SD)	1.19 (0.53)	4.13 (0.52)
Range	0.50-2.00	3.50-5.00
Number of participants achieving points, n (%)		
0.50 points	1 (13)	0 (0)
1.00 points	5 (62)	0 (0)
1.50 points	0 (0)	0 (0)
2.00 points	2 (25)	0 (0)
2.50 points	0 (0)	0 (0)
3.00 points	0 (0)	0 (0)
3.50 points	0 (0)	2 (25)
4.00 points	0 (0)	3 (38)
4.50 points	0 (0)	2 (25)
5.00 points	0 (0)	1 (13)

[a]It indicates that the stars achievement system ranged from 0 to 5 (including half points), reflecting the result achieved in the VR exergame gameplay (higher scores indicate more stars and better performance).

XSL•FO
RenderX

Table 6. Participants' self-reported gameplay experience during the "N" (No problem! questionnaires) step of the WON (Warm-up, Objective evaluation, No problem!) protocol, comparing session 5 and session 16, in a pilot study involving adolescents with mild intellectual disabilities (N=8). Data collected from a special school in Poland (October 2023 to May 2024) are presented as n (%). Results of the Wilcoxon signed-rank test comparing session 5 and session 16 are also shown.

Questions and responses	Session 5, n (%)	Session 16, n (%)	Comparison (session 5 vs session 16), *P* value
Question 1: Did you get used to the game quickly?			.01[a]
A lot	0 (0)	5 (62.5)	
A little	2 (25)	3 (37.5)	
No	6 (75)	0 (0)	
Question 2: Were the controls easy to use?			.01[a]
A lot	0 (0)	6 (75)	
A little	2 (25)	2 (25)	
No	6 (75)	0 (0)	
Question 3: Did the things you saw look real?			.15[b]
A lot	1 (12.5)	3 (37.5)	
A little	5 (62.5)	5 (62.5)	
No	2 (25)	0 (0)	
Question 4: Was the VR[c] goggles comfortable?			01[a]
A lot	1 (12.5)	7 (87.5)	.
A little	1 (12.5)	1 (12.5)	
No	6 (75)	0 (0)	
Question 5: Were you worried about putting on the VR goggles?			.02[a]
A lot	6 (75)	0 (0)	
A little	1 (12.5)	2 (25)	
No	1 (12.5)	6 (75)	
Question 6: Did it feel like you were in control?			.01[a]
A lot	0 (0)	7 (87.5)	
A little	2 (25)	1 (12.5)	
No	6 (75)	0 (0)	
Question 7: Did the way you moved look real?			.32[b]
A lot	2 (25)	7 (87.5)	
A little	5 (62.5)	1 (12.5)	
No	1 (12.5)	0 (0)	

[a]Statistically significant at $P<.05$.

[b]Not statistically significant at $P<.05$.

[c]VR: virtual reality.

XSL•FO
RenderX

Table 7. Trainer's assessment of participants' independence in gameplay during the "N" (NO problem! questionnaires) step of the WON (Warm-up, Objective evaluation, No problem!) protocol across sessions 5-16 in a pilot study involving adolescents with mild intellectual disabilities (N=8). Data collected from a special school in Poland (October 2023 to May 2024) are presented as median (IQR; Q1–Q3), mean (SD), and range for each session. The Wilcoxon signed-rank test performed to compare session 5 and session 16 revealed a statistically significant difference (*P*=.008) with a corresponding large effect size (r=0.89).

Variable	Session 5	Session 16
Independence in gameplay level, points[a]		
Median (IQR; Q1–Q3)	4.00 (0.00)	7.00 (1.00)
Mean (SD)	4.13 (0.35)	6.63 (0.52)
Range	4.00-5.00	6.00-7.00
Number of participants achieving points, n (%)		
1.00 points (total assistance)	0 (0)	0 (0)
2.00 points (maximum assistance)	0 (0)	0 (0)
3.00 points (moderate assistance)	0 (0)	0 (0)
4.00 points (minimal assistance)	7 (87.5)	0 (0)
5.00 points (supervision)	1 (12.5)	0 (0)
6.00 points (incomplete independence)	0 (0)	7 (87.5)
7.00 points (total independence)	0 (0)	1 (12.5)

[a]It refers to a 7-point scale (1-7) defined as follows: 7 points=total independence, 6 points=incomplete independence, 5 points=supervision, 4 points=minimal assistance, 3 points=moderate assistance, 2 points=maximum assistance, and 1 point=total assistance.

Table 8. Spearman rank-order correlation coefficients examining the relationship between participants' final independence levels and their final gameplay experience, in a pilot study involving adolescents with mild intellectual disabilities (N=8). Data collected from a special school in Poland (October 2023-May 2024) are shown. Statistical significance was assessed using the Spearman rank-order correlation test.

Gameplay experience questionnaire item	Spearman correlation coefficient, ρ	*P* value
Question 1: Did you get used to the game quickly?	0.94	.001[a]
Question 2: Were the controls easy to use?	0.91	.002[a]
Question 3: Did the things you saw look real?	0.69	.058[b]
Question 4: Were the VR goggles comfortable?	0.97	.001[a]
Question 5: Were you worried about putting on the VR goggles?	0.86	.006[a]
Question 6: Did it feel like you were in control?	0.94	.001[a]
Question 7: Did the way you moved look real?	0.85	.008[a]

[a]Statistically significant at *P*<.05.

[b]Not statistically significant at *P*<.05.

Discussion

Principal Results

This pilot study aimed to evaluate the effectiveness of the WISH and WON training protocols in enhancing the independence, performance, and overall gameplay experience of individuals with mild ID using a VR exergame. The findings suggest that a structured training methodology may be highly beneficial for this population. The results from the WISH protocol indicate its success in establishing foundational skills. The statistically significant improvements in participants' understanding of warm-up and imitation instructions (*P*=.009 and *P*=.007, respectively) with very large effect sizes (*r*=0.87 and *r*=0.91) suggest that the initial 4 sessions helped acclimate participants to the game's core mechanics and rules. This structured introduction appears to be a crucial prerequisite for successful, independent engagement. This is further supported by the significant increase in in-game exergame performance (*P*=.03). The protocol's emphasis on developing correct movement technique and preventing unhealthy posture (eg, slouching to catch a coin) appeared to be critical to ensuring both safety and proper skill acquisition.

The WON protocol appeared to successfully build upon this foundation to foster independence. We observed a statistically significant increase in participants' exergame performance from session 5 to session 16 (*P*=.008), with a large effect size (*r*=0.89). This finding suggests that the protocol's gradual reduction of trainer assistance may have effectively facilitated

XSL•FO
RenderX

a shift from supported engagement to autonomous performance. The correlation between increased independence and better exergame performance (ρ=0.91; *P*=.002) provides quantitative evidence that developing autonomy was associated with better performance outcomes. The clear progression in independence, from requiring minimal assistance in the initial sessions to achieving complete independence in the final sessions, underscores the protocol's potential in promoting self-reliance. This is especially important as the ability to manage one's own affairs and rely on personal effort and judgment is a key component of real-life independence.

The protocols also appeared to have a profound positive impact on the subjective user experience. A strong correlation was found between increased independence and a more positive overall gameplay experience (ρ=0.63; *P*<.05). This may indicate that the protocols not only improved objective skills but also built confidence and comfort in the virtual environment. This is consistent with previous research suggesting that adapting gameplay to individual capabilities is essential for a positive experience [11,30-32]. Our findings suggest that the OhShape! VR exergame, with its simple controls and full-body movement requirements, combined with the structured protocols, was well-suited for this population, potentially leading to significant improvements in perceived comfort, ease of controls, and a sense of control over the virtual avatar.

These findings have important implications for inclusive PE. The structured WISH and WON protocols may offer an evidence-based framework for integrating VR exergames into PE lessons for individuals with mild ID. By prioritizing procedural learning and independence, these protocols could help this population overcome a key barrier to PA.

Comparison With Previous Work

While a growing body of literature supports the use of VR exergames to promote PA and motor skills among individuals with ID, there remains a significant gap in research concerning structured, pedagogical methodologies for technology adoption. This study contributes to the field by not only reinforcing the potential benefits of VR exergaming but, more importantly, by proposing a concrete, evidence-based training framework for achieving independence.

Our findings appear consistent with previous work demonstrating that VR exergames are a highly engaging and motivating tool for individuals with ID, capable of increasing PA participation [33-35]. However, a key novelty of our research lies in the systematic approach taken. Most existing studies have focused on the outcomes of VR use, such as changes in physical fitness or energy expenditure, without detailing the structured process of teaching the technology itself. The WISH and WON protocols seek to explore ways to address this gap by offering a replicable, step-by-step methodology that aims to tackle the foundational challenges of task comprehension and motor learning in this population.

The systematic naming of our training methodologies as the WISH and WON protocols was a deliberate choice intended to enhance their clarity, memorability, and potential for widespread adoption. The use of acronyms as a tool for structuring complex information is a well-established practice across scientific disciplines, particularly in medicine and education, where it facilitates the memorization and application of multistep procedures. For example, the patient handover protocol described by Hanke et al [36] uses the ISBAR (Introduction, Situation, Background, Assessment, and Actions) acronym to ensure a structured and comprehensive communication process. Similarly, the names of our protocols serve as easy-to-recall frameworks for teachers or therapists, summarizing the core steps of each training phase. Given that the target population—individuals with mild ID—benefits significantly from clear, concise, and repeatable instructions, the acronyms WISH and WON provide a simple and effective mnemonic device.

In summary, while the therapeutic potential of VR exergaming for individuals with ID is well documented [37], this pilot study aims to provide a unique and essential contribution by offering a transferable scientific pattern. The WISH and WON protocols appear to serve as a practical teaching model that could be applied to various VR exergames, addressing the critical need for structured educational frameworks in this domain. Our findings may highlight that beyond simply providing access to technology, a well-designed pedagogical approach could be essential for cultivating true independence and maximizing the benefits of VR-based interventions.

Limitations

It is important to acknowledge the limitations of this study. First, the small sample size (N=8) inherently limits the generalizability of our findings. This small sample may also have influenced the statistical power, potentially affecting the significance of some results and making them not fully representative of the wider population of individuals with mild ID. As a pilot study, our primary aim was to test the feasibility and initial effectiveness of our protocols in a targeted group; thus, a larger and more representative sample was beyond the scope of this initial investigation. To address this in future research, replication with substantially larger sample sizes will be crucial, potentially through multisite collaborations, to increase generalizability and statistical robustness.

Second, the absence of a control group in this design is a significant limitation. This design prevents us from definitively isolating the effects of the WISH and WON protocols from other factors, such as natural learning or repeated practice. Furthermore, because participants were aware of their involvement in a novel intervention, the Hawthorne effect [38,39]—in which individuals modify their behavior in response to being observed—cannot be entirely ruled out as a contributing factor to the observed improvements. Consequently, while promising improvements were observed, we cannot conclusively state that they were solely caused by our intervention. This limitation was deemed acceptable for this pilot study, in which the primary focus was on assessing the feasibility and initial efficacy of the protocols in a small, targeted group rather than on direct comparison with traditional training methods. However, it is worth noting that the intervention sessions were integrated into the standard PE curriculum, occurring approximately once every 2 weeks from October to May and

XSL•FO
RenderX

interspersed with regular PE lessons. This intermittent schedule, rather than continuous daily exposure, may have partially mitigated the extent of natural learning or practice effects that might occur with more frequent, concentrated interventions. Despite these partial mitigations, our findings should be interpreted with caution regarding causality. To improve upon this, future research should implement randomized controlled trial designs to provide more robust evidence of the protocols' direct impact and to establish stronger causal links.

Third, the study's evaluation was conducted at a single institution—a special school complex in Poland. This geographic and institutional specificity limits the external validity of the findings, as the unique characteristics of this setting (eg, educational philosophy, available resources, specific student demographics, and cultural factors) may not be fully replicable elsewhere. To mitigate this in future research, the protocols should be replicated in diverse populations and settings, including different countries, school types, and therapeutic environments, to assess their generalizability across various contexts.

Fourth, the wide age range of participants (15-20 years) may introduce significant developmental variability within this population with ID. This age range was chosen partly on the premise that the most rapid phases of morphofunctional maturation [40] are generally less pronounced in late adolescence compared with younger childhood, which might provide a more stable baseline for evaluating the acquisition of new motor skills and independence through VR interventions. Nevertheless, this age variability still acts as a potential confounding factor that was not specifically controlled for or discussed in detail within this pilot study. While acknowledging this as a limitation, the homogeneity of the participants' mild intellectual disability level provided a baseline for consistent instruction. Future research should investigate narrower age cohorts or include age as a covariate in the analysis to better understand its potential influence on outcomes and tailor interventions accordingly.

Fifth, while the assessment of understanding instructions during the WISH protocol used a structured task sheet with predefined, detailed criteria for a 3-point rating scale, the data were collected by a single trainer. This reliance on a sole observer, despite the established rubric, introduces a potential for observer bias and limits the ability to assess interrater reliability. Consequently, the full objectivity of these specific findings cannot be fully ascertained. To enhance methodological rigor in future studies, it is crucial to incorporate multiple independent raters with demonstrated interrater agreement, using a comprehensive training protocol for raters to improve the objectivity and reliability of such observational data.

Sixth, the No problem! questionnaire used an adapted 7-level functional independence scale derived from the FIM levels. Although the FIM is an established tool, the specific adaptation of its levels and their application within our No problem! questionnaire for assessing exergame independence were not been subjected to separate validation or reliability measurements within the context of this study. Future research using this adapted methodology should conduct thorough psychometric evaluations to ensure its validity and reliability for the specific population and research context.

Finally, the novelty of VR exergaming in the study setting constitutes another limitation. In Poland, VR goggles are not yet widely used, and many participants had no previous experience with VR. Although this novelty likely contributed to high initial engagement, it also suggests that some observed improvements could be partly attributed to mere exposure to new technology rather than solely to the structured protocols. This limitation was partially mitigated by the WISH protocol's specific focus on familiarizing participants with the VR environment and the game's core mechanics. Future studies conducted in settings where VR technology is more commonplace, or those using a longer familiarization phase, could help to more clearly differentiate the effects of novelty from those of the intervention itself.

Despite these limitations, the findings from this pilot study may offer initial insights into the feasibility, acceptability, and potential benefits of the WISH and WON protocols. They may help inform the design of future, more robust, and controlled studies by highlighting areas for improvement and further investigation regarding participant engagement and performance within a VR exergame context for individuals with mild ID.

Implications and Future Work

In summary, this pilot study provides a unique and essential contribution by suggesting a transferable scientific pattern. The WISH and WON protocols appear to serve as a practical teaching model that could be applied to a variety of VR exergames, addressing the critical need for structured educational frameworks in this domain. Our findings highlight that, beyond simply providing access to technology, a well-designed pedagogical approach may be essential for cultivating true independence and maximizing the benefits of VR-based interventions.

For physical education teachers and therapists working with individuals with mild ID, our findings suggest a structured, phased approach to introducing VR exergames. These recommendations align with widely accepted principles of motor learning and skill acquisition [41-43]. This approach is particularly relevant given that children and adolescents with ID often exhibit slower motor learning and require more time and repetitions to acquire new tasks compared with typically developing peers [43]. Based on the WISH and WON protocols, key recommendations include foundational skill building (by WISH): beginning with a dedicated phase focused on familiarization with the VR environment, basic game mechanics, and explicit instruction on fundamental movements (eg, understanding warm-up cues, precise imitation of actions, maintaining correct posture). Crucially, warm-up exercises should incorporate movements and terminology directly related to the VR exergame, ensuring that participants are familiar with the required actions and verbal cues before donning the VR goggles for the main session. This initial phase, embodying the principle of teaching from easier to more difficult tasks, should prioritize task comprehension and motor learning over performance. Gradual independence fostering (by WON): systematically reducing assistance as participants demonstrate

XSL•FO
RenderX

mastery, progressing from full support to verbal cues and then to independent play, encouraging self-reliance and autonomous problem-solving within the game. This approach reflects the progressive nature of motor skill development and the fading of external support. Emphasis on procedural learning (by WISH and WON): recognizing that individuals with mild ID often benefit from clear, repetitive, and consistent instructions and that complex movements should be broken down into smaller, manageable steps. Integration of feedback and reinforcement (by WISH and WON): providing constructive feedback alongside positive reinforcement. Monitoring beyond performance (by WISH and WON): paying attention not only to in-game scores but also to indicators of independence, comfort with the VR technology, and overall enjoyment of the experience.

However, it is crucial to acknowledge the potential barriers to the widespread implementation of VR exergaming in educational or therapeutic settings. These include the significant initial cost of VR hardware and software, issues of accessibility for institutions with limited budgets or infrastructure, and the need for adequate technological literacy among both educators or therapists and participants. Addressing these practical challenges will be vital for the successful translation of these protocols into broader practice.

While this pilot study observed improvements in gameplay independence, the direct transferability of these skills to real-world daily tasks remains a hypothesis for future research. We speculate that by applying the WISH and WON protocols, teachers or therapists might be able to guide individuals with mild ID to achieve gameplay independence across various VR exergames involving similar full- or partial-body movements. This approach could potentially foster greater engagement in PA and may facilitate the translation of learned movements (eg, correct squat, no slouching) into practical daily life, thereby cultivating greater overall independence. Further studies are indeed needed to fully understand the long-term benefits of VR-based exercise on physical fitness and functional independence in individuals with ID.

Conclusions

This pilot study suggests that the structured WISH and WON training protocols may have the potential to enhance functional autonomy, exergame performance, and overall gameplay experience in individuals with mild ID.

Acknowledgments

The authors declare that generative artificial intelligence or large language models were used for language check only and were not used in any other part of the study or the analyses.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

JC and JM conceptualized the study. JC and JM recruited participants and conducted the study. JC managed data collection, conceptualized the protocols and educational concept, and prepared the video for projector imitation. JC also performed the statistical analysis and drafted the manuscript. JM supervised the study and its implementation.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Video of step “I” (Imitation) in the WISH (Warm-up, Imitation, Settings, Half-hour exergame session) protocol.
[MP4 File (MP4 Video), 126241 KB - games_v13i1e71823_app1.mp4]

Multimedia Appendix 2
Functional Independence Measure (FIM) questionnaire (original version).
[PDF File (Adobe PDF File), 1459 KB - games_v13i1e71823_app2.pdf]

References

1. Shogren K, Bradley V, Gomez S, Yeager MH, Schalock RL, Borthwick-Duffy WS, et al. Public policy and the enhancement of desired outcomes for persons with intellectual disability. Intellect Dev Disabil 2009;47(4):307-319. [doi: 10.1352/1934-9556-47.4.307] [Medline: 19650684]
2. Diagnostic and Statistical Manual of Mental Disorders. Arlington: American Psychiatric Association; 2013.
3. Pinborough-Zimmerman J, Satterfield R, Miller J, Bilder D, Hossain S, McMahon W. Communication disorders: prevalence and comorbid intellectual disability, autism, and emotional/behavioral disorders. Am J Speech Lang Pathol 2007;16(4):359-367. [doi: 10.1044/1058-0360(2007/039)] [Medline: 17971495]

XSL•FO
RenderX

4. Vuijk PJ, Hartman E, Scherder E, Visscher C. Motor performance of children with mild intellectual disability and borderline intellectual functioning. J Intellect Disabil Res 2010;54(11):955-965. [doi: 10.1111/j.1365-2788.2010.01318.x] [Medline: 20854287]
5. Boat TF, Wu JT. Mental Disorders and Disabilities Among Low-Income Children. Washington DC: National Academies Press; 2015.
6. Salaun L, Berthouze-Aranda SE. Physical fitness and fatness in adolescents with intellectual disabilities. J Appl Res Intellect Disabil 2012;25(3):231-239. [doi: 10.1111/j.1468-3148.2012.00659.x] [Medline: 22489034]
7. Hsu P, Chou H, Pan Y, Ju Y, Tsai C, Pan C. Sedentary time, physical activity levels and physical fitness in adults with intellectual disabilities. Int J Environ Res Public Health 2021;18(9):5033 [FREE Full text] [doi: 10.3390/ijerph18095033] [Medline: 34068668]
8. Fogel VA, Miltenberger RG, Graves R, Koehler S. The effects of exergaming on physical activity among inactive children in a physical education classroom. J Appl Behav Anal 2010;43(4):591-600 [FREE Full text] [doi: 10.1901/jaba.2010.43-591] [Medline: 21541146]
9. Kapidis P, Giannousi M, Vernadakis N, Papastergiou M, Zetou E, Antoniou P. Enhancing elementary school students' visual perceptual skills: a comparative study of Oculus Quest 2 exergames and conventional activities. Sport Mont 2024;22(2):139-146. [doi: 10.26773/smj.240719]
10. Vaghetti C, Monteiro-Junior R, Finco M. Exergames experience in physical education: a review. Phys Cult Sport Stud Res 2018;78(1):23-32. [doi: 10.2478/pcssr-2018-0010]
11. Wiemeyer J, Deutsch J, Malone LA, Rowland JL, Swartz MC, Xiong J, et al. Recommendations for the optimal design of exergame interventions for persons with disabilities: challenges, best practices, and future research. Games Health J 2015;4(1):58-62 [FREE Full text] [doi: 10.1089/g4h.2014.0078] [Medline: 26181682]
12. Lima JL, Axt G, Teixeira DS, Monteiro D, Cid L, Yamamoto T, et al. Exergames for children and adolescents with autism spectrum disorder: an overview. Clin Pract Epidemiol Ment Health 2020;16:1-6 [FREE Full text] [doi: 10.2174/1745017902016010001] [Medline: 32508964]
13. Martinez-Millana A, Michalsen H, Berg V, Anke A, Gil Martinez S, Muzny M, et al. Motivating physical activity for individuals with intellectual disability through indoor bike cycling and exergaming. Int J Environ Res Public Health 2022;19(5):2914 [FREE Full text] [doi: 10.3390/ijerph19052914] [Medline: 35270607]
14. Bartlett R. Quantitative motion analysis in sports biomechanics. In: Introduction to Sports Biomechanics: Analysing Human Movement Patterns. England: Routledge; 2014:135-195.
15. Pastel S, Chen C, Petri K, Witte K. Effects of body visualization on performance in head-mounted display virtual reality. PLoS One 2020;15(9):e0239226 [FREE Full text] [doi: 10.1371/journal.pone.0239226] [Medline: 32956420]
16. Desideri L, Lancioni G, Malavasi M, Gherardini A, Cesario L. Step-instruction technology to help people with intellectual and other disabilities perform multistep tasks: a literature review. J Dev Phys Disabil 2020;33(6):857-886. [doi: 10.1007/s10882-020-09781-7]
17. Sandjojo J, Gebhardt WA, Zedlitz AM, Hoekman J, den Haan JA, Evers AW. Promoting independence of people with intellectual disabilities: a focus group study perspectives from people with intellectual disabilities, legal representatives, and support staff. J Policy Pract Intellect Disabil 2018;16(1):37-52. [doi: 10.1111/jppi.12265]
18. Sandjojo J, Zedlitz AMEE, Gebhardt WA, Hoekman J, Dusseldorp E, den Haan JA, et al. Training staff to promote self-management in people with intellectual disabilities. J Appl Res Intellect Disabil 2018;31(5):840-850 [FREE Full text] [doi: 10.1111/jar.12440] [Medline: 29479785]
19. Jaarsma T, Klompstra L, Ben Gal T, Boyne J, Vellone E, Bäck M, et al. Increasing exercise capacity and quality of life of patients with heart failure through Wii gaming: the rationale, design and methodology of the HF-Wii study; a multicentre randomized controlled trial. Eur J Heart Fail 2015;17(7):743-748 [FREE Full text] [doi: 10.1002/ejhf.305] [Medline: 26139585]
20. Villada Castillo JF, Montoya Vega MF, Muñoz Cardona JE, Lopez D, Quiñones L, Henao Gallo OA, et al. Design of virtual reality exergames for upper limb stroke rehabilitation following iterative design methods: usability study. JMIR Serious Games 2024;12:e48900 [FREE Full text] [doi: 10.2196/48900] [Medline: 38206670]
21. Pirovano M, Surer E, Mainetti R, Lanzi PL, Alberto Borghese N. Exergaming and rehabilitation: a methodology for the design of effective and safe therapeutic exergames. Entertain Comput 2016;14:55-65. [doi: 10.1016/j.entcom.2015.10.002]
22. Chokshi K, Tedla J, Narayan A, Ganesan S, Reddy R. Functional Independence Measure (WeeFIM) reference values in Indian children aged 3-7 years: a cross-sectional study. Natl Med J India 2021;34(2):73-78. [doi: 10.4103/0970-258X.326748] [Medline: 34599115]
23. Msall ME, DiGaudio K, Duffy LC, LaForest S, Braun S, Granger CV. WeeFIM. Normative sample of an instrument for tracking functional independence in children. Clin Pediatr (Phila) 1994;33(7):431-438. [doi: 10.1177/000992289403300709] [Medline: 7955782]
24. Gold JI, Annick ET, Lane AS, Ho K, Marty RT, Espinoza JC. "Doc McStuffins: Doctor for a Day" virtual reality (DocVR) for pediatric preoperative anxiety and satisfaction: pediatric medical technology feasibility study. J Med Internet Res 2021;23(4):e25504 [FREE Full text] [doi: 10.2196/25504] [Medline: 33730687]

XSL•FO
RenderX

25. Kutt K, Drążyk D, Żuchowska L, Szelążek M, Bobek S, Nalepa GJ. BIRAFFE2, a multimodal dataset for emotion-based personalization in rich affective game environments. Sci Data 2022;9(1):274 [FREE Full text] [doi: 10.1038/s41597-022-01402-6] [Medline: 35672378]
26. Syropoulou A, Vernadakis N, Papastergiou M, Kourtessis T. The presence, enjoyment, mood experience, attitude and preference towards exergames scale in children with mild intellectual disability. Sport Sci Health 2021;18(1):203-210 [FREE Full text] [doi: 10.1007/s11332-021-00794-z]
27. Ciążyńska J, Maciaszek J, Worsk A, Bauza T. The effect of virtual reality exergames on general mood scale of summer sports camp participants with moderate intellectual disabilities. Acta kinesiol 2024;18(N4 2024):4-11 [FREE Full text] [doi: 10.51371/issn.1840-2976.2024.18.4.1]
28. Preibisch J, Tabriz N, Kaluschke M, Weyhe D, Uslar V. Virtual reality in clinical teaching and diagnostics for liver surgery: prospective cohort study. JMIR XR Spatial Comput 2024;1:e60383. [doi: 10.2196/60383]
29. Zhang T, Booth R, Jean-Louis R, Chan R, Yeung A, Gratzer D, et al. A primer on usability assessment approaches for health-related applications of virtual reality. JMIR Serious Games 2020;8(4):e18153 [FREE Full text] [doi: 10.2196/18153] [Medline: 33112243]
30. Wang Y, Wang X. Artificial intelligence in physical education: comprehensive review and future teacher training strategies. Front Public Health 2024;12:1484848 [FREE Full text] [doi: 10.3389/fpubh.2024.1484848] [Medline: 39583072]
31. Zhang C, Baalsrud Hauge J, Härenstam KP, Meijer S. Game experience and learning effects of a scoring-based mechanic for logistical aspects of pediatric emergency medicine: development and feasibility study. JMIR Serious Games 2021;9(1):e21988 [FREE Full text] [doi: 10.2196/21988] [Medline: 33704081]
32. Malone LA, Davlyatov GK, Padalabalanarayanan S, Thirumalai M. Active video gaming using an adapted gaming mat in youth and adults with physical disabilities: observational study. JMIR Serious Games 2021;9(3):e30672 [FREE Full text] [doi: 10.2196/30672] [Medline: 34435962]
33. Li X, Huang J, Kong Z, Sun F, Sit CHP, Li C. Effects of virtual reality–based exercise on physical fitness in people with intellectual disability: systematic review of randomized controlled trials. Games Health J 2023;12(2):89-99. [doi: 10.1089/g4h.2022.0168] [Medline: 36716183]
34. Qian J, McDonough D, Gao Z. The effectiveness of virtual reality exercise on individual's physiological, psychological and rehabilitative outcomes: a systematic review. Int J Environ Res Public Health 2020;17(11):4133 [FREE Full text] [doi: 10.3390/ijerph17114133] [Medline: 32531906]
35. Ryuh YJ, Chen CCJ, Pan Z, Gadke DL, Elmore-Staton L, Pan CY, et al. Promoting physical activity through exergaming in young adults with intellectual disabilities: a pilot study. Int J Dev Disabil 2022;68(2):227-233 [FREE Full text] [doi: 10.1080/20473869.2019.1605771] [Medline: 35309694]
36. Hanke LI, Schwoerer P, Huettl F, Vradelis L, Strelow K, Boedecker C, et al. Use of an immersive virtual reality application to educate medical students in patient handover: pilot study. JMIR Serious Games 2025;13:e73907 [FREE Full text] [doi: 10.2196/73907] [Medline: 40864895]
37. Berrezueta-Guzman S, Chen W, Wagner S. A therapeutic role-playing VR game for children with intellectual disabilities. : IEEE; 2025 Presented at: 2025 International Conference on Metaverse Computing, Networking and Applications (MetaCom); August 27-29, 2025; Seoul, Republic of Korea p. 389-396 URL: https://tinyurl.com/ynvx3d66 [doi: 10.1109/metacom65502.2025.00066]
38. Gasteiger N, van der Veer SN, Wilson P, Dowding D. Virtual reality and augmented reality smartphone applications for upskilling care home workers in hand hygiene: a realist multi-site feasibility, usability, acceptability, and efficacy study. J Am Med Inform Assoc 2023;31(1):45-60 [FREE Full text] [doi: 10.1093/jamia/ocad200] [Medline: 37846147]
39. Mira J, González M, Villalba C, Guerra L, Ramirez-Moya Y, Hernández J, et al. Improving hand hygiene skills using virtual reality: quasi-experimental study. J Med Internet Res 2025;27:e78882 [FREE Full text] [doi: 10.2196/78882] [Medline: 41057041]
40. Erofeeva MA, Ulyanova IV, Plakhotnikova IV, Kurilyuk YE, Egorov VA, Kochetkov IG. Reforming and developing socialization of children with limited abilities (mild intellectual disability). Electron J Gen Med 2019;16(2):em112. [doi: 10.29333/ejgm/108598]
41. Neto JLC, Silva EA, Guedes MS. Analysis of motor learning task in children with intellectual disability. Retos 2023;49:807-812. [doi: 10.47197/retos.v49.98355]
42. Spittle M. Motor Learning and Skill Acquisition: Applications for Physical Education and Sport. England: Bloomsbury Publishing; 2021:46-53.
43. Sullivan K, Kantak SS, Burtner PA. Motor learning in children: feedback effects on skill acquisition. Phys Ther 2008;88(6):720-732. [doi: 10.2522/ptj.20070196] [Medline: 18339797]

Abbreviations

FIM: Functional Independence Measure
ID: intellectual disabilities
IQ: intelligence quotient

XSL•FO
RenderX

ISBAR: Introduction, Situation, Background, Assessment, and Actions
PA: physical activity
PE: physical education
VR: virtual reality
WISH: Warm-up, Imitation, Settings, Half-hour exergame session
WON: Warm-up, Objective evaluation, No problem!

Edited by A Coristine; submitted 27.01.25; peer-reviewed by N Yein, J Garcia, R Garotti; comments to author 06.08.25; accepted 21.10.25; published 14.11.25.

Please cite as:
Ciążyńska J, Maciaszek J
Achieving Gameplay Independence in Virtual Reality Exergames for Individuals With Mild Intellectual Disabilities: Pilot Study
JMIR Serious Games 2025;13:e71823
URL: https://games.jmir.org/2025/1/e71823
doi:10.2196/71823
PMID:41236816

©Julia Ciążyńska, Janusz Maciaszek. Originally published in JMIR Serious Games (https://games.jmir.org), 14.11.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Effects of Balance-Based Exergame Training With Variable Difficulty on Balance and Spatiotemporal Gait Outcomes in Adults With Mild Cognitive Impairment: Randomized Controlled Trial

Aruba Saeed[1,2], MS; Imran Amjad[1,3], PhD; Imran Khan Niazi[3,4,5], PhD; Abdullah Alzahrani[5], PhD; Muhammad Shafiq[3], PhD; Heidi Haavik[3], PhD

[1]Riphah International University, Rawalpindi, , Pakistan

[2]Lahore University of Biological and Applied Sciences, Lahore, Pakistan

[3]Centre for Chiropractic Research, New Zealand College of Chiropractic, Auckland, , New Zealand

[4]Faculty of Health & Environmental Sciences, Health & Rehabilitation Research Institute, AUT University, Auckland, New Zealand

[5]Department of Health Rehabilitation Sciences, College of Applied Medical Sciences, Shaqra University, Riyadh, Saudi Arabia

Corresponding Author:
Imran Amjad, PhD
Riphah International University
RCRS Male Faculty, Al Mezan Campus, 274, Old supreme court building, Peshawar Road
Rawalpindi, 46000
Pakistan
Phone: 1 03324390125
Email: mianimran.pt@gmail.com

Abstract

Background: Exergame balance training integrates cognitive and motor challenges, potentially enhancing neuroplasticity, postural control, and gait stability in mild cognitive impairment (MCI). However, whether modulating the task difficulty of a balance-based exergame training may influence posture- and gait-related outcomes remains unknown.

Objective: We compared balance and gait improvements across exergame training groups performing exercise with different difficulty levels and a Wii Fit group in adults with MCI.

Methods: This 4-armed, parallel design, double-blinded, randomized clinical trial included 97 participants with MCI (Montreal Cognitive Assessment score=18-25). Participants were convenience-sampled from the Railway General Hospital, Rawalpindi, Pakistan, and randomized to one of 4 intervention groups: low-difficulty, moderate-difficulty, high-difficulty exergame training or a Wii Fit training group. Each participant completed 24 sessions (40 min, 3/week) supervised by physical therapists. Gait and balance were assessed using time up and go (TUG), cognitive time up and go (C-TUG), and the Gait & Balance mobile app at baseline and after 4 and 8 weeks. Although the calculated sample size was 80, 97 were recruited to offset attrition. Eighty-seven participants completed the study (94% adherence) (attrition: low-difficulty 1, moderate 3, high 2, Wii Fit 2; 10% total). Data were analyzed using mixed-model analysis of covariance with baseline values as covariates to assess time×group interactions. Bonferroni-adjusted post hoc comparisons revealed between-group differences.

Results: High-difficulty training showed the greatest TUG gains (–0.71, SD 0.32; *P*=.03, anteroposterior (AP) steadiness with eyes open (EO) on firm surface (0.04, SD 0.02; *P*=.04), step time variability head forward (HF; 0.06, SD 0.09; *P*=.02), walking speed HF (0.08, SD 0.04; *P*=.05), step time head turn (HT; –0.04, SD 0.02; *P*=.04), step time variability HT (–0.35, SD 0.09; *P*<.001), step length variability (–0.27, SD 0.13; *P*=.04), and walking speed HT (0.09, SD 0.04; *P*=.01) versus Wii Fit. Moderate-difficulty training improved AP steadiness EO firm (0.05, SD 0.02; *P*=.03) and reduced step time variability HT (–0.26, SD 0.09; *P*=.01). Low-difficulty training improved C-TUG (–1.61, SD 0.63; *P*=.01), AP steadiness EO firm (0.05, SD 0.02; *P*=.03), step time variability HF (–0.20, SD 0.09; *P*=.03), step time variability HT (–0.25, SD 0.09; *P*=.01), step length variability (–0.31, SD 0.12; *P*=.014), and walking speed HT (0.11, SD 0.04; *P*=.03). No significant differences observed between exergame difficulty groups (*P*>.05).

Conclusions: Balance-based exergame training improves balance and gait in adults with MCI, with no significant differences across difficulty levels, while the high- and low-difficulty training outperformed Wii Fit in several outcomes. High-difficulty training yielded the most consistent improvements in TUG, postural steadiness, gait variability, and walking speed. These results

XSL•FO
RenderX

support graded cognitive-motor exergaming as an effective strategy for enhancing postural control and walking stability in MCI, potentially aiding fall prevention and mobility preservation in aging populations.

Trial Registration: ClinicalTrials.gov NCT04959383; https://clinicaltrials.gov/study/NCT04959383

(JMIR Serious Games 2025;13:e74092) doi:10.2196/74092

KEYWORDS

balance; cognitive load; difficulty; exergaming; gait; mild cognitive impairment; walking

Introduction

Aging is associated with a decline in both physical and cognitive abilities [1], increasing the risk of falls, which is one of the leading causes of morbidity, mortality, and functional dependence among older adults [2]. Neuromuscular degeneration, sensory impairments, and executive function deficits contribute to balance and gait dysfunction, further elevating fall risk in older adults [3]. Importantly, age-related neurodegenerative changes can accelerate cognitive decline, with aging being one of the strongest risk factors for developing mild cognitive impairment (MCI) and dementia [3]. MCI is a clinical condition that represents an intermediate stage between normal cognitive aging and dementia [4]. It is defined by cognitive decline greater than expected for age and education, which may affect complex tasks but does not significantly impair independence [4-6]. Individuals with MCI are at an increased risk of progressing to dementia, a condition marked by more severe cognitive decline that interferes with daily functioning [4-6]. Beyond cognitive changes, MCI is increasingly recognized for its motor implications, with impairments in gait and balance being among the earliest motor manifestations [3,7] contributing to a more than 2-fold increase in fall risk and up to 4 times the rate of recurrent falls compared with cognitively healthy peers [8]. The risk is further amplified in dementia, where individuals experience a 2- to 8-fold greater likelihood of falls relative to their cognitively intact peers [9]. Gait alterations, including slower walking speed, increased stride variability, and reduced balance, are well-established predictors of future falls [10]. Given that 17.3% of adults aged 60 years and older are affected by MCI (95% CI 13.8%-20.8%) [11], and the annual incidence is 1%-2% in the general population [12]. The growing number of affected individuals highlights the urgent need for targeted interventions that address gait and balance to reduce fall risk and maintain functional independence [7].

Cognitive functions are integral to balance and gait, brain regions associated with specific cognitive processes, such as memory and executive functioning, also play a key role in motor control (eg, balance and gait) [13]. MCI is associated with atrophy and hypometabolism in the prefrontal cortex and hippocampus, as well as disrupted fronto-striatal and parietal networks [14]. These regions are crucial for executive function, memory, visuospatial processing, and motor planning [14]. Such neural alterations impair cognitive-motor integration, leading to deficits in dual-task performance and manifesting as balance instability and gait disturbances [15,16]. Motor and cognitive areas are functionally interconnected through extensive neural networks that integrate executive processing with locomotor control [13]. The prefrontal cortex, which governs higher-order cognitive functions such as attention, planning, and decision-making, communicates bidirectionally with subcortical structures, including the basal ganglia and limbic system [17], as well as locomotor centers such as the subthalamic locomotor region, mesencephalic locomotor region, and cerebellar locomotor region [18]. This integration allows cognitive processes to influence motor planning, postural adjustments, and adaptive responses to environmental demands [17,18]. When cognitive domains such as judgment, reaction time, and task switching become impaired, the prefrontal cortex's regulatory influence on these motor circuits is diminished [19,20]. Consequently, the efficiency of motor execution is disrupted, leading to deficits in gait initiation, reduced walking speed, impaired postural stability, and increased stride variability, which collectively heighten the risk of imbalance and falls [21]. Given the strong interplay between cognition and motor function, interventions targeting both domains simultaneously may offer substantial benefits, including enhanced cognitive processing, neuroplasticity, improved balance, gait speed, and functional mobility [22].

Exergaming is a combination of physical exercise with interactive gaming technology, in which users perform real-life body movements to control game actions through motion sensors, balance boards, or camera-based systems [23-25]. By blending structured physical activity with engaging digital environments, exergaming has gained prominence as a rehabilitation tool, offering an enjoyable and motivating way to promote regular exercise and therapeutic training [26]. Unlike traditional training, it delivers real-time multisensory feedback, demanding simultaneous visual attention, decision-making, and motor coordination [27]. This multimodal approach engages users in concurrent physical and cognitive tasks, which promote skilled learning, provide cognitive stimulation, and are associated with neuroplastic adaptations in key brain regions, including the frontal cortex, hippocampus, and parietal areas [24,28]. Importantly, exergaming has shown promise in improving postural stability, balance, and functional mobility in individuals with MCI [29]. The mechanism is likely multifactorial: exergames require continuous weight shifting and postural adjustments, which train the vestibular system and proprioceptive feedback loops, leading to improved balance control [30]. Simultaneously, responding to visual and auditory cues enhances sensorimotor integration and reaction speed [25,29]. The cognitive demands of exergaming, such as divided attention, decision-making, and task switching, activate executive function networks in the prefrontal cortex, which are strongly linked to gait regulation [25,29,30]. Repeated dual-task practice strengthens neural pathways between motor and cognitive centers (eg, frontal cortex, cerebellum, and basal

XSL•FO
RenderX

ganglia), fostering neuroplastic changes that support better gait speed, stride regularity, and stability [30]. Collectively, these adaptations help reduce fall risk and improve walking confidence in this population [24,26,31]. However, the role of task difficulty and graded cognitive-motor challenge in optimizing these balance and gait outcomes remains underexplored.

Integrating balance exercises into exergaming may enhance motor and cognitive function by engaging key neural structures such as the vestibular system, cerebellum, hippocampus, and prefrontal cortex [32]. Beyond these neural effects, balance training improves postural control, proprioceptive feedback, and lower-limb strength, leading to better weight shifting, stride regularity, and stability during walking [33]. These adaptations contribute to measurable improvements in gait speed, step length, and dynamic balance, which are critical for reducing fall risk in individuals with MCI [33,34]. Notably, the effectiveness of balance training is strongly linked to task complexity, as higher attention demands typically foster greater cognitive-motor engagement [35]. The "Guided Plasticity Facilitation Framework" suggests that continuously challenging tasks promote ongoing neural adaptations, strengthening cognitive-motor integration and improving coordination [24,36]. These benefits are further amplified when balance training is conducted on an unstable surface, which increases cognitive-motor engagement and enhances vestibular system activation compared with stable surfaces [37].

Although exergaming has been studied for fall prevention and postural control in MCI [29], the effect of graded cognitive-motor challenge in optimizing training effects on balance and gait remains unexplored. Addressing this gap is essential to better understand how exergaming can be optimized to enhance gait and balance performance, ultimately contributing to fall prevention strategies in MCI.

This study introduces a novel exergame with an adjustable difficulty setting designed to systematically modulate cognitive load and motor demands. Previous findings support the efficacy of this exergame in delivering graded cognitive challenges [38]. The difficulty levels were modified by adjusting the goal area and ball speed. To further enhance cognitive-motor interactions, training was conducted on a wobble board, introducing postural instability that necessitated continuous balance adjustments, thereby amplifying cognitive-motor engagement.

This randomized clinical trial (NCT04959383) was designed with preplanned outcomes, with our previous paper focusing on executive function improvements in response to exergaming [39]. This study examines a separate objective: comparing the effects of balance-based exergame training with progressive difficulty levels on unstable surfaces against both other experimental groups and with the exergame balance training on stable surfaces (Wii Fit). Given the greater cognitive-motor demands of progressively challenging tasks, we hypothesize that exergame training with increasing difficulty on unstable surfaces will result in greater improvements in balance and spatiotemporal gait parameters.

Methods

Design and Participants

This 4-arm, parallel design randomized clinical trial was conducted at Railway Hospital in Rawalpindi, Pakistan, from September 2021 to December 2022, with participants randomly allocated in a 1:1:1:1 ratio to one of the 4 intervention groups. The methodology follows the Consolidated Standards of Reporting Trials (CONSORT) 2010 [40] guidelines for randomized trials and the CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) checklist for reporting exergame interventions [41], with the completed checklists provided as Multimedia Appendices 1 and 2.

We have structured the description of both the intervention and control conditions using the Consensus on Exercise Reporting Template. The interventions aligned with the Helsinki Declaration of Ethical Principles for Medical Research. Participants were recruited through 2 medical camps and community advertisements, including radio announcements (FM 102.2) and printed flyers. Follow-up telephone calls were conducted to ensure participant compliance with the intervention protocol.

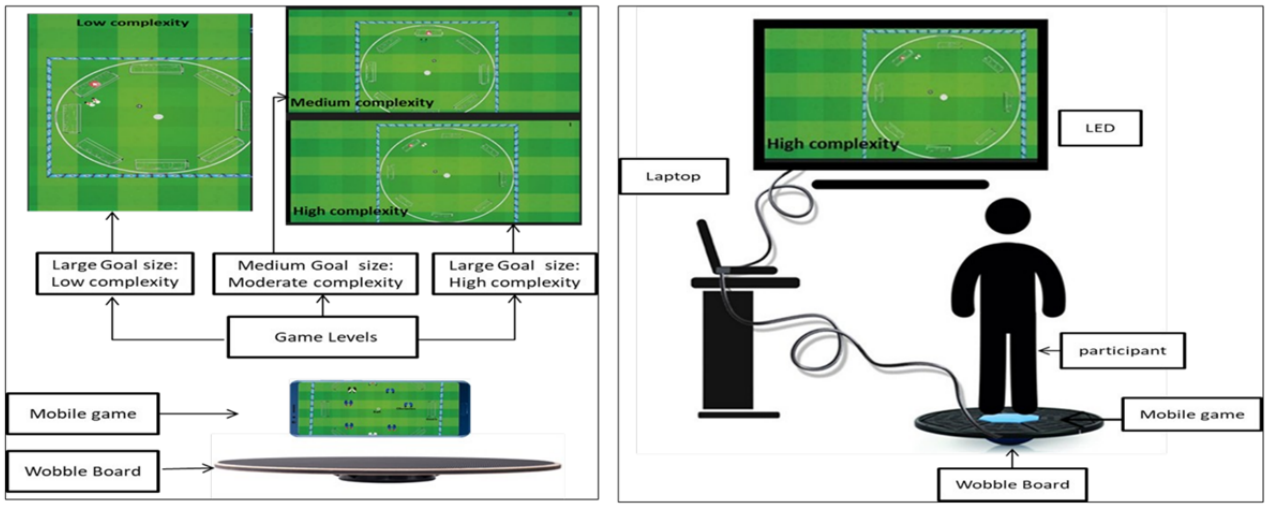
The participants were male and female aged 60-75 years who met the Petersen/Winblad criteria for MCI [42] and had a Montreal Cognitive Assessment (MoCA) score of 18-25 (Urdu Version) [43], consistent with more recent literature defining the MCI range, which slightly extends below the range (20-24) specified in the trial registry. Participants underwent a medical screening to confirm exercise safety. Those with controlled metabolic or cardiovascular conditions (eg, well-managed hypertension or type 2 diabetes) cleared by their physician were eligible. Additional inclusion criteria required participants to have the ability to read and write in Urdu, adequate vision for engaging in physical and exercise training, and a balance score of 20-25 (mild balance impairment category) on the Mini-Balance Evaluation Systems Test (Mini-BESTest), ensuring that participants had sufficient baseline balance capacity to safely perform exergame tasks. As the intervention involved the exergame technology, the ability to understand and follow the exergame requirements was considered necessary and thus served as an implicit eligibility criterion. Participants involved in any other exercise of >150 min/week, impaired mobility, history of dizziness, vertigo, clinically significant psychiatric, other neurological, unstable or severe cardiovascular, metabolic, or orthopedic conditions were excluded [44]. At baseline, demographic information (age and sex) and clinical measures (BMI and Mini-BESTest scores) were collected for all participants before randomization. Age and sex were recorded through participant self-report and confirmed against medical records where available. Height and weight were measured using a standardized stadiometer and calibrated digital weighing scale, and BMI was calculated as weight in kilograms divided by height in meters squared (kg/m^2). The Mini-BESTest was administered by trained physical therapists following standardized procedures to assess baseline balance function.

XSL•FO
RenderX

Ethical Considerations

The study protocol was reviewed and approved by the Riphah University Ethical Review Committee (approval number Riphah/RCRS/REC/1954) and was registered at ClinicalTrials.gov (NCT04959383). All participants were provided with detailed information regarding the study's objectives, procedures, potential risks, and benefits. They had the opportunity to ask questions, and written informed consent was obtained before participation, ensuring their voluntary involvement. To maintain privacy and confidentiality, all data were anonymized and deidentified before analysis. No identifiable images of participants are included in this manuscript. Participants did not receive any financial compensation for their participation. The overall process of participant enrolment, randomization, follow-up, and dropout was documented according to CONSORT guidelines, with the flow diagram presented in the Results section (Figure 1).

Figure 1. (A) Image of the exergaming interface used in the intervention group, illustrating the balance-based training tasks. (B) Experimental setup showing a participant performing exergame-based balance training. These images represent the intervention arm of a randomized controlled trial conducted in adults with mild cognitive impairment [39]. LED: light-emitting diode.



Sample Size

The sample size was calculated using G-Power software (Kiel University, DE) [45], based on a Cohen *f* effect size of 0.38 for MoCA [46]. Using a 95% confidence level (α=0.05) and 80% power, with 4 groups and one covariate, the minimum required sample size was 80 participants. Accounting for an expected attrition rate of 20%, a total of 97 patients were recruited.

Randomization and Blinding

Participants were selected through a nonprobability convenience sampling technique and randomized using a lottery method. To ensure allocation concealment, 100 sealed opaque envelopes (25 per group) containing prelabeled group assignments were prepared in advance by the principal investigator. Participants selected an envelope only after completing all preintervention assessments, preventing knowledge of allocation prior to baseline testing. The principal investigator generated the allocation sequence and prepared the envelopes, while the intervention administrator, who was not involved in baseline assessments, opened the envelopes and assigned participants to their respective groups. The intervention providers were not blinded due to the nature of the study [47]. Participants were blinded to the study hypothesis and were not informed about the other intervention groups. To ensure blinding, assessments were conducted by a single trained physical therapist who was not involved in the intervention. The assessor was unaware of the group allocation.

Development of the Intervention (Experimental Groups)

The balance-based exergame training used in this study was developed through a structured, theory-driven process. A prior EEG-based study demonstrated that each difficulty level elicited distinct acute cognitive demands, thereby validating the capacity of the training framework to deliver graded cognitive load. These findings provided the rationale for stratifying participants by difficulty level in the present trial [38]. The intervention design was guided by the Design, Dynamics, and Experience framework, which helped structure gameplay elements across design mechanics, real-time interactions, and user experience [48]. Although end users were not directly involved in development, the Design, Dynamics, and Experience framework supported the integration of meaningful cognitive-motor challenges. Additionally, the intervention was grounded in core neuroplasticity principles, specificity, repetition, intensity, and salience, and aligned with motor learning frameworks to promote cognitive-motor engagement and adaptive outcomes [49]. The manipulation of cognitive task variables, such as dual-task components, attention demands, and response inhibition, was guided by the theoretical foundation of the Challenge Point Framework and motor learning theory, which advocate for optimizing task difficulty relative to individual capacity to enhance learning and neuroplastic adaptation [50].

Weight-shifting was selected as the core motor task due to its safety and feasibility for individuals with MCI, its role as a foundational balance skill essential for gait initiation, and its suitability for precise cognitive-motor manipulation without

XSL•FO
RenderX

altering the primary motor task. This choice enabled controlled, progressive targeting of both balance and gait-related postural control in a clinically safe, evidence-informed manner. Details of the intervention parameters and training protocol are mentioned in Table 1, and intervention specifications are provided in Table 2.

Table 1. Intervention parameters and training protocol.

Parameter	Description
Intervention parameters [51]	
Frequency (F)	• Three supervised sessions on nonconsecutive weekdays (Monday, Wednesday, and Friday) per week for 8 weeks (total 24 sessions).
Intensity (I)	• Fixed for each difficulty group and Wii Fit group throughout the intervention.
Time (T)	• Forty minutes per session (5-minute warm-up, 30 minutes of exergame training, 5-minute cool-down).
Type (T)	• Experimental groups: Android-based football-themed exergame on a wobble board. • Control group: Wii Fit balance games (Soccer Heading, Penguin Slide, and Table Tilt).
Density (D)	• Each session comprised five 6-minute rounds with a 5-minute rest after the third round (work-to-rest ratio 6:1).
Training protocol [39,52]	
Instructions to participants	• All participants were thoroughly instructed at the start of the intervention regarding all these key aspects of the study protocol.
Setting and supervision	• All sessions were center-based and fully supervised by experienced physical therapists >2 years of experience in neurorehabilitation, who underwent specific training in exergame delivery at Pakistan Railway Hospital.
Delivery of sessions	• Sessions were conducted individually under supervision to ensure safety and correct execution.
Trial session	• Each participant received a trial session regarding the demonstration of the setup and training procedure.
Standardized exergame	• Exercises were standardized for each group (low, moderate, and high difficulty) and Wii Fit, not individually tailored.
Starting level decision rule	• Participants were assigned to an exergaming group based on randomization rather than performance.
Safety strategy	• A standing frame was positioned in front of each participant to ensure their safety during the intervention.
Adherence measurement	• Attendance logs were maintained for all 24 sessions, and missed sessions were tracked and completed.
Motivation strategies	• Therapists used verbal encouragement, progressive feedback, and goal setting (eg, achieving higher game scores) to motivate participants.
Home program component	• The intervention did not include a home-based program; all training was conducted in the hospital.
Assessment time Points	• Assessments were carried out at baseline, after 4 weeks, and 8 weeks.
Adverse events	• Throughout the clinical trial, adverse events, including dizziness, nausea, vertigo, and musculoskeletal pain, were monitored and documented in a therapist's log. Participants experiencing persistent symptoms were documented and subsequently excluded from the study.

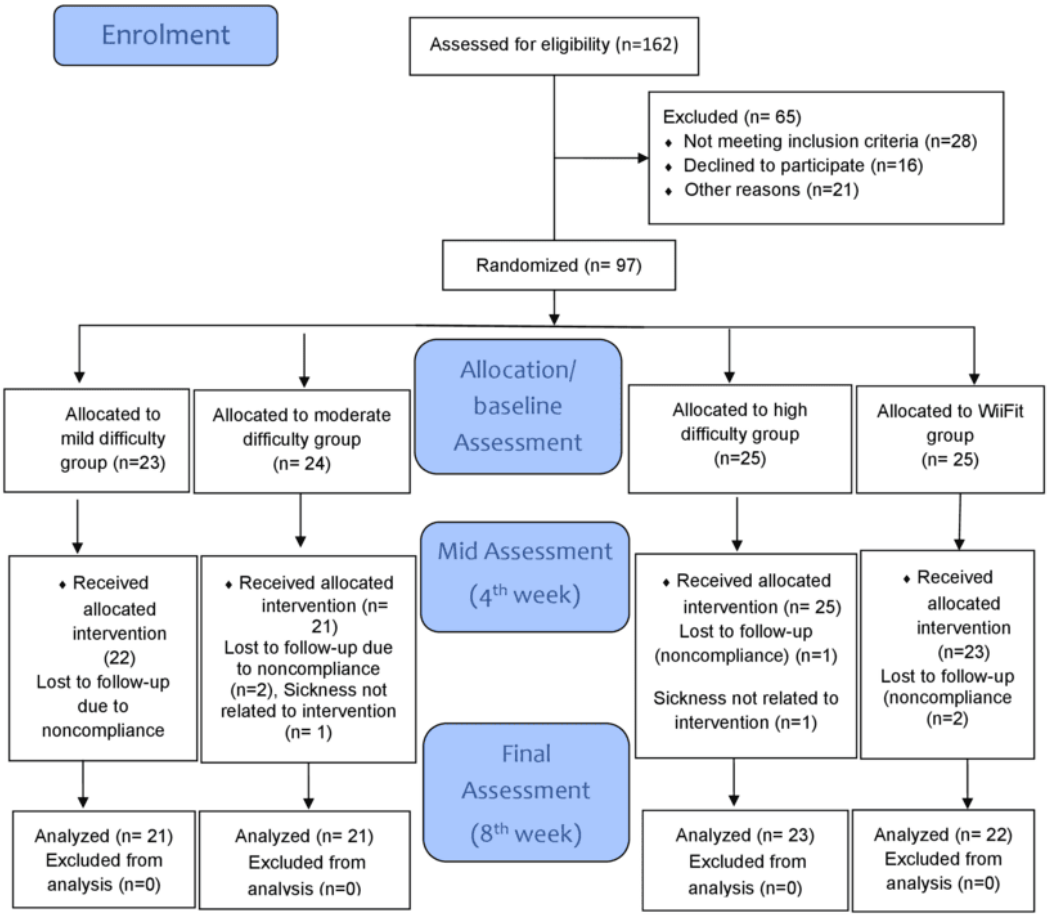
XSL•FO
RenderX

Table 2. Intervention details.

Groups	Experimental groups (low, moderate, and high difficulty)	Wii Fit group
Intervention protocol and training setup	• Customized wobble board, measuring 40 cm in diameter and 10 cm in height from the apex to the top [38,53], with a designed space at the center to hold the mobile. • The Android game was displayed on a 43" light-emitting diode (LED) screen. Experimental group game and setup are given in Figure ! [39].	Wii Fit Balance Board (WBB; Nintendo, Kyoto, Japan)
Games	• An Android-based football exergame featuring 3 predefined difficulty levels. The game includes 8 goal spots with dynamic obstacles in front of each goal spot. Criteria for difficulty levels • Low: • Goal spot size:1 unit in length • Ball speed: slow (the ball moves in 0.8 seconds) • Moderate: • Goal spot size: 0.8 units in length • Ball speed: medium (the ball moves in 0.6 seconds) • High: • Goal spot size: 0.6 units in length • Ball Speed: high (the ball moves in 0.4 seconds) [38,39].	Three balance training games: soccer heading, table tilt, and penguin slide
Intervention session plan (30 minutes)	• Total rounds of football games: 5 rounds • Games per round: 3 games • Duration per game: 2 minutes • Total duration per round: 6 minutes • Rest interval: 5 minutes to prevent fatigue after the 3rd round (after 18 minutes) [39].	• Soccer heading: 7 times (Time: 1 minute/game) • Penguin slide: 7 games (Time: 1 minute/game) • Table tilt: 8 games (Time: 2 minutes/game) • Rest interval: 5 minutes to prevent fatigue after 2 sets of games (after 14 minutes) [39].
Game Task	• The task is to score a goal on the highlighted target by shifting body weight on the wobbleboard in either direction to score more goals.	• Soccer heading: Hit the ball quickly and accurately while avoiding obstacles. • Penguin slide: The goal is to catch fish by tilting the iceberg while maintaining balance • Table tilt: The goal is to move all the balls into the holes within a set time limit
Procedure	• A novel android-based football-themed exergame, developed by Ghani et al [38], was used for the experimental groups. The game was specifically designed to provide graded cognitive-motor challenges through 3 predefined difficulty levels (low, moderate, and high). Participants stood on a custom-made wobble board and shifted their body weight to control the ball on the screen, aiming to score goals on highlighted targets while navigating dynamic obstacles. This setup created a dual-task environment, requiring balance control, postural adjustments, and visuomotor coordination. The training setup is illustrated in Figure 2.	• The Wii Balance Board's embedded electronic sensors detected subtle shifts in body weight and provided real-time visual feedback on the screen. The training consisted of 3 Wii Fit balance games—Soccer Heading, Table Tilt, and Penguin Slide. • In Soccer Heading, participants shifted weight to head virtual reality–based soccer balls while dodging obstacles; in Penguin Slide, they rocked side-to-side to catch fish; and in Table Tilt, they leaned in multiple directions to guide balls into holes, requiring precise balance control.

XSL•FO
RenderX

Figure 2. CONSORT (Consolidated Standards of Reporting Trials) flow diagram.



Intervention of the Control Group

The control condition consisted of motor-cognitive exergaming using a recreational exergame (Wii Fit), which was selected as an active comparator rather than a passive control because it is validated for improving balance and incorporates cognitive-motor engagement [54]. This alignment reflects the current classification framework that distinguishes exergame-based motor-cognitive training (serious exergame), designed for rehabilitation purposes, from motor-cognitive exergaming (Wii Fit), which is primarily developed for entertainment. The use of Wii Fit allowed comparison of a purposefully designed, theory-driven serious exergame intervention with a widely available, recreational exergame, while ensuring participant engagement, reducing dropout risk, and providing an appropriate control group intervention.

Measurement of Balance and Spatiotemporal Gait Parameters

Balance and gait outcomes were assessed by the evaluator using standardized clinical tools and a validated mobile app. The Mini-BESTest was administered as a screening measure of balance prior to enrolment [55]. Outcome measures included the timed up and go (TUG), cognitive time up and go (C-TUG), and the Gait & Balance (G&B) smartphone app. These instruments were selected for their well-established psychometric properties and clinical relevance in older adults [56-58]. Together, they provide a comprehensive evaluation of both static and dynamic balance, as well as spatiotemporal gait parameters, enabling a multidimensional assessment of mobility and fall risk.

Clinical Balance Measures

TUG

It is a valid and reliable test for measuring functional mobility and fall risk in adults with MCI intraclass correlation coefficient 0.757-0.988) [56,59]. The TUG assessment required the participants to stand from a seated position, walk a distance of 3 meters, turn around, walk back, and return to a sitting position, quickly but safely [56]. The total duration to complete the activity was measured in seconds. The shorter duration represented better functional mobility [56].

C-TUG

It has high validity and intrarater reliability (intraclass correlation coefficient 0.91-0.94) in older adults with and without dementia [60,61]. The participants stood up from a seated position, walked 3 meters, and returned to the chair while serially counting backward by subtracting 3 loudly from 20 [57]. The time to complete the C-TUG was measured in seconds.

Gait & Balance Smartphone App (G&B App)

It is a valid and reliable mobile phone app that assesses balance and gait parameters by using inertial data from accelerometers and gyroscopes embedded into the smartphone [62]. For assessment purposes, the same iPhone 7 (Apple Inc) was used for all participants to ensure consistency of sensor-based

XSL•FO
RenderX

measurements. The phone was strapped with the help of an elastic core stability belt within the customized phone case (Sports armband, Tech. Inc, Auckland, New Zealand) strapped in the lower lumbar spine, approximately L5/S1 level. The assessment protocol comprised 6 standardized tasks: 4 static balance tasks and 2 dynamic gait tasks. During the quiet stance, participants stood with feet hip-width apart and arms at their sides for 30 seconds while the app recorded postural sway in mediolateral (ML) and anteroposterior (AP) directions. Tasks included (1) standing on a hard surface with eyes open (FirmEO), (2) standing on a hard surface with eyes closed (FirmEC), (3) standing on a soft surface with eyes open (ComplaintEO), (4) standing on a soft surface with eyes closed (ComplaintEC). Dynamic gait tasks included (5) walking with looking forward (head forward [HF]), which involves walking while keeping the head and eyes facing straight ahead; (6) walking with head turns (HTs), which involves walking while actively turning the head side to side [62]. Each walking task was performed 4 times at a comfortable pace, with trials lasting 6 seconds. Auditory prompts ("ready, set, go," "rest," and "turn around") standardized task execution. Testing was discontinued if participants lost balance, opened their eyes during eyes closed (EC) tasks, or required assistance [62]. The app extracted metrics such as postural sway, walking speed, step time, step length, and variability and asymmetry of step timing and length. Validation studies have demonstrated moderate-to-excellent convergent validity of the app with 3D kinematic systems for postural stability (r=0.98), step time (r=0.97), walking speed (r=0.79), and step length (r=0.73) [63]. Test-retest reliability for these parameters was also moderate to excellent. However, validity and reliability were poor for variability and asymmetry metrics, a limitation that was also evident in gold-standard systems and is attributed to low between-subject variance and brief trial durations [63].

Important Changes to the Trial After Commencement

Few deviations from the initially registered clinical trial protocol were made to enhance methodological rigor and study quality.

1. We recruited 97 participants instead of the registered 90 to ensure adequate power in case of dropouts, although the original calculation required only 80 plus 20% for attrition.
2. Eligibility criteria were refined: the MoCA cut-off was set at <26 (MCI range 18-25) in line with current literature [64], and the Mini-BESTest replaced the Berg Balance Scale due to its broader assessment of anticipatory, reactive, sensory, and dynamic balance domains.
3. The intervention duration was increased from 30 to 40 minutes per session by adding warm-up and cool-down periods, ensuring alignment with best-practice exercise guidelines.

Statistical Analysis

Data analysis was performed using SPSS Statistics for Windows (version 26; IBM Corp). The Kolmogorov-Smirnov test was conducted to assess normality. For TUG and C-TUG, the data were normally distributed. However, the data were nonnormal for all of the variables of the G&B mobile app except for HF walking speed, HT walking speed, and HT step time. Log transformation was applied to normalize the data with a nonnormal distribution. The Box-M test was used to assess the assumption of homogeneity, with results >0.05 for all variables. A mixed-model analysis of covariance (ANCOVA) with baseline values set as covariates was used to account for potential baseline differences [65,66]. The analysis determined the main effects of both "groups" and "time" as well as interaction effects between "groups" (low, moderate, high difficulty, and Wii Fit groups) and "time" (baseline, 4 weeks, and 8 weeks). The assumption of sphericity was assessed through Mauchley test of sphericity. Post-hoc analysis with Bonferroni correction was applied to adjust the significance level for multiple comparisons and reduce the risk of type I error [66]. ANCOVA-adjusted mean differences between groups were reported to account for baseline scores, and Bonferroni-corrected pairwise comparisons (Bonferroni-adjusted *t* tests) were then conducted to identify which groups differed significantly. A statistical significance level of 0.05 was established to test the null hypothesis, and the effect size was computed using η^2 (partial eta-squared) interpreted according to conventional benchmarks (η^2 ; small=0.01, medium=0.06, large=0.14) [67].

Results

Overview

The calculated sample size was 80; however, to account for potential attrition, a total of 97 participants were initially recruited for the study. After 4 weeks (mid assessment), 92 participants completed the intervention (low difficulty: n=23; moderate difficulty: n=22; high difficulty: n=24; Wii Fit group: n=23), and by 8 weeks, 87 participants had completed the intervention (low difficulty: n=21; moderate difficulty: n=21; high difficulty: n=23; Wii Fit: n=22). This represented an overall attrition rate of 10%, well within the <20% threshold commonly cited for randomized controlled trials [68]. Completion adherence was 91.3% in the low-difficulty group, 87.5% in the moderate-difficulty group, 92.0% in the high-difficulty group, and 88.0% in the control group, indicating strong participant engagement across all arms. All participants randomized were included in the primary analysis according to their original assigned groups (intention-to-treat). Missing data due to dropout (10%) were addressed using the mixed model method, ensuring unbiased estimation and preservation of randomization integrity. The CONSORT flow diagram (Figure 2) illustrates participant recruitment, allocation, follow-up, and analysis, and demographic characteristics are provided in Table 3.

XSL•FO
RenderX

Table 3. Baseline demographic and clinical characteristics of participants across the groups.

Demographic variables and subcomponents	Exergaming group			
	Low difficulty	Moderate difficulty	High difficulty	Wii Fit
Sex				
Male, n (%)	20 (20.61)	17 (17.52)	22 (22.68)	17 (17.52)
Female, n (%)	3 (3.09)	7 (7.21)	3 (3.09)	8 (8.24)
Age (years), mean (SD)	72.52 (7.38)	64.83 (4.88)	69.72 (8.07)	72.32 (7.74)
BMI, mean (SD)	25.56 (5.71)	24.99 (2.95)	24.85 (3.77)	27.84 (5.36)
Mini BESTest[a], mean (SD)	23.87 (2.94)	23.58 (3.53)	24.20 (2.87)	23.17 (3.14)

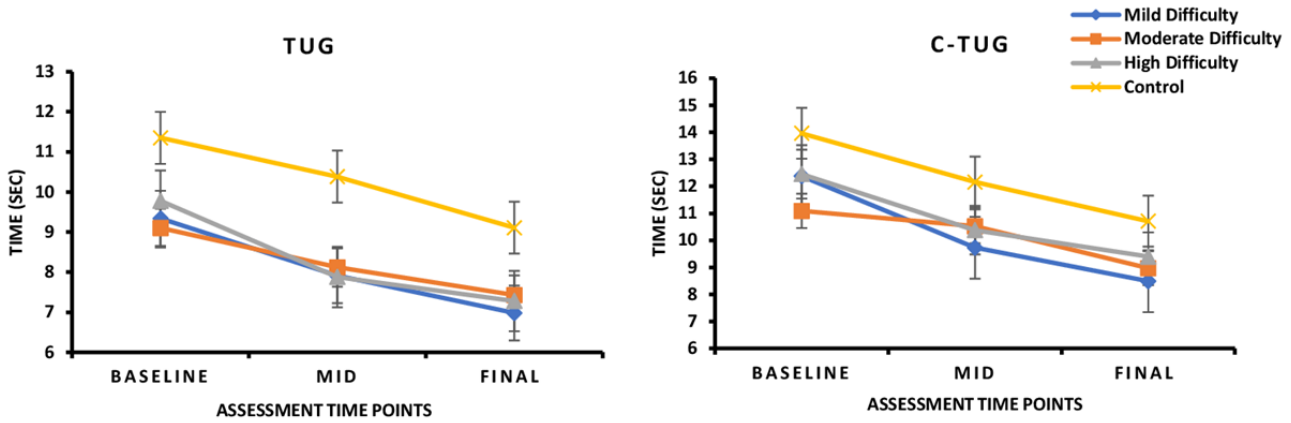
[a]Mini BESTest: Mini balance evaluation system.

Clinical Balance Measures

Interaction Effect (Groups × Time)

Significant group × time interactions were observed for TUG test ($F_{1,3}$=19.34, P<.001, η^2 =0.217) and the C-TUG test ($F_{1,3}$=21.24, P<.001, η^2 =0.233; Figure 3). These findings indicate that changes in performance over time varied significantly between intervention groups.

Figure 3. In this 8-week randomized controlled trial involving individuals with mild cognitive impairment, changes in TUG and C-TUG performance varied across intervention groups. Participants in the high-difficulty group demonstrated the greatest reduction in completion time for TUG, while participants in the low-difficulty group showed the greatest reduction in C-TUG time, indicating improved functional mobility in both groups compared with the Wii Fit control group (P<.05). C-TUG: cognitive time up and go; TUG: time up and go.



Post Hoc Pairwise Comparisons

Post hoc ANCOVA revealed that the high-difficulty group showed significantly greater improvements in TUG scores than the Wii Fit group (P<.05). For C-TUG, the low-difficulty group improved significantly more than the Wii Fit group (P<.05; Table S1 in Multimedia Appendix 3). These results suggest that higher difficulty training enhanced general mobility and dynamic balance, while lower difficulty training benefited mobility tasks involving greater cognitive engagement, such as dual-task walking.

Static Stability

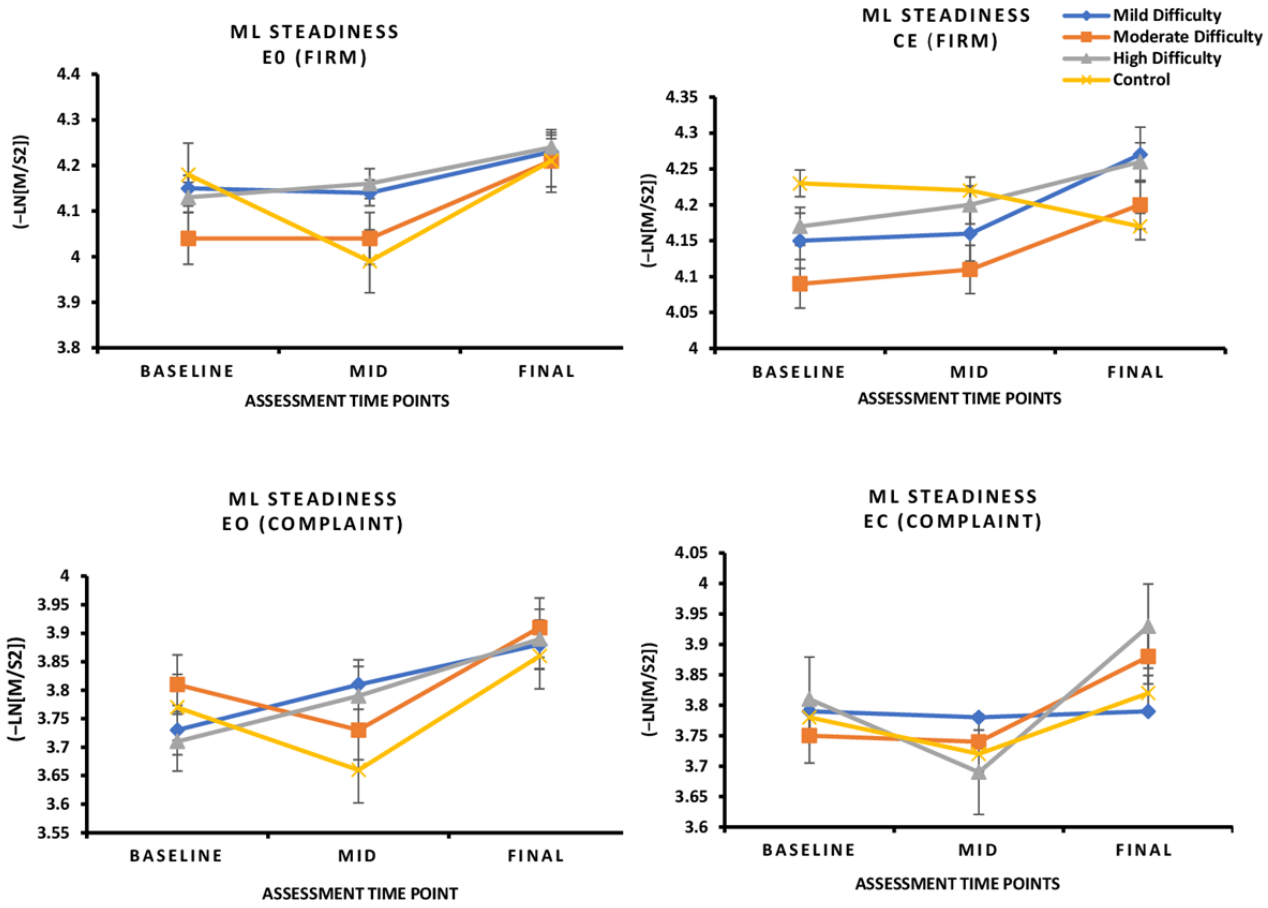
Interaction Effect (Groups × Time)

No significant group × time interaction effect was found for MLsteadiness during FirmEO, FirmEC, ComplaintEO, and ComplaintEC conditions (P>.05; Figure 4).

A significant group × time interaction was found for AP steadiness during the FirmEO condition ($F_{1,3}$=4.270, P=.04, η^2 =0.066; Figure 5).

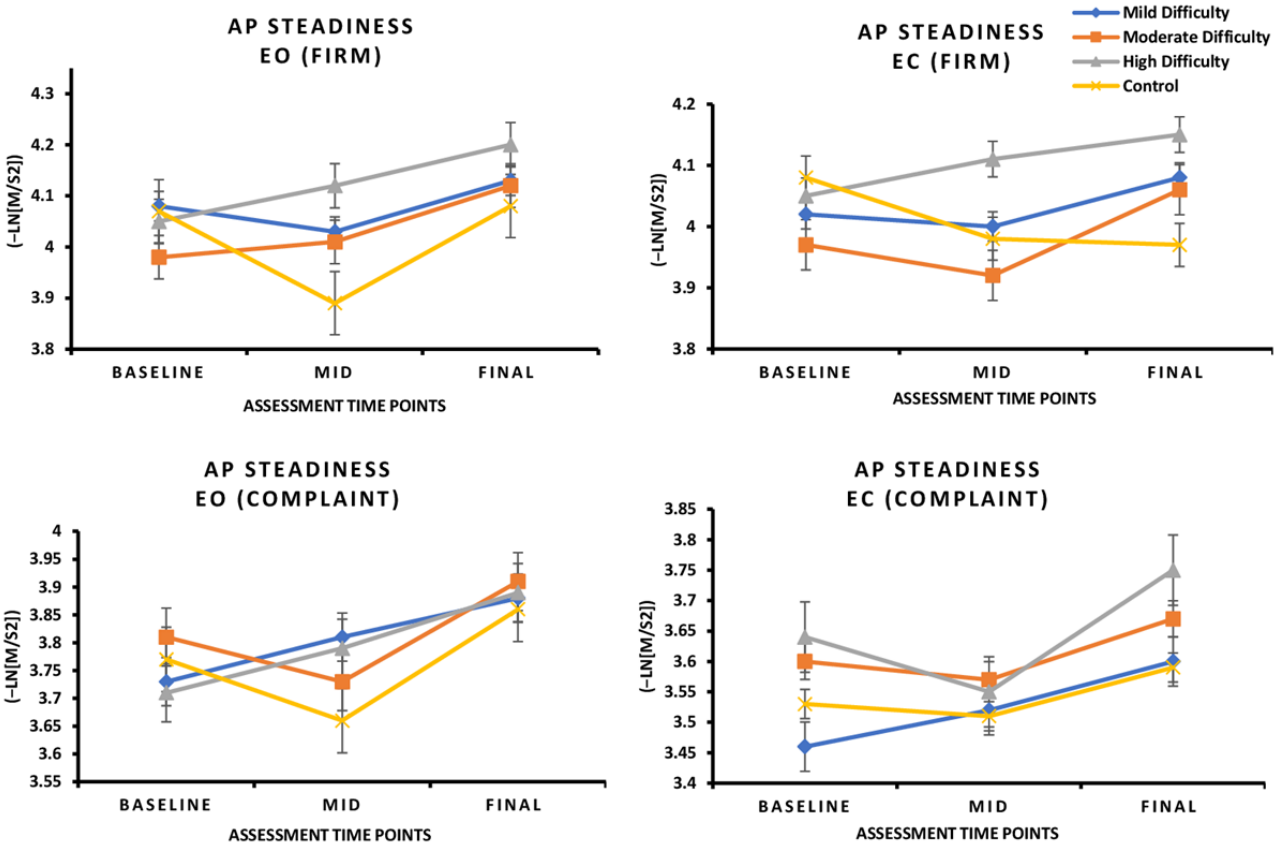
XSL•FO
RenderX

Figure 4. In this 8-week randomized controlled trial involving individuals with mild cognitive impairment, changes in ML steadiness under EO and EC conditions on firm and compliant surfaces did not vary across intervention groups. No significant between-group differences were observed in ML steadiness under either EO or EC conditions on firm or compliant surfaces (P>.05). EC: eyes closed; EO: eyes open; ML: mediolateral.



XSL•FO
RenderX

Figure 5. In this 8-week randomized controlled trial involving individuals with mild cognitive impairment, changes in AP steadiness under EO and EC conditions on firm and compliant surfaces varied across intervention groups. Participants in the low-, moderate-, and high-difficulty groups demonstrated a significant improvement in AP steadiness with EO on the compliant surface compared with the Wii Fit control group (P<.05). AP: anteroposterior; EC: eyes closed; EO: eyes open.



Post Hoc Pairwise Comparisons

Post hoc analysis showed that low-, moderate-, and high-difficulty groups all demonstrated significantly greater AP steadiness improvements compared with the Wii Fit group (P<.05; Table S2 in Multimedia Appendix 3). These findings indicate that balance control improved across all 3 exergame difficulty levels.

Spatiotemporal Gait Parameters

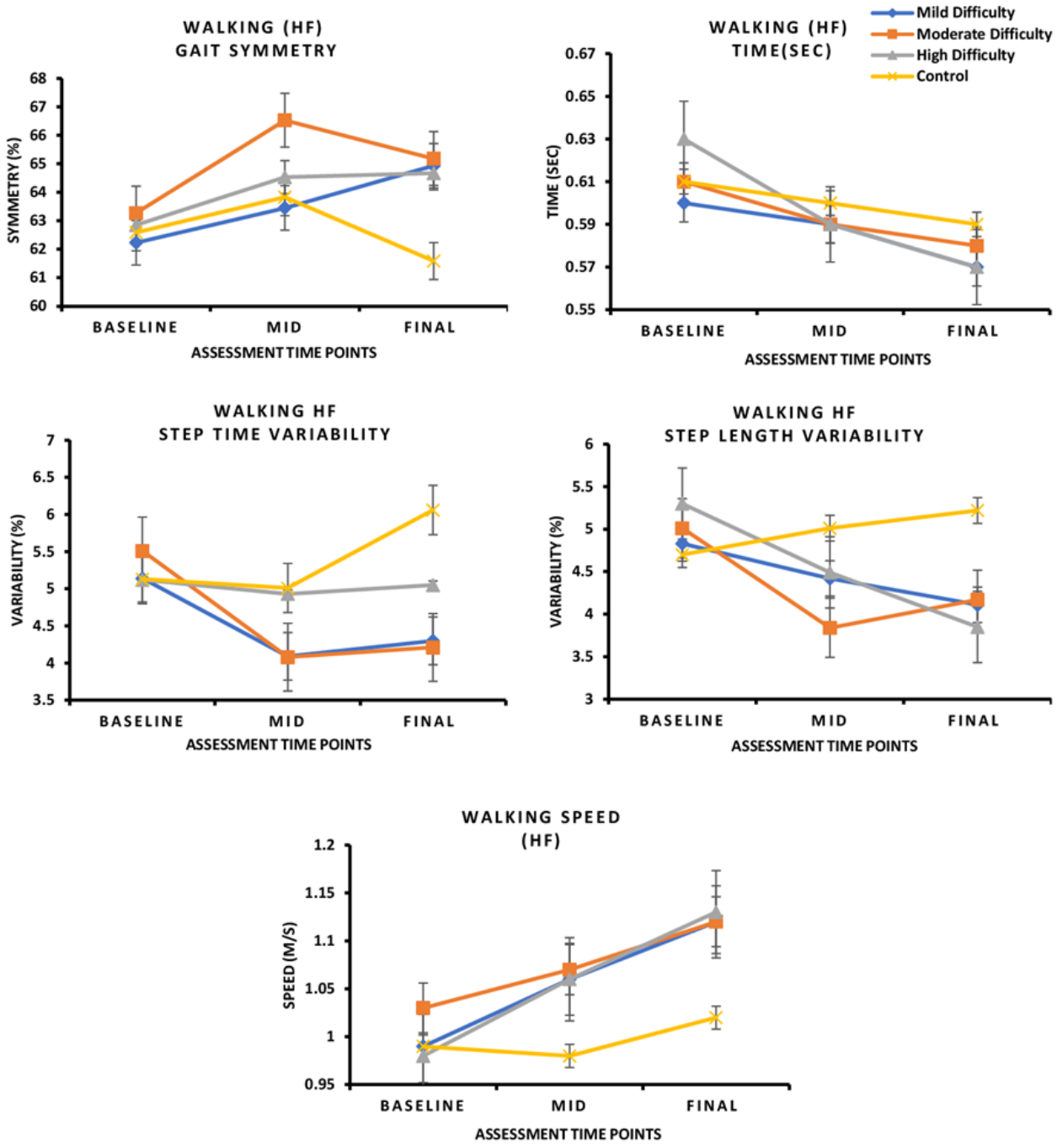
Interaction Effect (Groups × Time)

For the walking with HF task, significant interaction effects were observed for step time ($F_{1,3}$=6.117, P=.02, η^2=0.093) and step time variability ($F_{1,3}$=3.739, P=.05, η^2 =0.059; Figure 6).

For the walking with HT task, a significant interaction effect was found for step time variability ($F_{1,3}$=3.739, P=.06, η^2 =0.059; Figure 7). These results suggest that different levels of exergame difficulty influenced step timing and gait stability over time.

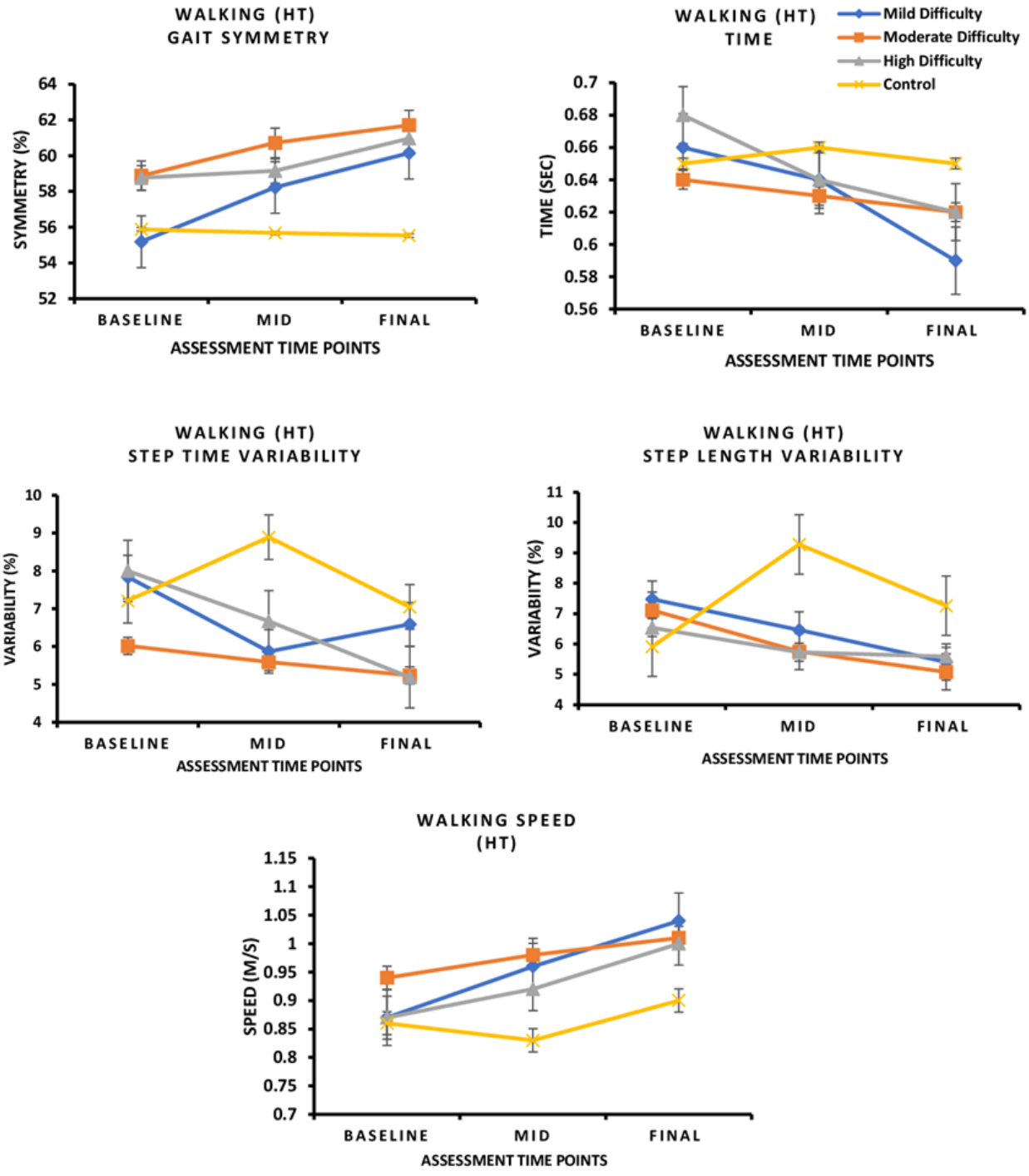
XSL•FO
RenderX

Figure 6. In this 8-week randomized controlled trial involving individuals with mild cognitive impairment, changes in walking parameters under HT conditions varied across intervention groups. Participants in the low- and high-difficulty groups showed a significant reduction in step time compared with the Wii Fit control group (P<.05). Moreover, the high-difficulty group demonstrated a significant reduction in step time variability and an improvement in walking speed relative to the Wii Fit group (P<.05). HF: head forward; HT: head turn.



XSL•FO
RenderX

Figure 7. In this 8-week randomized controlled trial involving individuals with mild cognitive impairment, changes in walking parameters under HT conditions varied across intervention groups. Step time variability was significantly reduced across all difficulty groups compared with the Wii Fit control group (P<.05). In addition, the high-difficulty group demonstrated a significant reduction in step time (P<.05), while both the low- and high-difficulty groups showed significant reductions in step time variability and improvements in walking speed relative to the Wii Fit group (P<.05). HT: head turn.



Post Hoc Pairwise Comparisons

Walking With HF

The low- and high-difficulty groups showed significantly greater improvements in step time variability compared with the Wii Fit group (P<.05). The high-difficulty group also demonstrated significant improvements in walking speed versus Wii Fit (P<.05; Table S3 in Multimedia Appendix 3).

Walking With HT

The high-difficulty group showed significant improvements in step time, step time variability, step length variability, and walking speed compared with the Wii Fit group (all P<.05). The moderate- and low-difficulty groups also improved step time variability, and the low-difficulty group demonstrated additional gains in walking speed compared with Wii Fit (P<.05; Table S3 in Multimedia Appendix 3).

XSL•FO
RenderX

Discussion

Principal Findings

To the best of our knowledge, the impact of balance-based exergame training involving graded cognitive-motor challenges on gait and balance functions in older adults with MCI has not been thoroughly assessed. The study showed significant improvement in dynamic balance (measured using TUG and C-TUG tests), as well as in AP steadiness with FirmEO condition. Improvement was also noted in gait parameters, including step time and step time variability under HF and HT conditions. The training was designed to enhance controlled weight-shifting in an exergaming environment, promoting both static and dynamic balance, alongside cognitive improvement. The effect of this intervention in enhancing various executive functions has already been reported, with more pronounced improvement in the high difficulty group [39].

Clinical Balance Measures

This study identified significant improvement in TUG and C-TUG tests among participants in the low and high-difficulty groups, while the moderate-difficulty group showed comparatively less progress. This pattern can be understood through several theoretical frameworks in motor learning and rehabilitation science.

The differing outcomes across difficulty levels reflect that cognitive and motor learning have distinct optimal challenge thresholds. In the low-difficulty group, tasks were cognitively undemanding, allowing participants to allocate attentional resources to motor refinement, resulting in significant gains in balance and mobility tasks but little measurable cognitive improvement. This aligns with Fitts and Posner cognitive stage of motor learning [69], where reduced cognitive load supports the acquisition of correct movement patterns. In contrast, the moderate-difficulty group may have occupied a zone of challenge that is cognitively demanding enough to divide attention and increase mental load, but not sufficiently intense to trigger the deeper engagement and neuroplastic motor adaptations. By contrast, the high-difficulty group provided intense cognitive-motor demands and error-driven feedback, fulfilling key principles of experience-dependent neuroplasticity (specificity, salience, and challenge) [49]. This higher level of demand required participants to integrate motor planning, postural control, and rapid corrective adjustments, translating into improvements across both cognitive parameters and speed-related gait and fall-risk outcomes.

These findings align with previous research emphasizing that the cognitively challenging training enhanced balance performance [70,71]. Virtual reality–based cognitive training has been shown to improve cognitive function and performance in both TUG [25,70,72] and C-TUG tests [70]. Exergaming facilitates sensorimotor integration by enhancing the brain's ability to process and integrate visual, vestibular, and proprioceptive inputs, leading to more efficient motor control [73]. The real-time visual feedback provided during exergaming aids in correct postural alignment and movement patterns, resulting in better coordination of muscle groups for specific motor tasks and ultimately contributing to improving TUG performance [72]. Moreover, exergaming on unstable surfaces introduces simultaneous multitasking, which enhances the brain's capacity to manage motor and cognitive tasks concurrently, an essential aspect of executive functions. Regular engagement in these graded cognitive-motor challenges has been shown to improve dual-task performance, as observed in the C-TUG [74].

Interestingly, no significant differences in TUG and C-TUG improvements were observed across different difficulty groups overall. However, the group comparison indicated that the high-difficulty training group showed a better TUG performance than the Wii Fit group, suggesting that higher cognitive demands may be needed to effectively enhance mobility outcomes. In contrast, for C-TUG, a significant improvement was found between the low-difficulty and Wii Fit groups, suggesting that dual-task mobility improved most when participants trained under conditions of manageable cognitive load. According to cognitive load theory [75], tasks that do not overwhelm working memory allow attentional resources to be shared across motor and cognitive domains, supporting dual-task integration. By contrast, moderate and high difficulty tasks may have exceeded participants' cognitive capacity, overloaded executive function, and limited gains in dual-task walking performance. An additional explanation may be related to attentional prioritization, whereby participants exposed to higher difficulty levels may have prioritized motor performance during training at the expense of cognitive performance. This would explain why improvements emerged primarily in single-task TUG but not in C-TUG performance, consistent with previous findings on prioritization effects in dual-task performance [76].

Static Stability

A significant group × time interaction effect was observed for AP steadiness under the FirmEO condition, indicating a marked difference in AP stability among various difficulty levels of exergame balance training. AP movements, such as forward and backward shifts, are generally easier to control, as they engage larger muscle groups, including quadriceps and hamstrings [77]. In contrast, ML movements, which involve side-to-side shifts, require more intricate control from smaller stabilizing muscles, along with greater reliance on proprioceptive and vestibular systems [77]. Additionally, AP steadiness tends to improve more readily due to the anatomical alignment of the foot, which is optimized for efficient movement control in the forward-backward plane. This finding is consistent with previous research, which suggests VR training significantly enhances AP stability in adults with MCI [78]. However, a study with contradictory findings reported great improvement in ML stability compared with AP stability [79]. This difference may be explained by the higher initial instability in the ML direction observed in that study, along with the training conducted on a stable surface that emphasized side-to-side movements, which could account for the greater enhancement of ML stability relative to AP stability.

A meta-analysis revealed that individuals with MCI demonstrate greater postural sway while standing with eyes open (EO) compared with healthy controls. This may be attributed to impairments in visual processing and visuospatial integration;

XSL•FO
RenderX

when these impairments are more serious, balance deficits become more pronounced, since vision is essential for achieving a stable posture [7]. The significant improvement in AP steadiness with EO condition observed in this study underscores the importance of visual feedback in postural control. With EO, participants continuously receive visual information about their body's position, enabling precise and timely postural adjustments relative to the environment. This visual input also enhances the integration of sensory information from different systems, facilitating more rapid learning and contributing to the observed improvement in AP balance during exergaming [80]. On the other hand, in the EC condition, individuals must rely heavily on proprioceptive and vestibular inputs to maintain balance. In individuals with MCI, these systems may be less effective due to cognitive decline affecting sensory integration and processing, further challenging postural control. This impaired integration can make it more challenging to maintain AP and ML steadiness with eyes closed [80].

The post-hoc analysis demonstrated no significant difference between difficulty groups. Static steadiness tasks, which primarily assess balance in a fixed position, may not be sensitive enough to detect improvements in balance during less challenging tasks, as these tasks do not fully reflect the dynamic balance demands required in higher difficulty levels. This may explain the lack of significant differences in EC steadiness across difficulty levels. However, participants in the low, moderate, and high-difficulty groups showed significant improvement in AP steadiness under the EO condition compared with the Wii Fit group. These results suggest that higher difficulty exergaming effectively improves AP steadiness, likely due to the increased physical and cognitive demands of such exercises, which promote greater focus, coordination, and overall balance stability [31].

Spatiotemporal Gait Parameters

A significant difference in improvement was observed among the groups after 8 weeks of intervention in step time and step time variability during waking under the HF conditions. Step time variability also differs across the groups under the HT condition, as indicated by mixed ANCOVA analysis. Gait is a multifaceted task requiring substantial attention, which involves physical capacity, balance, and cognitive control, including attention allocation, inhibition of distractions, and motor planning [81]. Integrating cognitive interventions, such as dual-tasking exergaming with cognitively challenging activities, may further enhance gait performance [33]. The literature supports these findings, showing significant improvement in gait speed with exergaming [31,33], and gait speed is widely recognized as one of the primary gait parameters positively influenced by exergaming [33]. Furthermore, step time variability serves as an important indicator of gait stability [33], suggesting that exergaming not only improves speed but also contributes to greater stability and control during walking.

The post-hoc analysis revealed differential improvements across groups in specific walking parameters under dual-task conditions. Notably, the high-difficulty condition yielded greater reductions in step time variability (in HF) and significantly faster walking speed (HF), as well as lower step time (HT), compared with the Wii Fit group. These results indicate that increased task challenge, combining cognitive load and physical demand, selectively enhanced motor control, consistency, and gait efficiency.

This aligns with findings from a recent randomized controlled trial and a meta-analysis, both of which reported that exergames incorporating greater postural and cognitive demands yield more pronounced improvements in gait and balance [32,82].

Furthermore, studies emphasize that exergames integrating cognitive-motor dual-tasking—especially in those with higher difficulty levels- produce gains in executive function and attentional control that mediate gait improvements [29]. Neurophysiological research has shown that increased frontal theta activity during higher-demand exergames reflects enhanced neural engagement of cognitive control networks essential to coordinate timing and balance [83]. These neurocognitive demands likely improved executive-motor integration, thereby supporting faster and more stable walking under HF/HT conditions [84].

Building on these findings, a differentiated pattern emerged across difficulty levels. The low- and high-difficulty exergaming groups achieved the most notable gains, though likely through different mechanisms. Low-difficulty tasks appeared to provide an accessible point for motor practice by reducing cognitive demands and minimizing dual-task interference, thereby enabling participants to consolidate core postural strategies and movement patterns [85,86]. By contrast, high-difficulty tasks placed greater demands on attention, motor planning, and adaptive control, which evidence suggests can stimulate neuromuscular activation, sensorimotor integration, and neuroplastic changes leading to improvements in gait variability and walking speed [32]. The moderate-difficulty condition, however, seemed to sit in a "grey zone," neither simple enough to free attentional resources for basic motor learning nor challenging enough to drive significant motor adaptation [87]. Together, these findings highlight the importance of carefully selecting the exergame task difficulty, starting with manageable tasks to consolidate fundamentals, then advancing to higher-demand tasks to trigger more complex motor adaptations and maximize therapeutic benefit [32,87]. Although task difficulty in this study was not individualized, the observed response patterns across groups are consistent with the principles of the optimal challenge point framework, which suggests that different levels of task demand may elicit distinct motor and cognitive adaptations [88].

The study findings suggest that graded cognitive-motor challenging exergames may be a valuable approach to targeting balance and gait deficits, emphasizing the potential for these training programs to optimize functional outcomes. The results also suggest that incorporating balance-focused exergame training with graded difficulty levels can enhance balance and gait performance in older adults with MCI, and potentially extend to other populations experiencing cognitive-motor impairments, such as individuals with other neurological conditions. These potential benefits highlight the clinical relevance and impact of this intervention [89]. Clinicians may consider incorporating such interventions alongside conventional

XSL•FO
RenderX

therapy to enhance dual-task performance and potentially reduce fall risk in this vulnerable and fall-prone population [90]. Because exergame difficulty can be adjusted, therapists can tailor challenges to individual capabilities and progressively modify tasks to maintain optimal engagement, making this approach suitable for use in outpatient clinics.

Limitations

The study had some limitations. First, the 8-week intervention duration may have restricted the overall impact of the training, and extending the intervention duration could potentially reveal more significant differences between various difficulty levels. Although adherence was closely monitored to ensure full exposure to the planned dose, longer interventions in future studies may reveal stronger or more sustained effects. Second, since the majority of the participants were male, the findings may not be fully generalizable to both sexes. Improvement in balance and gait differs between males and females [91], and this difference may persist when exposed to varying levels of exergaming difficulty. Recruiting a more sex-balanced sample is essential to enhance the generalizability of findings and to explore potential sex-based differences. Third, the C-TUG test used the same starting number and subtraction pattern throughout the study. This design may have introduced familiarity effects, potentially reducing the test's sensitivity to detect subtle changes in dual-task performance and cognitive-motor interference across groups. Moreover, cognitive task performance (ie, number of correct responses) was not recorded during the C-TUG, preventing evaluation of possible dual-task trade-offs where mobility improvements could occur at the expense of cognitive accuracy or vice versa. Future studies should vary the secondary task and record accuracy to capture dual-task interference better.

Fourth, the intervention used a fixed design, although tasks were structured at low, moderate, and high levels; no progression mechanism was applied across the 24 sessions. It is therefore possible that motor or cognitive learning reduced the internal load over time, limiting the challenge necessary to trigger optimal neuroplastic adaptations. Future research should integrate progressive overload principles to maintain optimal challenge. Fifth, no objective or subjective measures of task demand or internal loading metrics (eg, perceived exertion or cognitive load) were calculated, limiting the ability to confirm whether the intended challenge levels were achieved for each individual. Future studies should incorporate these assessments to strengthen intervention validation. Six, as the trial was powered on cognition rather than motor outcomes, it is possible that some gait or balance effects were underpowered; future trials may consider sample size estimation directly based on these parameters. Seven, while the G&B app is valid for core gait measures, its reliability for step time variability and asymmetry is limited. Findings related to these variables should therefore be viewed in light of their known measurement limitations.

Implications for Research

Future studies should test difficulty-graded exergames in broader clinical populations (eg, Parkinson disease and poststroke) and examine their long-term effects. Incorporating objective and subjective measures of task demand (eg, perceived exertion and cognitive load) will enable difficulty to be tailored to individual capacity and adjusted progressively, providing clearer insight into optimal dosing. Research should also evaluate how these interventions can be embedded into fall-prevention and healthy aging programs to maximize their impact on functional decline, independence, and accessibility across clinical, community, and home settings. The future work should also explore the integration of step-based training components alongside weight-shifting tasks to enhance gait adaptability, balance-gait synergy, and ecological validity, particularly in populations at risk of mobility decline.

Conclusions

In conclusion, the 8-week exergame balance training effectively improved balance and spatiotemporal gait measures, including TUG, C-TUG, and AP stability on a firm surface, as well as step time and step time variability during walking under HF and HT conditions. However, no significant differences in improvement were observed across different difficulty groups for clinical balance measures, stability, and gait parameters. Both high-difficulty and low-difficulty groups showed improvement in stability, step time, and step length variability. Moreover, step time and walking speed with HF and HT significantly improved in the high-difficulty group. These findings suggest that exergame-based interventions with varying difficulty levels may enhance balance and gait parameters in older adults with MCI.

Acknowledgments

The authors thank all of the patients and staff from the Rehabilitation Center of Railway Hospital, Rawalpindi, Pakistan, who contributed to this study. This study is funded by the National Institute of Health, Pakistan (SG-21-63).

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

IA, AS, and IKN contributed to the conceptualization of the study, while IA and AS were responsible for data curation and formal analysis. Funding acquisition was handled by IA. Investigation was conducted by AS and AI, and the methodology was developed by IA, AS, IKN, HH, and AA. Project administration was managed by IA and AS, and resources were provided by IA, IKN, HH,

XSL•FO
RenderX

and AA. Supervision was undertaken by IA. Validation was carried out by IA, IKN, MS, and AA, and visualization was performed by AS and AI. The original draft of the manuscript was prepared by AS, IA, IKN, AA, and MS, while all authors (AS, IA, IKN, AA, HH, and MS) contributed to the review and editing of the manuscript. All authors (AS, IA, IKN, AA, HH, and MS) agree to be accountable for all aspects of the work, ensuring that questions related to the accuracy or integrity of any part of the study are appropriately investigated and resolved.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT 2010 checklist.
[PDF File (Adobe PDF File), 55 KB - games_v13i1e74092_app1.pdf]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 3093 KB - games_v13i1e74092_app2.pdf]

Multimedia Appendix 3
Between-group differences in difficulty levels for clinical balance measurement, during quiet stance, and during walking.
[DOCX File , 46 KB - games_v13i1e74092_app3.docx]

References

1. Clouston SAP, Brewster P, Kuh D, Richards M, Cooper R, Hardy R, et al. The dynamic relationship between physical function and cognition in longitudinal aging cohorts. Epidemiol Rev 2013;35(1):33-50 [FREE Full text] [doi: 10.1093/epirev/mxs004] [Medline: 23349427]
2. Ambrose AF, Paul G, Hausdorff JM. Risk factors for falls among older adults: a review of the literature. Maturitas 2013;75(1):51-61. [doi: 10.1016/j.maturitas.2013.02.009] [Medline: 23523272]
3. Kim S, Kim H. Relationship between falls, cognitive decline, and dementia in older adults: insights from the Korean longitudinal study of aging, 2006-2020. Exp Gerontol 2024;194:112481 [FREE Full text] [doi: 10.1016/j.exger.2024.112481] [Medline: 38871235]
4. Smid J, Studart-Neto A, César-Freitas KG, Dourado MCN, Kochhann R, Barbosa BJAP, et al. Subjective cognitive decline, mild cognitive impairment, and dementia - syndromic approach: recommendations of the scientific department of cognitive neurology and aging of the Brazilian academy of neurology. Dement Neuropsychol 2022;16(3 Suppl 1):1-24 [FREE Full text] [doi: 10.1590/1980-5764-DN-2022-S101PT] [Medline: 36533160]
5. Petersen RC, Smith GE, Waring SC, Ivnik RJ, Tangalos EG, Kokmen E. Mild cognitive impairment: clinical characterization and outcome. Arch Neurol 1999;56(3):303-308. [doi: 10.1001/archneur.56.3.303] [Medline: 10190820]
6. Schwarzinger M, Dufouil C. Forecasting the prevalence of dementia. Lancet Public Health 2022;7(2):e94-e95 [FREE Full text] [doi: 10.1016/S2468-2667(21)00277-2] [Medline: 34998486]
7. Bahureksa L, Najafi B, Saleh A, Sabbagh M, Coon D, Mohler MJ, et al. The impact of mild cognitive impairment on gait and balance: a systematic review and meta-analysis of studies using instrumented assessment. Gerontology 2017;63(1):67-83 [FREE Full text] [doi: 10.1159/000445831] [Medline: 27172932]
8. Michel JP, Beattie BL, Martin FC, Walston JD. Oxford Textbook of Geriatric Medicine (3 edn). England, United Kingdom: Oxford University Press; 2018.
9. Fernando E, Fraser M, Hendriksen J, Kim CH, Muir-Hunter SW. Risk factors associated with falls in older adults with dementia: a systematic review. Physiother Can 2017;69(2):161-170 [FREE Full text] [doi: 10.3138/ptc.2016-14] [Medline: 28539696]
10. Pieruccini-Faria F, Sarquis-Adamson Y, Anton-Rodrigo I, Noguerón-García A, Bray NW, Camicioli R, et al. Mapping associations between gait decline and fall risk in mild cognitive impairment. J Am Geriatr Soc 2020;68(3):576-584. [doi: 10.1111/jgs.16265] [Medline: 31846071]
11. Pessoa RMP, Bomfim AJL, Ferreira BLC, Chagas MHN. Diagnostic criteria and prevalence of mild cognitive impairment in older adults living in the community: a systematic review and meta-analysis. Arch Clin Psychiatry (São Paulo) 2019;46(3):72-79. [doi: 10.1590/0101-60830000000197]
12. Petersen RC, Roberts RO, Knopman DS, Boeve BF, Geda YE, Ivnik RJ, et al. Mild cognitive impairment: ten years later. Arch Neurol 2009;66(12):1447-1455 [FREE Full text] [doi: 10.1001/archneurol.2009.266] [Medline: 20008648]
13. Booth V, Logan P, Harwood R, Hood V. Falls prevention interventions in older adults with cognitive impairment: a systematic review of reviews. Int J Ther Rehabil 2015;22(6):289-296. [doi: 10.12968/ijtr.2015.22.6.289]

XSL•FO
RenderX

14. Doi T, Makizako H, Shimada H, Park H, Tsutsumimoto K, Uemura K, et al. Brain activation during dual-task walking and executive function among older adults with mild cognitive impairment: a fNIRS study. Aging Clin Exp Res 2013;25(5):539-544. [doi: 10.1007/s40520-013-0119-5] [Medline: 23949972]
15. Koppelmans V, Silvester B, Duff K. Neural mechanisms of motor dysfunction in mild cognitive impairment and alzheimer's disease: a systematic review. J Alzheimers Dis Rep 2022;6(1):307-344 [FREE Full text] [doi: 10.3233/ADR-210065] [Medline: 35891638]
16. Holtzer R, Izzetoglu M. Mild cognitive impairments attenuate prefrontal cortex activations during walking in older adults. Brain Sci 2020;10(7):415 [FREE Full text] [doi: 10.3390/brainsci10070415] [Medline: 32630216]
17. Rosano C, Studenski SA, Aizenstein HJ, Boudreau RM, Longstreth WT, Newman AB. Slower gait, slower information processing and smaller prefrontal area in older adults. Age Ageing 2012;41(1):58-64 [FREE Full text] [doi: 10.1093/ageing/afr113] [Medline: 21965414]
18. Sarasso E, Filippi M, Agosta F. Clinical and MRI features of gait and balance disorders in neurodegenerative diseases. J Neurol 2023;270(3):1798-1807. [doi: 10.1007/s00415-022-11544-7] [Medline: 36577818]
19. Kearney FC, Harwood RH, Gladman JR, Lincoln N, Masud T. The relationship between executive function and falls and gait abnormalities in older adults: a systematic review. Dement Geriatr Cogn Disord 2013;36(1-2):20-35. [doi: 10.1159/000350031] [Medline: 23712088]
20. Ward C, Touchet E, Marfeo E, Ward N. Mobility and cognitive decline in older adults with cognitive impairment. In: Thomas AK, Gutchess A, editors. The Cambridge Handbook of Cognitive Aging: A Life Course Perspective. Cambridge, UK: Cambridge University Press; 2020:701-716.
21. Scherder E, Eggermont L, Visscher C, Scheltens P, Swaab D. Understanding higher level gait disturbances in mild dementia in order to improve rehabilitation: 'last in-first out'. Neurosci Biobehav Rev 2011;35(3):699-714. [doi: 10.1016/j.neubiorev.2010.08.009] [Medline: 20833200]
22. Zhu S, Sui Y, Shen Y, Zhu Y, Ali N, Guo C, et al. Effects of virtual reality intervention on cognition and motor function in older adults with mild cognitive impairment or dementia: a systematic review and meta-analysis. Front Aging Neurosci 2021;13:586999 [FREE Full text] [doi: 10.3389/fnagi.2021.586999] [Medline: 34025384]
23. Witherspoon L. ACSM information on exergaming. United States: American College of Sports Medicine; 2013.
24. Perrot A, Maillot P. Factors for optimizing intervention programs for cognition in older adults: the value of exergames. NPJ Aging 2023;9(1):4 [FREE Full text] [doi: 10.1038/s41514-023-00103-7] [Medline: 36991073]
25. Cai Z, Ma Y, Li L, Lu GZ. Effects of exergaming in older individuals with mild cognitive impairment and dementia: a systematic review and meta-analysis. Geriatr Nurs 2023;51:351-359. [doi: 10.1016/j.gerinurse.2023.03.028] [Medline: 37099867]
26. Chan JYC, Liu J, Chan ATC, Tsoi KKF. Exergaming and cognitive functions in people with mild cognitive impairment and dementia: a meta-analysis. NPJ Digit Med 2024;7(1):154 [FREE Full text] [doi: 10.1038/s41746-024-01142-4] [Medline: 38879695]
27. Kueider AM, Parisi JM, Gross AL, Rebok GW. Computerized cognitive training with older adults: a systematic review. PLoS One 2012;7(7):e40588 [FREE Full text] [doi: 10.1371/journal.pone.0040588] [Medline: 22792378]
28. Law LL, Barnett F, Yau MK, Gray MA. Effects of combined cognitive and exercise interventions on cognition in older adults with and without cognitive impairment: a systematic review. Ageing Res Rev 2014;15:61-75. [doi: 10.1016/j.arr.2014.02.008] [Medline: 24632497]
29. Zhao Y, Feng H, Wu X, Du Y, Yang X, Hu M, et al. Effectiveness of exergaming in improving cognitive and physical function in people with mild cognitive impairment or dementia: systematic review. JMIR Serious Games 2020;8(2):e16841 [FREE Full text] [doi: 10.2196/16841] [Medline: 32602841]
30. Swanenburg J, Wild K, Straumann D, de Bruin ED. Exergaming in a moving virtual world to train vestibular functions and gait; a proof-of-concept-study with older adults. Front Physiol 2018;9:988 [FREE Full text] [doi: 10.3389/fphys.2018.00988] [Medline: 30108511]
31. Ali N, Tian H, Thabane L, Ma J, Wu H, Zhong Q, et al. The effects of dual-task training on cognitive and physical functions in older adults with cognitive impairment; a systematic review and meta-analysis. J Prev Alzheimers Dis 2022;9(2):359-370. [doi: 10.14283/jpad.2022.16] [Medline: 35543010]
32. Schättin A, Arner R, Gennaro F, de Bruin ED. Adaptations of prefrontal brain activity, executive functions, and gait in healthy elderly following exergame and balance training: a randomized-controlled study. Front Aging Neurosci 2016;8:278 [FREE Full text] [doi: 10.3389/fnagi.2016.00278] [Medline: 27932975]
33. Zhang W, Low L, Gwynn JD, Clemson L. Interventions to improve gait in older adults with cognitive impairment: a systematic review. J Am Geriatr Soc 2019;67(2):381-391. [doi: 10.1111/jgs.15660] [Medline: 30506754]
34. Liu X, Chen MH, Yue GH. Postural control dysfunction and balance rehabilitation in older adults with mild cognitive impairment. Brain Sci 2020;10(11):873 [FREE Full text] [doi: 10.3390/brainsci10110873] [Medline: 33227910]
35. Bamidis PD, Fissler P, Papageorgiou SG, Zilidou V, Konstantinidis EI, Billis AS, et al. Gains in cognition through combined cognitive and physical training: the role of training dosage and severity of neurocognitive disorder. Front Aging Neurosci 2015;7:152 [FREE Full text] [doi: 10.3389/fnagi.2015.00152] [Medline: 26300772]

XSL•FO
RenderX

36. Joubert C, Chainay H. Aging brain: the effect of combined cognitive and physical training on cognition as compared to cognitive and physical training alone - a systematic review. Clin Interv Aging 2018;13:1267-1301. [doi: 10.2147/CIA.S165399] [Medline: 30057444]
37. Quesada PM, Geiger JT. Effects of cognitive task execution on stable and unstable surface balance. Cogent Eng 2017;4(1):1311440. [doi: 10.1080/23311916.2017.1311440]
38. Ghani U, Signal N, Niazi IK, Taylor D. Efficacy of a single-task ERP measure to evaluate cognitive workload during a novel exergame. Front Hum Neurosci 2021;15:742384 [FREE Full text] [doi: 10.3389/fnhum.2021.742384] [Medline: 34566610]
39. Saeed A, Niazi IK, Alzahrani A, Trager RJ, Haavik H, Amjad I. A difficulty based comparison of novel exergame balance training for cognitive functions in adults with mild cognitive impairment: a randomized trial. Ment Health Phys Act 2024;27:100637. [doi: 10.1016/j.mhpa.2024.100637]
40. Moher D, Hopewell S, Schulz K, Montori V, Gøtzsche PC, Devereaux P, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. BMJ 2010;340:c869 [FREE Full text] [doi: 10.1136/bmj.c869] [Medline: 20332511]
41. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of web-based and mobile health interventions. J Med Internet Res 2011;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
42. Jak AJ, Bondi MW, Delano-Wood L, Wierenga C, Corey-Bloom J, Salmon DP, et al. Quantification of five neuropsychological approaches to defining mild cognitive impairment. Am J Geriatr Psychiatry 2009;17(5):368-375 [FREE Full text] [doi: 10.1097/JGP.0b013e31819431d5] [Medline: 19390294]
43. Trzepacz PT, Hochstetler H, Wang S, Walker B, Saykin AJ, Alzheimer's Disease Neuroimaging Initiative. Relationship between the montreal cognitive assessment and mini-mental state examination for assessment of mild cognitive impairment in older adults. BMC Geriatr 2015;15:107 [FREE Full text] [doi: 10.1186/s12877-015-0103-3] [Medline: 26346644]
44. Eggenberger P, Wolf M, Schumann M, de Bruin ED. Exergame and balance training modulate prefrontal brain activity during walking and enhance executive function in older adults. Front Aging Neurosci 2016;8:66 [FREE Full text] [doi: 10.3389/fnagi.2016.00066] [Medline: 27148041]
45. Faul F, Erdfelder E, Lang A, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007;39(2):175-191. [doi: 10.3758/bf03193146] [Medline: 17695343]
46. Schwenk M, Sabbagh M, Lin I, Morgan P, Grewal GS, Mohler J, et al. Sensor-based balance training with motion feedback in people with mild cognitive impairment. J Rehabil Res Dev 2016;53(6):945-958 [FREE Full text] [doi: 10.1682/JRRD.2015.05.0089] [Medline: 28475201]
47. Stanmore EK, Mavroeidi A, de Jong LD, Skelton DA, Sutton CJ, Benedetto V, et al. The effectiveness and cost-effectiveness of strength and balance exergames to reduce falls risk for people aged 55 years and older in UK assisted living facilities: a multi-centre, cluster randomised controlled trial. BMC Med 2019;17(1):49 [FREE Full text] [doi: 10.1186/s12916-019-1278-9] [Medline: 30813926]
48. Walk W, Görlich D, Barrett M. Design, Dynamics, Experience (DDE): An Advancement of the MDA Framework for Game Design. In: Korn O, Lee N, editors. Game Dynamics. Cham, Switzerland: Springer, Cham; 2017.
49. Kleim JA, Jones TA. Principles of experience-dependent neural plasticity: implications for rehabilitation after brain damage. J Speech Lang Hear Res 2008;51(1):S225-S239. [doi: 10.1044/1092-4388(2008/018)] [Medline: 18230848]
50. Kafri M, Atun-Einy O. From motor learning theory to practice: a scoping review of conceptual frameworks for applying knowledge in motor learning to physical therapist practice. Phys Ther 2019;99(12):1628-1643. [doi: 10.1093/ptj/pzz118] [Medline: 31407003]
51. Herold F, Zou L, Theobald P, Manser P, Falck RS, Yu Q, et al. Beyond FITT: addressing density in understanding the dose-response relationships of physical activity with health-an example based on brain health. Eur J Appl Physiol 2025:1-29. [doi: 10.1007/s00421-025-05858-3] [Medline: 40569418]
52. Slade SC, Dionne CE, Underwood M, Buchbinder R. Consensus on exercise reporting template (CERT): explanation and elaboration statement. Br J Sports Med 2016;50(23):1428-1437. [doi: 10.1136/bjsports-2016-096651] [Medline: 27707738]
53. Binns E, Kerse N, Peri K, Cheung G, Taylor D. Combining cognitive stimulation therapy and fall prevention exercise (CogEx) in older adults with mild to moderate dementia: a feasibility randomised controlled trial. Pilot Feasibility Stud 2020;6:108. [doi: 10.1186/s40814-020-00646-6] [Medline: 32724661]
54. Zhang L, Guo J, Zhang J, Zhang L, Li Y, Yang S, et al. Interactive cognitive motor training: a promising approach for sustainable improvement of balance in older adults. Sustainability 2023;15(18):13407. [doi: 10.3390/su151813407]
55. Viveiro LAP, Gomes GCV, Bacha JMR, Carvas Junior N, Kallas ME, Reis M, et al. Reliability, validity, and ability to identity fall status of the berg balance scale, balance evaluation systems test (BESTest), Mini-BESTest, and Brief-BESTest in older adults who live in nursing homes. J Geriatr Phys Ther 2019;42(4):E45-E54. [doi: 10.1519/JPT.0000000000000215] [Medline: 30407272]
56. Rolenz E, Reneker JC. Validity of the 8-foot up and go, timed up and go, and activities-specific balance confidence scale in older adults with and without cognitive impairment. J Rehabil Res Dev 2016;53(4):511-518. [doi: 10.1682/JRRD.2015.03.0042] [Medline: 27532337]

XSL•FO
RenderX

57. Pohl PS, Gras LZ, Bosch PR, Ganley KJ, Mayer J. Dual task timed up-and-go for older adults with and without balance deficits. Phys Occup Ther Geriatr 2019;37(4):247-259. [doi: 10.1080/02703181.2019.1642975]
58. Shafi H, Awan WA, Olsen S, Siddiqi FA, Tassadaq N, Rashid U, et al. Sensors (Basel) 2023;23(24):9718 [FREE Full text] [doi: 10.3390/s23249718] [Medline: 38139564]
59. Puengtanom S, Suttanon P. Psychometric properties of falls risk assessment in older people with mild cognitive impairment and dementia: a systematic review. Vajira Med J 2018;62(4):335-344. [doi: 10.14456/vmj.2018.29]
60. Hofheinz M, Schusterschitz C. Dual task interference in estimating the risk of falls and measuring change: a comparative, psychometric study of four measurements. Clin Rehabil 2010;24(9):831-842. [doi: 10.1177/0269215510367993] [Medline: 20562166]
61. Lemke NC, Wiloth S, Werner C, Hauer K. Validity, test-retest reliability, sensitivity to change and feasibility of motor-cognitive dual task assessments in patients with dementia. Arch Gerontol Geriatr 2017;70:169-179. [doi: 10.1016/j.archger.2017.01.016] [Medline: 28182895]
62. Rashid U, Barbado D, Olsen S, Alder G, Elvira JLL, Lord S, et al. Validity and reliability of a smartphone app for gait and balance assessment. Sensors (Basel) 2021;22(1):124. [doi: 10.3390/s22010124] [Medline: 35009667]
63. Olsen S, Rashid U, Allerby C, Brown E, Leyser M, McDonnell G, et al. Smartphone-based gait and balance accelerometry is sensitive to age and correlates with clinical and kinematic data. Gait Posture 2023;100:57-64 [FREE Full text] [doi: 10.1016/j.gaitpost.2022.11.014] [Medline: 36481647]
64. Carson N, Leach L, Murphy KJ. A re-examination of Montreal Cognitive Assessment (MoCA) cutoff scores. Int J Geriatr Psychiatry 2018;33(2):379-388. [doi: 10.1002/gps.4756] [Medline: 28731508]
65. Twisk J, Bosman L, Hoekstra T, Rijnhart J, Welten M, Heymans M. Different ways to estimate treatment effects in randomised controlled trials. Contemp Clin Trials Commun 2018;10:80-85. [doi: 10.1016/j.conctc.2018.03.008]
66. de Boer MR, Waterlander WE, Kuijper LD, Steenhuis IH, Twisk JW. Testing for baseline differences in randomized controlled trials: an unhealthy research behavior that is hard to eradicate. Int J Behav Nutr Phys Act 2015;12:4 [FREE Full text] [doi: 10.1186/s12966-015-0162-z] [Medline: 25616598]
67. Richardson JT. Eta squared and partial eta squared as measures of effect size in educational research. Educ Res Rev 2011;6(2):135-147. [doi: 10.1016/j.edurev.2010.12.001]
68. Fewtrell MS, Kennedy K, Singhal A, Martin RM, Ness A, Hadders-Algra M, et al. How much loss to follow-up is acceptable in long-term randomised trials and prospective studies? Arch Dis Child 2008;93(6):458-461 [FREE Full text] [doi: 10.1136/adc.2007.127316] [Medline: 18495909]
69. Fitts PM, Posner MI. Human Performance. Belmont, CA: Brooks/Cole Publishing; 1967.
70. Htut TZC, Hiengkaew V, Jalayondeja C, Vongsirinavarat M. Effects of physical, virtual reality-based, and brain exercise on physical, cognition, and preference in older persons: a randomized controlled trial. Eur Rev Aging Phys Act 2018;15:10 [FREE Full text] [doi: 10.1186/s11556-018-0199-5] [Medline: 30305852]
71. Park J, Yim J. A new approach to improve cognition, muscle strength, and postural balance in community-dwelling elderly with a 3-D virtual reality kayak program. Tohoku J Exp Med 2016;238(1):1-8 [FREE Full text] [doi: 10.1620/tjem.238.1] [Medline: 26656425]
72. Yang JG, Thapa N, Park HJ, Bae S, Park KW, Park JH, et al. Virtual reality and exercise training enhance brain, cognitive, and physical health in older adults with mild cognitive impairment. Int J Environ Res Public Health 2022;19(20):13300 [FREE Full text] [doi: 10.3390/ijerph192013300] [Medline: 36293881]
73. Sakhare AR. A Virtual Reality Exergaming System to Enhance Brain Health in Older Adults at Risk for Alzheimer's Disease. Los Angeles, California: University of Southern California; 2021.
74. Phirom K, Kamnardsiri T, Sungkarat S. Beneficial effects of interactive physical-cognitive game-based training on fall risk and cognitive performance of older adults. Int J Environ Res Public Health 2020;17(17):6079 [FREE Full text] [doi: 10.3390/ijerph17176079] [Medline: 32825555]
75. Sweller J, Ayres P, Kalyuga S. Cognitive Load Theory. New York, NY, USA: Springer New York; 2011.
76. Plummer P, Altmann L, Feld J, Zukowski L, Najafi B, Giuliani C. Attentional prioritization in dual-task walking: effects of stroke, environment, and instructed focus. Gait Posture 2020;79:3-9 [FREE Full text] [doi: 10.1016/j.gaitpost.2020.03.013] [Medline: 32302930]
77. Nashner L. Balance function assessment and management. Ear Hear 2008;29(6):972. [doi: 10.1097/AUD.0b013e31818bbd1d]
78. Imaoka Y, Hauri L, Flury A, de Bruin ED. Linking cognitive functioning and postural balance control through virtual reality environmental manipulations. Front Aging Neurosci 2022;14:954050 [FREE Full text] [doi: 10.3389/fnagi.2022.954050] [Medline: 36118684]
79. Schwenk M, Sabbagh M, Lin I, Morgan P, Grewal GS, Mohler J, et al. Sensor-based balance training with motion feedback in people with mild cognitive impairment. J Rehabil Res Dev 2016;53(6):945-958. [doi: 10.1682/JRRD.2015.05.0089] [Medline: 28475201]
80. Sant'Anna P, de Oliveira Silva F, de Mello Alves Rodrigues AC, Plácido J, Ferreira JV, Meereis ECW, et al. Posturographic analysis of older adults without dementia and patients with alzheimer's disease: a cross-sectional study. Dement Neuropsychol 2019;13(2):196-202 [FREE Full text] [doi: 10.1590/1980-57642018dn13-020008] [Medline: 31285794]

XSL•FO
RenderX

81. Yogev-Seligmann G, Hausdorff JM, Giladi N. The role of executive function and attention in gait. Mov Disord 2008;23(3):329-42; quiz 472. [doi: 10.1002/mds.21720] [Medline: 18058946]
82. Pacheco TBF, de Medeiros CSP, de Oliveira VHB, Vieira ER, de Cavalcanti FAC. Effectiveness of exergames for improving mobility and balance in older adults: a systematic review and meta-analysis. Syst Rev 2020;9(1):163. [doi: 10.1186/s13643-020-01421-7] [Medline: 32682439]
83. Ghani U, Niazi I, Signal N, Kumari N, Amjad I, Haavik H, et al. The influence of multiple cognitive workload levels of an exergame on dorsal attention network connectivity at the source level. Physiol Behav 2024;284:114628. [doi: 10.1016/j.physbeh.2024.114628] [Medline: 38977259]
84. Liao Y, Chen I, Lin Y, Chen Y, Hsu W. Effects of virtual reality-based physical and cognitive training on executive function and dual-task gait performance in older adults with mild cognitive impairment: a randomized control trial. Front Aging Neurosci 2019;11:162 [FREE Full text] [doi: 10.3389/fnagi.2019.00162] [Medline: 31379553]
85. Adcock M, Fankhauser M, Post J, Lutz K, Zizlsperger L, Luft AR, et al. Effects of an in-home multicomponent exergame training on physical functions, cognition, and brain volume of older adults: a randomized controlled trial. Front Med (Lausanne) 2019;6:321 [FREE Full text] [doi: 10.3389/fmed.2019.00321] [Medline: 32047751]
86. Fang Q, Ghanouni P, Anderson SE, Touchett H, Shirley R, Fang F, et al. Effects of exergaming on balance of healthy older adults: a systematic review and meta-analysis of randomized controlled trials. Games Health J 2020;9(1):11-23. [doi: 10.1089/g4h.2019.0016] [Medline: 31800322]
87. Manser P, Herold F, de Bruin ED. Components of effective exergame-based training to improve cognitive functioning in middle-aged to older adults - a systematic review and meta-analysis. Ageing Res Rev 2024;99:102385 [FREE Full text] [doi: 10.1016/j.arr.2024.102385] [Medline: 38914262]
88. Guadagnoli MA, Lee TD. Challenge point: a framework for conceptualizing the effects of various practice conditions in motor learning. J Mot Behav 2004;36(2):212-224. [doi: 10.3200/JMBR.36.2.212-224] [Medline: 15130871]
89. Glatt RM, Patis C, Miller KJ, Merrill DA, Stubbs B, Adcock M, et al. Front Sports Act Living 2024;6:1449699. [doi: 10.3389/fspor.2024.1449699] [Medline: 39712081]
90. Lee K. Home-based exergame program to improve physical function, fall efficacy, depression and quality of life in community-dwelling older adults: a randomized controlled trial. Healthcare (Basel) 2023;11(8):1109. [doi: 10.3390/healthcare11081109] [Medline: 37107943]
91. Jovanov E, Frith K, Madhushri P, Hunter A, Coffey S, Milenkovic A. Gender differences in mobility of elderly: measurements and interventions to improve mobility. In: Legato M, editor. Principles of Gender-Specific Medicine. Cambridge, Massachusetts: Academic Press; 2017:639-653.

Abbreviations

ANCOVA: analysis of covariance
AP: anteroposterior
CONSORT: Consolidated Standards of Reporting Trials
CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
C-TUG: cognitive time up and go
EC: eyes closed
EO: eyes open
G&B: Gait & Balance
HF: head forward
HT: head turn
MCI: mild cognitive impairment
Mini-BESTest: Mini-Balance Evaluation Systems Test
ML: mediolateral
MoCA: Montreal Cognitive Assessment
TUG: time up and go

XSL•FO
RenderX

Edited by A Coristine; submitted 17.03.25; peer-reviewed by E Dove, P Manser, F Herold; comments to author 22.07.25; revised version received 11.08.25; accepted 15.09.25; published 18.11.25.

Please cite as:
Saeed A, Amjad I, Niazi IK, Alzahrani A, Shafiq M, Haavik H
Effects of Balance-Based Exergame Training With Variable Difficulty on Balance and Spatiotemporal Gait Outcomes in Adults With Mild Cognitive Impairment: Randomized Controlled Trial
JMIR Serious Games 2025;13:e74092
URL: https://games.jmir.org/2025/1/e74092
doi:10.2196/74092
PMID:

©Aruba Saeed, Imran Amjad, Imran Khan Niazi, Abdullah Alzahrani, Muhammad Shafiq, Heidi Haavik. Originally published in JMIR Serious Games (https://games.jmir.org), 18.11.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Effects of an Immersive Virtual Reality–Based Exercise Intervention on Psychological and Physiological Outcomes in College Students: Randomized Controlled Trial

Wenxi Liu[1], PhD; Daniel J McDonough[2], PhD; John Oginni[3], MS; Jessh Mavoungou[3], MS; Zan Gao[3], PhD

[1]Department of Physical Education, Shanghai Jiao Tong University, Shanghai, Shanghai, China

[2]School of Public Health, University of Minnesota-Twin Cities, Minneapolis, MN, United States

[3]Department of Kinesiology, Recreation, and Sport Studies, University of Tennessee at Knoxville, Knoxville, TN, United States

Corresponding Author:
Zan Gao, PhD
Department of Kinesiology, Recreation, and Sport Studies
University of Tennessee at Knoxville
328 HPER Building
1914 Andy Holt Avenue
Knoxville, TN, 37996-2700
United States
Phone: 1 865 974 7971
Email: zan@utk.edu

Abstract

Background: Physical inactivity is a major public health issue among college students, often exacerbated by academic pressures and lifestyle shifts. Traditional exercise interventions often face challenges with adherence due to low motivation and engagement. Immersive virtual reality (VR)–based exercise interventions may address these barriers by providing interactive and motivating experiences, yet empirical evidence regarding their psychological and physiological benefits remains scarce.

Objective: This study aims to evaluate the effects of a 4-week immersive VR-based exercise intervention on psychological and physiological health outcomes in college students.

Methods: A randomized controlled trial was conducted involving 36 college students randomized into either a VR exercise group (n=17, 47%) or a no-intervention control group (n=19, 53%). Participants in the VR group engaged in immersive VR cycling sessions (two 60-minute sessions weekly) using the VirZoom VR system, while the control group continued their normal routines. Psychological outcomes were assessed during preintervention and postintervention assessments, including exercise motivation (Behavioral Regulation in Exercise Questionnaire-2), mood states (Brunel Mood Scale), and depressive symptoms (Beck Depression Inventory). Physiological outcomes assessed were cardiovascular fitness (3-Minute Step Test) and body composition (bioelectrical impedance analysis). We used a 2-way repeated measures ANOVA to analyze the effects of the intervention.

Results: Significant time×group interactions indicated enhanced intrinsic motivation (P=.02; η^2=0.25); improved mood states with increased vigor (P=.01; η^2=0.18); and decreased confusion (P=.01; η^2=0.17), fatigue (P=.02; η^2=0.16), and tension (P=.003; η^2=0.24) in the VR group. Depressive symptoms were also significantly reduced (P=.03; η^2=0.14). Physiological outcomes showed significant improvements in the VR group, including decreased body fat percentage (P<.001; η^2=0.34) and enhanced cardiovascular fitness (P<.001; η^2=0.47) compared to the control group.

Conclusions: This study indicated that a 4-week immersive VR-based exercise intervention may confer short-term psychological and physiological benefits among college students compared to the no-intervention control group under COVID-19 pandemic–related constraints. Future studies should adopt active control designs and be conducted in real-world settings, incorporating objectively determined intensity monitoring and follow-up to further investigate effectiveness and real-world scalability.

Trial Registration: ClinicalTrials.gov NCT06902727; https://clinicaltrials.gov/study/NCT06902727

(JMIR Serious Games 2025;13:e75777) doi:10.2196/75777

XSL•FO
RenderX

KEYWORDS

body composition; cardiovascular fitness; depression; health; mood; motivation

Introduction

Physical inactivity has emerged as a major public health issue worldwide, linked to an increased risk of chronic diseases, such as obesity, cardiovascular diseases, and mental health disorders [1]. Despite widespread awareness of these risks, a significant portion of the global population remains inactive [2]. In particular, college students face unique challenges that contribute to physical inactivity, including the demands of academic life, transitioning to independent living, and balancing work and social commitments [3]. Research highlights a steep decline in physical activity (PA) levels among this population, with fewer than half meeting the American College of Sports Medicine–recommended guidelines of 150 minutes of moderate to vigorous PA weekly [4]. Addressing physical inactivity in this population is a public health priority, yet existing interventions have had limited success in sustaining long-term engagement [5]. The ongoing challenges of stress and mental health in higher education further underscore the need for interventions that promote both physical and mental well-being [6-8].

Virtual reality (VR)–based exercise represents a promising solution. By combining physical exertion with immersive and interactive virtual environments, VR-based exercise addresses common barriers to PA, such as monotony and a lack of motivation [9-11]. Immersive VR systems offer unique features that enhance the user experience, including customizable challenges, real-time feedback, and an increased sense of "presence," the feeling of "being there" in a virtual environment [12,13]. These elements not only increase enjoyment but also reduce perceived exertion and discomfort, facilitating greater adherence to exercise programs [14-16].

From a psychological perspective, VR aligns with the principles of self-determination theory (SDT), which emphasizes the importance of autonomy, competence, and relatedness in sustaining behavior change [17-19]. VR-based exercise can enhance intrinsic motivation by providing users with choices (autonomy), offering achievable challenges and performance feedback (competence), and creating opportunities for social interaction in multiplayer environments (relatedness), fostering a sense of personal accomplishment and enjoyment [20-22]. Moreover, VR-based exercise has demonstrated potential for improving mental health outcomes, including mood enhancement and reductions in depressive symptoms [10,23].

Despite the promising findings, high-quality experimental evidence on the effectiveness of VR-based exercise interventions in healthy populations remains limited. Previous studies have primarily focused on acute effects or clinical populations, leaving significant gaps in understanding the broader applicability of VR-based exercise [10,24-26]. Therefore, the purpose of this study was to examine the effects of a 4-week VR-based exercise intervention on health-related psychological (ie, motivation, mood, and depressive symptoms) and physiological (ie, body composition and cardiovascular fitness) outcomes among college students. This study hypothesized that intervention participants would have significant improvement in health-related psychological and physiological outcomes compared to a no-intervention control group after a 4-week VR-based exercise intervention. The findings of this study would provide insights into the potential of VR-based exercise to improve health-related outcomes in the young adult population.

Methods

Study Design

This study used a 2-arm randomized controlled trial to evaluate the efficacy of a 4-week VR-based exercise intervention on health-related psychological and physiological outcomes in college students. The use of a no-intervention control group was chosen to first establish the efficacy of this specific VR intervention before comparing it to other forms of PA. Participants were randomly assigned to either the intervention group or the control group, with assessments conducted at baseline and after the intervention (4 weeks).

Participants

Participants were recruited from a single large public university in the Midwestern United States via university-wide email newsletters and departmental listserves, targeting both undergraduate and graduate students. Recruitment and data collection took place between August 2020 and March 2021. We determined the required sample size using G*Power (version 3.1; Heinrich Heine University Düsseldorf) for a repeated measures ANOVA with a between-factor interaction. With an effect size f(V) of 0.25 (a medium effect), an α of .05, and power of 0.80, with 2 groups and 2 measurements and a correlation among repeated measures of 0.5, the required total sample size was 34. Our final analyzed sample of 36 exceeded this requirement.

The inclusion criteria were as follows: (1) college students aged between 18 and 35 years, (2) without diagnosed physical or mental disabilities, (3) willing to provide informed consent and complete the Physical Activity Readiness Questionnaire in English, and (4) no self-reported motion sickness symptoms during a brief VR familiarization and screening session. Participants were also screened for any musculoskeletal issues that would contraindicate participation in this study. Information on previous exercise experience was also collected.

Ethical Considerations

This study was approved by the institutional review board of the University of Minnesota Twin Cities (00010580). All procedures were conducted in accordance with the ethical standards of the institutional research committee and the Declaration of Helsinki 1964 and its later amendments. All participants provided written informed consent before participating in this study. The consent process included a detailed explanation of the study procedures, potential risks and benefits, the right to withdraw at any time without penalty, and contact information for the research team and the institutional

XSL•FO
RenderX

review board. All study data were deidentified and assigned unique participant ID numbers. Data were stored on password-protected computers accessible only to the research team. Hard copies of consent forms were stored separately from study data in locked filing cabinets. In addition, participants received a US $20 incentive gift card upon completing all assessments (details in Multimedia Appendices 1-4). This trial was registered on ClinicalTrials.gov (NCT06902727).

Measurements

Motion Sickness

Participants were asked to complete a 2-item survey on a 3-point Likert scale (1=not at all; 3=a lot) for assessing the symptoms of motion sickness. Two questions were included in the survey: (1) “When riding the VR bike, I feel sick to my stomach” (2) “When riding the VR bike, I feel dizzy.” The survey was administered during the preintervention screening session.

Demographics and Anthropometrics

Participants’ age, race or ethnicity, and sex were obtained via baseline self-administered surveys. Participants’ height was measured using the Seca stadiometer. Body weight and body fat (BF) percentage were measured by the Tanita BC-558 IRONMAN Segmental Body Composition Monitor. In addition, participants’ previous VR experiences were assessed at baseline.

PA Assessment

Participants’ free-living PA levels were assessed via the International Physical Activity Questionnaire (IPAQ) Short Form. The IPAQ Short Form assessed 3 specific types of activity: walking, moderate-intensity activities, and vigorous-intensity activities. Participants were asked to recall the time spent (in minutes) walking, performing moderate-intensity PA, and engaging in vigorous-intensity PA in the past 7 days. Regarding the IPAQ scoring, the weekly PA volume was computed by weighing each type of activity by its energy requirements defined in metabolic equivalents (METs) to yield a score in MET-minutes. The mean weekly PA levels were calculated as the study outcomes to examine the difference between the 2 groups at baseline.

Exercise Motivation

Participants’ general exercise motivation was assessed by the Behavioral Regulation in Exercise Questionnaire-2 (BREQ-2) [27]. The BREQ-2 has 19 items assessing the continuum of behavioral regulation in exercise, which includes amotivation, external regulation, introjected regulation, identified regulation, and intrinsic regulation. The BREQ-2 has demonstrated good reliability and validity in previous research.

Mood States

Participants’ mood states were measured using the Brunel Mood Scale (BRUMS) [28]. The BRUMS assesses mood across 6 subscales (anger, confusion, depression, fatigue, tension, and vigor) using a 24-item questionnaire, with responses ranging from 0 (not at all) to 4 (extremely). The BRUMS has shown good psychometric properties in adult populations.

Depressive Symptoms

The Beck Depression Inventory was used to assess participants’ depressive symptoms [29]. This is a 21-item, self-report rating inventory that measures characteristic attitudes and symptoms of depression. The total score was calculated to evaluate the severity of depressive symptoms.

Cardiovascular Fitness

Participants’ cardiovascular fitness (CRF) was assessed by the 3-Minute Step Test [30]. Participants stepped on and off a 30-cm plyometric box in time with a metronome set to 96 bpm for 3 minutes. Heart rate was measured immediately after the test via Omron Fingertip Oximeter. The change in heart rate from baseline to after the test served as the primary outcome for cardiovascular fitness.

Procedure

Recruitment was conducted via online channels, including university email lists and departmental newsletters. Interested college students contacted the investigator to schedule a preintervention screening session. To mitigate motion sickness risk, potential participants underwent a 15-minute VR biking session using the VirZoom VR bike system during which they engaged in both “Le Tour” and “Race Car” exergames. After the session, participants completed the 2-item motion sickness survey. Only those who completed the session without reporting motion sickness symptoms were enrolled in the study. Enrolled participants completed demographic surveys (age, sex, ethnicity, and previous VR experience) and underwent physiological and psychological assessments, including height, weight, body composition (BF percentage and BMI), CRF (step test), exercise motivation (BREQ-2), mood (BRUMS), and depressive symptoms (Beck Depression Inventory). Following baseline assessments, participants were randomized into the intervention or control group. The intervention group engaged in VR-based exercise sessions twice weekly for 4 weeks, while the control group maintained their usual activities. After 4 weeks, all participants returned for postintervention assessments identical to the baseline measures.

Randomization

Randomization was executed using a computer-generated sequence in Microsoft Excel, allocating participants in a 1:1 ratio to either the intervention or control group. This process ensured equal distribution of participants across both groups, minimizing selection bias.

Intervention Group

Participants in the intervention group engaged in an immersive VR cycling regimen using the commercial VirZoom VR exercise bike connected to a PlayStation VR system (Sony), which includes a head-mounted display. The intervention consisted of exercising for 1 hour per session, twice weekly, for 4 weeks. Each session included 2 VR exergames (demonstrated in Figure 1). The first game was “Le Tour,” a cycling race game through a virtual scenic mountain road featuring 15 timing gates. Participants aimed to meet various performance goals at each gate, such as maintaining specific pedaling speeds or drafting behind virtual cyclists. The game offered customizable workout

XSL•FO
RenderX

plans with 2 intensity modes—comfort (light to moderate) and intense (moderate to vigorous). Participants were encouraged to exercise at a moderate-intensity level, and perceived exertion was monitored. After 30 minutes of gameplay, the system automatically transitioned participants to the second game to maintain engagement and adherence. The second game was "Race Car," a high-speed racing game where participants pedal to control a virtual race car, aiming to surpass all virtual competitors. The game features a "ghost car" representing the fastest previous performance, encouraging participants to beat their own or others' records. Our previous study demonstrated that VirZoom VR cycling effectively elicited moderate to vigorous PA, with most of the exercise duration classified as vigorous intensity [11,20,31].

Figure 1. Participants engaging in immersive virtual reality–based exercise intervention using the VirZoom VR exercise bike connected to the PlayStation VR system.



Control Group

Participants in the control group were instructed to maintain their usual routine and daily activities, without engaging in any other PA interventions during the 4-week study period. We did not formally measure the PA levels of the control group throughout the study. Both groups received identical assessments at baseline and after intervention.

Intervention Fidelity

To ensure intervention fidelity, several strategies were implemented, including providing detailed instructions on survey completion, conducting preintervention tutorials and hands-on VR exercise sessions, and tracking attendance and adherence through exercise session logs. A Microsoft Excel–based tracking system reminded participants of upcoming assessments, while individualized schedules and email reminders were sent to intervention participants.

Statistical Analysis

Data were analyzed using SPSS Statistics (version 27.0; IBM Corp). Descriptive statistics (means, SDs, and frequencies) were calculated for demographic variables, physiological measures, and psychological outcomes. Baseline differences between the intervention and control groups were evaluated using independent 2-tailed *t* tests for continuous variables and chi-square tests for categorical variables to confirm effective randomization. A robust regression–based outlier detection procedure was used to identify outliers, and any identified outliers were excluded from the analyses. A 2-way repeated measures ANOVA was used to examine the changes in the following health-related physiological and psychological outcomes: exercise motivation (amotivation, external regulation, introjected regulation, identified regulation, and intrinsic regulation), mood states (anger, confusion, depression, fatigue, tension, and vigor), depressive symptoms (total depression score), and physiological measures (BF percentage and CRF).

Results

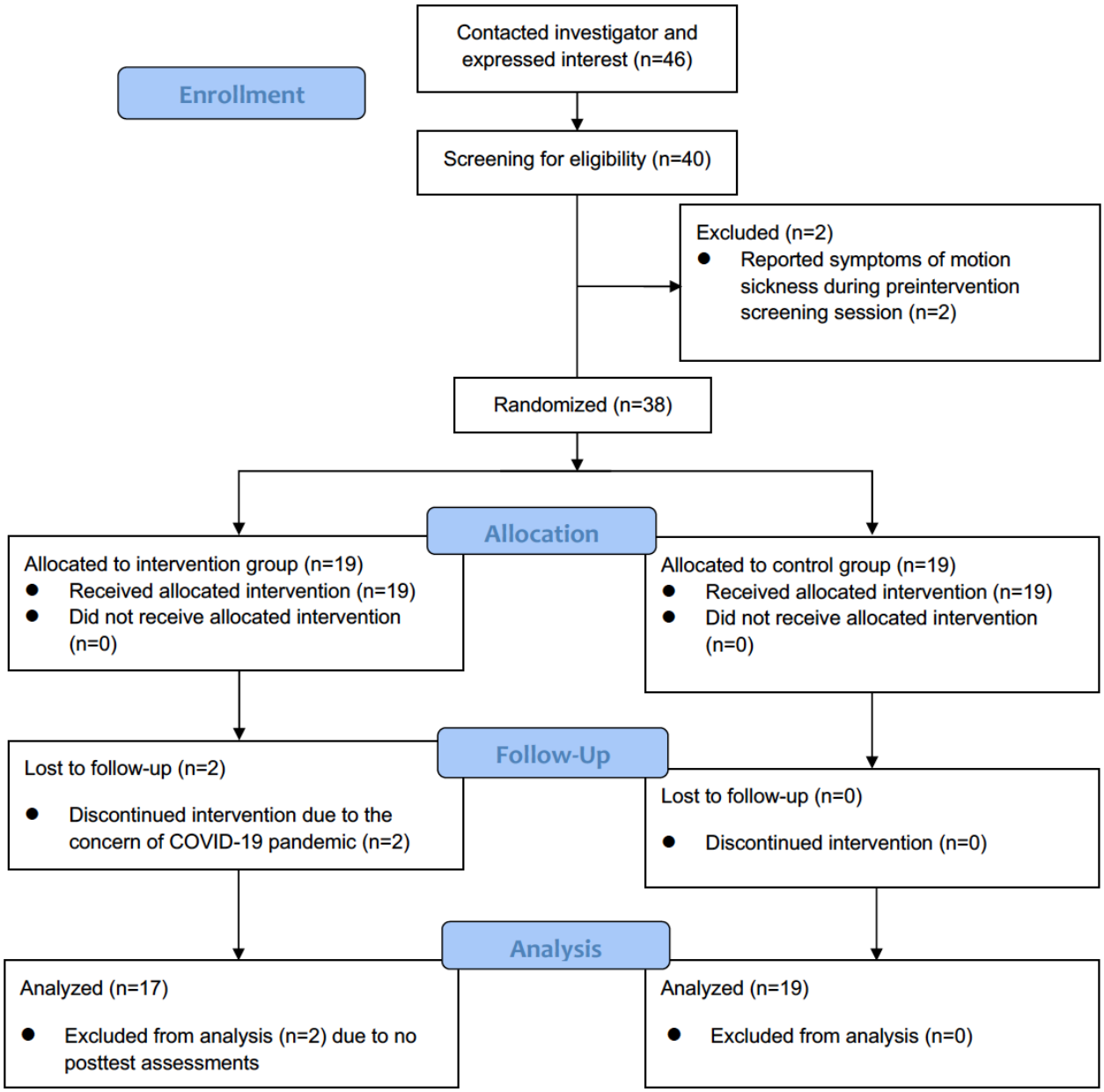
Participant Flow

Figure 2 demonstrates the participant flow of this study. At the beginning, 46 college students contacted the principal investigator and expressed interest in participation via email. Next, 40 (87%) students came to the laboratory for the prestudy screening session. College students who completed the screening session and reported no motion sickness symptoms were eligible

XSL•FO
RenderX

for this study. In total, 2 (5%) college students reported motion sickness symptoms and thus were excluded from this study. As a result, 38 (95%) eligible college students were enrolled in the study and randomized into intervention and control groups. Notably, 2 (5%) participants dropped out of the intervention group during the first week of intervention, expressing concerns about participating in an in-person study during the COVID-19 pandemic. Finally, 36 participants (intervention group: n=17, 47%; control group: n=19, 53%) completed baseline and postintervention assessments. Their data were entered for statistical analysis.

Figure 2. Consolidated Standards of Reporting Trials (CONSORT) flow diagram of participant flow.



Baseline Participant Characteristics

The descriptive statistics of participants' baseline demographic, physiological, and psychological outcomes by group are shown in Table 1. The chi-square test and the independent *t* test indicated that there were no statistically significant differences in demographic, physiological, and psychological characteristics between the intervention and control groups at baseline. Regarding previous VR experience, 44.4% (16/36) of the participants reported having previous VR experience, and 55.6% (20/36)reported no previous VR experience. The chi-square test of VR experience between 2 groups indicated there was no statistically significant group difference at baseline.

XSL•FO
RenderX

Table 1. Baseline group comparisons for study outcomes.

Characteristic	Intervention (n=17)	Control (n=19)	*P* value
Sex, n (%)			.54
Male	8 (47)	7 (37)	
Female	9 (53)	12 (63)	
Race or ethnicity, n (%)			.41
Asian	10 (59)	10 (53)	
White	6 (35)	5 (26)	
Other	1 (6)	4 (21)	
Age (y), mean (SD)	22.82 (2.51)	24.84 (4.14)	.09
Height (cm), mean (SD)	171.26 (7.21)	169.71 (7.35)	.53
Weight (kg), mean (SD)	74.38 (16.49)	73.56 (14.75)	.88
BMI (kg/m^2), mean (SD)	25.21 (4.36)	25.37 (3.69)	.91
Body fat percentage, mean (SD)	26.11 (7.14)	27.65 (5.18)	.46
Cardiovascular fitness (bpm), mean (SD)	113 (22.10)	115.95 (16.64)	.65
Physical activity (metabolic equivalent min/wk)	1152.06 (618.57)	1080.26 (935.04)	.79
Self-determination theory– based motivation, mean (SD)			
Amotivation	1.57 (0.64)	1.75 (0.64)	.41
External regulation	1.94 (0.71)	2.32 (0.86)	.17
Introjected regulation	3.33 (1.14)	2.86 (0.88)	.17
Identified regulation	4.00 (0.71)	3.74 (0.65)	.25
Intrinsic regulation	3.71 (0.94)	3.63 (0.86)	.81
Mood states, mean (SD)			
Anger	1.28 (0.57)	1.60 (1.09)	.30
Confusion	1.53 (0.75)	1.67 (0.97)	.63
Depression	1.41 (0.69)	1.62 (0.81)	.42
Fatigue	2.00 (0.54)	2.09 (1.01)	.74
Tension	1.85 (0.72)	1.84 (1.03)	.97
Vigor	2.74 (0.50)	2.51 (0.74)	.30
Depression score, mean (SD)	5.94 (5.68)	6.68 (6.25)	.71

Psychological Outcomes

The descriptive statistics for exercise motivation by group at baseline and 4 weeks are presented in Table 2. Regarding the 5 subscales of motivational regulation, the 2-way repeated measures ANOVA indicated significant main effect of time on amotivation ($F_{1,34}$=6.10; P=.02; η^2=0.15), external regulation ($F_{1,34}$=4.37; P=.04; η^2=0.11), identified regulation ($F_{1,34}$=6.55; P=.02; η^2=0.16), and intrinsic regulation ($F_{1,34}$=18.11; P<.001; η^2=0.35). The significant time×group interactions were observed on identified regulation ($F_{1,34}$=6.55; P=.02; η^2=0.16) and intrinsic regulation ($F_{1,34}$=11.21; P=.02; η^2=0.25). Post hoc analyses further indicated that intervention participants' identified regulation and intrinsic regulation significantly increased after the 4-week intervention compared to the control group.

The descriptive statistics for mood states are presented in Table 3. Regarding the 6 subscales of mood states, the 2-way repeated measures ANOVA indicated a significant main effect of time on anger ($F_{1,34}$=4.54; P=.04; η^2=0.12), fatigue ($F_{1,34}$=5.35; P=.03; η^2=0.14), and vigor ($F_{1,34}$=50.79; P<.001; η^2=0.6). The significant time×group interactions were observed on confusion ($F_{1,34}$=6.72; P=.01; η^2=0.17), fatigue ($F_{1,34}$=6.46; P=.02; η^2=0.16), tension ($F_{1,34}$=10.44; P=.003; η^2=0.24), and vigor ($F_{1,34}$=7.22; P=.01; η^2=0.18). Our data further indicated that intervention participants perceived less sense of confusion, fatigue, tension, and greater vigor compared to the control group at postintervention assessment.

For depressive symptoms, the results revealed a significant time×group interaction on the mean scores of depression symptom assessment ($F_{1,34}$=5.53; P=.03; η^2=0.14). In the intervention group, the mean total depression score decreased

XSL•FO
RenderX

from 5.94 (SD 5.68) at baseline to 3.24 (SD 4.28) after 4 weeks (*P*=.03). In contrast, in the control group, the mean total depression score increased from 6.68 (SD 6.25) at baseline to 7.21 (SD 6.75) after 4 weeks (*P*=.03).

Table 2. Descriptive statistics for self-determination theory–based exercise regulations by group.

	Baseline, mean (SD)	After 4 weeks, mean (SD)	*P* value
Amotivation			.05
Intervention	1.57 (0.64)	1.07 (0.15)	
Control	1.75 (0.64)	1.61 (0.89)	
External regulation			.21
Intervention	1.94 (0.71)	1.76 (0.56)	
Control	2.32 (0.86)	1.99 (0.91)	
Introjected regulation			.06
Intervention	3.33 (1.14)	3.67 (0.99)	
Control	2.86 (0.88)	3.05 (0.79)	
Identified regulation			.02
Intervention	4.00 (0.71)	4.59 (0.33)	
Control	3.74 (0.65)	3.74 (0.61)	
Intrinsic regulation			.02
Intervention	3.71 (0.94)	4.81 (0.26)	
Control	3.63 (0.86)	3.76 (0.61)	

Table 3. Descriptive statistics for mood subscales by group.

Mood subscale and group	Baseline, mean (SD)	After 4 weeks, mean (SD)	*P* value
Anger			.26
Intervention	1.28 (0.57)	1.01 (0.06)	
Control	1.59 (1.09)	1.51 (0.95)	
Confusion			.01
Intervention	1.53 (0.75)	1.15 (0.18)	
Control	1.67 (0.97)	1.84 (0.97)	
Depression			.06
Intervention	1.41 (0.69)	1.01 (0.06)	
Control	1.62 (0.81)	1.63 (0.87)	
Fatigue			.02
Intervention	2.00 (0.54)	1.44 (0.40)	
Control	2.09 (1.01)	2.12 (1.01)	
Tension			.003
Intervention	1.85 (0.72)	1.35 (0.52)	
Control	1.84 (1.03)	2.00 (1.11)	
Vigor			.01
Intervention	2.74 (0.50)	4.01 (0.50)	
Control	2.51 (0.74)	3.09 (0.88)	

Physiological Outcomes

The descriptive statistics for BMI, BF percentage, and CRF by group at baseline and 4 weeks are presented in Table 4. The results indicated that there was a significant main effect of time on CRF ($F_{1,34}$=20.65; *P*<.001; η^2=0.38). However, there was no significant main effect of time on BMI ($F_{1,34}$=1.48; *P*=.23; η^2=0.04) and BF percentage ($F_{1,34}$=3.07; *P*=.09; η^2=0.08). In

XSL•FO
RenderX

addition, significant time×group interactions were observed on BF percentage ($F_{1,34}$=17.26; P<.001; η^2=0.34) and CRF ($F_{1,34}$=30.05; P<.001; η^2=0.47). Post hoc analyses further indicated that intervention participants' BF percentage and CRF had significant improvement as compared to the control group at the postintervention assessment.

Table 4. Descriptive statistics for physiological outcomes by group.

Physiological outcomes	Baseline, mean (SD)	After 4 weeks, mean (SD)	*P* value
BMI (kg/m^2)			.55
Intervention	25.21 (4.36)	25.00 (4.05)	
Control	25.52 (3.71)	24.87 (5.11)	
Body fat percentage			<.001
Intervention	26.11 (7.14)	24.51 (6.98)	
Control	27.28 (5.09)	28.16 (4.60)	
Cardiovascular fitness (bpm)			<.001
Intervention	113 (22.10)	96.12 (16.48)	
Control	115.19 (16.64)	117.53 (16.56)	

Discussion

Principal Findings

This study demonstrated that a 4-week immersive VR exercise program significantly improved certain psychological and physiological outcomes among college students. Participants in the VR group reported higher intrinsic and identified motivation, more positive mood states, and fewer depressive symptoms while also achieving measurable reductions in BF percentage and better CRF compared with the control group. Overall, the VR-based exercise intervention produced meaningful health outcome changes, suggesting that immersive engagement in PA may amplify exercise effects even with a relatively brief exposure.

Comparison to Prior Work

The findings of this study align with and extend a growing body of evidence that immersive VR exercise can enhance motivation, engagement, and health outcomes. Earlier studies demonstrated that immersive and game-based exercise elicits higher enjoyment, perceived competence, and self-efficacy than traditional exercise and that these qualities translate into greater adherence and positive affect [15,20-22]. More recently, Merola et al [32] reported that a VR high-intensity interval boxing game elicited vigorous heart rate responses while maintaining intrinsic motivation comparable to traditional circuit training, and Boots et al [33] found that active VR games, such as *Beat Saber* and *Gorilla Tag*, produced light to moderate PA levels with high enjoyment among youth. A qualitative study in China showed that young adults with overweight or obesity viewed VR exergames as enjoyable and motivating, although concerns about cost and cybersickness remained [34]. Complementing these findings, a 12-week trial observed that stationary cycling with VR significantly reduced depressive symptoms and produced higher remission rates than non-VR exercise among adults with mild to moderate depression [35]. Likewise, systematic reviews suggest that VR interventions not only alleviate psychological distress but also foster positive mental health [36].

The sustained improvements in motivation and mood observed in this study build upon those acute findings and can be understood through the lens of SDT. The immersive VR environment in our intervention may provide autonomy through user-controlled difficulty and task selection; competence through real-time feedback and visible progress; and relatedness through simulated social features, such as "ghost riders." Collectively, these elements may have supported basic psychological needs, fostering intrinsic motivation and self-determined exercise behavior; however, this mechanism was not directly measured in this study. Previous studies have shown that when these needs are met, individuals experience enhanced enjoyment, persistence, and well-being [18,19]. The concurrent improvements in intrinsic regulation, vigor, and reduced depressive symptoms observed here likely reflect this motivational pathway. This is consistent with recent findings from Zhang et al [35], who reported that a 12-week VR cycling program significantly reduced depressive symptoms and achieved higher remission rates than non-VR exercise in adults with mild to moderate depression. Together, these results may point to the potential of immersive VR exercise to simultaneously promote mental health and motivation through satisfaction of psychological needs.

From a mental health perspective, our results are consistent with recent evidence indicating that VR interventions not only alleviate psychological distress but also enhance positive functioning. A systematic review of VR interventions for mental health found that immersive environments improve both symptom reduction and well-being by promoting engagement, relaxation, and self-efficacy. The review emphasized that VR-based "positive technologies" can foster emotional resilience and flourishing, aligning with our observed increases in intrinsic motivation and mood [36]. Furthermore, Chen et al [34] found that young adults with overweight or obesity in China viewed VR exergames as highly acceptable and beneficial for motivation and psychological well-being, despite minor concerns about cost or cybersickness. These qualitative insights parallel the high adherence and enjoyment reported in our study,

XSL•FO
RenderX

underscoring VR's ability to overcome motivational barriers commonly faced in traditional exercise.

Physiologically, the improvements in cardiovascular fitness and body composition observed in our study corroborate previous findings that VR exercise can elicit moderate to vigorous intensities comparable to conventional aerobic training. Merola et al [32] observed mean heart rates exceeding 160 bpm during VR boxing, confirming that immersive games can meet intensity thresholds necessary for cardiovascular benefits. Our previous study suggested that VR cycling improved energy expenditure and adherence among university students compared with stationary cycling alone [20,31]. These convergent findings suggest that immersive engagement and attentional distraction during VR exercise may reduce perceived exertion and encourage sustained effort, leading to greater physiological adaptation within shorter time frames [37,38].

Beyond acute fitness outcomes, the integration of VR into PA programs may provide long-term behavioral advantages. Studies have shown that enjoyment and presence in VR are strong predictors of continued participation [39,40]. This aligns with the SDT perspective that affective experiences and perceived autonomy facilitate habit formation and sustained PA behavior. Therefore, immersive VR exercise may not only improve short-term outcomes but also act as a bridge toward more autonomous and lasting engagement in health-promoting activity.

Limitations

Several limitations should be noted. First, the sample comprised healthy, technology-tolerant college students, which may limit generalizability to less active or older populations. Second, individuals prone to motion sickness were excluded, potentially biasing results toward those more comfortable with VR environments. Third, physiological outcomes were measured using bioelectrical impedance and a submaximal step test, which are less precise than gold-standard laboratory assessments. Fourth, the weekly exercise dose in our intervention (2×60 min; 120 min/week) was below the American College of Sports Medicine recommendation of 150 minutes per week of moderate to vigorous PA; therefore, the observed physiological changes may be overestimated due to less accurate measurements, and the results should be interpreted with caution. Fifth, we used a brief, nonvalidated 2-item screen for cybersickness susceptibility rather than a standardized instrument (eg, Simulator Sickness Questionnaire). This may reduce replicability and limit comparability with previous VR studies. Sixth, the control group did not engage in an active exercise condition, so the contribution of nonspecific effects, such as attention or novelty (Hawthorne effect), cannot be ruled out. In addition, we did not objectively track the control group's PA during the intervention period, which may limit the certainty that they maintained usual routines and potentially bias the observed changes. Seventh, although the BREQ-2 showed improvements in motivational regulation, we did not measure SDT-based basic psychological needs (ie, autonomy, competence, and relatedness), indicating that the underlying mechanism warrants further investigation. Eighth, this study was conducted during the COVID-19 pandemic and adopted a no-intervention control arm; therefore, we could not fully disentangle the VR-specific effects from the pandemic-related contextual changes. As shown, the control group exhibited elevated depressive symptoms, increased BF percentage, and poorer CRF over time, consistent with real-world COVID-19 pandemic stressors and restricted activity opportunities. Consequently, the observed time×group interactions were likely inflated due to a combination of VR intervention effects and concurrent deterioration in the control group rather than a pure effect against a stable baseline. The findings should be interpreted with caution. Finally, the intervention was delivered in a supervised, in-person laboratory setting, which ensured safety and study fidelity under the COVID-19 pandemic context but might have increased adherence and engagement; consequently, the observed effects might not fully be generalized to unsupervised real-world settings.

Future Directions

Future research should address several key questions to advance the field of VR-based exercise. First, long-term studies are needed to determine whether the psychological and physiological benefits observed in our 4-week intervention are sustained for months or years and whether VR exercise can promote lasting behavior change and exercise adherence. Second, comparative effectiveness trials with active control groups (eg, traditional cycling and video-guided exercise) are necessary to isolate the unique benefits of VR immersion and determine the added value of VR technology beyond general exercise effects. Third, dose-response studies should investigate the optimal frequency, duration, and intensity of VR exercise sessions for maximizing health benefits while maintaining engagement and minimizing adverse effects such as motion sickness. Fourth, research should explore strategies to make VR exercise more accessible and tolerable for individuals prone to motion sickness, such as gradual exposure protocols or technological improvements. Fifth, studies should examine the effectiveness of VR exercise in more diverse populations, including different age groups, clinical populations, and individuals with varying levels of baseline fitness and technological literacy. Sixth, future studies should adopt validated and objective measures to assess physiological outcomes, such as dual-energy x-ray absorptiometry for body composition and the maximal oxygen uptake test for CRF. Finally, research should investigate the mechanisms underlying the psychological benefits of VR exercise, such as the roles of presence, flow, distraction, and social features, to inform the design of more effective interventions.

Conclusions

This study indicated that a 4-week immersive VR-based exercise may confer short-term psychological and physiological benefits among college students compared to a no-intervention control group under COVID-19 pandemic–related constraints. However, the supervised context, pandemic-era control deterioration, the untested SDT mechanism, and other design constraints temper causal inference and ecological validity. Future studies should adopt active control designs and be conducted in real-world settings, incorporating objectively determined intensity

XSL•FO
RenderX

monitoring and follow-up to further investigate the effectiveness and real-world scalability of VR exercise interventions.

Acknowledgments

The authors extend their sincere thanks to all participants for their time and commitment to this study during the COVID-19 pandemic. The authors are grateful to the University of Minnesota Twin Cities for authorizing and facilitating the conduct of this work under pandemic-related constraints. The authors also thank colleagues from the Physical Activity Epidemiology Laboratory for their valuable support and guidance throughout the project.

During manuscript preparation, ChatGPT (OpenAI) was used solely for grammar, spelling, and fluency editing of author-written text. The authors verified and edited all outputs and are fully accountable for the content. No generative artificial intelligence was used for study design, data collection, data analysis, interpretation, reference selection, or figure creation, and no confidential or identifiable data were provided to the tool.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: WL (lead), ZG (supporting)
Data curation: WL (lead)
Investigation: WL (lead)
Methodology: WL (lead), DJM (supporting), ZG (supporting)
Project administration: WL (lead), ZG (supporting)
Resources: ZG (lead)
Supervision: ZG (lead)
Writing—original draft: WL (lead)
Writing—review and editing: WL (lead), DJM (supporting), JO (supporting), JM (supporting), ZG (supporting)

Conflicts of Interest

None declared.

Multimedia Appendix 1
Study questionnaires.
[PDF File (Adobe PDF File), 147 KB - games_v13i1e75777_app1.pdf]

Multimedia Appendix 2
Study recruitment flyer.
[PDF File (Adobe PDF File), 404 KB - games_v13i1e75777_app2.pdf]

Multimedia Appendix 3
Consent form for participation.
[PDF File (Adobe PDF File), 207 KB - games_v13i1e75777_app3.pdf]

Multimedia Appendix 4
CONSORT Checklist. CONSORT: Consolidated Standards of Reporting Trials.
[PDF File (Adobe PDF File), 1136 KB - games_v13i1e75777_app4.pdf]

References

1. Ekelund U, Sanchez-Lastra MA, Dalene KE, Tarp J. Dose-response associations, physical activity intensity and mortality risk: a narrative review. J Sport Health Sci 2024 Jan;13(1):24-29 [FREE Full text] [doi: 10.1016/j.jshs.2023.09.006] [Medline: 37734548]

XSL•FO
RenderX

2. Herrmann SD, Conger SA, Willis EA, Ainsworth BE. Promoting public health through the 2024 compendium of physical activities: strategies for adults, older adults, and wheelchair users. J Sport Health Sci 2024 Nov;13(6):739-742 [FREE Full text] [doi: 10.1016/j.jshs.2024.05.013] [Medline: 38825318]
3. Buckworth J, Nigg C. Physical activity, exercise, and sedentary behavior in college students. J Am Coll Health 2004;53(1):28-34. [doi: 10.3200/JACH.53.1.28-34] [Medline: 15266727]
4. American College of Sports Medicine. ACSM's Guidelines for Exercise Testing and Prescription. 10th edition. Philadelphia, PA: Wolters Kluwer; 2018.
5. Tremblay MS, Duncan MJ, Kuzik N, Silva DA, Carson V. Towards precision 24-hour movement behavior recommendations-the next new paradigm? J Sport Health Sci 2024 Nov;13(6):743-748 [FREE Full text] [doi: 10.1016/j.jshs.2024.05.003] [Medline: 38734180]
6. Auerbach RP, Alonso J, Axinn WG, Cuijpers P, Ebert DD, Green JG, et al. Mental disorders among college students in the World Health Organization World Mental Health Surveys. Psychol Med 2016 Oct;46(14):2955-2970 [FREE Full text] [doi: 10.1017/S0033291716001665] [Medline: 27484622]
7. Lipson SK, Lattie EG, Eisenberg D. Increased rates of mental health service utilization by U.S. college students: 10-year population-level trends (2007-2017). Psychiatr Serv 2019 Jan 01;70(1):60-63 [FREE Full text] [doi: 10.1176/appi.ps.201800332] [Medline: 30394183]
8. Wang X, Hegde S, Son C, Keller B, Smith A, Sasangohar F. Investigating mental health of US college students during the COVID-19 pandemic: cross-sectional survey study. J Med Internet Res 2020 Sep 17;22(9):e22817 [FREE Full text] [doi: 10.2196/22817] [Medline: 32897868]
9. Gao Z, Chen S, Pasco D, Pope Z. A meta-analysis of active video games on health outcomes among children and adolescents. Obes Rev 2015 Sep;16(9):783-794. [doi: 10.1111/obr.12287] [Medline: 25943852]
10. Valenzuela T, Okubo Y, Woodbury A, Lord SR, Delbaere K. Adherence to technology-based exercise programs in older adults: a systematic review. J Geriatr Phys Ther 2018;41(1):49-61. [doi: 10.1519/JPT.0000000000000095] [Medline: 27362526]
11. Zeng N, Pope Z, Gao Z. Acute effect of virtual reality exercise bike games on college students' physiological and psychological outcomes. Cyberpsychol Behav Soc Netw 2017 Jul;20(7):453-457. [doi: 10.1089/cyber.2017.0042] [Medline: 28715263]
12. Slater M. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. Philos Trans R Soc Lond B Biol Sci 2009 Dec 12;364(1535):3549-3557 [FREE Full text] [doi: 10.1098/rstb.2009.0138] [Medline: 19884149]
13. Cummings JJ, Bailenson JN. How immersive is enough? A meta-analysis of the effect of immersive technology on user presence. Media Psychol 2015 May 15;19(2):272-309. [doi: 10.1080/15213269.2015.1015740]
14. Plante TG, Aldridge A, Bogden R, Hanelin C. Might virtual reality promote the mood benefits of exercise? Comput Human Behav 2003 Jul;19(4):495-509. [doi: 10.1016/s0747-5632(02)00074-2]
15. Gao Z, Zeng N, McDonough DJ, Su X. A systematic review of active video games on youth's body composition and physical activity. Int J Sports Med 2020 Aug;41(9):561-573 [FREE Full text] [doi: 10.1055/a-1152-4959] [Medline: 32375184]
16. Stewart TH, Villaneuva K, Hahn A, Ortiz-Delatorre J, Wolf C, Nguyen R, et al. Actual vs. perceived exertion during active virtual reality game exercise. Front Rehabil Sci 2022;3(4):887740 [FREE Full text] [doi: 10.3389/fresc.2022.887740] [Medline: 36189005]
17. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000 Jan;55(1):68-78. [doi: 10.1037//0003-066x.55.1.68] [Medline: 11392867]
18. Deci EL, Ryan RM. Self-determination theory: a macrotheory of human motivation, development, and health. Can Psycho 2008 Aug;49(3):182-185. [doi: 10.1037/a0012801]
19. Ryan RM, Deci EL. Self-Determination Theory: Basic Psychological Needs in Motivation, Development, and Wellness. New York, NY: Guilford Publications; Nov 06, 2018.
20. Zeng N, Liu W, Pope ZC, McDonough DJ, Gao Z. Acute effects of virtual reality exercise biking on college students' physical responses. Res Q Exerc Sport 2022 Sep;93(3):633-639 [FREE Full text] [doi: 10.1080/02701367.2021.1891188] [Medline: 34663191]
21. Ng Y, Ma F, Ho F, Ip P, Fu K. Effectiveness of virtual and augmented reality-enhanced exercise on physical activity, psychological outcomes, and physical performance: A systematic review and meta-analysis of randomized controlled trials. Comput Human Behav 2019 Oct;99:278-291. [doi: 10.1016/j.chb.2019.05.026]
22. Annesi JJ, Mazas J. Effects of virtual reality-enhanced exercise equipment on adherence and exercise-induced feeling states. Percept Mot Skills 1997 Dec;85(3 Pt 1):835-844. [doi: 10.2466/pms.1997.85.3.835] [Medline: 9399288]
23. Liu W, McDonough DJ, Gao Z. Comparing college students' mood states among immersive virtual reality, non-immersive virtual reality, and traditional biking exercise. PLoS One 2024 Jan;19(11):e0311113-e0311154 [FREE Full text] [doi: 10.1371/journal.pone.0311113] [Medline: 39570986]
24. Garcia-Agundez A, Reuter C, Becker H, Konrad R, Caserman P, Miede A, et al. Development of a classifier to determine factors causing cybersickness in virtual reality environments. Games Health J 2019 Dec;8(6):439-444. [doi: 10.1089/g4h.2019.0045] [Medline: 31295007]

XSL•FO
RenderX

25. Shaw LA, Wünsche BC, Lutteroth C, Marks S, Callies R. Challenges in virtual reality exergame design. In: Proceedings of the 16th Australasian User Interface Conference. 2015 Presented at: AUIC '15; January 27-30, 2015; Sydney, Australia p. 61-68 URL: https://www.cs.auckland.ac.nz/~christof/publications/ShawEtAl2015-ExergameDesign.pdf
26. Cioffi R, Lubetzky AV. BOXVR versus guided YouTube boxing for stress, anxiety, and cognitive performance in adolescents: a pilot randomized controlled trial. Games Health J 2023 Jun;12(3):259-268 [FREE Full text] [doi: 10.1089/g4h.2022.0202] [Medline: 36745402]
27. Markland D, Tobin V. A modification to the behavioural regulation in exercise questionnaire to include a measure of amotivation. J Sport Exerc Psychol 2004;26(2):191-196. [doi: 10.1123/jsep.26.2.191]
28. Terry PC, Lane AM, Fogarty GJ. Construct validity of the profile of mood states — adolescents for use with adults. Psychol Sport Exerc 2003 Apr;4(2):125-139. [doi: 10.1016/s1469-0292(01)00035-8]
29. Beck AT, Steer RA, Brown GK. Beck Depression Inventory: Manual. 2nd edition. San Antonio, TX: Psychological Corporation; 1996.
30. McArdle WD, Katch FI, Katch VL. Exercise Physiology: Nutrition, Energy, and Human Performance. 8th edition. Philadelphia, PA: Lippincott Williams & Wilkins; 2015.
31. Liu W, Zeng N, Pope ZC, McDonough DJ, Gao Z. Acute effects of immersive virtual reality exercise on young adults' situational motivation. J Clin Med 2019 Nov 12;8(11):1947 [FREE Full text] [doi: 10.3390/jcm8111947] [Medline: 31726672]
32. Merola P, Cardoso MB, Barreto G, Chagas MC, Farias Oliveira Saunders L, Saunders B, et al. Virtual reality high-intensity interval training exergaming compared to traditional high-intensity circuit training among medical students: pilot crossover study. JMIR Serious Games 2025 Jan 07;13:e63461 [FREE Full text] [doi: 10.2196/63461] [Medline: 39773879]
33. Boots B, Berg D, Hewitt E, Naugle K, Naugle K. Physical activity and enjoyment in active virtual reality games in youth: comparative analysis of Gorilla Tag and Beat Saber. JMIR Serious Games 2025 Apr 01;13:e66593 [FREE Full text] [doi: 10.2196/66593] [Medline: 40168642]
34. Chen Y, Guan B, Zhang Y, Lee SC, Liu JY, Li S, et al. Acceptability of and willingness to use virtual reality exergames for weight loss among young adults with overweight or obesity in China: qualitative study. JMIR Serious Games 2025 May 01;13:e66998 [FREE Full text] [doi: 10.2196/66998] [Medline: 40311124]
35. Zhang N, Hong C, Wang Y, Chen H, Zhu Y, Da M, et al. Stationary cycling exercise with virtual reality to reduce depressive symptoms among people with mild to moderate depression: randomized controlled trial. J Med Internet Res 2025 Jul 15;27:e72021 [FREE Full text] [doi: 10.2196/72021] [Medline: 40663793]
36. Li Pira G, Aquilini B, Davoli A, Grandi S, Ruini C. The use of virtual reality interventions to promote positive mental health: systematic literature review. JMIR Ment Health 2023 Jul 06;10:e44998 [FREE Full text] [doi: 10.2196/44998] [Medline: 37410520]
37. Goudman L, Jansen J, Billot M, Vets N, De Smedt A, Roulaud M, et al. Virtual reality applications in chronic pain management: systematic review and meta-analysis. JMIR Serious Games 2022 May 10;10(2):e34402 [FREE Full text] [doi: 10.2196/34402] [Medline: 35536641]
38. Ochi G, Kuwamizu R, Fujimoto T, Ikarashi K, Yamashiro K, Sato D. The effects of acute virtual reality exergaming on mood and executive function: exploratory crossover trial. JMIR Serious Games 2022 Sep 28;10(3):e38200 [FREE Full text] [doi: 10.2196/38200] [Medline: 36169992]
39. Mouatt B, Smith AE, Mellow ML, Parfitt G, Smith RT, Stanton TR. The use of virtual reality to influence motivation, affect, enjoyment, and engagement during exercise: a scoping review. Front Virtual Real 2020 Dec 23;1:1-23. [doi: 10.3389/frvir.2020.564664]
40. Lemmens JS. Persistence and pleasure in VR: enhancing exercise endurance and enjoyment through virtual environments. Psychol Sport Exerc 2023 Nov;69:102494. [doi: 10.1016/j.psychsport.2023.102494] [Medline: 37665929]

Abbreviations

BF: body fat
BREQ-2: Behavioral Regulation in Exercise Questionnaire-2
BRUMS: Brunel Mood Scale
CRF: cardiovascular fitness
IPAQ: International Physical Activity Questionnaire
MET: metabolic equivalent
PA: physical activity
SDT: self-determination theory
VR: virtual reality

XSL•FO
RenderX

Edited by A Coristine; submitted 10.04.25; peer-reviewed by B Lai, AV Lubetzky, M Tusconi; comments to author 16.07.25; revised version received 12.11.25; accepted 13.11.25; published 15.12.25.

Please cite as:
Liu W, McDonough DJ, Oginni J, Mavoungou J, Gao Z
Effects of an Immersive Virtual Reality–Based Exercise Intervention on Psychological and Physiological Outcomes in College Students: Randomized Controlled Trial
JMIR Serious Games 2025;13:e75777
URL: https://games.jmir.org/2025/1/e75777
doi:10.2196/75777
PMID:41307454

©Wenxi Liu, Daniel J McDonough, John Oginni, Jessh Mavoungou, Zan Gao. Originally published in JMIR Serious Games (https://games.jmir.org), 15.12.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Effects of Competitive Exergaming–Based Esports on Older Adults in Hong Kong: Nonrandomized Controlled Pilot Study

Ka Man Leung[1*], MPhil, PhD; Yuchen Shi[2*], BBA, MA, PhD

[1]Department of Health and Physical Education, Education University of Hong Kong, Tai Po, China (Hong Kong)

[2]Department of Business and Strategy, Lee Shau Kee School of Business and Administration, Hong Kong Metropolitan University, Ho Man Tin, China (Hong Kong)

[*]all authors contributed equally

Corresponding Author:
Yuchen Shi, BBA, MA, PhD
Department of Business and Strategy
Lee Shau Kee School of Business and Administration
Hong Kong Metropolitan University
30 Good Shepherd Street
Ho Man Tin
China (Hong Kong)
Phone: 852 29686026
Email: ashi@hkmu.edu.hk

Abstract

Background: Hong Kong faces a rapidly aging population, with many older adults not meeting recommended physical-activity levels and struggling to maintain long-term exercise adherence. Exergaming offers an accessible, technology-supported way to promote health conditions while providing immediate feedback and task variability among older adults.

Objective: This study aimed to evaluate the effects of an 8-week competitive exergaming intervention on functional fitness, cognition, loneliness, physical activity (PA) enjoyment, and quality of life among community-dwelling older adults in Hong Kong.

Methods: We conducted a pragmatic, nonrandomized, 2-group pilot with community participants allocated to either a competitive exergaming group (EG) or a passive control group (CG). EG attended 16 instructor-led 90-minute sessions playing Fitness Boxing (Nintendo Switch), including head-to-head bouts and a bracketed tournament. Outcomes were assessed at baseline and postintervention. Primary analyses used repeated-measure analysis of covariance (time: pre and post; group: EG and CG) adjusted for age, sex, education, marital status, employment, financial status, and housing. Partial η^2 was used to quantify effect sizes. Given the pilot nature, *P* values are unadjusted for multiplicity and interpreted cautiously.

Results: A total of 48 older adults completed assessments (EG: n=24; mean age 69.50, SD 4.77 years; CG: n=24; mean age 71.50, SD 6.74 years). Significant group × time interactions favored EG for lower-body strength (30-second chair stand: $F_{1, 48}$=12.39; *P*<.001; partial η^2=0.22), aerobic endurance (2-minute step: $F_{1, 48}$=4.89; *P*=.03; η^2=0.10), and PA enjoyment (Physical Activity Enjoyment Scale: $F_{1, 48}$=9.36; *P*<.001; η^2=0.18). For the Number Comparison Test (processing speed), the group × time interaction was not significant (*P*=.08); however, an exploratory main effect of group indicated higher performance in EG across time (*P*=.04). Executive function (Trail Making Test parts A and B), loneliness, and Short Form-36 Health Survey subscales showed no significant effects.

Conclusions: Competitive exergaming was feasible and produced small to moderate improvements in lower-body strength, aerobic endurance, and enjoyment of PA. Cognitive effects were inconclusive and should not be overinterpreted given the nonrandomized design, passive control, small sample, and multiple outcomes. Future randomized trials with active comparators and longer duration are warranted.

(JMIR Serious Games 2025;13:e77655) doi:10.2196/77655

KEYWORDS

esports participation; exergaming; health promotion; intervention study; older adults

XSL•FO
RenderX

Introduction

Hong Kong's aging population poses a considerable public health challenge. The overwhelming majority of older adults in Hong Kong do not meet the minimum standards for physical activity (PA) levels recommended by the World Health Organization (WHO) [1]. According to the Census and Statistics Department of the Hong Kong Special Administrative Region (HKSAR) [2], the number of individuals aged 65 years or older is expected to increase to 2.16 million by 2031 and 2.56 million by 2041. Being physically inactive and having a sedentary lifestyle contribute to physiological, psychological, and cognitive health problems. The financial burden of age-related health issues poses social challenges, including increased pressure on health care and social welfare systems [3]. Due to the effects of net medical inflation, population growth, and aging (assuming service enhancement continues at historical levels), Hong Kong's recurrent social welfare and health expenditure as a percentage of nominal GDP is projected to increase from US $56.9 billion and US $52.4 billion in 2014 and 2015 to US $523.3 billion in 2041 and US $563.6 billion in 2042. A structural financial deficit may then occur within a decade.

In examining the aging population and its social implications, studies have suggested various types of PA to promote active aging. Theoretically, "exergaming" refers to technology-mediated play that requires bodily movement and yields exercise-like exertion [4]. "Esports" denotes organized competitive digital gaming characterized by structured competition, skill, information technology, and interpersonal interaction [5,6]. We use the term "competitive exergaming" to denote the intersection—exergames delivered with explicit competitive structures (eg, head-to-head play, scoring, and brackets) in supervised, organized sessions. This sits at the amateur and community end of the esports spectrum; esports is commonly defined as organized competitive digital gaming that can range from community tournaments to professional leagues. Our program implemented 2-player bouts throughout training and a double-elimination tournament, thereby satisfying the competitive and organized attributes associated with esports while retaining the physical demands of exergaming. Framing the intervention as competitive exergaming thus clarifies both the active ingredient we are testing and the leisure community context in which it occurred for older adults. In this study, when the competitive mode of play and interactive elements are integrated with physical and cognitive efforts in an exercise gaming environment, we propose competitive exergaming as one form of esports.

The acceptance and recognition of esports are reflected in the International Olympic Committee's Olympic esports initiatives, which include organizing an esports forum in 2018 and creating the Olympic Esports Series in 2023. Esports is a fast-growing global industry that is particularly prominent in China and Hong Kong [7]. Since 2000, esports has exploded in popularity in China. Chinese esports revenue accounted for more than 18% of global esports revenue (US $13.63 million) in 2018 [8]. Capitalizing on this trend, the HKSAR government proposed esports as a new sector for economic development in 2017 and engaged in several initiatives to promote the development of the esports industry [9]. For example, the HKSAR government provided more than HK$100 (US $12.85) million to Cyberport, a digital technology community in Hong Kong, to promote esports by organizing esports events, empowering the esports industry, developing esports talent, making esports mainstream, and promoting a positive public image [10].

Unlike traditional video gaming, which is often associated with negative consequences, such as excessive screen time, addiction, and sedentary behaviors [11], exergaming offers substantial benefits, including improved physical fitness and range of motion [12], increased stress relief and motivation [13], and enhanced social connectedness and interaction [13]. In this context, competitive exergaming has been studied for its contributions to health and wellness goals among older adults [14]. Monteiro Pereira et al [14] indicated that esports engagement encouraged active lifestyle habits and did not contribute to health problems, contradicting the conclusions of some other studies regarding esports participation, although an association between intense esports practice and mental health problems was observed in certain cases. Moreover, exergames have shown benefits for older adults' physical function and social outcomes, including improved functional capacity and adherence [15] and reduced loneliness in some contexts [16]. Systematic reviews and meta-analyses indicate exergaming can enhance physical and cognitive outcomes in specific populations and designs [17,18]. Importantly, competitive or gamified features (eg, leaderboards, tournaments) can heighten enjoyment and motivation, which are key determinants of adherence in later life [19,20]. However, few community-based programs have deliberately combined exergaming with formal competition for older adults, leaving a gap in whether "competitive exergaming" offers incremental value for enjoyment and functional gains compared with usual activity. Some esports participants report supplementary off-game physical training, but the on-game activity is typically sedentary; in contrast, exergaming itself is physically active and is the modality evaluated in this study.

To date, the literature examining the effects of esports on health and well-being has focused mostly on young populations, especially young adults (eg, students and collegiate players) [14,21] and competitive esports athletes [22-24]. Moreover, a scoping review investigating the associations between esports participation and health conditions revealed that all participants of the 33 included articles were young adults aged 16-27 years [14]. While the effects of esports on older adults have started to be studied in recent years, participation in competitive exergaming–based esports remains a novel form of PA for older adults and is less investigated. A recent meta-analysis included 47 studies and revealed that exergame training could benefit older adults in multiple aspects of physical functioning, such as balance, upper body strength, lower body strength, aerobic endurance, and gait [25]. However, the competitive and interactive features of esports were absent in these studies. In Hong Kong, Leung and Chu [26] conducted a qualitative study on the perception of esports (eg, e-swimming and e-car racing games) among middle-aged and older adults and demonstrated that the physical benefits (improvements in attentiveness,

XSL•FO
RenderX

strategic thinking, and memory function), psychological benefits (cognitive improvements and sense of achievement), and social benefits (social interactions and sense of belonging) of esports participation could be explained using the theory of planned behavior, in which the intentions and attitudes underlying actions influence achievement-driven behavior. Their subsequent qualitative study [27] further examined the esports perceptions and experiences of middle-aged and older adult participants in Hong Kong using a social marketing approach. That study hypothesized that a community-based esports intervention meeting the "five Ps" criteria (product, price, promotion, place, and people) would encourage older adults to participate in esports, increase their PA levels, and improve their health. Although the video games used in those 2 studies were equipped with competitive and interactive elements, the seated, sedentary mode of play meant they could not be considered exergaming-based esports. Based on the aforementioned studies, a clear research gap exists in examining the benefits of competitive exergaming–based esports participation on older adults, including its effects on physical and psychological health. Moreover, previous studies have suggested that older adults could gain a greater sense of enjoyment by engaging in competitive forms of PA [28].

To encourage active aging, this study implemented a competitive exergaming–based esports intervention embedded with competition elements for older adults. This study investigated the effects of competitive exergaming–based esports participation on the physical fitness, cognitive function, and selected psychological attributes (ie, loneliness, PA enjoyment, and quality of life [QoL]) of older adults. We hypothesized that competitive exergaming–based esports participation would significantly improve physical function, cognitive attributes, and psychological health in older adults compared to nonparticipation.

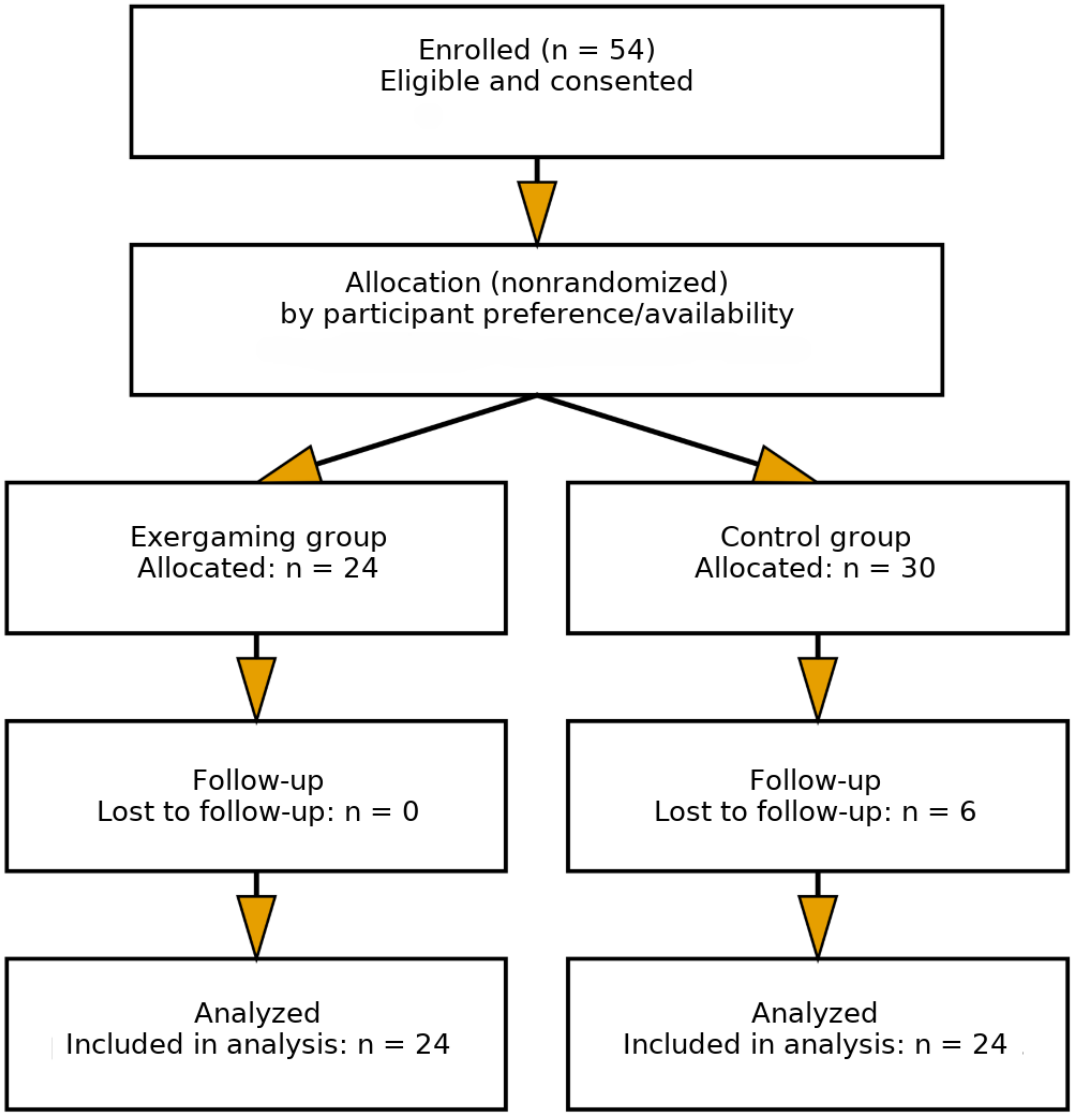
Methods

Participants

A convenience sampling strategy was used to recruit participants (Figure 1). Older adults from 2 branches (Kwun Tong and Hing Tin) of a neighborhood older adult care center in Hong Kong were targeted for recruitment between June and August 2021. All recruited participants were assigned to groups stratified by sex and experience with competitive exergaming–based esports activities. The inclusion criteria were (1) being aged 60 years or older, (2) living independently in the community, (3) having no cognitive impairment, (4) not participating in a structured PA program in the 6 months preceding the study (ie, not attending supervised exercise, sport, or fitness classes ≥1 per week for ≥30 minutes), which was confirmed with a brief screening questionnaire at enrollment, and (5) earning passing scores on the Timed Up-and-Go Test [29] and the Abbreviated Mental Test [30]. Participants were required to complete the Timed Up-and-Go Test within 20 seconds. A minimum score of 6 points on the Abbreviated Mental Test was required to qualify for this study. Participant evaluations were carried out by research assistants from The Education University of Hong Kong, who received a 1-day training at the university 1 month prior to recruitment and evaluation. All research assistants were instructed to fully understand the meaning of each question for the measures used in this study. They were responsible for answering any questions raised by participants when completing the questionnaire in both the pretest and posttest.

XSL•FO
RenderX

Figure 1. Participant recruitment flowchart.



Measures

Functional Fitness

Functional fitness was evaluated using the Senior Fitness Test developed and validated by Rikli and Jones [31]. The Senior Fitness Test assesses older adults' ability to perform normal daily activities independently. The test comprises 6 parts. First, the back scratch test measures the distance between the middle fingers of both hands when extended, one over the shoulder and one up the back. Each participant completed 2 trials, and the best distance was recorded. Second, a chair sit-and-reach test requires participants to sit on the edge of a chair with one leg bent and the other leg extended straight in front. The test measures the distance between the extended middle finger and the tip of the toe of the extended leg. Each participant completed 2 trials for each leg, and the best distances were recorded. Third, an 8-foot up-and-go test involved participants standing from a seated position, walking as quickly as possible around a cone placed 8 feet away, and returning to their seats. Each participant completed 2 trials, and the shortest time to finish was recorded. Fourth, a 30-second chair stand test requires participants to repeatedly stand up from a fully seated position as many times as possible within 30 seconds. Each participant completed 2 trials, and the number of stands completed within the time frame was recorded. Fifth, a 30-second arm curl test requires participants to lift a dumbbell with their dominant hand as often as possible within 30 seconds. Each participant completed a practice trial followed by a test trial, and the number of repetitions completed was recorded. Sixth, a 2-minute step test involved participants marching in place, raising each knee to the level of a mark on the wall corresponding to midway between their iliac crest and patella, for 2 minutes. Each participant was allowed a practice trial followed by a test trial, and the number of full steps completed in 2 minutes was recorded. These tests are reliable (intraclass correlation

XSL•FO
RenderX

coefficient=0.80-0.98) and valid, as demonstrated in content, construct, and criterion analyses [31].

Cognitive Function

To assess cognitive function, 2 tests were administered—the Number Comparison Test (NCT) [32] and the Trail Making Test (TMT) [33]. Both the NCT and TMT have been widely used as cognitive tools for examining older adults' attention and executive functions [34,35]. The NCT requires participants to identify nonmatching digit pairs by marking an "X" on them within 90 seconds. Scores are calculated by subtracting the number of incorrect markings from the total number of correct markings. The TMT consists of 2 parts. For part A, participants are asked to draw a line connecting the numbers 1 through 25, which are scattered on a page, in ascending order as quickly as possible. Part B requires participants to draw a continuous line alternating between numbers (1-13) and letters (A-L) in ascending and alphabetical order (1-A-2-B, etc). The time taken to complete each part is recorded.

Loneliness

Loneliness was measured using a short version of the University of California, Los Angeles (UCLA) Loneliness Scale [36], which has been psychometrically developed and evaluated for older adults in various languages [37]. This scale consists of 8 items evaluating whether participants feel a lack of social contact and assessing their level of loneliness (eg, "I feel isolated from others"). Items are scored on a 4-point scale from 1 (never) to 4 (always), with total scores ranging from 8 to 32. Higher scores indicate a greater level of loneliness. This scale demonstrated satisfactory internal consistency in this study, with a Cronbach α of 0.81.

Enjoyment of PA

The level of enjoyment of PA was measured using the Chinese version of the Physical Activity Enjoyment Scale (PACES), which was developed by Chung and Leung [38]. This measure has been validated as a suitable tool to evaluate older adults' satisfaction with PA [39]. The scale comprises 8 items scored on a 7-point scale from 1 ("I enjoy it") to 7 ("I dislike it"). A total of 4 of the items (items 1, 4, 5, and 7) are reverse coded. Higher scores indicate greater enjoyment of PA. The scale demonstrated excellent reliability and internal consistency in a previous study (Cronbach α=0.92) [39]. The mean value of the 8 item scores was used as the overall PA enjoyment score.

Quality of Life

The Short Form-36 Health Survey (SF-36) is widely used to assess health-related QoL among older adults [40]. In this study, 2 subscales—Social Functioning (2 items) and Emotional Well-being (5 items)—were used. These subscales assess, respectively, motivation to participate in social activities and the frequency of experiencing certain emotional states (feeling nervous, depressed, peaceful, happy, and calm) over the past 4 weeks. Items are rated on a 5-point Likert scale, with higher scores indicating a better state of health. The reliability and validity of the SF-36 have been verified in Chinese participants [41]. In this study, the SF-36 Social Functioning and Emotional Well-being subscales demonstrated acceptable internal consistency, with Cronbach α values of approximately 0.63 and 0.83, respectively [41].

Intervention

The intervention programs were delivered from July to September 2021 by registered instructors and a research assistant from The Education University of Hong Kong. Sessions were delivered in small groups at 2 neighborhood older adult care centers. Participants trained in pairs using the game's 2-player mode for head-to-head play. The final 3 sessions implemented a double-elimination bracket. This format operates competitive exergaming in a community and leisure setting for older adults. A registered instructor supervised technique, safety, and progression. All training sessions were carried out at the 2 neighborhood older adult care center branches (Kwun Tong and Hing Tin) in Hong Kong. The program comprised 16 training sessions, with 2 sessions per week, each lasting 90 minutes. Participants trained in pairs using the game's 2-player mode for head-to-head play. Each 90-minute session comprised a 10-minute warm-up, 70 minutes of game-based practice and competitive bouts (with short rests), and a 5-minute cooldown. The fitness video game Fitness Boxing, developed by Imagineer (Tokyo, Japan) for the Nintendo Switch console, was selected for participants to play. Fitness Boxing was chosen because it requires multiple physical and cognitive skills. This rhythm-based boxing game combines physical exercise with music and requires players to perform movements by following on-screen cues and musical rhythms. Moreover, its unique 2-player mode creates competitiveness between players by comparing the points they score during pairwise competition. Participants used the console's Joy-Con motion controllers to perform basic boxing moves, including punching, ducking, weaving, and stepping. All exergaming group (EG) participants were asked to take part in the training sessions at the same time. The exercise intensity was designed to range from low to moderate. A Fitness Boxing competition was arranged during the final 3 sessions. After the intervention, participants were expected to have a basic understanding of competitive exergaming–based esports and basic boxing skills. This was a nonrandomized controlled intervention; in line with JMIR Publications' policy that registration is required for randomized trials and not required for nonrandomized designs, the study was not prospectively registered. Instead, this nonrandomized controlled intervention was reported in accordance with the Transparent Reporting of Evaluations with Nonrandomized Designs (TREND) statement; a completed TREND checklist is provided in Multimedia Appendix 1. The protocol was approved by the Human Research Ethics Committee of The Education University of Hong Kong (reference number 2019-2020-0468) prior to enrollment (the full trial protocol is provided in Multimedia Appendix 2).

Procedures

We applied a nonrandomized controlled trial design in this study. We selected a nonrandomized design because we anticipated that prospective participants would have strong preferences for participating in the EG over the control group (CG), thus making randomization infeasible. Prior to the intervention, an instructional session was provided to all

XSL•FO
RenderX

participants to explain the details of the program. Participants were required to sign consent forms before the commencement of the intervention. After all participants provided written informed consent, they were divided into 2 groups: a competitive exergaming–based esports group (EG) and a control group (CG). A total of 54 older adults participated in our study. All participants measured their height, weight, and body fat percentage using a Tanita body composition analyzer (Model BC-418) and completed the aforementioned fitness tests. EG participants then began the 8-week intervention immediately after the pretest. A qualified boxercise instructor was recruited to teach and train EG participants in the necessary skills (ie, straight punch, hook, uppercut, mixed punches, weaving, and stepping) for playing the game, with assistance from a research assistant. Finally, a Fitness Boxing competition was held in the final 3 sessions using a double-elimination bracket. After the 8-week intervention, all participants completed the outcome measures again following the same procedures as in the pretest. All tests were conducted at the community centers. CG participants were instructed to maintain their normal daily activities during the study period. Group allocation was not randomized. Participants' preferences and practical considerations were taken into account; those keen to join the exergaming program were allocated to the EG on a first-come, first-serve basis. Once the EG vacancies were fully filled, the remaining participants were served as the CG.

Statistical Analysis

We conducted repeated measures analysis of covariance (ANCOVA) for each outcome, with time (baseline, post) as the within-participant factor and group (exergaming vs control) as the between-participant factor. We included the following covariates to adjust for baseline imbalances: age, sex, education, marital status, employment status, financial status, and housing. The group × time interaction was the effect of primary interest. With 2 time points, the sphericity assumption is not applicable. We examined model assumptions (normality of residuals, homogeneity of variances, and homogeneity of regression slopes) and observed no material violations. Analyses were run in IBM SPSS Statistics (version 27) using the general linear model (repeated measures) procedure, type III sums of squares, α=.05, and partial η^2 as the effect-size index.

Ethical Considerations

Data collection was approved by the Human Research Ethics Committee of The Education University of Hong Kong (reference number 2019-2020-0468). Outcome assessors and data entry staff were blinded to treatment allocation. To maintain blinding, the research assistants who conducted the preintervention and postintervention assessments were not involved in delivering the intervention and were not informed of participants' group assignments. Participants were instructed not to reveal their group to the assessors. Similarly, data entry was performed using coded identifiers that did not indicate group membership. All participants provided written informed consent and were assured that their personal information would remain confidential. Participants were compensated at the end of the study with a HK $50 (US $6.43) supermarket voucher. This study adheres to TREND guidelines; a completed TREND checklist is provided in Multimedia Appendix 1.

Results

Participants were asked to complete the questionnaire after completing the intervention program. All sociodemographic questions were self-generated. A total of 54 older adults participated in our study (n=24 in EG; n=30 in CG). However, 6 participants did not complete the posttest; data from the remaining 48 participants (n=24 in EG and n=24 in CG) were included in our analysis. We used complete-case analyses without imputation; intention-to-treat analysis was not performed and is acknowledged as a limitation. A post hoc power analysis showed that this sample size provided 80% power at P<.05 to detect the intervention's average effect size of partial η^2=0.14. Participant sociodemographic details are presented in Table 1.

XSL•FO
RenderX

Table 1. Sociodemographic characteristics of participants.

Characteristic	Esports (n=24)	Control (n=24)
Age in years, mean (SD)	69.50 (4.77)	71.50 (6.74)
Height (cm), mean (SD)	155.63 (8.61)	158.70 (7.13)
Weight (kg), mean (SD)	54.12 (10.90)	57.61 (9.49)
BMI (kg/m^2), mean (SD)	22.18 (3.21)	22.85 (3.40)
Sex, n		
Female	6	6
Male	18	18
Education, n		
Primary	3	2
Secondary	18	12
Tertiary or above	3	10
Marriage, n		
Married	5	6
Single	14	16
Cohabitation	5	2
Employment, n		
Housewife	3	3
Retired	21	21
Financial status, n		
Low	8	4
Medium	15	18
High	1	2
Housing types, n		
Private housing	22	19
Home ownership scheme housing	2	5

The mean and SD values for physical fitness, cognitive function, and psychological measures at pretest and posttest are presented in Table 2. A total of 48 older adults (n=24 in EG and n=24 in CG) completed the study. There were no significant differences between the EG and CG in any outcome measure at baseline (all $P>.05$). A repeated-measures ANCOVA was conducted to examine group differences from pretest to posttest. In this model, time (pretest vs posttest) was treated as a within-participant factor and group (EG vs CG) as a between-participant factor. In addition, all participant demographic characteristics, including age, sex, education level, marital status, employment status, financial status, and housing type, were included as covariates to control for potential confounding influences. This approach ensured that any group differences at posttest were adjusted for initial differences and background factors. Table 2 is descriptive and shows the unadjusted means (SD) for each outcome at baseline (T1) and postintervention (T2) for both groups (exergaming vs control). An independent samples 2-tailed *t* test was conducted to indicate whether the groups were different before the intervention.

XSL•FO
RenderX

Table 2. Summary of measures at baseline (T1) and postintervention (T2).

	T1		Esports (n=24)		Control (n=24)	
	t test (*df*)	*P* value	T1 (pre), mean (SD)	T2 (post), mean (SD)	T1 (pre), mean (SD)	T2 (post), mean (SD)
Functional fitness						
Back scratch—left (cm)	0.44 (47)	.66	–1.33 (10.69)	–1.48 (9.42)	–2.69 (10.80)	–2.35 (10.70)
Back scratch—right (cm)	0.55 (47)	.58	4.32 (7.64)	3.25 (7.55)	3.06 (8.17)	2.54 (7.22)
Chair sit-and-reach—left (cm)	0.73 (47)	.47	4.70 (11.47)	5.40 (12.35)	2.44 (10.05)	6.99 (9.33)
Chair sit-and-reach—right (cm)	0.67 (47)	.51	4.89 (10.95)	6.52 (10.91)	2.85 (10.24)	7.23 (9.89)
8-foot up-and-go (s)	–0.99 (47)	.33	5.94 (1.12)	4.96 (0.98)	6.31 (1.45)	5.36 (1.36)
Chair stand-up (frequency)	–1.11 (47)	.27	14.08 (3.91)	19.79 (5.37)	15.46 (4.69)	16.38 (5.07)
Arm curl (frequency)	–1.22 (47)	.23	10.58 (24.01)	20.46 (7.49)	16.71 (5.35)	19.71 (6.24)
2-minute step (frequency)	–0.70 (47)	.49	82.38 (22.80)	96.79 (20.69)	87.00 (22.99)	88.71 (17.38)
Cognitive function						
Number comparison (s)	–0.41 (47)	.68	31.13 (9.44)	33.00 (10.62)	32.25 (9.31)	38.21 (8.09)
TMT[a] part A (s)	–0.14 (47)	.89	74.46 (33.64)	61.18 (15.36)	75.76 (30.03)	71.56 (30.15)
TMT part B (s)	–0.80 (47)	.43	44.58 (19.42)	45.24 (18.35)	49.36 (21.85)	45.76 (17.36)
ULS-8[b] (score)	0.77 (47)	.45	1.91 (0.54)	1.89 (0.50)	1.79 (0.54)	1.69 (0.50)
PACES[c] (score)	–1.17 (47)	.25	5.26 (0.85)	5.74 (0.85)	5.59 (1.09)	5.44 (1.10)
Quality of life						
Social functioning (score)	0.85 (47)	.40	78.56 (10.23)	80.45 (9.95)	80.12 (9.87)	79.67 (10.14)
Emotional well-being (score)	–1.10 (47)	.28	82.34 (8.76)	83.12 (8.45)	81.98 (9.12)	81.45 (9.00)

[a]TMT: Trail Making Test.

[b]ULS-8: University of California, Los Angeles Loneliness Scale.

[c]PACES: Physical Activity Enjoyment Scale.

All inferential tests reported in Table 3 are adjusted for age, sex, education, marital status, employment status, financial status, and housing. In particular, lower-body strength (30-second chair stand test) improved significantly more in the EG than in the CG (group × time: $F_{1, 48}$=12.39; P<.001; partial η^2=0.22). Similarly, aerobic endurance (2-minute step test) showed a greater improvement in the EG compared to the CG ($F_{1, 48}$=4.89; P=.03; η^2=0.10), again favoring the EG.

XSL•FO
RenderX

Table 3. Summary of repeated-measures (group × time) analysis of covariance controlling for confounders.

	Esports, mean (SD)	Control, mean (SD)	Group (*P* value)	T1 (baseline), mean (SD)	T2 (postintervention), mean (SD)	Time (*P* value)	Group × time (*P* value)
Functional fitness							
Back scratch—left (cm)	1.73 (1.99)	1.55 (2.04)	.95	1.66 (1.47)	1.62 (1.49)	<.01	.62
Back scratch—right (cm)	3.49 (1.49)	3.34 (1.53)	.95	3.80 (1.15)	3.03 (1.08)	.049	.61
Chair sit-and-reach—left (cm)	4.95 (2.09)	4.81 (2.09)	.96	3.57 (1.56)	6.19 (1.56)	.38	.15
Chair sit-and-reach—right (cm)	5.62 (2.02)	5.12 (2.02)	.86	3.871 (1.56)	6.88 (1.50)	.93	.29
8-foot up-and-go (s)	5.58 (0.21)	5.71 (0.21)	.65	6.12 (0.16)	5.16 (0.15)	.83	.80
Chair stand-up (frequency)	16.74 (0.87)	16.11 (0.87)	.62	14.77 (0.62)	18.08 (0.75)	.06	<.01
Arm curl (frequency)	18.08 (1.05)	18.29 (1.00)	.89	16.02 (0.79)	20.35 (0.98)	.52	.26
2-minute step (frequency)	89.37 (3.89)	88.06 (3.89)	.82	84.69 (3.34)	92.75 (2.81)	.86	.03
Cognitive function							
Number comparison (s)	31.24 (1.59)	36.05 (1.59)	.04	31.69 (1.26)	35.60 (1.24)	.66	.08
TMT[a] part A (s)	69.92 (4.19)	71.56 (4.19)	.79	75.11 (3.85)	66.37 (2.68)	.77	.13
TMT part B (s)	46.83 (2.51)	45.64 (2.51)	.74	46.97 (2.43)	45.5 (2.06)	.32	.68
ULS-8[b] (score)	1.90 (0.10)	1.75 (0.10)	.34	1.85 (0.08)	1.79 (0.07)	.66	.62
PACES[c] (score)	5.47 (0.18)	5.55 (0.18)	.77	5.43 (0.13)	5.59 (0.14)	.15	<.01
Quality of life							
Social functioning (score)	79.25 (9.80)	80.40 (10.05)	.48	81.10 (9.50)	79.50 (10.20)	.08	.75
Emotional well-being (score)	83.50 (8.30)	82.10 (8.95)	.30	84.20 (8.10)	81.55 (9.05)	.16	.70

[a]TMT: Trail Making Test.

[b]ULS-8: University of California, Los Angeles Loneliness Scale.

[c]PACES: Physical Activity Enjoyment Scale.

The intervention also had a positive effect on participants' enjoyment of PA. There was a significant group × time interaction for PA enjoyment (PACES score: $F_{1,\ 48}$=9.36; *P*<.001; η^2=0.18), indicating a higher enjoyment of PA after the program for the EG relative to the CG. Besides, the repeated measures ANCOVA also noted the effect of group was significant of EG on cognitive functioning, measured by the NCT ($F_{1,\ 48}$=4.51; *P*=.04; η^2=0.09).

In addition to these interaction effects, significant main effects were observed for certain measures. Upper-body flexibility (back scratch test) improved over time in both groups, as evidenced by significant main effects of time for both the left shoulder reach ($F_{1,\ 48}$=10.61; *P*<.001; η^2=0.20) and right shoulder reach ($F_{1,\ 48}$=4.07; *P*=.049; η^2=0.09). This indicates that shoulder flexibility increased from pretest to posttest for participants overall. The group × time interaction for the NCT was not statistically significant (*P*=.08); thus, there is no evidence that the intervention improved NCT performance more than the control over time. An exploratory main effect of group (*P*=.04) indicated that, when averaging across time points, EG participants performed better than CG participants; the time main effect was not significant (*P*=.66). Given multiple outcomes and the pilot design, this group difference should be interpreted cautiously and not as proof of a cognitive benefit attributable to the intervention.

No other outcome showed a statistically significant between-group difference or interaction effect. Upper-body strength (30-second arm curl test) did not significantly differ between groups at posttest. The 8-foot up-and-go test (agility and dynamic balance) improved modestly in both groups, but the group × time interaction was not significant, indicating similar gains for EG and CG. Likewise, there were no significant group differences in lower-body flexibility (chair sit-and-reach test) aside from the overall time effect noted in the back scratch, and no between-group improvements were observed in BMI. In terms of cognitive function, executive function as assessed by the TMT (parts A and B) did not differ significantly between the EG and CG (no significant group or interaction effects, *P*>.05 for both parts). For psychological outcomes, loneliness scores (UCLA Loneliness Scale) decreased slightly in both groups but showed no significant group difference or interaction (*P*>.05). Similarly, none of the SF-36 quality-of-life subscales showed a significant improvement or between-group difference after the intervention (all *P*>.05).

XSL•FO
RenderX

Discussion

Overview

This study investigated the effects of an 8-week competitive exergaming–based esports intervention on physical fitness, cognitive functions, and psychological attributes among community-dwelling older adults. Our findings revealed that the EG scored higher than the CG on the following four outcome measures: (1) lower-body strength (chair stand test), (2) aerobic endurance (2-minute step test), (3) cognitive functioning (NCT), and (4) PA enjoyment (PACES). However, no differences were observed between groups in flexibility, upper-body strength and endurance, BMI, or loneliness. These results partially support the hypothesis that greater lower-body strength, agility and dynamic balance, aerobic endurance, cognitive functioning, and PA enjoyment would be observed in competitive exergaming–based esports participants compared to their peers who did not engage in competitive exergaming–based esports.

Despite the statistically significant improvements observed in the EG, the magnitude of changes was modest in absolute terms. The effect sizes for the significant group differences were in the small to moderate range (partial $\eta^2 \approx 0.10$-0.22). For example, on the chair stand-up test, the EG performed only a few more repetitions on average than the CG at posttest. Similarly, the gain of about 14 steps in the 2-minute step test for the EG, while significantly higher than the CG's change, represents a moderate improvement. These modest effect sizes suggest that, from a clinical perspective, the benefits, although measurable, were not very large over the 8-week period. Caution is therefore warranted in interpreting the clinical significance of the findings. It is possible that longer or more intensive training would be required to translate these statistical improvements into more pronounced functional gains. Nonetheless, even small improvements in strength or endurance can be meaningful in an older population, as they may contribute to better functional capacity and reduce the risk of disability over time.

Improvements in Physical Health

Significant improvements in lower-body strength and aerobic endurance were observed in the EG. These improvements were attributed to the physical intensity required to play Fitness Boxing during the 8-week intervention. Instructors taught various skills, including different punches (hooks, uppercuts, and mixed combinations), as well as ducking, weaving, and stepping, and participants practiced these skills throughout the intervention. Specific training sessions focusing on advanced skill combinations were organized to gradually increase the intensity of practice. The total activity time in this competitive exergaming–based esports program exceeded the WHO's PA guideline of a minimum of 150 minutes of moderate-intensity PA per week [1]. This level of activity likely accounts for the observed improvements in physical attributes among older adults in the EG. Moreover, our results align with those of a systematic review indicating that older adults who played exercise-based or motion-sensing games exhibited improvements in physical functions [42]. Another study on the effects of exergames and video game training on older adults' physical functioning also demonstrated significant enhancements in lower-limb strength and balance among participants in an exergame training group [42]. Our findings are consistent with those of research using exergames involving motor function and balance in rehabilitation, suggesting the positive effects of competitive exergaming–based esports participation on promoting older adults' physical health [43]. In this study, **agility** (8-foot up-and-go) improved to a similar extent in both groups, suggesting a general practice or learning effect on this quickness or balance task rather than a unique benefit of the intervention. Notably, our intervention did **not** significantly improve upper-body strength or endurance as measured by the arm curl test. This outcome is understandable because the Fitness Boxing game, while it involves arm movements, does not provide progressive resistance to build muscle strength. Participants were punching the air rather than lifting weights or hitting a heavy bag, so the stimulus for upper-body strength gains was minimal. Without targeted resistance training for the arms, even an active game is unlikely to produce measurable improvements in arm strength. Likewise, measures of flexibility (eg, chair sit-and-reach) did not differ between groups; any small gains in flexibility (such as the improved back scratch test scores over time) occurred in both groups, likely due to general daily activities or repeated testing. In summary, the exergame program proved effective for improving lower-body physical fitness components that are linked to the game's activity profile, whereas fitness components not strongly engaged by the game (upper-body strength and flexibility) showed no differential change.

Unchanged in Cognitive Functions

Our null findings on the Trail Making Test performance are consistent with the task-specificity of executive training. Fitness Boxing heavily loads visuomotor tracking and reaction speed but offers little explicit set-shifting or working-memory updating—the core demands of TMT part B. Executive improvements from exergames are most evident when games embed dual-tasking, inhibition, and switching mechanics (eg, cybercycling with virtual competitors or exergames designed with executive challenges), whereas physically engaging but cognitively simple titles seldom transfer to TMT-like outcomes [44-46]. This pattern is also reflected in network meta-analytic evidence showing that interactive, cognitively stimulating video game formats outperform low–cognitive-load comparators on processing speed and global cognition in older adults [47].

Dose and population also matter. Meta-analyses in clinical cohorts (mild cognitive impairment or dementia) report larger cognitive effects, particularly when interventions extend beyond ~12 weeks and incorporate targeted cognitive components, whereas short, low–cognitive-load programs in cognitively healthy samples often yield small or null changes [48,49]. Our 8-week esports program, centered on a reflex-oriented title, therefore may have been insufficiently long or specific to move executive outcomes. A recent randomized controlled trial in dementia further underscores this point—exergaming with explicit cognitive engagement produced superior executive gains versus aerobic activity alone [50].

Finally, measurement issues can mask subtle treatment effects. Short-interval retesting on neuropsychological batteries,

XSL•FO
RenderX

including TMT variants, produces robust practice (retest) effects that complicate detection of small true changes without large samples or reliable-change methods [51,52]. Together, these literatures align with our finding of no cognitive benefit: without explicit executive-load mechanics, adequate duration, and sensitive analytic approaches, exergaming may leave TMT-like measures unchanged in healthy older adults. These findings are consistent with those of other studies showing that exergame training positively affects older adults' cognitive functioning [42,53]. For instance, older adults trained using Nintendo Wii exergames demonstrated substantial improvements on the NCT and TMT parts A and B [53]. Moreover, Hou and Li [42] found that an exergaming training intervention enhanced cognitive functioning—specifically verbal memory, top-down attention, and learning efficiency—in older adults. One plausible explanation for the significant improvements in cognitive function observed in our study is the complexity of the exergaming. By providing dynamic game content that combines both cognitive and physical exercises during play [53], the Nintendo Switch Fitness Boxing game exemplified a cognitively demanding exergame, as it required players to continually monitor the screen and adjust their movements to respond appropriately. This complicated play style requires significant cognitive effort and is likely responsible for the cognitive benefits observed in older adults. Neurological mechanisms may also underlie these cognitive improvements. In our study, the exergame's combination of physical and mental demands (often referred to as "dual-task" training) could have strengthened neural processes related to quick decision-making and visuomotor coordination, leading to the observed improvement in the number comparison task. Engaging in dual-task activities that simultaneously challenge the mind and body can activate prefrontal executive function networks and promote visual-motor integration, leading to improved cognitive processing. Over time, such stimulation may induce neuroplastic changes that enhance information processing speed and attention in older adults [54,55]. In other words, the combined mental and physical challenge of exergaming likely strengthened neural pathways associated with executive function in our participants, contributing to their gains in cognitive performance.

Improvements in PA Enjoyment

One of the noteworthy outcomes of this study is the significant increase in enjoyment of PA reported by the EG. Participants in the EG had a higher postintervention PACES score, reflecting that they found the PA intrinsically more enjoyable after engaging in the Fitness Boxing program. These results are consistent with studies using other exercise-based video games on different exergaming technology platforms, such as Nintendo Wii [56] and Microsoft Kinect [42], as well as a study using interactive video dance games [56]. Lee et al [19] reported that enjoyment is the primary reason individuals play exergames. Unlike traditional video games, where participants sit and merely manipulate a controller, Fitness Boxing is an exercise-based game that requires players to be active by engaging in a boxing match against human or computer opponents—an enjoyable form of PA. Additionally, the Fitness Boxing competition for EG participants at the end of the program may explain the increase in enjoyment. This suggestion is consistent with the findings of Vorderer et al [57] regarding factors affecting video game enjoyment, which indicated that competition is a key element driving players' enjoyment. Therefore, the Fitness Boxing competition in the final 3 sessions likely enhanced the overall experience and contributed to the players' enjoyment of PA. On the other hand, the intervention's impact on other psychological measures was limited. Feelings of loneliness did not significantly change in the EG compared to the CG. Initially, we hypothesized that participating in a group-based esports activity might enhance social connectedness and reduce loneliness. While participants did interact with each other and the instructors during the program, the duration (8 weeks) and frequency of meetings (twice weekly) may not have been sufficient to produce a measurable reduction in loneliness, especially if participants already had moderate or low loneliness at baseline. It is also possible that more explicit social components (eg, team-based play or off-session social gatherings) are needed to influence loneliness through an exercise program.

Similarly, there were no observable improvements in health-related QoL as measured by the SF-36. QoL is a broad outcome that encompasses physical, emotional, and social well-being. Given the relatively short intervention and the fact that many QoL domains (eg, general health perceptions, social functioning) were not directly targeted by our program, it is not surprising that SF-36 scores remained stable. Often, significant changes in QoL in older adults are reported in longer-term interventions or those that produce large changes in functional ability or mental health. In our study, the improvements in physical function, while meaningful, may not have been large enough over 8 weeks to translate into perceptible changes in participants' day-to-day QoL.

Limitations and Future Research

This study is the first to implement a competitive exergaming–based esports intervention program to assess physical fitness, cognitive functions, and psychological attributes among community-dwelling older adults. Nevertheless, the study is not without limitations. First, participants were recruited through convenience sampling from only 2 community centers for older adults, and most participants lived near these centers and preferred to participate with friends from the same center. This situation may have introduced bias in participant recruitment. Unmeasured factors (such as motivation or social engagement levels) could have influenced who chose to join the esports program. This nonrandom sampling and group allocation could have led to selection bias and potentially inflated the intervention effects. Because group allocation followed participant preference on a first-come, first-served basis, the EG may have comprised more motivated or socially engaged individuals at baseline. This nonrandom allocation increases the risk that unmeasured factors, rather than the intervention, contributed to observed differences. The findings are framed as preliminary signals requiring confirmation in randomized trials. Second, the predominance of female participants (36/48, 75%) limits the generalizability of the results. Notably, enjoyment of PA may differ by sex; with so few male participants, we could not assess whether older male participants derive similar enjoyment and benefits from

XSL•FO
RenderX

exergaming as older female participants. Future research should use random sampling from diverse locations to enhance generalizability. Expanding the sample size and including multiple centers would also improve generalizability. Third, no follow-up assessments were conducted to monitor older adults' continued competitive exergaming–based esports participation and any changes in outcomes over time. Because competitive exergaming–based esports is still a novel form of PA for older adults in Hong Kong, researchers should conduct longitudinal studies with follow-up assessments to examine retention in competitive exergaming–based esports participation and the sustainability of its health effects. Fourth, this study only examined the influence of competitive exergaming–based esports on older adults' health, and the mechanisms behind these benefits (eg, the role of competition) remain unexamined; future studies should investigate how the competitive nature of esports contributes to health outcomes. The fifth limitation is that the control group did not engage in any alternative activity; they were instructed to continue their usual routines, which led to a Hawthorne effect. While this allows us to attribute changes to the intervention, it also means the social interaction and novelty were only experienced by the EG. These nonspecific factors can elevate EG members' enjoyment (and sometimes performance) independent of the specific active ingredients of exergaming. Besides, participants may behave differently if they understand that they are being observed. In this study, both groups were stationed in different centers where they were members. Participants therefore felt more comfortable interacting in the familiar environment and timing of their own older adult care center, which may reduce their awareness of being observed and hence minimize the Hawthorne effect. Additionally, outcome assessors were blinded so they were less aware of the research aims and further reduced bias. A more rigorous design in the future might use an active control group (eg, a traditional exercise program or a social activity group) to account for placebo or expectancy effects and to isolate the specific contribution of the exergaming format. Another limitation could be the relatively small sample size (n=48), limiting the statistical power to detect smaller effects and preventing any detailed subgroup analyses (eg, comparing outcomes for male vs female participants). A larger sample would provide more confidence in the stability of the results and allow subgroup analyses (eg, by age range or fitness level). Moreover, we did not implement intention-to-treat procedures, which is a limitation; estimates could be affected by differential loss to follow-up (6/30 in CG and 0/24 in EG). Future randomized controlled trials should prespecify and apply intention-to-treat analyses based on the intervention design in this study. Finally, with many end points, the risk of type I error is nontrivial. We report unadjusted *P* values and treat findings as hypothesis-generating. Future confirmatory trials should prespecify a primary end point and control multiplicity.

Conclusion

In this nonrandomized pilot, competitive exergaming produced small to moderate gains in lower-body strength, aerobic endurance, and PA enjoyment, while cognitive effects were inconclusive. This examination of the benefits of competitive exergaming–based esports participation on older adults' physical and psychological health outcomes has several practical implications. The improvement in lower-body strength can contribute to better functional mobility. Even a small increase in the ability to rise from a chair or perform repeated standing movements can enhance an older adult's independence in daily activities (eg, making it easier to climb stairs or get up from a low seat without assistance). Improved aerobic endurance, as reflected in the step test, suggests that participants in the exergaming program built greater stamina. This could translate into an increased capacity for walking longer distances, participating in community events, or simply maintaining energy levels throughout the day. Importantly, these physical gains were achieved through a mode of exercise that participants found enjoyable, which addresses one of the common barriers to exercise in older people—lack of interest or motivation. The enjoyment aspect implies that competitive exergaming–based exercise could lead to higher adherence rates if implemented in community centers or senior exercise programs, thereby potentially yielding long-term health benefits beyond the duration of the study. The fact that this was achieved through a fun PA is noteworthy—it suggests a dual benefit of exergaming for both body and mind. Geriatric health practitioners and recreation program planners might consider incorporating exergaming, or "active video gaming," sessions into their regular offerings for older adults. Such programs can be a novel way to engage older people who might be reluctant to join traditional exercise classes, thereby broadening participation in PA among this population. However, given selection bias, including a passive control and multiple outcomes, these results should be viewed as preliminary. Randomized, adequately powered trials with active comparators and game mechanics that directly load executive processes are warranted.

Conflicts of Interest

None declared.

Multimedia Appendix 1
TREND checklist.
[PDF File (Adobe PDF File), 285 KB - games_v13i1e77655_app1.pdf]

Multimedia Appendix 2
The trial protocol.
[PDF File (Adobe PDF File), 105 KB - games_v13i1e77655_app2.pdf]

XSL•FO
RenderX

References

1. Global recommendations on physical activity for health. World Health Organization. 2010. URL: https://www.who.int/publications/i/item/9789241599979 [accessed 2025-11-27]
2. 2018 edition of "Hong Kong Annual Digest of Statistics". Census and Statistics Department of the HKSAR. 2018. URL: https://www.censtatd.gov.hk/en/press_release_detail.html?id=4531 [accessed 2025-11-27]
3. Report of the working group on long-term fiscal planning. Financial Service and the Treasury Bureau, The Government of the Hong Kong Special Administrative Region of the People's Republic of China. URL: https://www.fstb.gov.hk/en/treasury/pub_finance/docs/english_report_online_version.pdf [accessed 2024-12-20]
4. Oh Y, Yan S. Defining exergames and exergaming. In: Conference: Meaningful Play. 2010 Presented at: Proceedings of Meaningful Play 2010; October 21-23, 2010; Michigan URL: https://www.researchgate.net/publication/230794344_Defining_exergames_exergaming
5. Formosa J, O'Donnell N, Horton EM, Türkay S, Mandryk RL, Hawks M, et al. Definitions of esports: a systematic review and thematic analysis. Proc ACM Hum-Comput Interact 2022;6:1-45 [FREE Full text] [doi: 10.1145/3549490]
6. Tang D, Sum R, Li M, Ma R, Chung P, Ho R. What is esports? A systematic scoping review and concept analysis of esports. Heliyon 2023;9(12):e23248 [FREE Full text] [doi: 10.1016/j.heliyon.2023.e23248] [Medline: 38149186]
7. Zhao Y, Lin Z. Umbrella platform of Tencent eSports industry in China. J Cult Econ 2020;14(1):9-25. [doi: 10.1080/17530350.2020.1788625]
8. Report on promotion of e-sports development in Hong Kong. Office of the Government Chief Information Officer (OGCIO), HKSAR.: Hong Kong SAR Government; 2018. URL: https://www.info.gov.hk/gia/general/201803/28/P2018032800307.htm [accessed 2025-11-27]
9. Report on promotion of e-sports development in Hong Kong. Hong Kong SAR Government.: Hong Kong Cyberport Management Company Limited; 2018. URL: https://www.info.gov.hk/gia/general/201803/28/P2018032800307.htm [accessed 2025-11-27]
10. Biegun J, Edgerton J, Roberts L. Measuring problem online Video gaming and its association with problem gambling and suspected motivational, mental health, and behavioral risk factors in a sample of university students. Games Cult 2020;16(4):434-456 [FREE Full text] [doi: 10.1177/1555412019897524]
11. Santos IKD, Medeiros RC, Medeiros JA, Sena DCS, Cobucci RN, Oliveira RS, et al. Active Video games for improving mental health and physical fitness-an alternative for children and adolescents during social Isolation: an overview. Int J Environ Res Public Health 2021;18(4):1641 [FREE Full text] [doi: 10.3390/ijerph18041641] [Medline: 33572159]
12. Osmanovic S, Pecchioni L. Beyond entertainment: motivations and outcomes of video game playing by older adults and their younger family members. Games Cult 2015;11(1-2):130-149 [FREE Full text] [doi: 10.1177/1555412015602819]
13. Kappen DL, Mirza-Babaei P, Nacke LE. Older adults' physical activity and exergames: a systematic review. Int J Hum-Comput Interact 2018;35(2):140-167. [doi: 10.1080/10447318.2018.1441253]
14. Monteiro Pereira A, Costa JA, Verhagen E, Figueiredo P, Brito J. Associations between esports participation and health: a scoping review. Sports Med 2022;52(9):2039-2060 [FREE Full text] [doi: 10.1007/s40279-022-01684-1] [Medline: 35486374]
15. Júnior JLADS, Biduski D, Bellei EA, Becker OHC, Daroit L, Pasqualotti A, et al. A bowling exergame to improve functional capacity in older adults: Co-design, development, and testing to compare the progress of playing alone versus playing with peers. JMIR Serious Games 2021;9(1):e23423 [FREE Full text] [doi: 10.2196/23423] [Medline: 33512319]
16. Li J, Erdt M, Chen L, Cao Y, Lee S, Theng Y. The social effects of exergames on older adults: systematic review and metric analysis. J Med Internet Res 2018;20(6):e10486 [FREE Full text] [doi: 10.2196/10486] [Medline: 29954727]
17. Zhao Y, Feng H, Wu X, Du Y, Yang X, Hu M, et al. Effectiveness of exergaming in improving cognitive and physical function in people with mild cognitive impairment or dementia: systematic review. JMIR Serious Games 2020;8(2):e16841 [FREE Full text] [doi: 10.2196/16841] [Medline: 32602841]
18. Yang C, Han X, Jin M, Xu J, Wang Y, Zhang Y, et al. The effect of video game-based interventions on performance and cognitive function in older adults: Bayesian network meta-analysis. JMIR Serious Games 2021;9(4):e27058 [FREE Full text] [doi: 10.2196/27058] [Medline: 34967759]
19. Lee S, Kim W, Park T, Peng W. The psychological effects of playing exergames: a systematic review. Cyberpsychol Behav Soc Netw 2017;20(9):513-532 [FREE Full text] [doi: 10.1089/cyber.2017.0183] [Medline: 28837353]
20. Bravo-Aparicio J, Trillo-Charlín I, Avendaño-Coy J, Beltran-Alacreu H. Effectiveness of gamification on enjoyment and satisfaction in older adults: systematic review and meta-analysis. JMIR Aging 2025;8:e72559 [FREE Full text] [doi: 10.2196/72559] [Medline: 40505143]
21. Andre TL, Walsh SM, Valladao S, Cox D. Physiological and perceptual responses to a live collegiate esports tournament. Int J Exerc Sci 2020;13(6):1364-1379 [FREE Full text] [doi: 10.70252/kneb6696]
22. DiFrancisco-Donoghue J, Werner W, Douris PC, Zwibel H. Esports players, got muscle? Competitive video game players' physical activity, body fat, bone mineral content, and muscle mass in comparison to matched controls. J Sport Health Sci 2022;11(6):725-730 [FREE Full text] [doi: 10.1016/j.jshs.2020.07.006] [Medline: 32711155]
23. Kari T, Karhulahti V. Do e-athletes move? A study on training and physical exercise in elite e-sports. Int J Gaming Comput-Mediat Simul 2016;8(4):53-66. [doi: 10.4018/ijgcms.2016100104]

XSL•FO
RenderX

24. Gündoğdu S, Çolak Ö, Doğan EA, Gülbetekin E, Polat Ö. Assessment of mental fatigue and stress on electronic sport players with data fusion. Med Biol Eng Comput 2021;59(9):1691-1707 [FREE Full text] [doi: 10.1007/s11517-021-02389-9] [Medline: 34216320]
25. Hai L, Hou HY, Zhou C, Li HJ. The effect of exergame training on physical functioning of healthy older adults: a meta-analysis. Games Health J 2022;11(4):207-224 [FREE Full text] [doi: 10.1089/g4h.2021.0173] [Medline: 35653720]
26. Leung K, Chu W. eSports participation among Hong Kong middle-aged and older adults: a qualitative study. Hum Behav Emerg Technol 2023;2023:1-11 [FREE Full text] [doi: 10.1155/2023/6798748]
27. Leung KM, Chu W. Designing an eSports intervention for middle-aged and older adults in Hong Kong: social marketing approach. PLoS One 2023;18(4):e0284504 [FREE Full text] [doi: 10.1371/journal.pone.0284504] [Medline: 37104392]
28. Leung K, Chung P, Hagger M. The effects of light volleyball intervention programme in improving selected physical and psychological attributes of older adults in Hong Kong. Int J Sport Exerc Psychol 2018;18(1):1-12 [FREE Full text] [doi: 10.1080/1612197x.2018.1462231]
29. Podsiadlo D, Richardson S. J Am Geriatr Soc 1991;39(2):142-148 [FREE Full text] [doi: 10.1111/j.1532-5415.1991.tb01616.x] [Medline: 1991946]
30. Chu LW, Pei CK, Ho MH, Chan PT. Validation of the abbreviated mental test (Hong Kong version) in the elderly medical patient. Hong Kong Med J 1995;1(3):207-211 [FREE Full text]
31. Rikli RE, Jones CJ. Senior Fitness Test Manual. Illinois: Human Kinetics; 2013.
32. Ackerman PL, Cianciolo AT. Cognitive, perceptual-speed, and psychomotor determinants of individual differences during skill acquisition. J Exp Psychol Appl 2000;6(4):259-290 [FREE Full text] [doi: 10.1037//1076-898x.6.4.259] [Medline: 11218338]
33. Wang RY, Zhou JH, Huang YC, Yang YR. Reliability of the Chinese version of the trail making test and stroop color and word test among older adults. Int J Gerontol 2018;12(4):336-339. [doi: 10.1016/j.ijge.2018.06.003]
34. Perrot A, Maillot P, Hartley A. Cognitive training game versus action videogame: effects on cognitive functions in older adults. Games Health J 2019;8(1):35-40 [FREE Full text] [doi: 10.1089/g4h.2018.0010] [Medline: 30376364]
35. Suzuki H, Sakuma N, Kobayashi M, Ogawa S, Inagaki H, Edahiro A, et al. Normative data of the trail making test among Urban community-dwelling older adults in Japan. Front Aging Neurosci 2022;14:832158 [FREE Full text] [doi: 10.3389/fnagi.2022.832158] [Medline: 35693348]
36. Xu S, Qiu D, Hahne J, Zhao M. Psychometric properties of the short-form UCLA Loneliness Scale (ULS-8) among Chinese adolescents. Medicine 2018;97(38):e12373 [FREE Full text] [doi: 10.1097/md.0000000000012373]
37. Alsubheen SA, Oliveira A, Habash R, Goldstein R, Brooks D. Systematic review of psychometric properties and cross-cultural adaptation of the University of California and Los Angeles loneliness scale in adults. Curr Psychol 2021:1-15 [FREE Full text] [doi: 10.1007/s12144-021-02494-w] [Medline: 34785877]
38. Chung PK, Leung KM. Psychometric properties of eight-item physical activity enjoyment scale in a Chinese population. J Aging Phys Act 2018:1-6 [FREE Full text] [doi: 10.1123/japa.2017-0212] [Medline: 29722637]
39. Huffman MK, Christ SL, Ferraro KF, Klenosky DB, Marceau K, Amireault S. Questions to measure enjoyment of and satisfaction with physical activity: are they appropriate for use in an older population? Innov Aging 2021;5(4):igab041 [FREE Full text] [doi: 10.1093/geroni/igab041] [Medline: 34988294]
40. Hobart JC, Williams LS, Moran K, Thompson AJ. Quality of life measurement after stroke: uses and abuses of the SF-36. Stroke 2002;33(5):1348-1356 [FREE Full text] [doi: 10.1161/01.str.0000015030.59594.b3] [Medline: 11988614]
41. Zhang Y, Qu B, Lun SS, Guo Y, Liu J. The 36-item short form health survey: reliability and validity in Chinese medical students. Int J Med Sci 2012;9(7):521-526 [FREE Full text] [doi: 10.7150/ijms.4503] [Medline: 22991490]
42. Hou HY, Li HJ. Effects of exergame and video game training on cognitive and physical function in older adults: a randomized controlled trial. Appl Ergon 2022;101:103690 [FREE Full text] [doi: 10.1016/j.apergo.2022.103690] [Medline: 35066398]
43. Smith ST, Schoene D. The use of exercise-based videogames for training and rehabilitation of physical function in older adults: current practice and guidelines for future research. Aging Health 2012;8(3):243-252. [doi: 10.2217/ahe.12.30]
44. Anderson-Hanley C, Arciero PJ, Brickman AM, Nimon JP, Okuma N, Westen SC, et al. Exergaming and older adult cognition: a cluster randomized clinical trial. Am J Prev Med 2012;42(2):109-119 [FREE Full text] [doi: 10.1016/j.amepre.2011.10.016] [Medline: 22261206]
45. Jiang J, Guo W, Wang B. Effects of exergaming on executive function of older adults: a systematic review and meta-analysis. PeerJ 2022;10:e13194 [FREE Full text] [doi: 10.7717/peerj.13194] [Medline: 35433124]
46. Stojan R, Voelcker-Rehage C. A systematic review on the cognitive benefits and neurophysiological correlates of exergaming in healthy older adults. J Clin Med 2019;8(5):734 [FREE Full text] [doi: 10.3390/jcm8050734] [Medline: 31126052]
47. Yang C, Han X, Jin M, Xu J, Wang Y, Zhang Y, et al. The effect of video game-based interventions on performance and cognitive function in older adults: Bayesian network meta-analysis. JMIR Serious Games 2021;9(4):e27058 [FREE Full text] [doi: 10.2196/27058] [Medline: 34967759]
48. Cai X, Xu L, Zhang H, Sun T, Yu J, Jia X, et al. The effects of exergames for cognitive function in older adults with mild cognitive impairment: a systematic review and meta-analysis. Front Neurol 2024;15:1424390 [FREE Full text] [doi: 10.3389/fneur.2024.1424390] [Medline: 39081342]

XSL•FO
RenderX

49. Zhao Y, Feng H, Wu X, Du Y, Yang X, Hu M, et al. Effectiveness of exergaming in improving cognitive and physical function in people with mild cognitive impairment or dementia: systematic review. JMIR Serious Games 2020;8(2):e16841 [FREE Full text] [doi: 10.2196/16841] [Medline: 32602841]
50. Wu S, Ji H, Won J, Jo EA, Kim YS, Park JJ. The effects of exergaming on executive and physical functions in older adults with Dementia: randomized controlled trial. J Med Internet Res 2023;25:e39993 [FREE Full text] [doi: 10.2196/39993] [Medline: 36881445]
51. Duff K. One-week practice effects in older adults: tools for assessing cognitive change. Clin Neuropsychol 2014;28(5):714-725 [FREE Full text] [doi: 10.1080/13854046.2014.920923] [Medline: 24882553]
52. Goldberg TE, Harvey PD, Wesnes KA, Snyder PJ, Schneider LS. Practice effects due to serial cognitive assessment: implications for preclinical Alzheimer's disease randomized controlled trials. Alzheimers Dement (Amst) 2015;1(1):103-111 [FREE Full text] [doi: 10.1016/j.dadm.2014.11.003] [Medline: 27239497]
53. Wadhwa G, Walia S. Effect of Nintendo Wii brain training games on perceptual speed, working memory and spatial orientation in young adults. Int J Neurol Phys Ther 2018;4(1):7. [doi: 10.11648/j.ijnpt.20180401.12]
54. Hughes TF, Flatt JD, Fu B, Butters MA, Chang CC, Ganguli M. Interactive video gaming compared with health education in older adults with mild cognitive impairment: a feasibility study. Int J Geriatr Psychiatry 2014;29(9):890-898 [FREE Full text] [doi: 10.1002/gps.4075] [Medline: 24452845]
55. Ismail NA, Hashim HA, Ahmad YH. Physical activityexergames among older adults: a scoping review. Games Health J 2022;11:1-17 [FREE Full text] [doi: 10.1089/g4h.2021.0104]
56. Xu W, Xing QW, Zhu JD, Liu X, Jin P. Effectiveness of an extended-reality interactive learning system in a dance training course. Educ Inf Technol 2023;28(12):16637-16667. [doi: 10.1007/s10639-023-11883-6]
57. Vorderer P, Hartmann T, Klimmt C. Explaining the enjoyment of playing video games: the role of competition. Pittsburgh, PA: Carnegie Mellon University; 2003 Presented at: Proceedings of the Second International Conference on Entertainment Computing; Dec 10, 2003; Pittsburgh, PA p. 1-9. [doi: 10.1145/958720.958735]

Abbreviations

ANCOVA: analysis of covariance
CG: control group
EG: exergaming group
HKSAR: Hong Kong Special Administrative Region
NCT: Number Comparison Test
PA: physical activity
PACES: Physical Activity Enjoyment Scale
QoL: quality of life
SF-36: Short Form-36 Health Survey
TMT: Trail Making Test
TREND: Transparent Reporting of Evaluations with Nonrandomized Designs
UCLA: University of California, Los Angeles
WHO: World Health Organization

Edited by A Coristine, R Baranyi; submitted 17.05.25; peer-reviewed by W Zhu, D Tang; comments to author 29.08.25; revised version received 28.10.25; accepted 12.11.25; published 15.12.25.

Please cite as:
Leung KM, Shi Y
Effects of Competitive Exergaming–Based Esports on Older Adults in Hong Kong: Nonrandomized Controlled Pilot Study
JMIR Serious Games 2025;13:e77655
URL: https://games.jmir.org/2025/1/e77655
doi:10.2196/77655
PMID:

©Ka Man Leung, Yuchen Shi. Originally published in JMIR Serious Games (https://games.jmir.org), 15.12.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Immersive Virtual Reality–Assisted Therapy for Distressing Voices in Psychosis: Qualitative Study of Participants' and Therapists' Experiences in the Challenge Trial

Mads Juul Christensen[1,2], MSc; Matilde Poulsen Rydborg[1], MSc; Rikke Jørgensen[2,3], PhD; Cecilie Dueholm Nielsen[1], MD; Jan Mainz[1,2,4], PhD; Imogen H Bell[5,6], PhD; Neil Thomas[7,8], PhD; Lisa Charlotte Smith[9,10], PhD; Lise Sandvig Mariegaard[9], MSc; Thomas Ward[11,12], PhD; Merete Nordentoft[10,13], DMSc; Louise Birkedal Glenthøj[9,14], DMSc; Ditte Lammers Vernal[1,2], PhD

[1]Psychiatry, Aalborg University Hospital, North Denmark Region, Brandevej 5, Aalborg, Denmark

[10]Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark

[11]Department of Psychology, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, United Kingdom

[12]South London and Maudsley NHS Foundation Trust, London, United Kingdom

[13]Research Centre for Mental Health (CORE), Mental Health Centre Copenhagen, Copenhagen University Hospital – Mental Health Services CPH, Copenhagen, Denmark

[14]Department of Psychology, University of Copenhagen, Copenhagen, Denmark

[2]Department of Clinical Medicine, Aalborg University, Aalborg, Denmark

[3]Unit for Psychiatric Research, Psychiatry, Aalborg University Hospital, North Denmark Region, Aalborg, Denmark

[4]Department of Health Economics, University of Southern Denmark, Odense, Denmark

[5]Orygen, The National Centre of Excellence in Youth Mental Health, Parkville, Australia

[6]Centre for Youth Mental Health, The University of Melbourne, Parkville, Australia

[7]Centre for Mental Health and Brain Sciences, Swinburne University of Technology, Melbourne, Australia

[8]Alfred Hospital, Melbourne, Australia

[9]VIRTU Research Group, Mental Health Services CPH, Copenhagen University Hospital – Mental Health Services CPH, Copenhagen, Denmark

Corresponding Author:
Mads Juul Christensen, MSc
Psychiatry, Aalborg University Hospital, North Denmark Region, Brandevej 5, Aalborg, Denmark

Abstract

Background: Immersive virtual reality–assisted therapy (VRT) is a relational therapy for distressing voices in psychosis. Like AVATAR therapy (AT), VRT centers on therapist-facilitated dialogues with a digital avatar representing a voice. Unlike AT, VRT uses immersive virtual reality (VR). While participant experiences of AT have been explored, therapist perspectives remain unexamined, and for VRT, neither participant nor therapist experiences have been studied. Understanding these perspectives is essential to inform optimization of therapy, future research, and implementation.

Objective: The objective of this qualitative study was to explore both trial participants' and therapists' experience of VRT in the Challenge trial.

Methods: Semistructured interviews were conducted with 10 trial participants and 8 therapists across the 3 Challenge trial sites. Trial participants were purposively sampled to ensure site representation and variation in voice-hearing duration. Individual interviews were conducted with trial participants, while therapists participated in site-based groups with 2‐3 in each. Interviews were audio-recorded, transcribed, and subjected to reflexive thematic analysis from a critical realist position. Coding and theme development were inductive. People with lived experience were invited to an initial focus group for topic guide development and a later theme validation workshop. Reporting followed the Standards of Reporting Qualitative Research (SRQR).

Results: A total of five overarching themes were generated: (1) using technology to meet the voice, (2) a different approach to voice-hearing and treatment, (3) on a tight schedule, (4) a toolbox for transformation, and (5) a price to pay. Trial participants and therapists generally found VRT acceptable, appropriate, and feasible. Highlights included the acknowledging approach to the voice(s), facilitation of engagement with the voice(s), and opportunity to share the otherwise private experience of voice-hearing. Externalizing and embodying the voice(s) in VR-supported avatar role-plays was seen as a key affordance. Positive outcomes included increased trial participant empowerment and self-worth, enabling or improving voice dialogue, new understanding of voice intentions, and changes in voice frequency or content. Challenges included instances of participant anxiety, exhaustion, or suboptimal sense of (voice) presence; adverse voice reactions; technological malfunctions and limitations to avatar design;

XSL•FO
RenderX

measurement insensitivity; tensions between assertiveness and compassion; difficulties with reproducing negative voice content; and the demanding nature of the therapy and the nontraditional skills required of therapists.

Conclusions: The study provided comprehensive insights into trial participants' and therapists' experiences of VRT in the Challenge trial. Findings share several similarities with qualitative research on other relational therapies for distressing voices and highlight VRT's potential for positive change. Key considerations for future research and implementation include monitoring anxiety and voice reactions, ensuring operational reliability of hardware and software, and addressing the additional effort required by therapists, which may be unsustainable in routine practice. As a demanding intervention, the successful implementation of VRT will require adequate training, supervision, and structural support.

(JMIR Serious Games 2025;13:e77920) doi:10.2196/77920

KEYWORDS

virtual reality; virtual reality–assisted therapy; AVATAR therapy; voice-hearing; auditory verbal hallucinations; psychosis; schizophrenia; implementation outcomes; acceptability; feasibility; semi-structured interviews

Introduction

Hearing voices (auditory verbal hallucinations) is a phenomenon occurring in the general population, across various disorders, and commonly associated with psychosis or schizophrenia spectrum disorder diagnoses [1-4]. Voices may be highly distressing [5], especially due to appraisals of the meaning and intent of voices [6] and negative content [7,8].

Within psychosis, an estimated 1 in 3 individuals have a suboptimal pharmacological treatment response [9,10] with side effects being common [11]. Cognitive behavioral therapy for psychosis (CBTp) is recommended as a first-line psychological treatment in several national guidelines [e.g., 12] and shows a small-to-moderate effect on hallucinations [13]. However, CBTp has traditionally focused on psychosis more broadly rather than hearing voices specifically, limiting its sensitivity [14].

More targeted approaches have evolved that focus on voices more specifically [15-17]. Relational therapies are a promising approach for hearing voices, aiming to modify the interaction patterns and dynamics between the hearer and the voice through experiential dialogue with voice-related identities [18]. These include Avatar Therapy (AT) [19,20], Relating Therapy (RT) [21-25], and Talking with Voices (TwV) [26-28]. AT engages an individual in computer-assisted on-screen dialogues with a digital avatar representing the most distressing voice. In the AVATAR trial, therapists alternated between real-time voice transformation to role-play the avatar and speaking in their own voice, targeting therapeutic goals such as power, control, experiential disengagement, and compassion [29]. The AVATAR2 trial demonstrated that both brief (Avatar brief version, AV-BRF) and extended (Avatar extended version, AV-EXT) AT effectively reduced voice severity and distress [30]. AV-BRF focused on exposure, assertiveness, and self-esteem. AV-EXT emphasized understanding voices within the context of the individual's life history [31] and allowed greater scope to address broader treatment targets, including trauma [30,32]. In an early value assessment, the use of AT has been recommended to collect real-world evidence and cost-effectiveness [33].

Virtual reality (VR) offers a promising avenue for improving treatment outcomes in mental health [34] and psychosis [35]. In VR, the central concept of immersion refers to the extent to which users feel mentally and emotionally absorbed within virtual environments, and key influencing factors include an immersive system, a sense of presence, and ecological validity [34]. The immersive system—the technical aspects of the VR equipment and software—can influence sense of presence (henceforth presence) [36]—the subjective feeling of "being there" in a virtual environment [37]. Despite terminological ambiguity [37],presence may enhance therapeutic responses [38,39], when delivered optimally [40]. In nonimmersive AT, the concept of presence has been adapted to "sense of voice presence" (henceforth voice presence) to more accurately capture the subjective experience of dialoguing with the voice [41]. Voice presence was consistently observed during AVATAR sessions, with an interaction between reduced anxiety and strong presence being implicated in outcomes [41]. As presence is stronger in VR than nonimmersive formats [34], adding VR to AT may optimize outcomes. Thus, a variant of AT using VR and a separate software called virtual reality–assisted therapy (VRT) for distressing voices has been developed and shown promising results [42]. The recent Challenge trial demonstrated VRT's efficacy in reducing voice severity posttreatment [43], with impact of avatar features and "sense of spatial presence" (henceforth spatial presence; how real it felt to be in the virtual room with the avatar) on treatment outcomes reported elsewhere [44]. A modified VRT version has been explored in a Challenge substudy (K Rasmussen, unpublished data, 2025).

VRT avatar dialogues can evoke a range of emotional responses [45], with findings suggesting that negative emotions decrease while positive emotions increase as sessions progress [46]. These emotional shifts may be driven by changes in relational dynamics, as reflected in evolving discourse patterns [47]. Anxiety levels are initially high [46,48], with early sessions often perceived as stressful [48]. The initial phase is confrontational, and withdrawal and dropout rates are higher initially than in later phases [29,49]. Although adverse emotional responses are expected, too intense anxiety or distress may reduce tolerability. Indeed, for 29% of the Challenge trial participants, adjustments were made to the pace or duration of exposures in response to reports of anxiety [43]. Assessing unwanted events—such as therapy prolongation or symptom exacerbation—is essential for ensuring that therapies remain safe [50-52].

XSL•FO
RenderX

Digital mental health interventions risk a research-to-practice gap [53,54], making evaluation of implementation outcomes crucial for translation into routine care [55,56]. AT studies have reported acceptability [29,57] and feasibility [30,58], and a VRT study has addressed both [48]. However, uptake in routine care remains uncertain, as clinical and implementation research differ [59]. While dedicated implementation studies are needed, qualitative research within clinical trials can prove an initial window into lived experience and stakeholder perspectives [60], informing future implementation efforts [61-63]. Incorporating service-user perspectives is particularly valuable for optimizing therapy [64-66]. Implementation science outcomes, particularly acceptability, appropriateness, and feasibility [56], may provide a useful basis for interpreting experiences of VRT and identifying key considerations for implementation.

Participants' experiences of AT have been explored qualitatively in AVATAR1 and AVATAR2. The former found AT acceptable and beneficial, though some participants reported strong emotional reactions and trauma-related memories that did not disrupt therapy [57]. The latter, currently being prepared for publication, examined direct early work with verbatim voice content and the role of the therapeutic alliance [67]. In contrast, no equivalent studies have explored the experiences of VRT participants.

Therapists' experiences have not yet been explored in either AT or VRT. Considering dissemination and implementation, it is crucial to account for the perspectives of clinicians delivering these therapies [68]. This is especially important for VR-based interventions, where mental health professionals have raised concerns regarding technological limitations, potential side effects, and increased time and workload demands [69]. Furthermore, as demonstrated in both AT [20,29] and VRT [49], conducting avatar dialogues requires a nontraditional therapeutic skill set, particularly the "actor-like" components that pose unique challenges.

This study aimed to explore the experiences of trial participants and therapists with VRT in the Challenge trial and to interpret the findings through the lens of implementation outcomes.

Methods

Design and Setting

This qualitative study was part of Challenge, a Danish multicenter randomized controlled trial (RCT) on VRT for distressing voices (November 2020–June 2024). The VRT intervention comprised 7 sessions over 12 weeks (1 design session and 6 avatar dialogue sessions), inspired by the original AT protocol [29,70], with 2 booster sessions between weeks 12 and 24. Avatar dialogues were delivered using a head-mounted display (HMD), with participant and therapist present in the same location. A panic button allowed participants to remove the avatar. Elsewhere, more details are reported on Challenge's methodology [70] and outcomes [43]. Secondary analyses of trial participants' childhood trauma configurations and their relationships with voice-related distress are reported in a separate study [71].

Methodologically oriented toward pragmatism [72-74], this study adopted a practical and flexible approach to explore participants' and therapists' experiences. Semistructured interviews [75] were conducted at the 3 trial sites: a research unit in Copenhagen and 2 outpatient clinics in Aalborg and Esbjerg offering Early Intervention Services (OPUS) [76] or Flexible-Assertive Community Treatment (F-ACT) [77]. Reporting adheres to the standards for reporting qualitative research (SPQR) [78].

Trial Participant Selection

In this study, both voice-hearing individuals (trial participants) and therapists are considered participants. To differentiate, "trial participants" refers specifically to those receiving VRT. A multistage, purposeful sampling strategy [79] was used to ensure variation in voice-hearing duration and site representation, with interviews conducted within a year of therapy completion (mean 7.55 months from baseline). Participants were contacted by phone and therapists by email.

Sample size was guided by the concept of information power [80] and balanced with practical constraints. Information power was appraised throughout the interview process, leading to the conclusion of recruitment after interviewing 10 trial participants and 8 therapists. At this point, authors MJC and MPR judged the sample to hold sufficient information to elucidate the study's aims. There is no minimum sample size for thematic analysis [81]. However, a model of data saturation has recently been proposed [82], published only after the current study's interview phase. Although saturation was not a guiding principle due to debates about its compatibility with reflexive thematic analysis [83], the study's sample falls within the range estimated to be sufficient for achieving saturation [84].

Of the 21 trial participants invited, 10 were included, while the remaining were unavailable (n=5), unreachable (n=3), excluded due to acute psychosis (n=1), or declined without explanation (n=3). All trial participants met the Challenge trial's inclusion criteria, including ages 18 years and older, a schizophrenia spectrum disorder diagnosis (excluding schizophrenia simplex and schizotypal disorder; *International Classification of Diseases-10*), persistent auditory hallucinations (≥3 months, Scale for the Assessment of Positive Symptoms [SAPS] score ≥3), in Danish psychiatric care, able to consent, stable antipsychotic regimen ≥4 weeks, insufficient response to current or ≥2 past antipsychotics if no current antipsychotic regimen [70]. Of the 11 eligible therapists, 8 still active in the trial agreed to participate; 3 were no longer involved or unreachable.

Individual Interview and Data Collection

Trial participants were interviewed individually, while therapists were interviewed in site-based groups, with 2 - 3 in each. Interview guides were developed in mid-2022 based on (1) a lived experience focus group with trial participants (n=4) not involved in the current qualitative interviews and (2) two focus groups with therapists (n=6) from two trial sites. After feedback from research members (3 of 6 responded) and external VR experts (1 of 3 responded), a pilot trial participant interview was conducted and later included among the final 10. While the guides were continuously revised, only minor changes were

XSL•FO
RenderX

made to improve clarity, ensuring consistency across interviews (see Multimedia Appendices 1 and 2 for English versions).

Trial participants could bring a peer or case manager, resulting in one interview being with the presence of nonparticipants. Interviews were audio-recorded and transcribed with GoodTape AI [85]. Author MPR proofread, anonymized, and verified the transcripts for accuracy. Trial participant interviews (March 2023-October 2023) lasted 51:34-112:46 minutes (mean 86:03 min), and therapist group interviews (April 2023-February 2024) lasted 140:21–148:06 minutes (mean 145:11 min).

Background information on trial participants was obtained from the main trial records and supplemented during the interview, while therapist information was collected during the interviews and through follow-up email correspondence.

Research Team and Reflexivity

The research team had clinical and, in most cases, research experience with psychotic disorders, with backgrounds in nursing, clinical psychology, or medicine. Most were involved in the Challenge trial or employed at participating mental health organizations. Data collection was conducted by team members experienced in clinical interviews, though with varying qualitative research experience. The first author led the analysis, supervised by author RJ, a qualitative research specialist. Reflexivity was an ongoing process, addressing how previous relationships, roles, and disciplinary backgrounds might shape interpretation. RJ, who was not involved in the trial, provided independent oversight to strengthen the reflexive approach and to prevent trial-involved team members from making interpretations not fully supported by the data. A total of 2 authors who were also therapist interviewees did not participate in analysis. Some interviewers were professionally acquainted with participants. Notable dependent relationships included: (1) the first author having conducted a research assessment of a trial participant as part of the Challenge RCT; (2) the last author being both a therapist interviewee and manager of the authors MJC, MPR, and CDN; and (3) interviewers being colleagues—near or distant—of interviewed therapists. These relationships and associated hierarchical dynamics were openly discussed regarding their impact on data collection and analysis. Reflexivity was supported through notes, memos, collaborative coding, and critical discussions of diverging interpretations, enabling a reflective and iterative analytic process.

Data Analysis

Data were analyzed thematically, informed by reflexive thematic analysis [86-90], as it emphasizes researcher reflexivity and treats subjectivity as a resource. A critical realist position was adopted to locate and make sense of trial participants' and therapists' descriptions of their experiences of VRT. Critical realism theorizes an independent truth as possible but unreachable due to each individual's different locatedness and perspectives [91]. Coding and theme development were inductive to remain grounded in participants' accounts, while recognizing that it was inevitably shaped by underlying theoretical assumptions [92], including (but not limited to) the research team's optimism about VRT's efficacy [43] and intention to interpret themes in relation to implementation outcomes [56]. The 6-stage analysis progressed from data familiarization, open coding, and initial theme identification (stages 1 - 3) to theme refinement and presentation (stages 4 - 6). In stages 1 - 2, authors MJC, MPR, and CDN familiarized themselves with the data and coded transcripts, with participant interviews coded by authors MJC and CDN, and therapist interviews by MJC and MPR. Stage 3 involved developing initial themes separately for trial participant (authors MJC and CDN) and therapist (authors MJC and MPR) transcripts. Stage 4 involved the reinvestigation and refinement of initial separate themes. Stages 3 - 4 were repeated across the entire material, integrating trial participant and therapist materials (authors MJC and MPR). Information power was deemed sufficient as no new insights were generated in the final trial participant interview, and all available therapists had been interviewed. Coding was managed using Nvivo (version 14; Lumivero), and codes were collapsed into themes using online whiteboard Miro. Stage 5 involved clearly defining and naming the final themes. The processes in stages 3 - 5 involved ongoing discussions among authors MJC, MPR, and CDN, supervised by RJ. Stage 6 involved writing up the analysis, with participant quotes used for illustration (see the Thematic Analysis subsection in the Results section). For theme validation, all Challenge experiment group participants from one trial site (n=44) were invited to a workshop to discuss the findings (see Multimedia Appendix 3).

Ethical Considerations

The Challenge RCT was approved by the Committee on Health Research Ethics of the Capital Region of Denmark and the Danish Data Protection Agency (Project ID: H-19086621). This study adhered to the Declaration of Helsinki, with no additional approval required. All participants received written information and provided written informed consent. Confidentiality was maintained; however, participants were informed that any disclosure of significant risks during interviews would be shared with their care teams. All interview data were anonymized before analysis to protect participant privacy. No compensation was provided.

Results

Demographics

The sample included 10 trial participants (Participants 1-10) who had completed six (n=1) or all seven (n=9) VRT sessions, and eight therapists (Therapists 1-8) from the Challenge trial across sites (see Table 1).

XSL•FO
RenderX

Table . Background information of interviewed Challenge-virtual reality–assisted therapy (VRT) trial participants (demographic, clinical, and trial-related characteristics) and therapists (demographic and professional characteristics).

Characteristics	n (%)	Mean (SD)
Participants		
Age (years)	__[a]	34.5 (13.6)
Gender		
Men	3 (30)	—
Women	6 (60)	—
Other	1 (10)	—
Ethnicity		
White	10 (100)	—
Black or mixed Black	0 (0)	—
South Asian or mixed South Asian	0 (0)	—
Diagnosis (*ICD-10*[b])		
DF20.0 Paranoid schizophrenia	5 (50)	—
DF20.3 Undifferentiated schizophrenia	2 (20)	—
DF20.9 Schizophrenia, unspecified	1 (10)	—
DF22.8 Other persistent delusional disorders	1 (10)	—
DF28 Other nonorganic psychotic disorders	1 (10)	—
Recruitment site		
Capital Region of Denmark	4 (40)	—
North Denmark Region	4 (40)	—
Region of Southern Denmark	2 (20)	—
Level of education		
Primary school	4 (40)	—
Secondary	2 (20)	—
Vocational	2 (20)	—
Further education (<2.5 y)	2 (20)	—
Higher education (>2.5 y)	0 (0)	—
Employment		
Full time or part time	0 (0)	—
Student	1 (10)	—
On sick leave	1 (10)	—
Disability retirement (or retired)	5 (50)	—
Unemployed	3 (30)	—
Duration of voice-hearing (years)	—	16.7 (10.43)[c]
Number of voices		
1	4 (40)	—
2 - 5	4 (40)	—
>5	2 (20)	—
Self-reported experience with VR[d] (range 1 - 7; with 1="beginner" and 7="expert")	—	1.3 (1.9)
Months from trial baseline assessment to qualitative interview	—	7.55 (2.73)

XSL•FO
RenderX

Characteristics	n (%)	Mean (SD)
PSYRATS-AH-Total[e] (baseline)	—	33.9 (2.84)
Therapists		
Age (years)	—	46.3 (9.3)
Gender	—	
Women	8 (100)	—
Ethnicity		
White	8 (100)	—
Education		
Psychologist	5 (62.5)	—
Nurse	2 (25)	—
Medical doctor	1 (12.5)	—
Recruitment site		
Capital Region of Denmark	2 (25)	—
North Denmark Region	3 (37.5)	—
Region of Southern Denmark	3 (37.5)	—
Years of professional experience before trial entry		
Working within one's profession	—	17.1 (9.2)
Working in psychiatry	—	15.1 (7.6)
Working with psychotic disorders	—	13.9 (8.3)
Conducting therapy	—	13.3 (6.9)
Number of challenge therapies commenced	—	13.6 (6.3)

[a]Not applicable.

[b]*ICD- 10*: *International Classification of Diseases*.

[c]Voice-hearing durations used for sampling were based on percentiles from the Challenge trial: 0.05%=1.75 years (n=0), 0.1%=4 years (n=2), 0.25%=7.5 years (n=2), 0.5%=14 years (n=1), 0.75%=20 years (n=2), 0.9%=30 years (n=2), 0.95%=35.5 years (n=1). No trial participants had <2 yrs of AVH (0.05 percentile) and completed their last therapy session within a year, making it impossible to sample from this percentile.

[d]VR: virtual reality.

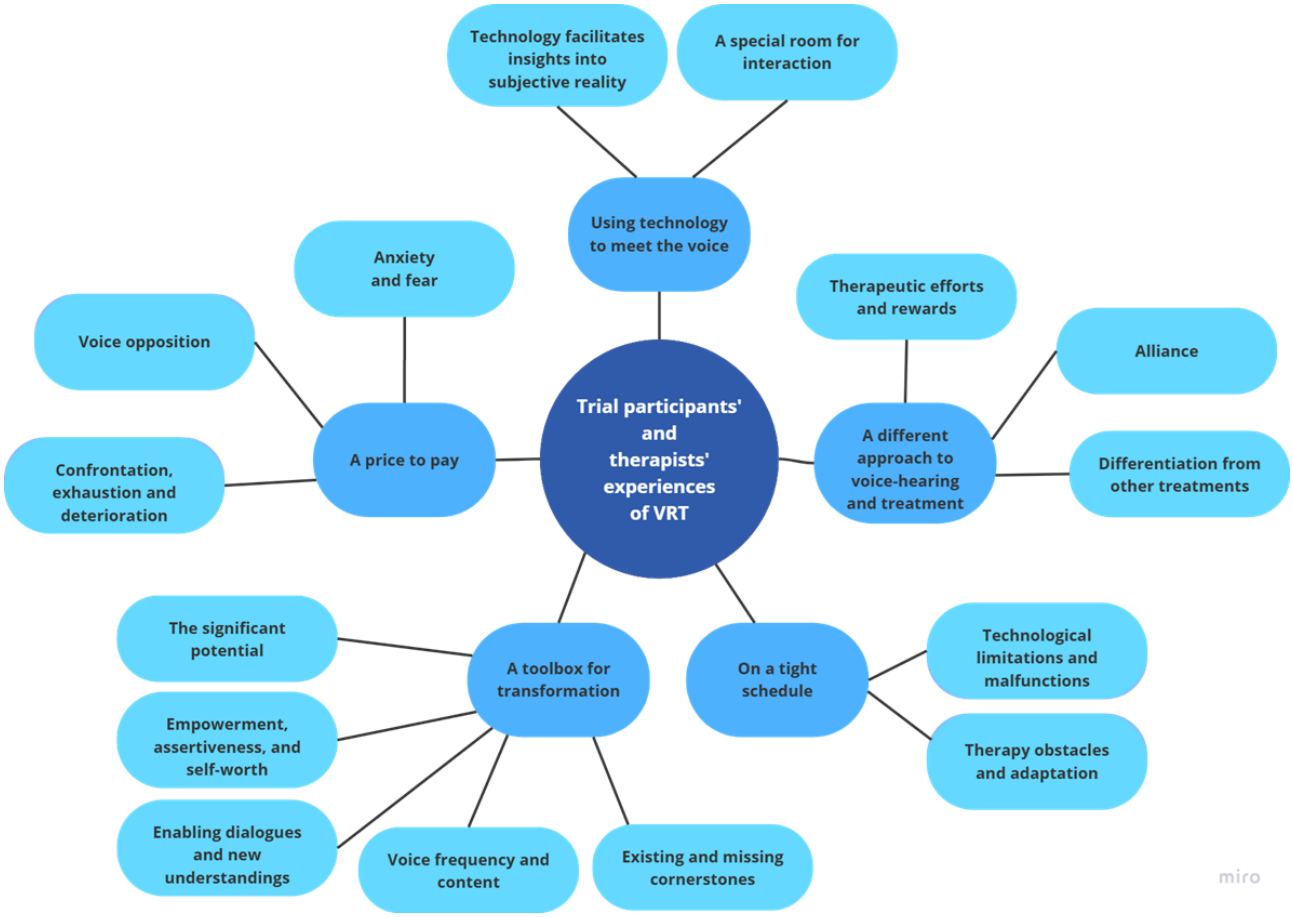
[e]PSYRATS-AH-Total: Psychotic Symptom Rating Scales-Auditory Hallucinations-Total.

Thematic Analysis

Thematic analysis resulted in five overarching themes: (1) a different approach to voice-hearing and treatment, (2) using technology to meet the voice, (3) on a tight schedule, (4) a toolbox for transformation, and (5) a price to pay. Within each theme, subthemes were separated out (see Figure 1) with illustrative quotes provided for each (see Table 2).

XSL•FO
RenderX

Figure 1. Thematic map illustrating themes and their associated subthemes generated through the thematic analysis of Challenge virtual reality–assisted therapy (VRT) trial participant and therapist interviews.



XSL•FO
RenderX

Table . Outputs of the thematic analysis, showing themes and subthemes derived from Challenge-virtual reality–assisted therapy (VRT) trial participant and therapist interviews, with illustrative quotes for each.

Theme and subthemes	Illustrative quote—trial participants	Illustrative quote—therapists
Using technology to meet the voice		
Technology facilitates insights into subjective reality	You got a bit more of a tangible experience of it [in VR], than if we had just sat and talked face to face (…) because it was his voice, his image. And I spoke directly to him [the voice] while looking at him. (…) It felt a bit more realistic. [Participant 6]	You've got that aspect of observing the dynamic between the patient and the avatar and experiencing it live [in VR]. (…) You definitely get a lot of information that you wouldn't if we were just sitting across from each other at a desk, looking each other in the eyes. [Therapist 5]
A special room for interaction	It's strange. For many years, I've been unaccustomed to having a dialogue with my voices. But the last two times [in VR], I've sat and talked with the voice, with him. I haven't been able to do so for 20 years. [Participant 8]	I'm well aware that I'm not wearing a VR headset. Still, I feel like I'm in the room. I can follow along on my screen, and I can see what the patient sees. I'm there as the avatar and as myself (…). I have a sense - and often talk with the patients about it - that I'm going in there with them. [Therapist 5]
A different approach to voice-hearing and treatment		
Therapeutic efforts and rewards	You're working with someone who is actively trying to support you, but at the same time (…) is supposed to sabotage you. It's kind of strange because the therapist has to be both hero and villain. [Participant 10]	The biggest challenge is figuring out how we can create a change in the [voice] relationship (…). Even though we have a manual, we end up thinking very individually about each person's history. There are many things to consider. [Therapist 4]
Alliance	[The therapist] is skilled. At putting herself in my shoes. Really listening. Giving input. I have really been able to use it for a lot. Even if it had just been conversations with her, I would still have been able to benefit greatly. But it has just become even better because there was an avatar. [Participant 6]	We were in this together very quickly. (…) We met in a shared work environment, in a shared work project, and it [the alliance] was very easy to establish with the patients. It didn't take long. [Therapist 1]
Differentiation from other treatments	You just need to take some medication for that, right?' That's kind of what it [previous treatment] has been like. (…) I've talked about what they [the voices] say and so on. But not talked about what I could do to make it better. I think that's what I've gotten here, right? [Participant 6]	Quite a few spontaneously say: 'It's nice that there's someone who wants to listen to it', or 'who is so open in asking about the experience', because it can be something that's hard to make space for elsewhere. [Therapist 5]
On a tight schedule		
Technological limitations and malfunctions	I think it [reaction to technology problems] depends a lot on how it's handled. In general, [the therapist] is extremely professional and super sweet. We just continued talking while she tried to get it up and running and she made a joke and such. So, it wasn't like you got annoyed or angry. It was just delayed a bit. [Participant 2]	(…) There was a glitch in her headset. She could hear what she had said herself, playing on a loop. (…) It probably took 10 - 15 seconds before I realized something was wrong. It worsened her voices for a short period after the session: She heard both the voice and herself in a loop. Unfortunately, the next two times, the equipment still didn't work properly. Not the same issue, but we couldn't get started at all. [Therapist 7]
Adaptations made	We just had to stop a few times and do some breathing exercises and take off our shoes and get grounded and things like that because it was really, really hard. [Participant 5]	I've had a couple of times where it was clear that we couldn't do it today. Instead, I had to help them with something else. And then schedule a new appointment. There were a few times where I thought it would have been unethical if I had just said, 'You can talk to your case manager about that next week, now we're putting on the VR headset. [Therapist 8]
A toolbox for transformation		

XSL•FO
RenderX

Theme and subthemes	Illustrative quote—trial participants	Illustrative quote—therapists
The significant potential	When he [the voice] says things like: 'If you don't swallow a pen', or 'if you don't cut yourself', or 'if you don't overdose, your family will die.' Then, I go in and say: 'Isn't it really you who's afraid of what might happen to my family?' 'Yes, that's it.' 'But are you sure my family will die?' No, he wasn't sure about that. In that way, it has helped. [Participant 8]	It [the voice] was her enemy before because it always talked condescendingly, and she had perceived it as if it didn't want anything good for her. But then it came to her indirectly - what its actual function was - and then she could begin to see it differently and view it as a friend instead. It just needed to express things in a different way. [Therapist 2]
Empowerment, assertiveness, and self-worth	I think a lot of it is having the tools to sort of set them [the voices] straight. I'm not exactly sure if they can understand it, but it's about the fact that it's also given me better self-esteem because now I have all these tools for it. [Participant 9]	They [the participants] realize: 'We're sitting here openly talking about the voice, the voice doesn't like it, but nothing happens.' In some way, the voice's power diminishes because it becomes less threatening - they realize that the voice can't do everything it claimed it would. [Therapist 7]
Enabling dialogues and new understandings	It is less aggressive. It has its periods where it can really escalate if it thinks something very dangerous is about to happen. Now, I can almost tell them to calm down. Then, it's as if it tries to speak a bit more decently. [Participant 4]	When you have that approach that the voice is actually there to help you - it's just been doing it in the wrong way. That's something that has made them [the participants] start thinking differently about the voice: 'Okay, it might always have been discouraging, but it actually did it with good intentions. It just came out the wrong way.' [Therapist 2]
Voice frequency and content	It's not so much with insults anymore. Now, we can have a regular dialogue. (...) Instead of calling me 'a big fat idiot', then 'that might not have been so well done' or 'you need to do this differently.' It doesn't have those negative words in it. (…) I can feel that I'm less negative myself as well, so it has a positive effect on me that my voice is less negative. [Participant 3]	It [the voice] has started giving compliments instead, or it has just turned into something like: 'Remember your glasses when you go out the door', or 'you just spoke wrong when you were crocheting'. From being completely critical to suddenly being an extra attention reminder. [Therapist 8]
Existing and missing cornerstones	Unless you're deeply psychotic, you know that you're sitting in a hospital and have a therapist sitting right there (…). I mean, she [the therapist] doesn't stop existing just because I can't see her. I think, no matter what, you have to accept the premise that this is something we're playing. [Participant 4]	You are missing a cornerstone in the treatment if you cannot get the patient to buy into the premise. We've had someone who said: 'I know that this isn't real. But if I'm going to get something out of this, I also know that I must pretend that it's real'. And he benefited greatly from it. It's fair enough if they say they know it's not reality. [Therapist 1]
A price to pay		
Anxiety and fear	It was very anxiety-provoking at first. I almost had a panic attack because it was very intense. Or not intense, but it was very overwhelming. To see him [the voice]. [Participant 8]	Patients could be very anxious at the beginning. More than we might have expected. Also about entering VR. We must spend a bit more time at the beginning to go at a pace that people can keep up with. Some are very afraid of their voices. Consequently, we cannot proceed as quickly as in the manual. [Therapist 4]
Voice opposition	It got worse. A lot. Mainly because he [the voice] asked me to stop attending, and I chose to keep talking. I talked to [the therapist] about it and my [regular] therapist in psychiatry, and eventually, I decided to keep going. It definitely affected my daily life quite a bit for a while because it got really bad, and I was hospitalized. [Participant 2]	Well, I have also encountered voices that resort to screaming when they can't get through with criticizing and make them [participants] stop the therapy (…). We've worked with this in VR (…): 'I'll scream until you stop this therapy. Until you do what I say. [Therapist 4]
Confrontation, exhaustion, and deterioration	I could have some slightly worse days because it could be quite draining, energy-wise, to participate. It could lead to a few negative days. Those are aftereffects of having worked with my voices. The voices could be very negative and macabre. (…) It happens when I'm under pressure. [Participant 3]	She [the participant] experienced that her father [the voice] was holding her in a vice. Literally, and it went on for three weeks. The supervisor and I thought we should stop (…). I didn't dare to continue, but she wanted to go on. And it ended with the voice disappearing. A fantastic process, but it was definitely intense. [Therapist 3]

XSL•FO
RenderX

Theme 1: Using Technology to Meet the Voice

VR enabled participants to share the typically private experience of voice-hearing and offered a new way to interact with the voice(s) through the avatar.

Technology Facilitates Insights Into Subjective Reality

Several trial participants described giving the voice a face and a body as an essential element of the therapy:

> *I desperately wanted something different than a silhouette. Because, I have no control over a silhouette. A face is easier to control. It's easier to have a conversation with a face than just something floating around.* [Trial Participant 3]

Some had a clear idea of how the voice(s) looked, while others did not. A few were guided by the voice(s) in designing the avatar, but most experienced that the voice(s) would interfere with the process. For one trial participant, this raised a dilemma: whether to create the avatar based on their own perception or to let the voice dictate its appearance:

> *Very quickly I came to terms with the fact that the avatar should be made the way I saw the voice, not the way the voice saw itself.* [Trial Participant 9]

A high degree of resemblance between the voice and avatar was generally reported, though complete convergence was rare. Most trial participants found some level of similarity essential:

> *That makes the premise easier to accept. I wouldn't be able to accept it at all if it were, like, a little girl sitting there and I was supposed to accept it as my voice. I hear an older man, so it's not, like, a little baby. That would be completely ridiculous. So, I think it's extremely important that you're allowed to design it yourself.* [Trial Participant 2]

However, some trial participants felt that a less exact match made them feel safer:

> *If it resembles too much, then I would just start hearing the voice instead of what [the therapist] says. It would be a bit hard to know who's speaking, if it were completely accurate. So, I felt it was actually a bit comforting that it wasn't completely accurate during those first times I spoke. But it could have been more accurate in the later sessions.* [Trial Participant 10]

Achieving insights into what the voice looked like and merely seeing it in front of them was described as helpful or facilitating relief by some trial participants, with one participant mentioning that visualizing it provided a point in space to direct their attention. In some cases, the avatar was found to resemble a known persona from the trial participant's life, as noted by a therapist:

> *I could clearly see who this avatar resembled. Luckily, she [the trial participant] also realized it herself at one point and said, 'Oh, it's me. This is a part of me, isn't it?' Already at that moment, something shifted for her. It was interesting to witness because I don't think we would have necessarily reached that realization if we hadn't been working together on shaping this avatar.* [Therapist 6]

Through VR, therapists gained insights into the subjective world of voice-hearing, experiencing firsthand how intense, human-like, and overwhelming the voices could feel. It was as if they were granted access to a private, intimate space, allowing them to witness the unfolding dynamics and participants' real-time reactions—informing their therapeutic approach:

> *They are there with their voice, even though they know it's VR. You get a completely different insight into the subjective world of living with a voice compared to just asking about it.* [Therapist 4]

Some therapists noted that collaboratively designing the avatar helped trial participants articulate their perceptions, acting as a shortcut to establishing shared understanding. Most trial participants echoed that the VR encounter facilitated a dialogue that was not obtainable in real-world settings:

> *It [VR] meant that we could somehow get around that therapist-patient barrier, which I think is very clear in psychiatry. Because [the therapist] had to take on the role of the voice. This way, she was kind of invited into a world that I would otherwise have a hard time giving her access to. You would never have that kind of dynamic conversation if you were talking about your symptoms. She kind of got to experience them.* [Trial Participant 2]

Audio recordings allowed trial participants to revisit their responses and progress, with some finding them helpful for reflection. Others shared recordings or avatar images with relatives, peers, or professionals. While this could provoke anxiety about others' reactions, it also served as a new tool to communicate and explain the inner experience of voice-hearing, fostering shared understanding:

> *It's not just inside my head anymore. It's kind of out in the open. My loved ones have been let in on what I struggle with.* [Trial Participant 6]

A Special Room for Interaction

VR was seen as creating a distinct therapeutic space, where interacting with the voice through the avatar helped focus attention and reduce distractions. Both trial participants and therapists valued this immersive environment, which offered opportunities for deeper engagement while also presenting challenges:

> *It was transgressive [being in VR], because suddenly you're in another place - even though your brain knows you're sitting in a treatment room that doesn't look like that. It's also about giving up control – about not knowing what's going on around you.* [Trial Participant 3]

Highlighting the unique experience of creating and interacting with a visual representation of the voice, some trial participants described how the avatar's tangibility, presence, and direct gaze brought the voice "to life"—as if it were a real person in VR:

> *It just became so real. I usually spend a lot of energy thinking, 'No, it's not real, it's just in your head.' But*

XSL•FO
RenderX

> *then it suddenly felt even more real when it was right there. (…) It feels much more real in VR.* [Trial Participant 5]

Most trial participants did not distinguish between the avatar and the voice. Instead, as reflected in emotional and physical impulses like wanting to hit or confront the avatar, most experienced the interaction as a direct encounter with the voice:

> *Due to me being inside this 'zone bubble' it was as if we, in fact, sat in front of each other and talked.* [Trial Participant 6]

However, while some trial participants became fully immersed, forgetting time and surroundings, others remained highly aware of being in the therapist's office rather than a different world, for example, due to VR graphics:

> *It is so different that it feels a bit unreal (…). You're made of flesh and blood… And when you're wearing the glasses [HMD], it's like you're in a cartoon.* [Trial Participant 1]

On one hand, VR graphics enhanced concentration for some and reassured others that it was not reality, making it less frightening:

> *It actually made me feel safe. Because I knew that if something went wrong, it could just be taken away. Click the [panic] button if it gets too much. Just take off the glasses [HMD].* [Trial Participant 10]

On the other hand, some trial participants had to actively "accept the terms," as the setup or the realism of the interaction did not come naturally to them. Consequently, some approached the intervention with ambivalence:

> *When you know it's not real… but you have to pretend… that it is real, you have to try to… trick your own body. I know that the conversation with the voice is actually just the psychologist. So, I kind of have to try to convince myself that it's not the psychologist.* [Trial Participant 10]

Potentially, having to pretend could hinder immersion, yet the awareness that VR was not real also offered reassurance:

> *The feeling I get when I can't distinguish between reality and non-reality - then I get really scared. I didn't feel that in the same way with this [VR], because I could already sense that it wasn't real.* [Trial Participant 5]

The mix of structured and unstructured role-play was challenging for some. While lack of scripted lines was anxiety-provoking for a few, others found open dialogue more authentic. VR helped some participants engage more fully, even those initially skeptical, as noted by a therapist:

> *That he had to put on the VR glasses [HMD] and engage with what he saw, it just made him go along with the game. Then it was just a completely different dialogue we had.* [Therapist 8.]

Theme 2: A Different Approach to Voice Hearing and Treatment

This theme highlighted therapists' efforts and rewards in delivering VRT, the strong therapeutic alliance fostered, and participants' appreciation for explicitly acknowledging and addressing voice hearing.

Therapeutic Efforts and Rewards

Therapists were thoughtful in delivering high-quality VRT, striving to portray authentic avatars, carefully timing VR exposure for optimal impact, and tailoring interventions by connecting symptoms to life experiences—all while upholding high personal expectations of their performance:

> *Having to go in and play a role also requires that you can clearly remember what exactly you said last time – where the dialogue ended. So, on days when it wasn't possible for me to go back and either listen to or read what I had written, I wasn't entirely satisfied.* [Therapist 7]

Therapists described VRT's delivery as a steep learning curve needing thorough training and supervision. Managing avatar dialogue improvisation, monitoring exposure and reactions through the HMD, and adhering to the manual simultaneously was a demanding task. Particularly, reproducing negative voice content in avatar dialogues could be uncomfortable:

> *I also found that really difficult at first, because it's just miles away from how I usually speak to the patients. You would never say those kinds of things to patients in a regular setting. But that's what it takes for the dialogue to be relevant and believable.* [Therapist 5]

Remembering the intention of accurately reproducing voice content helped therapists step into the avatar role, though it remained challenging. Direct engagement with the content connected therapists to the reality of the voice, and thereby to the person:

> *The harshest voices could cross some of my own boundaries, in terms of what I had to say. Having to say: 'You whore child.' That didn't roll easily off my tongue, but that was what the voice said. It's about finding the language that the voice uses. There's no exposure to it if it's some kind of light-version. That part hasn't necessarily gotten any easier. (…) But it is a role you're playing, after all.* [Therapist 8]

Therapists could feel vigilant and nervous delivering VR dialogues, aware of the intensity and potential risks involved:

> *You want to deliver something… really good.* [Therapist 5]
>
> *Yes, and maybe you also become a little nervous, because I have a sense that you could deliver something that could cause harm, right? Because it's so intense when you're in there [in VR]. There's a lot at stake in there, so you're on the ball.* [Therapist 4]

XSL•FO
RenderX

Conversely, therapists were inspired to continue by witnessing participants' breakthroughs, experiencing a sense of accomplishment and motivation from positive outcomes:

> *That's such empowerment, right? To witness it in the patients is truly fantastic. Once you've experienced it a few times, you get a bit of a taste for it. And think, it's okay [reproducing negative voice content].* [Therapist 4]

All therapists reported that delivering VRT demanded extra effort—not only to effectively integrate VR into therapy but also to handle the technology and its malfunctions:

> *I think we did it 200 percent and it has worn me out. I also took on a lot of responsibility for the technical aspects as well.* [Therapist 2]

Alliance

Most trial participants and therapists felt the therapeutic alliance formed faster than in usual treatment. Therapists highlighted VR's unique role in creating a shared experience by referring to the concept of "the common third." This concept from social pedagogy refers to "an activity or an experience they have together which feels unique in a positive way" [93] or "an arena for equal sharing and participation" [94]:

> *It's a common third: 'Now we go in here and do this together. Then we go out and talk about the experience.* [Therapist 4]

Most trial participants echoed the value of the shared activity and collaborative approach in VR, particularly during avatar creation, which was often a powerful experience:

> *The way you collaborate, that's what works for me-doing it together. If I had made the avatar myself, I'm not sure I would have gotten as far.* [Trial Participant 3]

Therapists were key in enabling shared experiences, with their role-shifting, guidance, and support during avatar dialogues seen as essential to therapeutic collaboration. While some trial participants felt self-conscious speaking to an avatar controlled by a nearby therapist, others found the therapist's dual role comforting but odd:

> *The fact that she [the therapist] was both there as a challenger and a support, was strange, but in a way reassuring to know that the person who is challenging me right now is also one of my biggest supporters throughout this process.* [Trial Participant 9]

While participants valued common therapeutic factors like the therapist's qualities, VR appeared to add a distinct dimension to the alliance. It fostered a deeper, shared understanding between therapist and participant—beyond what could be attributed to therapist skill or receiving a novel talk therapy alone:

> *It [using VR] meant a lot. That she [the therapist] kind of got to have a bit more of a real image… a dynamic image… of the world that I carry with me. That it wasn't just me sitting there, having to talk about what had happened since last time, but that we could, in a way, create what was happening together.* [Trial Participant 2]

The combination of VRT's shared experiences and common therapeutic factors made trial participants feel safe and confident in the therapist—marking a notable contrast to some participants' prior experiences with psychiatric care:

> *There was just some kind of respect and understanding that I haven't had elsewhere. I've had treatment providers with 25 years of experience, but they had no idea how it was like [hearing voices].* [Trial Participant 4]

Differentiation From Other Treatments

For trial participants, the main motivation for joining the Challenge trial was the desire for an alternative to medication—driven by past experiences of limited effectiveness, intolerable side effects, or a preference for gaining tools to manage voices rather than focusing solely on "symptom treatment." Most trial participants felt they worked alongside the therapist to understand and manage the voice(s), valuing being treated as more than "just a number." One therapist recalled a participant finding the participatory and active nature of VRT helpful:

> *It's not just about adjusting medication. That's more something others do to him [the trial participant]. But if he gets something he can use and work with himself, then he gains a sense of control.* [Therapist 8]

Unlike previous experiences, trial participants appreciated the explicit focus on voice-hearing and found avatar dialogues to be a novel, tangible way to engage with the voice(s). By externalizing a voice, participants and therapists were able to engage with it collaboratively, in contrast to earlier treatments positioning the voice(s) as a problem within the individual. Many valued that the voice(s) were treated as real experiences rather than merely symptoms to eliminate:

> *For the first time ever, it [the voice] has been treated as if it were real. Psychiatry is really good at immediately saying, 'But it's just a symptom of your schizophrenia.' And then approaching it from that perspective.* [Trial Participant 2]

Therapists echoed the positives to focusing on voice-hearing and exploring it in VR, which provided structure, a shared goal, and rapid insight into participants' experiences. They were impressed by participants' ability to stay engaged despite challenges. Many participants expressed relief and satisfaction at finally being asked in-depth about the voice(s)—something they felt was often overlooked in other treatments. While sharing voice-hearing experiences could feel transgressive for some, therapists noted that VRT often made such disclosure easier, helping to normalize the experience. For some trial participants, this openness reduced feelings of shame or self-stigma:

> *What this therapy has taught me is quite a lot, because I know I'm not the only one who has been part of it. All of those who have taken part… they all hear*

XSL•FO
RenderX

voices. It's not just me. That has made it less of a taboo for me to talk about it. [Trial Participant 9]

Theme 3: On a Tight Schedule

This theme explored technological challenges and the adaptations therapists made during therapy.

Technological Limitations and Malfunctions

Trial participants and therapists noted that avatar design limitations—including lack of options for attire, hairstyles, body shapes, or nonhuman features—sometimes reduced how well avatars represented the voice. As a result, therapists occasionally had to adapt planned dialogues to fit the available designs. Besides these limitations, therapists faced challenges learning and navigating the hardware and software, compounded by frequent technical errors:

> *There have been really, really many technical problems. It has been a strong presence for all the therapists. It has disrupted therapy incredibly much. (…) When there's no sound, or no image, or something is wrong - then we can't do it.* [Therapist 4]

Therapists found the technology malfunctions frustrating, especially when they left participants distressed or prevented crucial exposure work:

> *All the technical stuff, I would almost say, that was probably the hardest part. Because everything was so unfamiliar to us. And we didn't have anyone right next to us to help.* [Therapist 2]

Despite therapists' frustration, most interviewed trial participants seemed unaffected by technical issues, with only one finding them stressful. Therapists noted trial participants and relatives were generally understanding, viewing errors as part of a new treatment. Some therapists even saw these moments as opportunities to build rapport, occasionally reversing roles as trial participants helped with the technology, fostering a relaxed atmosphere.

> *It's a bit annoying when it doesn't work right away. Which technology sometimes doesn't do. I mean, sometimes we went half an hour over time. Because the equipment just had its problems every now and then. That didn't really matter to me, to be honest.* [Trial Participant 2]

Therapy Obstacles and Adaptations

Therapists valued the short therapy format for boosting participants' motivation and adherence but often found the schedule too tight. Delays from technical issues, participant needs, and organizational demands required frequent adaptations. While the manual provided helpful structure, it lacked specific guidance for complex cases, prompting creative solutions and sessions to run over time. Such complexities often included high anxiety, delusions, voice dynamics, or unresolved trauma:

> *I remember a case with a lot of traumas. After seven sessions, it just felt really short, because we couldn't get through it all, and the voice was very much connected to the trauma. She [the trial participant] would fall back into these traumas when the voice said things related to the trauma. It was also a person from those traumas that she heard speaking to her. I really wish I had had more time.* [Therapist 4]

Furthermore, organizational constraints sometimes limited therapists' ability to deliver optimal VRT. Competing tasks and multiple cases led to compromises in preparation, with some therapists working after hours to stay ready. The lack of scheduled buffer time for technical issues caused sessions to run over, which a trial participant accepted but noted may be unfeasible in standard care due to stricter scheduling (Participant 2).

Theme 4: A Toolbox for Change

This theme highlighted the changes experienced and reflections on therapeutic targets, mechanisms, and prerequisites influencing outcomes.

The Significant Potential

Although not all trial participants experienced improvement, therapists were optimistic about the therapy's potential, with one calling it "revolutionary" and "life-changing" for some. A trial participant expressed a similar sentiment:

> *I used to lie there when going to sleep, thinking: 'I really hope I don't wake up again'. Now, it's completely the opposite. Now, I go to bed looking forward to waking up the next day.* [Trial Participant 4]

Despite some participants feeling ambivalent about changing or letting go of the voice(s), fearing what the future might bring, most welcomed the changes:

> *I can feel that I have a better life now. (…) I can go shopping more easily and I can go out and walk and stuff.* [Trial Participant 7]

However, some therapists expressed disappointment over measurements not being able to detect changes achieved through VRT, including improved communication with the voice, brief relief, or shifts in tone or content. VRT also brought subtle or indirect changes, including increased energy, enabling participants to pursue goals like obtaining a driver's license, starting school, or building a family:

> *Things happen - maybe one voice among several disappears, or subtle shifts occur in what the person can now decide for themselves, something they couldn't before. These changes can be really difficult to capture.* [Therapist 5]

Empowerment, Assertiveness, and Self-Worth

Assertive communication, boundary-setting, and disrupting the power balance helped trial participants regain control, a change therapists found moving. Most felt more equal or superior to the voice—an unusual shift sometimes met with fear of setbacks. Some became less fearful, more accepting, or able to negotiate with the voices, leading for some to improved relationships or self-worth:

XSL•FO
RenderX

> *I have become more confident in myself because, before, one of the things I was really hard on myself about was that I couldn't figure out how to stand my ground. But through the process [VRT], I've learned that I can do that.* [Trial Participant 9]

Recognizing that the voice is neither dangerous nor powerful may help reduce anxiety, as a trial participant suggested:

> *As I've become less afraid, his power has also diminished. Just automatically. (…) Because as I'm less afraid of him, then I get less upset when he says mean things. So, he has less power to make me sad.* [Trial Participant 2]

In trauma-related cases, a therapist highlighted the usefulness for the avatar to say new, constructive things in VR—potentially improving the relationship, if applied to the voice:

> *I experience, when we talk about trauma, that it [VR] really can make a difference. Having the avatar specifically say some of these things, like, 'It was me who was completely wrong about that. And not you.' Or, 'I don't have that power today. I can see that now.' To separate the past from the present: 'I was a big grown man back then, and you were a little girl. You couldn't do anything. So, I had the power then.* [Therapist 4]

Enabling Dialogues and New Understandings

Many participants saw the ability to engage in dialogue with the avatar—and to transfer this to the voice(s)—as a key outcome of therapy, transforming it from an unresponsive, 1-way communication:

> *Before, it was just those same sentences they used all the time, and no matter what I answered, it would just be degrading. Now, we can actually sometimes have a conversation; with some of them I can talk.* [Trial Participant 4]

Besides enabling dialogue, therapists aimed, when feasible, to foster a new understanding of the voice and its intentions:

> *Maybe it warns me when I'm in situations that are hard to handle. I mean, it's about getting a different understanding of what it stems from.* [Therapist 6]

Some trial participants began to view the voice(s) as having positive intentions but poor communication skills, making them feel more human-like. This shift helped some disengage from the demeaning messages and adopt more accepting, compassionate responses, improving their relationship. However, not all participants embraced this more positive interpretation, as a therapist noted:

> *I've also had several patients who have said that they simply can't see that there could be any meaning to it. They're really angry at their voice because it has ruined their whole life, and they can't see anything positive about it (…) and that's something you have to acknowledge.* [Therapist 3]

Voice Frequency and Content

Some trial participants reported a reduction in voice frequency to varying degrees, with fluctuations occurring within the same person, ranging from minutes to hours, depending on the day:

> *Now I'm getting used to it a bit, right? But at first, I thought there must be something completely crazy going on, since there were no voices—that I had been killed or something like that—but there wasn't.* [Trial Participant 8]

Therapists observed changes in voice content, including voices asking for participants' opinions, becoming less self-assured, adopting a neutral, observational tone, or even becoming more positive. Some trial participants echoed such changes:

> *They might say, 'Oh, I think that guy is spying on you and keeping an eye on you.' And then they ask, 'What do you think?' I'd say: 'I don't think anyone is really trying to watch me.' 'Oh no, they could actually see that.'* [Trial Participant 8]

Existing and Missing Cornerstones

Therapists reflected on factors influencing therapeutic outcomes, noting that some trial participants saw no improvement, especially when voices lacked a relational aspect or power dynamic, making them harder to engage or shift through therapy:

> *It [the voice] was like continuing in the same vein, with three fixed phrases. We just couldn't get it to stop. It might also be me, maybe I approached it the wrong way, but it felt very stuck to me* [Therapist 4.]

Barriers perceived by therapists included participants reluctant to disclose voice experiences, lack of emotional reactions, or difficulties with (voice) presence. Untreated trauma and delusions could sometimes impair progress. However, those individuals that therapists judged to have more limited cognitive ability seemed to do just as well as those judged to be of higher intelligence.

Therapists suggested that positive outcomes were more likely when trial participants showed inner motivation, persistence through challenges, and readiness to form a new relationship with the voice(s). Factors like being further along in recovery, support from others, completing homework, and making life changes outside therapy also contributed. Especially, therapists viewed engagement, (voice) presence, and immersion as deeply interconnected and essential to the therapeutic process:

> *I think immersion already begins when we create the voice, right? That they engage, that they relate to things like eye color - but if they just say, 'just pick one,' or something like that. That always signals a little bit of a bad start, right? In terms of their immersion* [Therapist 8.]

Multiple factors were suggested by both trial participants and therapists to influence presence. Beyond graphics, a trial participant highlighted their therapist's endeavors:

> *If it was something we had talked about three sessions ago–she [the therapist] could just make a little, like, move, to bring it in. And then she could create some*

XSL•FO
RenderX

> *kind of complete picture of him [the voice], which made it easier for me to believe in the premise* [Trial Participant 2.]

Theme 5: A Price to Pay

This theme highlights the adverse effects during and between sessions, including anxiety, negative voice reactions, exhaustion, and general symptom worsening.

Anxiety and Fear

The first avatar encounter was intense and anxiety-provoking for most trial participants, triggering strong emotions (eg, anger, crying, or panic) and physical reactions (eg, nausea, hypertension, sweating, or flight). Therapists were often surprised by the intensity of these responses, which, despite participants being informed that anxiety was expected, at times led to emotional flooding and required adaptations or delays.

> *When they enter VR for the first time and have to meet the avatar, I think many of them react strongly. The exposure is really, really high, so you can clearly see that this is really uncomfortable.* [Therapist 6]

Most trial participants echoed that the first encounters in the initial phase could provoke anxiety:

> *The closer it [the avatar] came, the more I couldn't control what was happening inside me. I became furious, angry, and sad all at once. And anxious. Very anxious. Panicked. I was shaking. I collapsed in the chair. It was either run or fight.* [Participant 4]

One trial participant ascribed the anxiety to a loss of control and the unaccustomed situation:

> *In the beginning, it caused anxiety. Because a bit of my control was taken from me. (…) Suddenly, you're sitting face-to-face. I don't normally do that with my voice.* [Trial Participant 3]

Even trial participants familiar with VR reacted to avatar dialogues with emotions like "shock and surprise" (Trial Participant 10) or "sadness and anger" resulting in using the panic button (Trial Participant 9). This suggests that it may not be VR per se, but the specific way it is applied that elicits such reactions, as one therapist reflected:

> *It [VR] really does something to some of the patients. It's almost like they tear the glasses off the first time. So, it's really a powerful tool to be able to use.* [Therapist 2]

Therapists viewed co-presence in the room as essential for providing timely support and ensuring safety during anxiety. The immersive format did not seem to disrupt the therapeutic bond; in fact, trial participants valued the therapist's physical presence as reassuring, with a participant noting it helped them avoid feeling "trapped" in VR:

> *If I had been sitting there all alone, I think it would have been really difficult for me to handle. But when [the therapist] was there and my contact person was there, it was a bit easier to handle the anxiety and all the emotions that came up while sitting inside [VR].* [Trial Participant 8]

Voice Opposition

Most trial participants reported that the voice(s) reacted negatively to their involvement in the therapy—some dismissing it, urging withdrawal, or becoming angry or condescending. Others expressed skepticism, resisted the avatar design process, or disrupted the VR dialogues:

> *It's fraud and nonsense'; 'You can't use it for anything'; 'It's just people who want to play with you; 'They don't want anything good for you'; 'You'll never get rid of us.' It's been things like this all the way through. They [the voices] really tried to talk it [VRT participation] down so I wouldn't believe it.* [Trial Participant 6]

Voices interfering could complicate therapy:

> *The therapist might tell the avatar to be kinder to me, but the voice might say, 'I don't want to,' at the same time. I find it really confusing.* [Trial Participant 5]

Standing up to the voice could have consequences. A trial participant questioned whether posttherapy hospitalizations were linked to VRT, while another required hospitalization during therapy but later resumed and completed the sessions:

> *At first, he [the voice] was totally indifferent. He thought it was a ridiculous trial and that it was a waste of both our time: 'But if that was what I wanted to do, then I should just do it, because I always did what I wanted.' But then, during therapy, he didn't want me to show up anymore and became very threatening and negative. This eventually led to me being hospitalized.* [Trial Participant 2]

Despite interference or opposition from voices, trial participants generally stood up to the voices and continued therapy:

> *I fought against it, at least. I said, 'Now. I have to get through this." Then I told the voices, 'Now you need to be quiet.* [Trial Participant 7]

Confrontation, Exhaustion, and Deterioration

As the avatar encounters could be demanding, some trial participants were exhausted afterward:

> *After each session, I had to go home and lie in bed for half an hour to recover. Because it was really transgressive to go in and talk to the voice in VR and work with it. It takes energy.* [Trial Participant 10]

Confronting the avatar often led to postsession exhaustion, with responses ranging from sleepiness to hyperarousal, insomnia, or worsened symptoms. Some trial participants had experienced similar responses to past stressors:

> *There's always a small price to pay afterward, and that's just how it is. (…) Fatigue and lack of energy, along with increased symptom activity. (…) When I'm stressed, that comes.* [Trial Participant 5]

Although both therapists and participants were familiar with symptoms worsening from new activities, VRT seemed to carry an additional cost for some. Though already accustomed to using medication as a coping strategy, a participant took sedatives after each session, on advice from nontrial staff, to

XSL•FO
RenderX

manage anticipated symptom flare-ups. The intensity and confrontational style of VRT were seen by some therapists as particularly taxing, affecting both participants and therapists alike.

> *You're putting much more pressure on the patient than you do with other treatment methods. You can also clearly see that the patient gets worse. (…) That's one of the burdens as a therapist.* [Therapist 1]

Therapists identified the intensity of VRT as the primary reason for dropout, with some participants opting for less confrontational methods to manage the voice(s):

> *It was too confrontational. The person preferred to adjust their medication rather than being in such a therapeutic process. They thought it was simply too intense.* [Therapist 5]

Discussion

Principal Findings

This study explored 10 trial participants' and 8 therapists' experiences with VRT within the Challenge RCT, generating five overarching themes: (1) using technology to meet the voice, (2) a different approach to voice-hearing and treatment, (3) on a tight schedule, (4) a toolbox for transformation, and (5) a price to pay. Findings share several similarities with qualitative research on AT [57,67], RT [23,24], and TwV [27,28,95], as elaborated below.

A total of 3 implementation outcomes (acceptability, appropriateness, and feasibility) are used to interpret and contextualize findings beyond the RCT. Following conceptual distinctions [56], acceptability is perceived satisfaction with VRT in terms of content, complexity, and comfort; appropriateness is VRT usefulness and relevance; and feasibility is perceptions of whether VRT could be conducted in practice, considering enablers and barriers. Table 3 summarizes the main learnings and implications.

XSL•FO
RenderX

Table . Summary of discussion points highlighting the main findings of identified themes within implementation outcomes (acceptability, appropriateness, and feasibility) and their clinical implications.

Implementation outcome	Findings	Clinical implications
Acceptability		
Content	• (+)[a] Acknowledging approach. • (+) VR:[b] Embodiment and externalization. • (+) VR: Meaningful space for dialogue. • (-)[c] Suboptimal avatar design. • (-) Suboptimal (voice) presence. • (-) Discomfort reproducing content. • (+) Deeper insight into voice-hearing. • (+) Strong alliance.	• Voice(s) can be acknowledged as real phenomena, not symptoms. • Voice(s) can be externalized and made more tangible in VR. • Voice(s) can be dialogued with as avatar(s) in VR. • Voice(s) may have highly specific appearance(s). • To some, challenges remain with establishing (voice) presence in VR. • Reproducing verbatim content is hard, but necessary and meaningful • Avatar dialogues may deepen therapists' understanding of voices. • Shared activity or "common third" may support alliance-building.
Comfort	• (-) Malfunctions, disruptive and frustrating. • (-) Anxiety, exhaustion, and deterioration. • (-) Voice reactions.	• Operational reliability of software and hardware should be ensured. • Individualized starting point and continuous tolerability monitoring. • Voice(s) may oppose therapy.
Complexity	• (-) Demanding therapy for participants. • (-) Role-play and interfering voices. • (-) Nontraditional therapeutic skills.	• Attention should be given to introduction and monitoring. • Consider direct voice communication to create space for therapy. • Training, supervision, and standardized materials are essential.
Appropriateness		
Usefulness	• (+) Positive changes. • (-) Measurement insensitivity to changes. • (+) Individual means and targets. • (+) Sharing avatar picture or recording.	• Changes can be transformative, but also subtle and multifaceted. • Individualized meaningful outcomes could be discussed. • Treatment focus should be individualized. • Encourage sharing to facilitate normalization and reduce stigma.
Relevance	• (+) Embodiment and externalization. • (+) Avatar dialogues. • (-) Assertiveness versus compassion.	• Voice(s) are more tangible when given a face. • VR-supported assertiveness practice. • Flexibility in whether compassion comes from therapist or avatar.
Feasibility		
—[d]	• (+) Therapists' motivation and efforts. • (+) Participants' motivation and resilience. • (-) Participants' engagement and reactions. • (-) Limited organization resources. • (-) Technological malfunctions.	• VRT[e] requires dedicated therapists as many elements will be new. • Consider whether the timing is right. • Sensitivity for optimal exposure. • Organizational arrangements should allow sufficient time for therapy delivery, including handling of diversions or unexpected disruptions.

[a](+): positive aspect of virtual reality-assisted therapy (VRT).

[b]VR: virtual reality.

[c](-): area of concern.

[d]Not applicable.

[e]VRT: Virtual reality–assisted therapy.

XSL•FO
RenderX

Acceptability

Areas of high acceptance included the novel approach to working with voices, the specific focus on voice-hearing, and the collaborative relationship (themes 1 - 2). Contrastingly, technology malfunctions, limited resources, and participant discomfort were areas of concern (themes 3 and 5).

Content

The Acknowledging Approach

Trial participants appreciated that voices were acknowledged and approached as a real phenomenon rather than mere symptoms requiring medication (theme 2). This resonated with their motivation to seek alternatives to antipsychotics and reflects the broader tendency of preferring psychological therapies over pharmacological treatments [96]. It also echoes findings from TwV, where discouragement with previous treatment outcomes was highlighted as a key motivator for engagement [28]. Therapists' acknowledgment of voices may serve as an important first step in reducing stigma, counteracting perceived stigma from mental health providers that can "get under the skin" of service users and contribute to disempowerment [97]. The acknowledging approach was seen as trust-generating (theme 2) and mirrors the central normalizing tenet within the Hearing Voices Movement [16]. RT and TwV approach the voice(s) similarly, though typically without AT's and VRT's confrontational style [18]. Highlighting this commonality is important, as it supports the case for implementing a broader range of talk-based therapies for voices in routine practice, where dialogical engagement with voices remains relatively uncommon [98].

VR Affordances, Avatar Features, and Presence

VR was valued for creating a unique space to interact with voices (theme 1), echoing findings that VR can offer a meaningful environment for practicing therapeutic techniques, such as in agoraphobia treatment [66] or mindfulness practice in mood and anxiety disorders [99]. Similarly, VR-CBTp for paranoia supports direct engagement with paranoia-eliciting scenarios and affords greater exposure time than standard CBTp [100]. Trial participants emphasized the need for avatars to be realistic enough for believability, but not so lifelike as to threaten their sense of safety (theme 1). This aligns with research suggesting that cognitive awareness of VR's unreality can facilitate engagement [101,102] and help maintain a sense of safety, even when physical reactions occur [66]. Likewise, AT accounts show that avatar accuracy influences initial reactions, ranging from anxiety to disconnection, with discontinuation attributed to lack of realism [67]. While therapists identified (voice) presence as crucial for engagement and outcomes (theme 4), secondary Challenge RCT analyses showed that the sense of (spatial) presence was unrelated to outcome [44]. This underscores the need for clearer presence terminology and sensitivity in measurement [37], as well as further investigation into potential interaction effects, such as anxiety reduction [41]. Some trial participants felt they were meeting the voice itself in VR (theme 1), reflecting findings on presence and avatar realism in other studies [41,46,57]. Others struggled with (voice) presence or connecting the avatar to the voice (theme 4), mirroring findings from AT [67]. Given the bidirectional relationship between emotions and presence [38,39], and the role of (voice) presence in AT outcomes [41], VRT delivery must balance optimizing (voice) presence with avoiding excessive anxiety [40] to optimize outcomes.

Reproducing Voice Content

Therapists felt discomfort reproducing negative voice content but recognized its importance in avatar authenticity (theme 2), empowering trial participants to reclaim disempowering words [29]. Similarly, in a study on attitudes toward voice simulation training, therapists noted that while authentic simulations could cause distress, they were viewed as essential to preserve authenticity and avoid trivializing the experience [103]. Without proper hands-on training, this discomfort could lead to avoidance in routine care. The broader literature suggests that training in exposure techniques is associated with greater use of exposure therapy [104], and that therapist anxiety can cause clinicians to avoid or minimize treatments that elicit anxiety in both patients and themselves [105]. Difficulty and discomfort with reproducing negative voice content has been reported across AT [67], RT [24], and TwV [28]. In VRT and AT, this task falls primarily on the therapist, in TwV it rests with the hearer, and in RT, it may be shared between both through role-plays. This distinction is important, as responsibility for reproducing voice content may influence the therapy's acceptability for those tasked with carrying it out. Furthermore, TwV findings show that dialogical work can burden both participants, due to trauma's aftermath, and therapists, who bear witness to it [95], which highlights the need for extra care for both parties. Ensuring an open, nonjudgmental approach to all voice content and obtaining informed consent about what is used are therefore crucial considerations for future implementation.

Deeper Insight Into Voice-Hearing

Therapists appreciated the structured intervention (theme 2), but creative adaptations (theme 3) were often relevant to maintain a degree of individual flexibility. This underscores the importance of a flexible, formulation-driven approach aligned with individual needs [106-108] and preferred treatment outcomes [109]. Therapists also valued the deeper insight into voice-hearing offered by the intervention (themes 1 - 2), mirroring AT participants identifying how therapists developed a distinct depth of understanding via the avatar that could make them become an ally [67]. Given the distinction and divide between therapists' second-hand clinical knowledge and first-hand lived experience of voices [103], avatar dialogues may help bridge the gap, thereby deepening therapists' subjective understanding of voices. In this way, VRT could function both as a gateway for targeted intervention [34] and as a form of professional skills training, similar to voice simulation training [110].

Therapeutic Alliance

Trial participants and therapists found VRT rapidly fostered a strong therapeutic alliance (theme 2), echoing findings from AT [57,67]. A similar observation has been reported in TvW, where the therapists describe voice work as having a distinct intimate quality [95], and participants emphasize the importance

XSL•FO
RenderX

of the facilitator's personal qualities [27,28]. Some trial participants contrasted the trusting therapeutic relationship to past treatment experiences (theme 2), mirroring AT participants highlighting the importance of core therapeutic skills and person-centered flexibility [67]. As VR supported a process of joint exploration and collaboration (theme 1), it may enhance alliance-building by enabling shared engagement through a tangible, visible representation of the voice—what therapists described as a "common third" [93,94]. Given the therapeutic alliance's established links to engagement and symptom reduction [111], but mixed evidence regarding collaboration or formulation [112], further research is warranted into VR's potential to enhance alliance. Similarly, other research has shown that using smartphones to record voice-hearing experiences can accelerate rapport-building [113], suggesting that digital technologies can enhance therapeutic relationships by making internal experiences more tangible and communicable.

Comfort

Frustrating Malfunctions

Therapists found technology malfunctions to be highly disruptive (theme 3), occurring in 48% of Challenge RCT therapies [43]. While trial participants were largely unaffected, consistent with AT reports [29], therapists experienced emotional strain and increased workload. These issues may hinder implementation as they confirm clinicians' concerns that VR increases time and workload demands [69], compounding job stressors to the already demanding context of emotional labor [114,115].

Anxiety

As expected in exposure-based therapy, emotional reactions and symptom increases were common, particularly during early avatar encounters (theme 5). Trial participants reported a range of emotional responses (themes 1, 4, and 5), consistent with previous research [45,47]. More positive and less negative feelings were reported as therapy progressed, mirroring other findings [46]. Initial avatar encounters, however, could trigger anxiety (theme 5), likely contributing to the 29% of Challenge RCT therapies requiring exposure adjustments and most withdrawals occurring after the first VR session [43]. Withdrawal in the initial confronting phase has been observed across VRT [49] and AT [29], with emotional experiences becoming temporarily more difficult [67]. In TwV, hearers and voices emphasized the need for a "safe base" for dialogue work, as the process was anxiety-provoking and required trust and timing [27,28]. Anxiety tends to decrease across AT [41] and VRT [46,49], and 79% of Challenge RCT participants completed all sessions [43]. This indicates general tolerability, consistent with AT findings where emotional intensity of high voice presence was well tolerated [41,67]. VR may intensify expected anxiety (theme 5), highlighting the importance of finding a "sweet spot" where anxiety is neither too low nor overwhelming [66]. Establishing an appropriate starting point and allowing adequate time for preparation, monitoring, and calibration appears essential for optimizing outcomes [40].

Voice Reactions

Most trial participants reported negative voice reactions (theme 5). In Challenge RCT, voice worsening peaked at session three, with 37% reporting it [43], suggesting early but transient symptom exacerbation similar to trauma-focused treatments [52]. However, a trial participant attributed a hospitalization to increased voice severity (theme 5) but later resumed and completed therapy. Some participants linked worsening to general stress or the initiation of new treatments (theme 5), consistent with previous research [116,117]. Compared to a VR anxiety trial in which only 5% reported voice worsening [118], voice reactions may depend on the therapeutic application of VR. Notably, only a small AT subgroup experienced voice worsening [29], mirroring participants' own perspectives [57] with one noting increased aggression of voices only after completing therapy [67].

Complexity

Demanding Therapy

Some trial participants found VRT cognitively and emotionally demanding (theme 5), reflecting the intense emotional expressions often encountered in AT [45] and mirroring accounts from TwV highlighting the intensive nature of engaging with voices [28]. The emotional toll of avatar dialogues has even been presented as a reason for dropout in AT [67].

Nontraditional Skills

Therapists struggled to juggle technology, manual adherence, voice-content reproduction, and exposure monitoring (themes 2 - 3). They also needed to draw on complex, nontraditional skills to deliver avatar dialogues, consistent with previous descriptions of the demanding, "actor-like" approach [20,29,49]. While therapists reported gaining confidence and skills over time, particularly by building a repertoire of responses for avatar dialogues, they described VRT delivery as a steep learning curve (theme 2). Importantly, this was despite their high levels of qualification and psychiatric experience (see Table 1). Echoing research on the varied techniques used in avatar interactions [119], this underscores that therapeutic competency in VRT requires more than a "plug and play" approach. Supporting therapist confidence in working with voice-hearing—through supervision [120] and standardized materials [121]—should be central to future implementation efforts. Specifically, it seems warranted to incorporate structured role-plays and simulation-based training into both initial preparation and ongoing supervision, as also noted in TwV [95].

Appropriateness

Most trial participants and all therapists viewed VRT as a useful and relevant way of working with voices, indicating high appropriateness.

Usefulness

Positive Changes

Most trial participants, supported by the therapists' observations, reported gaining tools to manage voices or noted positive changes like increased control or new ways of relating to the

XSL•FO
RenderX

voice(s) (theme 4). This echoes therapeutic targets [29] and accounts of positive shifts in voices in AT [67], voice dialogue as a powerful enabler of positive change in TwV [27], and changes within the individual, the voice, or their communication observed in RT [23,24]. Therapists noted that subtle but significant changes—like new opportunities for dialogue with voices or shifts in their content or frequency—could be missed by standard assessments (theme 4). This highlights the need for more sensitive measures [14], as even minor changes can be life-changing, illustrated by a VRT case experiencing 15 minutes daily without voices [122] and the diverse ways resources for moving forward can manifest following TwV [28]. However, voices could remain distressing or unchanged (themes 4 - 5)—a point also reflected in TwV, where some changes were constrained by the lasting impact of past adversity [28]. In AT, most voices persisted, but their impact was substantially reduced for many [67], prompting debate about what constitutes a successful outcome [14]. Future research should explore whether digital approaches like VRT and AT produce similar or distinct outcomes [123,124], and whether they enhance sharing, to better assess the cost–benefit of investing in such technologies.

Individual Means and Targets

Most trial participants experienced changes in voice communication that extended somewhat into daily life (theme 4), mirroring shifts in communication between participants and avatars during therapy [96] but also difficulties translating changes to voices in AT [67]. Some trial participants continued practicing assertive communication with the voice(s), while others focused on self-understanding or making sense of the voice(s) (theme 4). This supports the exploration of AT in both brief and extended versions [30,58], tailoring the intervention to individual needs and therapeutic targets [29]. It also underscores the need to study which therapeutic approaches work best for different people at different times, since a poor fit can hinder engagement, as indicated in AT [67]. For example, whether exerting power over voices or fostering greater understanding and acceptance proves more helpful may inform decisions on which relational therapy is best suited to a given voice-hearer, as noted elsewhere [27].

Sharing Avatar Picture or Recording

Most trial participants valued sharing the avatar picture or audio recording to help others understand how they experience voice-hearing (theme 1). In Challenge RCT, 77% had used either material, 65% had shared the picture, and 41% had shared the audio recording [43]. Sharing could foster feelings of acknowledgment or normalization (theme 2), extending previous findings showing satisfaction with using AT recordings [57]. This is significant given the high prevalence and detrimental effects of stigma in this population, as well as the difficulties addressing it [125-127]. Sharing is not unique to VRT or AT as it also occurs and brings a sense of relief in nondigital therapies like RT [23,24]. However, digital tools may have an advantage, as experiences of hearing voices can be shared concretely through images or audio (theme 1) and, thus, help support living openly with voice-hearing without the burden of shame or stigma [16].

Relevance

Embodiment and Externalization

Most trial participants and all therapists considered VRT a relevant way to work with voices (themes 1 - 2). Particularly, dialoguing with the avatar as a proxy for the voice was valued, echoing the appreciation of direct voice dialogue in TwV [27]. Trial participants valued the avatar's embodiment of the voice, which offered a tangible and direct means of communication (theme 1). "Giving the voice a face" was seen as powerful by some, highlighting the value of externalizing an internal experience. This feature is central to AT [20], where focusing on the avatar's face has been shown to enhance assertiveness by allowing participants to direct communication toward something specific and externalized [67]. However, it contrasts with direct dialogue work in TwV and voice role-plays in RT. In RT, role-playing a voice can be challenging to some because it requires uniting voice and body [24], which is the opposite of the externalization in VRT or AT. However, limitations in avatar design (theme 3) sometimes made it harder for trial participants to relate to the avatar as they would to the voice, a challenge also noted in AT [67]. This finding aligns with previous research linking greater voice characterization to increased behavioral engagement [128] but contrasts with secondary Challenge RCT analyses indicating that auditory and visual resemblance of avatars shows only weak, non-significant trends toward reducing voice frequency [44].

Avatar Dialogues

Therapists found assertiveness and self-worth training highly relevant and were encouraged by participant progress, reflecting findings of changing avatar interactions [47] and increased positive emotions during interactions [46]. However, therapists noted that missing cornerstones (eg, suboptimal (voice) presence or life instability) could render the dialogical approach less relevant (theme 4), mirroring findings from AT that its suitability depends on the mental state and the stage of recovery [67]. Therapists observed that participants often showed remarkable resilience during avatar dialogues, which may go unnoticed in other settings but is likely supported by the context [29], and used this observed strength to empower participants in standing up to the voice(s). This aligns with the AT participants' perspectives, where determination and gradual improvements facilitated engagement [67]. Likewise, most trial participants found avatar dialogues useful for practicing assertiveness (theme 4), despite sometimes feeling confused, anxious, or awkward, similar to role-play experiences in RT [24]. Notably, some trial participants and therapists experienced that shifting from assertiveness to compassion toward the voice could be inappropriate given its negative impact (theme 3). Therapists were mindful of this tension and carefully weighed when to encourage compassion. As compassion-focused therapies for distressing voices are emerging [129,130], further research is needed on how best to navigate such ambivalence or tension that might arise from compassion blocks [131,132] or trauma-informed frameworks [133,134], e.g., when the voice represents a former perpetrator. Clarifying which therapeutic elements can be flexibly adapted while maintaining manual

XSL•FO
RenderX

fidelity is key to guiding implementation and tailoring therapy to individual preferences [135].

Feasibility

Therapist Barriers

Therapists emphasized that thorough training, feedback, and supervision were essential for feasible VRT delivery (themes 2 - 4), echoing research pointing to the multifaceted nature of AT requiring experienced therapists [29]. General flexibility and particularity in addressing voices directly, handling technological issues, and adapting to individual needs were also required. Therapists' motivation and effort supported feasibility (theme 2), but they stressed that participants needed to feel safe and ready to change, often requiring stability in other life areas. Addressing voice-related trauma was sometimes difficult within the current framework, highlighting the need for trauma-focused approaches [133,136] and the potential of the developmentally focused AV-EXT to more fully address trauma and broader therapeutic goals [30,32].

Participant Barriers

For some trial participants, therapy engagement and feasibility were impacted by difficulties with (voice) presence (themes 3 - 4), potentially compromising the outcome [34]. Likewise, emotional disconnection was reported by 2 AT participants who discontinued before the avatar dialogues, as the avatar did not match the voice adequately [67]. Furthermore, anxiety and (fear of) voice reactions could disrupt sessions and therapy progression, mirroring therapists' accounts of situations where the participant or voice(s) would not engage with the intervention in TwV [95]. To mitigate this, a sensitive, individualized approach is essential [29], with tailored exposure and gradual progression to avoid overwhelming responses [40].

Organizational Barriers

Limited organizational resources and external workload constrained therapists' time for training, preparation, and customization (theme 3). Therapists' extra effort (theme 2) may not be sustainable in routine care, underscoring the need for implementation strategies that address financial support and structural adjustments [137]. The 7-session VRT format is notably shorter than the 16-session minimum for CBTp recommended by guidelines [12], leaving little buffer for the frequent technological malfunctions that disrupted delivery (theme 3). While the usability of a similar digital psychotherapeutic tool may be rated highly under controlled conditions [138], the risk of malfunctions should be considered when evaluating such technologies for real-world use (theme 3). In addition, some trial participants were unable to rehearse avatar recordings at home due to lack of digital devices, highlighting digital exclusion in this population [139]. As there are potential benefits of digital interventions for self-management of symptoms [140], this digital exclusion highlights the research-to-practice gap in digital mental health [53,54].

Strengths and Limitations

The study's strengths include its comprehensive exploration of both VRT trial participants' and therapists' experiences. Contextualizing the findings within implementation outcomes enhances their relevance for future research, clinical trials, and real-world applications.

Several limitations should be noted. No interviews were conducted with dropouts, limiting insights into reasons for disengagement or discontinuation. However, it is reasonable to assume parallels with AT dropout accounts [67]. Relatedly, those who consented to participate might have had more positive experiences. Furthermore, trial participants' voices were not interviewed, though this might have offered valuable insights, as explored in TwV [27].

Involvement of people with lived experience was limited to designing the topic guide and validating themes. Greater involvement throughout the research process, including data collection and analysis, might have added important perspectives, challenged implicit assumptions, and enriched interpretations [141]. Co-produced qualitative research on AT [67] and digital interventions [65,66] or more integrated involvement in clinical trials [64] has shown promising potential.

The time between treatment cessation and interviews may have introduced recall bias and memory deficits, potentially leading to biased or inaccurate reporting. However, the sampling strategy prevented the selection of only the most recent completers.

Future Directions

Advancing VRT requires careful planning for sustainable, potentially commercial delivery models. Ensuring the operational reliability of hardware and software, therapist usability, and user tolerability will be essential, alongside regulatory compliance for integration into routine care. Dedicated implementation studies are needed.

Key VRT challenges include calibrating exposure to balance long-term benefits with the risk of flooding or dropout, with real-time physiological feedback being explored in a pilot trial led by author LBG. Optimizing (voice) presence, or developing strategies to manage suboptimal (voice) presence, remains central for balancing therapeutic impact and safety. Interviewing discontinuers could inform factors influencing engagement, as explored in AT [67].

Careful investment in therapist training, supervision, and standardized materials seems critical for sustaining fidelity and confidence in VRT delivery. However, as the level of experience held by therapists in the Challenge RCT may not be available in routine practice, it is important to examine whether VRT can be effectively delivered by less specialized clinicians, given appropriate training and supervision.

Future research should clarify how VRT compares to other therapies, including its effects on alliance-building and the facilitation of externalization. VRT's potential for adolescents should be investigated among voice-hearers, with a study currently underway by author DLV. Also, research should clarify what constitutes meaningful outcomes of VRT and establish greater sensitivity in outcome measurement. A forthcoming review will map methodologies for assessing voice content [142].

XSL•FO
RenderX

Investigating therapy experiences across diverse, underrepresented groups is essential. The AVATAR2 qualitative research program includes studies on therapy dropouts and Black individuals' experiences of AT [143].

Conclusion

This qualitative study provides comprehensive insights into trial participants' and therapists' experience of VRT in the Challenge trial. Several positive aspects of VRT were highlighted, including its validating approach to the voice(s), facilitation of engagement with the voice(s), and the opportunity to share the otherwise private experience of voice-hearing. Embodying the voice(s) in immersive VR-supported avatar role-plays was seen as a key affordance. The collaborative therapist-participant relationship was highly acceptable. Reported positive outcomes included improved voice dialogue, enhanced control, and increased well-being, alongside nuanced changes and experiences not easily captured by quantitative measurements.

VRT was generally perceived as acceptable, appropriate, and feasible. However, several challenges were identified. Content-related issues included instances of suboptimal sense of (voice) presence, limitations to avatar design, and difficulties with reproducing negative voice content. Comfort was affected by technological malfunctions, participant anxiety, exhaustion, and adverse voice reactions. Complexity arose from the demanding nature of the therapy and the nontraditional skills required of therapists. Usefulness could be concealed by measurement insensitivity. Relevance was challenged by tensions between assertiveness and compassion-focused approaches. Furthermore, feasibility seemed constrained by the need for therapists to exert "extra effort" (eg, troubleshooting technology and working beyond scheduled hours), which may be unsustainable in routine care without substantial organizational support (eg, protected time and dedicated technical assistance). Future implementation will require adequate training, supervision, and organizational support.

In conclusion, trial participants' and therapists' experiences support VRT as a promising - but challenging - intervention for distressing voices.

Acknowledgments

The authors thank all participants and stakeholders in the Challenge trial, with special appreciation to those who participated in the interviews and the theme validation workshop. The authors used ChatGPT to reduce word count, remove redundancy, and improve readability. All content was reviewed and edited by the authors, who take full responsibility for the final manuscript.

Funding

This study was supported by the same grants as the Challenge trial: Innovation Fund Denmark (8088-00023B), Independent Research Fund Denmark (2096-00012B), Innovation Fund North Denmark Region (2022-0010), Psychiatry Research Fund North Denmark Region (1-45-72-3778-24), and The ML Jørgensen and Gunnar Hansen Fund (2022-0019). Funders had no role in study design, data collection, analysis, interpretation, or manuscript preparation.

Data Availability

The datasets generated and analyzed during this study are not publicly available due to the protection of participant anonymity and privacy, but the corresponding author may be contacted to discuss access upon reasonable request.

Authors' Contributions

Conceptualization: MJC, RJ, DLV
Methodology: MJC, RJ, DLV
Investigation: MJC, MPR, CDN
Formal analysis: MJC, MPR, CDN
Supervision: RJ
Writing – original draft: MJC, MPR (supporting)
Writing – review & editing: MPR, RJ, DLV, JM, IHB, NT, LCS, LSM, TW, MN, LBG, CDN
Validation: MJC, MPR, RJ, CDN, JM, IHB, NT, LCS, LSM, TW, MN, LBG, DLV
Approval of final manuscript: MJC, MPR, RJ, CDN, JM, IHB, NT, LCS, LSM, TW, MN, LBG, DLV

Conflicts of Interest

MBG and DLV are involved in other funded studies on virtual reality (VR)-therapy and have received funding for teaching Challenge-VRT by the Danish VR-firm Khora. None declared by the other authors.

Multimedia Appendix 1
Interview guide for trial participants.
[PDF File, 379 KB - games_v13i1e77920_app1.pdf]

XSL•FO
RenderX

Multimedia Appendix 2
Interview guide for therapists.
[PDF File, 319 KB - games_v13i1e77920_app2.pdf]

Multimedia Appendix 3
Theme validation workshop.
[PDF File, 319 KB - games_v13i1e77920_app3.pdf]

References

1. Baumeister D, Sedgwick O, Howes O, Peters E. Auditory verbal hallucinations and continuum models of psychosis: a systematic review of the healthy voice-hearer literature. Clin Psychol Rev 2017 Feb;51:125-141. [doi: 10.1016/j.cpr.2016.10.010] [Medline: 27866082]
2. Kråkvik B, Larøi F, Kalhovde AM, et al. Prevalence of auditory verbal hallucinations in a general population: a group comparison study. Scand J Psychol 2015 Oct;56(5):508-515. [doi: 10.1111/sjop.12236] [Medline: 26079977]
3. Waters F, Allen P, Aleman A, et al. Auditory hallucinations in schizophrenia and nonschizophrenia populations: a review and integrated model of cognitive mechanisms. Schizophr Bull 2012 Jun;38(4):683-693. [doi: 10.1093/schbul/sbs045] [Medline: 22446568]
4. Köhler-Forsberg O, Madsen T, Behrendt-Møller I, Nordentoft M. The 10-year trajectories of auditory hallucinations among 496 patients with a first schizophrenia-spectrum disorder: Findings from the OPUS cohort. Schizophr Res 2022 May;243:385-391. [doi: 10.1016/j.schres.2021.06.033] [Medline: 34272121]
5. Johns LC, Kompus K, Connell M, et al. Auditory verbal hallucinations in persons with and without a need for care. Schizophr Bull 2014 Jul;40 Suppl 4(Suppl 4):S255-S264. [doi: 10.1093/schbul/sbu005] [Medline: 24936085]
6. Tsang A, Bucci S, Branitsky A, et al. The relationship between appraisals of voices (auditory verbal hallucinations) and distress in voice-hearers with schizophrenia-spectrum diagnoses: A meta-analytic review. Schizophr Res 2021 Apr;230:38-47. [doi: 10.1016/j.schres.2021.02.013]
7. Larøi F, Thomas N, Aleman A, et al. The ice in voices: Understanding negative content in auditory-verbal hallucinations. Clin Psychol Rev 2019 Feb;67:1-10. [doi: 10.1016/j.cpr.2018.11.001] [Medline: 30553563]
8. Larøi F, Sommer IE, Blom JD, et al. The characteristic features of auditory verbal hallucinations in clinical and nonclinical groups: state-of-the-art overview and future directions. Schizophr Bull 2012 Jun;38(4):724-733. [doi: 10.1093/schbul/sbs061] [Medline: 22499783]
9. Leucht S, Leucht C, Huhn M, et al. Sixty years of placebo-controlled antipsychotic drug trials in acute schizophrenia: systematic review, bayesian meta-analysis, and meta-regression of efficacy predictors. AJP 2017 Oct 1;174(10):927-942. [doi: 10.1176/appi.ajp.2017.16121358]
10. Gotfredsen DR, Wils RS, Hjorthøj C, et al. Stability and development of psychotic symptoms and the use of antipsychotic medication - long-term follow-up. Psychol Med 2017 Sep;47(12):2118-2129. [doi: 10.1017/S0033291717000563] [Medline: 28382874]
11. Rognoni C, Bertolani A, Jommi C. Second-generation antipsychotic drugs for patients with schizophrenia: systematic literature review and meta-analysis of metabolic and cardiovascular side effects. Clin Drug Investig 2021 Apr;41(4):303-319. [doi: 10.1007/s40261-021-01000-1] [Medline: 33686614]
12. NICE. Psychosis and schizophrenia in adults: prevention and management: NICE; 2014.
13. Berendsen S, Berendse S, van der Torren J, Vermeulen J, de Haan L. Cognitive behavioural therapy for the treatment of schizophrenia spectrum disorders: an umbrella review of meta-analyses of randomised controlled trials. EClinicalMedicine 2024 Jan;67:102392. [doi: 10.1016/j.eclinm.2023.102392] [Medline: 38274116]
14. Thomas N. What's really wrong with cognitive behavioral therapy for psychosis? Front Psychol 2015;6:323. [doi: 10.3389/fpsyg.2015.00323] [Medline: 25870572]
15. Thomas N, Hayward M, Peters E, et al. Psychological therapies for auditory hallucinations (voices): current status and key directions for future research. Schizophr Bull 2014 Jul;40 Suppl 4(Suppl 4):S202-S212. [doi: 10.1093/schbul/sbu037] [Medline: 24936081]
16. Corstens D, Longden E, McCarthy-Jones S, Waddingham R, Thomas N. Emerging perspectives from the hearing voices movement: implications for research and practice. Schizophr Bull 2014 Jul;40 Suppl 4(Suppl 4):S285-S294. [doi: 10.1093/schbul/sbu007] [Medline: 24936088]
17. Dellazizzo L, Giguère S, Léveillé N, Potvin S, Dumais A. A systematic review of relational-based therapies for the treatment of auditory hallucinations in patients with psychotic disorders. Psychol Med 2022 Aug;52(11):2001-2008. [doi: 10.1017/S003329172200143X] [Medline: 35855651]
18. Thomas N, Ward T, Longden E, et al. Relational therapies for people who hear voices: operationalisation and current status of an emergent group of psychological therapies. Oxford University Press [Forthcoming] 2025. [doi: 10.1093/schbul/sbaf142]
19. Leff J, Williams G, Huckvale M, Arbuthnot M, Leff AP. Avatar therapy for persecutory auditory hallucinations: What is it and how does it work? Psychosis 2014 Jun;6(2):166-176. [doi: 10.1080/17522439.2013.773457] [Medline: 24999369]

XSL•FO
RenderX

20. Craig T, Ward T, Rus-Calafell M. AVATAR therapy for refractory auditory hallucinations. In: Brief Interventions for Psychosis: A Clinical Compendium: London: Springer; 2016. [Medline: 27386617]
21. Hayward M, Jones AM, Bogen-Johnston L, Thomas N, Strauss C. Relating therapy for distressing auditory hallucinations: a pilot randomized controlled trial. Schizophr Res 2017 May;183(137–142):137-142. [doi: 10.1016/j.schres.2016.11.019]
22. Lincoln TM, Schlier B, Müller R, et al. Reducing distress from auditory verbal hallucinations: a multicenter, parallel, single-blind, randomized controlled feasibility trial of relating therapy. Psychother Psychosom 2024;93(5):328-339. [doi: 10.1159/000539809] [Medline: 39168112]
23. Deamer F, Hayward M. Relating to the speaker behind the voice:what Is changing? Front Psychol 2018;9:11. [doi: 10.3389/fpsyg.2018.00011] [Medline: 29422879]
24. Hayward M, Bogen-Johnston L, Deamer F. Relating Therapy for distressing voices: Who, or what, is changing? Psychosis 2018 Apr 3;10(2):132-141. [doi: 10.1080/17522439.2018.1469037]
25. Hayward M, Pilny M, Lincoln T. Relating Therapy for distressing voices: a treatment protocol. Psychol Psychother 2025 May 6. [doi: 10.1111/papt.12595] [Medline: 40329660]
26. Longden E, Corstens D, Bowe S, et al. A psychological intervention for engaging dialogically with auditory hallucinations (Talking With Voices): A single-site, randomised controlled feasibility trial. Schizophr Res 2022 Dec;250:172-179. [doi: 10.1016/j.schres.2022.11.007] [Medline: 36423442]
27. Middleton K, Cooke A, May R. “It allowed us to let our pain out”: perspectives from voice-hearers and their voices on the ‘talking with voices’ approach. Psychosis 2024 Apr 2;16(2):131-143. [doi: 10.1080/17522439.2022.2141840]
28. Longden E, Branitsky A, Jones W, Peters S. When therapists talk to voices: perspectives from service-users who experience auditory hallucinations. Psychol Psychother 2023 Dec;96(4):967-981. [doi: 10.1111/papt.12489] [Medline: 37551953]
29. Ward T, Rus-Calafell M, Ramadhan Z, et al. AVATAR therapy for distressing voices: a comprehensive account of therapeutic targets. Schizophr Bull 2020 Sep 21;46(5):1038-1044. [doi: 10.1093/schbul/sbaa061] [Medline: 32372082]
30. Garety PA, Edwards CJ, Jafari H, et al. Digital AVATAR therapy for distressing voices in psychosis: the phase 2/3 AVATAR2 trial. Nat Med 2024 Dec;30(12):3658-3668. [doi: 10.1038/s41591-024-03252-8] [Medline: 39468363]
31. Longden E, Corstens D, Escher S, Romme M. Voice hearing in a biographical context: a model for formulating the relationship between voices and life history. Psychosis 2012 Oct;4(3):224-234. [doi: 10.1080/17522439.2011.596566]
32. Garety P, Edwards CJ, Ward T, et al. Optimising AVATAR therapy for people who hear distressing voices: study protocol for the AVATAR2 multi-centre randomised controlled trial. Trials 2021 May 25;22(1):366. [doi: 10.1186/s13063-021-05301-w] [Medline: 34034792]
33. Digital health technologies to help manage symptoms of psychosis and prevent relapse in adults and young people: early value assessment health technology evaluation. NICE. 2024. URL: www.nice.org.uk/guidance/hte17 [accessed 2025-11-13]
34. Bell IH, Pot-Kolder R, Rizzo A, et al. Advances in the use of virtual reality to treat mental health conditions. Nat Rev Psychol 2024;3(8):552-567. [doi: 10.1038/s44159-024-00334-9]
35. Lan L, Sikov J, Lejeune J, et al. A systematic review of using virtual and augmented reality for the diagnosis and treatment of psychotic disorders. Curr Treat Options Psych 2023;10(2):87-107. [doi: 10.1007/s40501-023-00287-5]
36. Diemer J, Alpers GW, Peperkorn HM, Shiban Y, Mühlberger A. The impact of perception and presence on emotional reactions: a review of research in virtual reality. Front Psychol 2015;6:26. [doi: 10.3389/fpsyg.2015.00026] [Medline: 25688218]
37. Grassini S, Laumann K. Questionnaire measures and physiological correlates of presence: a systematic review. Front Psychol 2020;11:349. [doi: 10.3389/fpsyg.2020.00349] [Medline: 32265769]
38. Riva G, Mantovani F, Capideville CS, et al. Affective interactions using virtual reality: the link between presence and emotions. Cyberpsychol Behav 2007 Feb;10(1):45-56. [doi: 10.1089/cpb.2006.9993] [Medline: 17305448]
39. Riva G, Mantovani F. Being there: understanding the feeling of presence in a synthetic environment and its potential for clinical change. In: Eichenberg C, editor. Virtual Reality in Psychological, Medical and Pedagogical Applications: IntechOpen; 2012:3-34. [doi: 10.5772/46411]
40. Ling Y, Nefs HT, Morina N, Heynderickx I, Brinkman WP. A meta-analysis on the relationship between self-reported presence and anxiety in virtual reality exposure therapy for anxiety disorders. PLoS ONE 2014;9(5):e96144. [doi: 10.1371/journal.pone.0096144] [Medline: 24801324]
41. Rus-Calafell M, Ward T, Zhang XC, Edwards CJ, Garety P, Craig T. The role of sense of voice presence and anxiety reduction in AVATAR therapy. J Clin Med 2020 Aug 25;9(9):2748. [doi: 10.3390/jcm9092748] [Medline: 32854387]
42. Smith LC, Mateos AC, Due AS, et al. Immersive virtual reality in the treatment of auditory hallucinations: A PRISMA scoping review. Psychiatry Res 2024 Apr;334:115834. [doi: 10.1016/j.psychres.2024.115834] [Medline: 38452499]
43. Smith LC, Vernal DL, Mariegaard LS, et al. Immersive virtual reality-assisted therapy targeting persistent auditory verbal hallucinations in patients diagnosed with schizophrenia spectrum disorders in Denmark: the Challenge assessor-masked, randomised clinical trial. Lancet Psychiatry 2025 Aug;12(8):557-567. [doi: 10.1016/S2215-0366(25)00161-0] [Medline: 40617245]
44. Glenthøj LB, Mariegaard LS, Smith LC, et al. Impact of avatar features and presence on treatment outcomes in virtual reality-assisted therapy for auditory hallucinations. Schizophr Res 2025 Nov 13;287:1-2. [doi: 10.1016/j.schres.2025.11.004] [Medline: 41237437]

XSL•FO
RenderX

45. Hudon A, Lammatteo V, Rodrigues-Coutlée S, et al. Exploration of the role of emotional expression of treatment-resistant schizophrenia patients having followed virtual reality therapy: a content analysis. BMC Psychiatry 2023 Jun 12;23(1):420. [doi: 10.1186/s12888-023-04861-2] [Medline: 37308864]
46. Augustin E, Beaudoin M, Giguère S, Ziady H, Phraxayavong K, Dumais A. The relationship between sense of presence, emotional response, and clinical outcomes in virtual reality-based therapy for treatment-resistant schizophrenia: an exploratory correlational study. J Pers Med 2024 Jun 8;14(6):614. [doi: 10.3390/jpm14060614] [Medline: 38929835]
47. Beaudoin M, Potvin S, Machalani A, et al. The therapeutic processes of avatar therapy: a content analysis of the dialogue between treatment-resistant patients with schizophrenia and their avatar. Clin Psychol Psychother 2021 May;28(3):500-518. [doi: 10.1002/cpp.2556] [Medline: 33484042]
48. Dellazizzo L, Potvin S, Phraxayavong K, Dumais A. One-year randomized trial comparing virtual reality-assisted therapy to cognitive-behavioral therapy for patients with treatment-resistant schizophrenia. NPJ Schizophr 2021 Feb 12;7(1):9. [doi: 10.1038/s41537-021-00139-2] [Medline: 33580033]
49. Percie du Sert O, Potvin S, Lipp O, et al. Virtual reality therapy for refractory auditory verbal hallucinations in schizophrenia: A pilot clinical trial. Schizophr Res 2018 Jul;197:176-181. [doi: 10.1016/j.schres.2018.02.031] [Medline: 29486956]
50. Linden M, Schermuly-Haupt ML. Definition, assessment and rate of psychotherapy side effects. World Psychiatry 2014 Oct;13(3):306-309. [doi: 10.1002/wps.20153] [Medline: 25273304]
51. Hoppen TH, Lindemann AS, Morina N. Safety of psychological interventions for adult post-traumatic stress disorder: meta-analysis on the incidence and relative risk of deterioration, adverse events and serious adverse events. Br J Psychiatry 2022 Aug 12:1-10. [doi: 10.1192/bjp.2022.111] [Medline: 35959698]
52. Burger SR, Hardy A, van der Linden T, et al. The bumpy road of trauma-focused treatment: posttraumatic stress disorder symptom exacerbation in people with psychosis. J Trauma Stress 2023 Apr;36(2):299-309. [doi: 10.1002/jts.22907] [Medline: 36719408]
53. Mohr DC, Weingardt KR, Reddy M, Schueller SM. Three problems with current digital mental health research... and three things we can do about them. PS (Wash DC) 2017 May 1;68(5):427-429. [doi: 10.1176/appi.ps.201600541]
54. Graham AK, Lattie EG, Powell BJ, et al. Implementation strategies for digital mental health interventions in health care settings. Am Psychol 2020 Nov;75(8):1080-1092. [doi: 10.1037/amp0000686] [Medline: 33252946]
55. Grimshaw JM, Eccles MP, Lavis JN, Hill SJ, Squires JE. Knowledge translation of research findings. Implement Sci 2012 May 31;7(1):50. [doi: 10.1186/1748-5908-7-50] [Medline: 22651257]
56. Proctor E, Silmere H, Raghavan R, et al. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. Adm Policy Ment Health 2011 Mar;38(2):65-76. [doi: 10.1007/s10488-010-0319-7] [Medline: 20957426]
57. Rus-Calafell M, Ehrbar N, Ward T, et al. Participants' experiences of AVATAR therapy for distressing voices: a thematic qualitative evaluation. BMC Psychiatry 2022 May 24;22(1):356. [doi: 10.1186/s12888-022-04010-1] [Medline: 35610590]
58. Craig TK, Rus-Calafell M, Ward T, et al. AVATAR therapy for auditory verbal hallucinations in people with psychosis: a single-blind, randomised controlled trial. Lancet Psychiatry 2018 Jan;5(1):31-40. [doi: 10.1016/S2215-0366(17)30427-3] [Medline: 29175276]
59. Bauer MS, Kirchner J. Implementation science: What is it and why should I care? Psychiatry Res 2020 Jan;283:112376. [doi: 10.1016/j.psychres.2019.04.025] [Medline: 31036287]
60. O'Cathain A, Thomas K, Drabble S, Rudolph A, Hewison J. What can qualitative research do for randomised controlled trials? A systematic mapping review. Trials 2013 Nov;14(S1). [doi: 10.1186/1745-6215-14-S1-O52]
61. Berry N, Lobban F, Emsley R, Bucci S. Acceptability of interventions delivered online and through mobile phones for people who experience severe mental health problems: a systematic review. J Med Internet Res 2016 May 31;18(5):e121. [doi: 10.2196/jmir.5250] [Medline: 27245693]
62. Berry N, Lobban F, Bucci S. A qualitative exploration of service user views about using digital health interventions for self-management in severe mental health problems. BMC Psychiatry 2019 Jan 21;19(1):35. [doi: 10.1186/s12888-018-1979-1] [Medline: 30665384]
63. Bucci S, Lewis S, Ainsworth J, et al. Digital interventions in severe mental health problems: lessons from the Actissist development and trial. World Psychiatry 2018 Jun;17(2):230-231. [doi: 10.1002/wps.20535] [Medline: 29856569]
64. Owrid O, Richardson L, Allan S, et al. "There's no us vs. them, it's just us": a creative approach to centring lived experience within the AVATAR2 trial. BMC Psychiatry 2024 Nov 15;24(1):807. [doi: 10.1186/s12888-024-06268-z] [Medline: 39543539]
65. Greenwood KE, Gurnani M, Ward T, et al. The service user experience of SlowMo therapy: a co-produced thematic analysis of service users' subjective experience. Psychol Psychother 2022 Sep;95(3):680-700. [doi: 10.1111/papt.12393] [Medline: 35445520]
66. Bond J, Kenny A, Pinfold V, et al. A safe place to learn: peer research qualitative investigation of gameChange virtual reality therapy. JMIR Serious Games 2023 Jan 16;11:e38065. [doi: 10.2196/38065] [Medline: 36645707]
67. Rutter-Eley E, Jamieson-Craig T, Garety P, et al. 'It felt good to be able to say that out loud': a peer research qualitative study exploring voice content work and the therapeutic alliance in AVATAR therapy: OSF Preprint posted online on May 13, 2025 URL: https://osf.io/preprints/osf/jk68v_v1 [accessed 2025-11-14]

XSL•FO
RenderX

68. McHugh RK, Barlow DH. The dissemination and implementation of evidence-based psychological treatments. A review of current efforts. Am Psychol 2010;65(2):73-84. [doi: 10.1037/a0018121] [Medline: 20141263]
69. Bin S, Alrashdi DH, Whitehead T, Riches S, Drini E. Mental health professionals' attitudes towards virtual reality therapies: a systematic review. J technol behav sci 2025. [doi: 10.1007/s41347-025-00499-x]
70. Smith LC, Mariegaard L, Vernal DL, et al. The CHALLENGE trial: the effects of a virtual reality-assisted exposure therapy for persistent auditory hallucinations versus supportive counselling in people with psychosis: study protocol for a randomised clinical trial. Trials 2022 Sep 13;23(1):773. [doi: 10.1186/s13063-022-06683-1] [Medline: 36100943]
71. Christensen MJ, Olesen AV, Ward T, et al. Psychological and behavioral mechanisms linking childhood trauma profiles to voice-related distress in schizophrenia spectrum disorders: preregistration for latent class, mediation, and conditional process analyses in the Challenge trial. OSF. 2025 Aug 11. URL: https://doi.org/10.17605/OSF.IO/S6DTF [accessed 2025-11-18]
72. Dewey J. What does pragmatism mean by practical? J Philos Psychol Sci Methods 1908 Feb 13;5(4):85. [doi: 10.2307/2011894]
73. Ramanadhan S, Revette AC, Lee RM, Aveling EL. Pragmatic approaches to analyzing qualitative data for implementation science: an introduction. Implement Sci Commun 2021 Jun 29;2(1):70. [doi: 10.1186/s43058-021-00174-1] [Medline: 34187595]
74. Allemang B, Sitter K, Dimitropoulos G. Pragmatism as a paradigm for patient-oriented research. Health Expect 2022 Feb;25:38-47. [doi: 10.1111/hex.13384] [Medline: 34748689]
75. Kvale S, Brinkmann S. Interview: Det Kvalitative Forskningsinterview Som Håndværk [Book in Danish]: Hans Reitzels Forlag; 2015.
76. Petersen L, Jeppesen P, Thorup A, et al. A randomised multicentre trial of integrated versus standard treatment for patients with a first episode of psychotic illness. BMJ 2005 Sep 17;331(7517):602. [doi: 10.1136/bmj.38565.415000.E01]
77. Nielsen CM, Hjorthøj C, Killaspy H, Nordentoft M. The effect of flexible assertive community treatment in Denmark: a quasi-experimental controlled study. Lancet Psychiatry 2021 Jan;8(1):27-35. [doi: 10.1016/S2215-0366(20)30424-7]
78. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: a synthesis of recommendations. Acad Med 2014 Sep;89(9):1245-1251. [doi: 10.1097/ACM.0000000000000388] [Medline: 24979285]
79. Palinkas LA, Horwitz SM, Green CA, Wisdom JP, Duan N, Hoagwood K. Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. Adm Policy Ment Health 2015 Sep;42(5):533-544. [doi: 10.1007/s10488-013-0528-y] [Medline: 24193818]
80. Malterud K, Siersma VD, Guassora AD. Sample size in qualitative interview studies. Qual Health Res 2016 Nov;26(13):1753-1760. [doi: 10.1177/1049732315617444]
81. Wutich A, Beresford M, Bernard HR. Sample sizes for 10 types of qualitative data analysis: an integrative review, empirical guidance, and next steps. Int J Qual Methods 2024 Jan;23. [doi: 10.1177/16094069241296206]
82. Naeem M, Ozuem W, Howell K, Ranfagni S. Demystification and actualisation of data saturation in qualitative research through thematic analysis. Int J Qual Methods 2024 Jan;23. [doi: 10.1177/16094069241229777]
83. Braun V, Clarke V. To saturate or not to saturate? Questioning data saturation as a useful concept for thematic analysis and sample-size rationales. Qual Res Sport Exerc Health 2021 Mar 4;13(2):201-216. [doi: 10.1080/2159676X.2019.1704846]
84. Hennink M, Kaiser BN. Sample sizes for saturation in qualitative research: a systematic review of empirical tests. Soc Sci Med 2022 Jan;292:114523. [doi: 10.1016/j.socscimed.2021.114523] [Medline: 34785096]
85. Good Tape — Fast, secure and accurate transcription. URL: https://goodtape.io/ [accessed 2025-04-07]
86. Braun V, Clarke V. Conceptual and design thinking for thematic analysis. Qualitative Psychology 2022;9(1):3-26. [doi: 10.1037/qup0000196]
87. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
88. Clarke V, Braun V. Thematic analysis. J Posit Psychol 2017 May 4;12(3):297-298. [doi: 10.1080/17439760.2016.1262613]
89. Braun V, Clarke V. Toward good practice in thematic analysis: Avoiding common problems and be(com)ing a knowing researcher. Int J Transgend 2023 Jan 25;24(1):1-6. [doi: 10.1080/26895269.2022.2129597]
90. Braun V, Clarke V. Reflecting on reflexive thematic analysis. Qualitative Research in Sport, Exercise and Health 2019 Aug 8;11(4):589-597. [doi: 10.1080/2159676X.2019.1628806]
91. Braun V, Clarke V. Successful Qualitative Research: A Practical Guide for Beginners: Sage Publications Ltd; 2013.
92. Braun V, Clarke V. One size fits all? What counts as quality practice in (reflexive) thematic analysis? Qual Res Psychol 2021 Jul 3;18(3):328-352. [doi: 10.1080/14780887.2020.1769238]
93. Hatton K. Social Pedagogy in the UK: Theory and Practice: Russell House Publishing Ltd; 2013.
94. Ryynänen S, Nivala E. Inequality as a social pedagogical question. International Journal of Social Pedagogy 2019;7(1). [doi: 10.14324/111.444.ijsp.2019.v7.1.008]
95. Longden E, Branitsky A, Jones W, Peters S. "It's like having a core belief that's able to speak back to you": Therapist accounts of dialoguing with auditory hallucinations. Psychol Psychother 2022 Mar;95(1):295-312. [doi: 10.1111/papt.12373] [Medline: 34762756]

XSL•FO
RenderX

96. Morrison AP. Should people with psychosis be supported in choosing cognitive therapy as an alternative to antipsychotic medication: A commentary on current evidence. Schizophr Res 2019 Jan;203:94-98. [doi: 10.1016/j.schres.2018.03.010]
97. Wang K, Link BG, Corrigan PW, Davidson L, Flanagan E. Perceived provider stigma as a predictor of mental health service users' internalized stigma and disempowerment. Psychiatry Res 2018 Jan;259:526-531. [doi: 10.1016/j.psychres.2017.11.036]
98. Burr C, Schnackenberg JK, Weidner F. Talk-based approaches to support people who are distressed by their experience of hearing voices: A scoping review. Front Psychiatry 2022;13:983999. [doi: 10.3389/fpsyt.2022.983999] [Medline: 36299547]
99. Blackmore R, Giles C, Tremain H, et al. Examining the use of virtual reality to support mindfulness skills practice in mood and anxiety disorders: mixed methods study. J Med Internet Res 2024 Dec 6;26:e45640. [doi: 10.2196/45640] [Medline: 39641990]
100. Jeppesen UN, Vernal DL, Due AS, et al. Virtual reality-based versus standard cognitive behavioral therapy for paranoia in schizophrenia spectrum disorders: a randomized controlled trial. Nat Med 2025 Aug 13. [doi: 10.1038/s41591-025-03880-8] [Medline: 40804323]
101. Slater M, Gonzalez-Liencres C, Haggard P, et al. The ethics of realism in virtual and augmented reality. Front virtual real 2020 Mar 3;1. [doi: 10.3389/frvir.2020.00001]
102. Holgersen G, Nordgreen T, ten Velden Hegelstad W, Bircow Elgen I. Views of young people with psychosis on using virtual reality assisted therapy. A qualitative study. Early Intervention Psych 2023 Apr;17(4):361-367. [doi: 10.1111/eip.13331]
103. Hignett R, Curran R, Fisher HL, Riches S. Psychologists' experience of working with people who hear voices and their views on the idea of voice simulation training: A qualitative study. Psychol Psychother 2024 Sep;97(3):562-581. [doi: 10.1111/papt.12540] [Medline: 38970412]
104. Langthorne D, Beard J, Waller G. Therapist factors associated with intent to use exposure therapy: a systematic review and meta-analysis. Cogn Behav Ther 2023 Jul;52(4):347-379. [doi: 10.1080/16506073.2023.2191824] [Medline: 36974519]
105. Speers AJH, Bhullar N, Cosh S, Wootton BM. Correlates of therapist drift in psychological practice: A systematic review of therapist characteristics. Clin Psychol Rev 2022 Apr;93:102132. [doi: 10.1016/j.cpr.2022.102132] [Medline: 35316672]
106. Pain CM, Chadwick P, Abba N. Clients' experience of case formulation in cognitive behaviour therapy for psychosis. British J Clinic Psychol 2008 Jun;47(2):127-138. [doi: 10.1348/014466507X235962]
107. Easden MH, Kazantzis N. Case conceptualization research in cognitive behavior therapy: A state of the science review. J Clin Psychol 2018 Mar;74(3):356-384. [doi: 10.1002/jclp.22516] [Medline: 28776663]
108. Rainforth M, Laurenson M. A literature review of case formulation to inform mental health practice. J Psychiatr Ment Health Nurs 2014 Apr;21(3):206-213. [doi: 10.1111/jpm.12069] [Medline: 23551415]
109. Longden E, Branitsky A, Sheaves B, Chauhan N, Morrison AP. Preferred treatment outcomes in psychological therapy for voices: a comparison of staff and service-user perspectives. Psychosis 2024 Apr 2;16(2):107-117. [doi: 10.1080/17522439.2023.2215298]
110. Bradshaw T, Blakemore A, Wilson I, Fitzsimmons M, Crawford K, Mairs H. A systematic review of the outcomes of using voice hearing simulation in the education of health care professionals and those in training. Nurse Educ Today 2021 Jan;96:104626. [doi: 10.1016/j.nedt.2020.104626] [Medline: 33157364]
111. Bourke E, Barker C, Fornells-Ambrojo M. Systematic review and meta-analysis of therapeutic alliance, engagement, and outcome in psychological therapies for psychosis. Psychol Psychother 2021 Sep;94(3):822-853. [doi: 10.1111/papt.12330] [Medline: 33569885]
112. Shattock L, Berry K, Degnan A, Edge D. Therapeutic alliance in psychological therapy for people with schizophrenia and related psychoses: A systematic review. Clin Psychology and Psychoth 2018 Jan;25(1). [doi: 10.1002/cpp.2135]
113. Moore E, Williams A, Bell I, Thomas N. Client experiences of blending a coping-focused therapy for auditory verbal hallucinations with smartphone-based ecological momentary assessment and intervention. Internet Interv 2020 Mar;19:100299. [doi: 10.1016/j.invent.2019.100299] [Medline: 31890641]
114. Jeung DY, Kim C, Chang SJ. Emotional labor and burnout: a review of the literature. Yonsei Med J 2018 Mar;59(2):187-193. [doi: 10.3349/ymj.2018.59.2.187] [Medline: 29436185]
115. Andela M, Truchot D, Van der Doef M. Job stressors and burnout in hospitals: the mediating role of emotional dissonance. Int J Stress Manag 2016 Aug 1;23(3):298-317. [doi: 10.1037/str0000013]
116. Georgiades A, Almuqrin A, Rubinic P, Mouhitzadeh K, Tognin S, Mechelli A. Psychosocial stress, interpersonal sensitivity, and social withdrawal in clinical high risk for psychosis: a systematic review. Schizophr 2023;9(1). [doi: 10.1038/s41537-023-00362-z]
117. Muddle S, Jones B, Taylor G, Jacobsen P. A systematic review and meta‐analysis of the association between emotional stress reactivity and psychosis. Early Intervention Psych 2022 Sep;16(9):958-978. [doi: 10.1111/eip.13247]
118. Freeman D, Rosebrock L, Waite F, et al. Virtual reality (VR) therapy for patients with psychosis: satisfaction and side effects. Psychol Med 2023 Jul;53(10):4373-4384. [doi: 10.1017/S0033291722001167]
119. O'Brien C, Rus‐Calafell M, Craig TK, et al. Relating behaviours and therapeutic actions during AVATAR therapy dialogue: An observational study. British J Clinic Psychol 2021 Nov;60(4):443-462. [doi: 10.1111/bjc.12296]
120. McCluskey A, de Vries J. Help, I can't help: A qualitative study of psychiatric nurse's perspectives on their care for people who hear voices. Psychiatric Ment Health Nurs 2021 Apr;28(2):138-148. [doi: 10.1111/jpm.12642]

XSL•FO
RenderX

121. Bogen‐Johnston L, deVisser R, Strauss C, Hayward M. A qualitative study exploring how practitioners within early intervention in psychosis services engage with service users' experiences of voice hearing? Psychiatric Ment Health Nurs 2020 Oct;27(5):607-615. [doi: 10.1111/jpm.12612]
122. Vernal DL, Nordentoft M, Christensen MJ, et al. Status and clinical experiences from the challenge trial – a randomized controlled trial investigating virtual reality-based therapy for auditory hallucinations. World Social Psychiatry 2023;5(1):71-76. [doi: 10.4103/wsp.wsp_4_23]
123. Alderson-Day B, Jones N. Understanding AVATAR therapy: who, or what, is changing? Lancet Psychiatry 2018 Jan;5(1):2-3. [doi: 10.1016/S2215-0366(17)30471-6] [Medline: 29175275]
124. Hayward M. Continuing the conversation about AVATAR therapy. Lancet Psychiatry 2018 Mar;5(3):196. [doi: 10.1016/S2215-0366(18)30041-5] [Medline: 29482759]
125. Vilhauer RP. Stigma and need for care in individuals who hear voices. Int J Soc Psychiatry 2017 Feb;63(1):5-13. [doi: 10.1177/0020764016675888] [Medline: 27815515]
126. Colizzi M, Ruggeri M, Lasalvia A. Should we be concerned about stigma and discrimination in people at risk for psychosis? A systematic review. Psychol Med 2020 Apr;50(5):705-726. [doi: 10.1017/S0033291720000148] [Medline: 32063250]
127. Burke E, Wood L, Zabel E, Clark A, Morrison AP. Experiences of stigma in psychosis: A qualitative analysis of service users' perspectives. Psychosis 2016 Apr 2;8(2):130-142. [doi: 10.1080/17522439.2015.1115541]
128. Ward T, Lister R, Fornells-Ambrojo M, et al. The role of characterisation in everyday voice engagement and AVATAR therapy dialogue. Psychol Med 2021 Apr 8;52(16):1-8. [doi: 10.1017/S0033291721000659] [Medline: 33827728]
129. Heriot-Maitland C, McCarthy-Jones S, Longden E, Gilbert P. Compassion focused approaches to working with distressing voices. Front Psychol 2019;10:152. [doi: 10.3389/fpsyg.2019.00152] [Medline: 30774614]
130. Heriot-Maitland C, Bell T. Compassion-focused chairwork for voice-hearing relationships, body triggers and motivational states. Psychol Psychother 2025 May 11. [doi: 10.1111/papt.12600] [Medline: 40350786]
131. Carvalho D, Marques CC, Pinto AM, Martins MJ, Leal I, Castilho P. Fears of compassion scales in psychosis: confirmatory factor analysis and psychometric properties. Curr Psychol 1979. [doi: 10.1007/s12144-021-01979-y]
132. Gilbert P. Explorations into the nature and function of compassion. Curr Opin Psychol 2019 Aug;28:108-114. [doi: 10.1016/j.copsyc.2018.12.002] [Medline: 30639833]
133. Hardy A, Keen N, van den Berg D, et al. Trauma therapies for psychosis: A state-of-the-art review. Psychol Psychother 2024 Mar;97(1):74-90. [doi: 10.1111/papt.12499] [Medline: 37795877]
134. Hardy A. Pathways from trauma to psychotic experiences: a theoretically informed model of posttraumatic stress in psychosis. Front Psychol 2017;8:697. [doi: 10.3389/fpsyg.2017.00697] [Medline: 28588514]
135. Berry C, Baloc AI, Fowler D, Jones AM, Hazell CM, Hayward M. The psychological therapy preferences of patients who hear voices. Psychosis 2023 Jan 2;15(1):1-16. [doi: 10.1080/17522439.2022.2095000]
136. Brand RM, McEnery C, Rossell S, Bendall S, Thomas N. Do trauma-focussed psychological interventions have an effect on psychotic symptoms? A systematic review and meta-analysis. Schizophr Res 2018 May;195:13-22. [doi: 10.1016/j.schres.2017.08.037] [Medline: 28844432]
137. Powell BJ, McMillen JC, Proctor EK, et al. A compilation of strategies for implementing clinical innovations in health and mental health. Med Care Res Rev 2012 Apr;69(2):123-157. [doi: 10.1177/1077558711430690] [Medline: 22203646]
138. Brander M, Egger ST, Hürlimann N, et al. Virtual reality human-human interface to deliver psychotherapy to people experiencing auditory verbal hallucinations: development and usability study. JMIR Serious Games 2021 Jun 1;9(2):e26820. [doi: 10.2196/26820] [Medline: 33769295]
139. Watson A, Mellotte H, Hardy A, Peters E, Keen N, Kane F. The digital divide: factors impacting on uptake of remote therapy in a South London psychological therapy service for people with psychosis. J Ment Health 2022 Nov 2;31(6):825-832. [doi: 10.1080/09638237.2021.1952955]
140. Dennard S, Patel R, Garety P, Edwards C, Gumley A. A systematic review of users experiences of using digital interventions within psychosis: a thematic synthesis of qualitative research. Soc Psychiatry Psychiatr Epidemiol 2025 Feb;60(2):275-303. [doi: 10.1007/s00127-024-02692-4] [Medline: 38802509]
141. Happell B, Gordon S, Bocking J, et al. How did I not see that? Perspectives of nonconsumer mental health researchers on the benefits of collaborative research with consumers. Int J Mental Health Nurs 2018 Aug;27(4):1230-1239. [doi: 10.1111/inm.12453]
142. Christensen MJ, Vernal DL, Edwards CJ, Loizou S, Allen S, Ward T. Assessing the content of voice-hearing among clinical and non-clinical voice-hearers: a scoping review protocol. OSF 2024. [doi: 10.17605/OSF.IO/89TWE]
143. AVATAR2 research programme. OSF. URL: https://osf.io/y5dm6/ [accessed 2025-04-07]

Abbreviations

AT: Avatar therapy
AV-BRF: Avatar brief version
AV-EXT: Avatar extended version
RCT: randomized controlled trial

XSL•FO
RenderX

RT: relating therapy
TwV: talking with voices
VR: virtual reality
VRT: virtual reality-assisted therapy

Edited by A Coristine; submitted 22.05.25; peer-reviewed by L Jenner, T Craig; revised version received 30.09.25; accepted 30.09.25; published 01.12.25.

Please cite as:
Christensen MJ, Rydborg MP, Jørgensen R, Nielsen CD, Mainz J, Bell IH, Thomas N, Smith LC, Mariegaard LS, Ward T, Nordentoft M, Glenthøj LB, Vernal DL
Immersive Virtual Reality–Assisted Therapy for Distressing Voices in Psychosis: Qualitative Study of Participants' and Therapists' Experiences in the Challenge Trial
JMIR Serious Games 2025;13:e77920
URL: https://games.jmir.org/2025/1/e77920
doi:10.2196/77920

© Mads Juul Christensen, Matilde Poulsen Rydborg, Rikke Jørgensen, Cecilie Dueholm Nielsen, Jan Mainz, Imogen H Bell, Neil Thomas, Lisa Charlotte Smith, Lise Sandvig Mariegaard, Thomas Ward, Merete Nordentoft, Louise Birkedal Glenthøj, Ditte Lammers Vernal. Originally published in JMIR Serious Games (https://games.jmir.org), 1.12.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Kangaroo Stimulation Game in Tracheostomized Intensive Care–Related Dysphagia: Interventional Feasibility Study

Marjolein Jansen[1], RN; Ingrid D van Iperen[1], MSc; Anke Kroner[2], MD, PhD; Raphael Hemler[3], MD; Esther Dekker-Holverda[4], BSc; Peter E Spronk[1,2], MD, PhD

[1]Expertise Center for Intensive Care Rehabilitation Apeldoorn (ExpIRA), Gelre Hospitals, Apeldoorn, The Netherlands

[2]Department of ICU, Gelre Hospitals, A.Schweitzerlaan 31, Apeldoorn, The Netherlands

[3]Department of ENT, Gelre Hospitals, Apeldoorn, The Netherlands

[4]Department of Speech and Language Pathology, Gelre Hospitals, Apeldoorn, The Netherlands

Corresponding Author:
Peter E Spronk, MD, PhD
Expertise Center for Intensive Care Rehabilitation Apeldoorn (ExpIRA), Gelre Hospitals, Apeldoorn, The Netherlands

Abstract

Background: Dysphagia is common in intensive care unit (ICU) patients. Using surface electromyography (sEMG) signals as biofeedback training exercises might offer a promising path to improving swallowing function. The Rephagia biofeedback system uses sEMG to assess muscle strength, stamina, and timing of the swallowing action.

Objectives: The aim of this study was to evaluate the feasibility of the Rephagia system in ICU patients with dysphagia.

Methods: This feasibility study included patients admitted to a 14-bed mixed medical-surgical ICU. All patients underwent a new tracheostomy placement during ICU stay due to persistent aspiration and ICU-acquired weakness, accompanied by verified dysphagia. Following Rephagia training, patients completed a questionnaire assessing comprehension, satisfaction, and motivation. Swallowing characteristics were assessed via mean sEMG peak values during exercise.

Results: Twenty patients with a mean age of 69.4 (SD 8.2) years were included. The means of sEMG values at the beginning of a measurement were not significantly different at baseline versus everyone's last measurement (52 µV [23 µV] vs 57 µV [22 µV]; *P*=.50). The means of sEMG values obtained at the end of a measurement were not significantly different at baseline versus everyone's last measurement (56 µV [18 µV] vs 59 µV [23 µV]; *P*=.62). However, dysphagia improved in all patients. Patients understood the importance of the game in relation to their swallowing problems (16/80, 89%), which kept them motivated to participate in the training sessions (9/18, 50%).

Conclusions: The Rephagia biofeedback system for stimulating swallowing actions in tracheotomized ICU patients with dysphagia is feasible. No relation was found between clinical improvement in swallowing function and sEMG signals.

(JMIR Serious Games 2025;13:e60685) doi:10.2196/60685

KEYWORDS

dysphagia; swallowing; intensive care; ICU; swallowing disturbance; kangaroo stimulation; game; feasibility study; surface electromyography; training; exercise; Rephagia biofeedback; muscle strength; stamina; timing; tracheostomy; clinical; feasibility

Introduction

Swallowing disturbances occur in up to 30% of hospitalized patients with a length of stay longer than 48 hours [1]. Swallowing in general and the oral, pharyngeal, and esophageal phases of this process require a complex interaction between the neuromuscular system in the tongue, floor of mouth, pharynx, and larynx [2]. This intricate physiological course of muscle events and interactions may become disturbed in patients treated in the intensive care unit (ICU). Swallowing impairment or dysphagia is not uncommon in ICU patients with a reported prevalence of up to 62% of those who were recently extubated [3]. Dysphagia in ICU patients is probably caused in part by muscle weakness, sensory neuropathy, and cognitive disturbance due to critical illness but possibly also in part by inadvertent trauma to laryngeal structures due to endotracheal intubation [4]. Dysphagia is associated with prolonged length of hospital stay and the composite end point of increased risk of pneumonia, reintubation, or death. This significantly affects the patient's well-being, ICU resources, and health care costs [5]. It has also been hypothesized that certain interventions used to prevent aspiration, such as cuffed tracheostomies, might in the long run actually maintain or worsen dysphagia due to disturbance of normal laryngeal anatomy and sensibility [6].

The prevalence of dysphagia is frequently underestimated and interventions to improve swallowing function are not evident [7,8]. Most frequently used interventions include exercises such as repetitive swallowing, chin tuck against resistance, and

XSL•FO
RenderX

respiratory exercises [9]. Recently, electrical stimulation of the pharyngeal muscle groups has been shown in one center to be promising, especially in neurological patients [10]. Muscle activity can be observed as surface electromyography (sEMG), although its interpretation is complex [11]. In parallel, it is intriguing to consider the use of sEMG signals as biofeedback training exercises to affect different parts of the swallowing process including laryngeal lifting, oral transit time, and excursion of the hyoid bone. Indeed, this approach has been shown to be feasible in patients with several diseases including Parkinson disease and those who experienced a stroke [12-14]. The Rephagia biofeedback system uses sEMG to assess muscle strength, stamina, and timing of the swallowing action and visualizes this using a graph. Furthermore, the sEMG signal from the patient may be used in an interactive game using various exercises in which timing and strength of the swallowing action are trained.

We hypothesized that the Rephagia biofeedback system using an interactive game to evaluate and stimulate swallowing is feasible and potentially useful to evaluate dysphagia over time in ICU patients.

Methods

Setting

This study was designed as an uncontrolled prospective effectiveness and feasibility study between August 2017 and September 2021 and was undertaken in the 14-bed mixed medical-surgical ICU of Gelre Hospitals Apeldoorn, a nonacademic university–affiliated teaching hospital. The team of ICU care professionals including the speech and language professionals and ear-nose-throat consultants did not change during the study.

Patients

Patients were included if they had a new tracheostomy (Tracoe cannula with cuff) placed during ICU stay due to persistent aspiration and ICU-acquired weakness and did not have a history of swallowing impairment or other dysphagia symptoms. Patients were considered for inclusion as soon as they were in the process of weaning from the ventilator and could be detached from the ventilator for at least 2 consecutive hours. Patients were excluded if language barriers existed, or if permanent cognitive impairment was present. In addition, patients with severe visual impairment were excluded.

Ethical Considerations

The study was approved by the institutional ethical review board (no. TCO 16.10) in accordance with institutional and national guidelines for research involving human participants. Ethical approval covered all aspects of the study, including patient recruitment, data collection, and analysis. Eligible patients or, if applicable, their legal representatives were verbally informed about the study and given the opportunity to ask questions. After being invited to participate, the patients and their legal representatives received an information pamphlet detailing the study procedures. They were given at least 30 minutes to consider participation before signing an informed consent form. Patients were allowed to cease participating at any time if they did not want to continue. All patient data were anonymized to ensure confidentiality. Identifiable information was not included in the study records or publications. Patients were observed closely while assessing swallowing and the test was paused or discontinued if swallowing safety was compromised or at the request of the participant. No financial or other compensation was provided to the participants.

Procedure

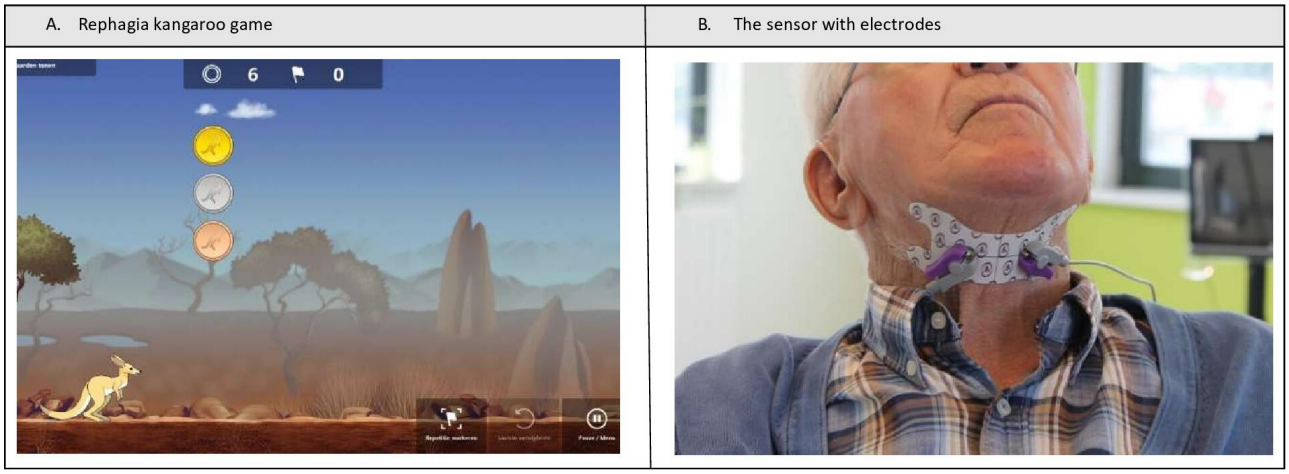
If there was any suspicion of dysphagia, the speech and language pathologist (SLP) was contacted and initiated the flexible endoscopic evaluation of swallowing (FEES), which was performed by the ear, nose, and throat specialist together with the SLP. All patients underwent the standardized FEES examination using the Murray Secretion Scale [15] and the Rosenbek Penetration Aspiration Scale (PAS) [16] on thin liquid, thick liquid, and regular food if safe. For thin liquid, water with blue dye was used and for thick liquid, regular yoghurt was used.

Dysphagia was judged to be present if Murray Secretion Scale score was 2 or higher and PAS score was higher than 2 in thin, modified liquid or regular food [17]. If FEES demonstrated dysphagia or aspiration, patients were eligible for enrollment in the next step of the study. After the FEES, Rephagia training was started as soon as possible. All participating patients were asked after the first Rephagia training to fill out a questionnaire with questions concerning understanding, satisfaction, and motivation using a modified tool (Table S1 in Multimedia Appendix 1).

Subsequently, daily training and exercise was performed using the Rephagia kangaroo game (Figure 1A). Settings (bolus size, bolus consistency, and how many swallowing actions) of the exercise were defined by an experienced SLP, using a specified protocol as described previously [1]. The daily Rephagia swallowing exercises were supervised by a trained ICU nurse. In addition, all patients were evaluated once weekly by an SLP. Parameters related to acquired sEMG in relation to clinical parameters such as swallow timing and strength were stored for later analysis. The daily training session was paused if one of the following safety impairments occurred: coughing, wet voice, and decrease of oxygen saturation of 3% or more on the finger pulse oximeter. The training sessions were terminated if improvement of swallowing had not occurred after 3 weeks of training sessions, or if the patient had been successfully decannulated.

XSL•FO
RenderX

Figure 1. The Rephagia kangaroo game (A) and the sensor with electrodes (B).



Swallowing Characteristics and Rephagia System

Besides the initial FEES analysis, swallowing characteristics were assessed using the sEMG data provided by the Rephagia biofeedback (SilverFit). These data may be used as a tool to characterize swallowing action using the sEMG peaks in individual patients. It is a noninvasive but sensitive method to identify a swallowing action. An electrode with a sensor for the sEMG signals was attached to the submandibular region to assess the swallowing characteristics (Figure 1B). The sensor was connected to a laptop with the Rephagia software via Bluetooth. The signals from the sensor were used in an interactive game. Before the start of the actual game, the sEMG signals were calibrated for each individual patient and each test by obtaining 3 normal swallows to obtain an individual baseline value. In the game, the patient received feedback of the swallow activity in the form of a hopping kangaroo. The administered test consisted of repetitions where the patient was asked to swallow when the kangaroo stopped at a coin and waited for the patient to swallow.

The swallows were repeated 10 - 25 times on individual basis using bolus volume and type of food prescribed by an experienced SLP. In this study, swallow characteristics were obtained using the mean peak values of swallow actions of the participants obtained over the course of an exercise. The first 3 swallow actions and the last 3 swallow actions were used to correct for individual variation on the one hand and to gain insight in individual swallowing characteristics and changes in the individual swallowing characteristics during the exercise on the other hand. The first 3 swallow actions are described as the beginning of a measurement and the last 3 swallow actions are described as the end of a measurement. Normal values of sEMG were previously described in healthy individuals as mean 82 µV (SD 33 µV) [18].

Statistical Analysis

Descriptive statistics were generated for all outcome measures. Data distribution was tested for normality using the Shapiro-Wilk test. Accordingly, variables were described as mean and SD or as median and IQR. Categorical variables were described using frequencies and percentages. The analysis of the comparison of baseline swallowing characteristics with the swallowing characteristics of the last swallowing measurement was conducted using the paired 2-tailed *t* test or the Wilcoxon signed rank test depending on the normality of data distribution. The change in swallowing characteristic within a measurement at baseline versus the last measurement was also analyzed using either the paired 2-tailed *t* test or the Wilcoxon signed rank test.

Individual changes in swallowing characteristics were analyzed by calculating the percentage of measurements higher or lower than the baseline measurements. Data were analyzed with the statistical package statistical package SPSS (version 25; IBM Corp), Excel (version 2202; Microsoft Corp), and R (version 4.0.4; R Foundation for Statistical Computing). A *P* value of <.05 was considered significant.

Results

Patients

After screening 58 patients, 20 patients were formally included. Thirty-eight patients could not be included because of the following reasons: no informed consent (N=4); too late for inclusion (N=4); no FEES possible due to the lack of trained ear, nose, and throat specialist or SLP present (N=4); preexistent swallowing problems; cognitive, neurological, or vision problems; or delirium (N=11). Fifteen patients were subsequently excluded from study inclusion after the FEES because no signs of dysphagia were found. All outcome parameters were recorded prior to participation at baseline until end of the study (Table 1). Mean age at baseline of the included patients was 69.4 (SD 8.2) years. The mean ICU length of stay was 52.0 (SD 21.9) days. In 12 patients, the PAS score for their ability to swallow thin liquid was assessed. All 12 patients demonstrated scores higher than 2 on the PAS, classifying as dysphagia. In 18 patients, the PAS score for their ability to swallow thick liquid was assessed. Seventeen patients obtained scores higher than 2 and 1 patient obtained a score of 1. The Murray Secretion Scale score was assessed in 17 patients. Two patients obtained a normal rating, 3 patients had secretions outside the laryngeal vestibule that were cleared with spontaneous swallows, 4 patients had deeply pooled secretions, and 8 patients had secretions in the laryngeal vestibule that were not cleared. The comprehensive PAS scores and Murray Secretion Scale scores are shown in Table S1 in Multimedia

XSL•FO
RenderX

Appendix 2. One month after the end of the study, patient charts were reviewed. There was no occurrence of pneumonia, ICU readmission, and reintubation. All patients were successfully decannulated. Patients were decannulated based on clinical observations made by registered nurses and assessments by SLPs. Two patients died during the ICU admission, not related to swallowing abnormalities.

Table . Patient characteristics[a].

Characteristics		Values
Sex, n (%)		
	Male	13 (65)
	Female	7 (35)
Age (years), mean (SD)		69.4 (8.2)
Reason for admission, n (%)		
	Medical	16 (80)
	Emergency surgery	3 (15)
	Elective surgery	1 (5)
Hospital length of stay (days), mean (SD)		74.3 (23.0)
ICU[b] length of stay (days), mean (SD)		52.0 (21.9)
Number of practice moments, median (IQR)		4.0 (3.0-6.2)
Confirmed infection during hospital stay, n (%)		
	No	2 (10)
	Yes	18 (90)
SAPS II [c] score, mean (SD)		44.2 (14.7)
Mechanical ventilation at admission[d], n (%)		
	No	13 (65)
	Yes	7 (35)
Mechanical ventilation within 24 hours, n (%)		
	No	4 (20)
	Yes	16 (80)
Hours on mechanical ventilator, median (IQR)		978.5 (638.7-1257.8)
Days witch tracheostomy[e], median (IQR)		31.0 (22.4-95.1)
APACHE II score, mean (SD)		19.9 (7.2)
APACHE III score, mean (SD)		76.1 (28.8)
Predicted mortality (%)[f], mean (SD)		34.1 (22.5)
Died, n (%)		
	No	18 (90)

[a]Variables with a normal distribution are described as mean (SD) and variables with a nonnormal distribution are described as median (IQR). For the categorical variable, count (n) and percentage (%) are presented.

[b]ICU: intensive care unit.

[c]SAPSII: Simplified Acute Physiology Score.

[d]Use of a ventilator at the moment of intensive care admission or immediately (within 15 minutes) thereafter.

[e]Number of days from the placement of the tracheostomy until decannulation.

[f]Predicted mortality score according to the APACHE IV prediction model (in percentages).

Swallow Characteristics

No safety-related events occurred during the training sessions. The total duration of the exercise with Rephagia training was 20 minutes per session. The median number of practice sessions was 4 [3-6] per patient. Details for the participating patients are shown in Table 1. The means of sEMG values at the beginning of a measurement were not significantly different at baseline versus everyone's last measurement (52 µV [23 µV] vs 57 µV [22 µV]; *P*=.50). Similarly, the means of sEMG values obtained

XSL•FO
RenderX

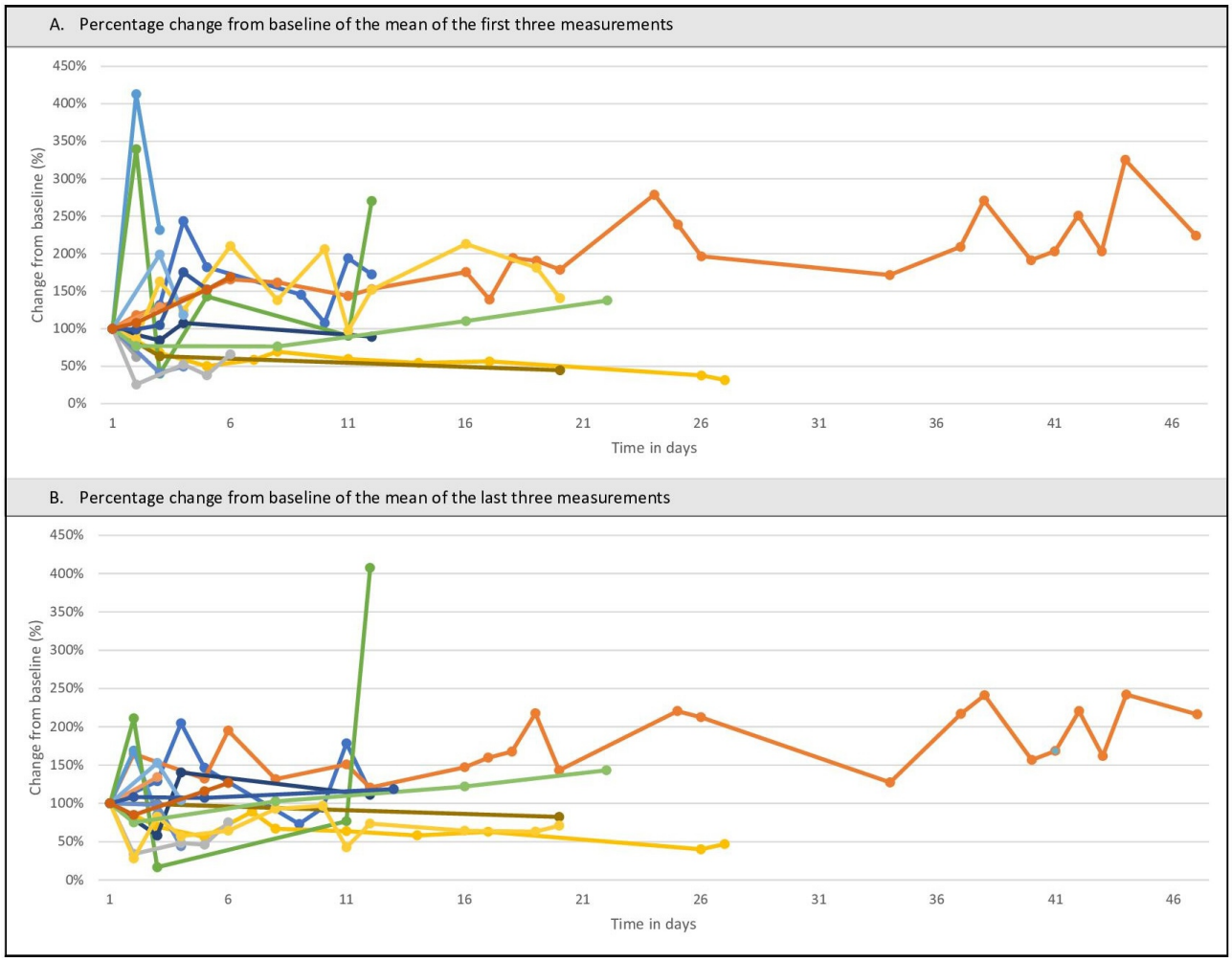
at the end of a measurement were not significantly different at baseline versus everyone's last measurement (56 µV [18 µV] vs 59 µV [23 µV]; *P*=.62).

The means of the sEMG values at the beginning of a measurement were not significantly different from the means of the sEMG values at the end of a measurement when looking at baseline compared with everyone's last measurement (3.7 µV vs 2.6 µV; *P*=.72).

In 2 patients, the PAS score was assessed during a follow-up FEES after decannulation. The PAS scores of the first patient were 8 on both thin and thick liquid and 1 after decannulation on both consistencies. The PAS scores of the second patient were 7 on thin liquid, 3 on thick liquid, and 1 after decannulation on both consistencies.

In Figure 2 all individual sEMG signals are depicted. The means of the swallows during each practice moment are shown as a percentual change from the patients' own baseline, which are plotted against the time in days. No clear trend can be seen in these graphs.

Figure 2. The percentage change from baseline of the surface electromyography data of the mean swallows of the (A) first 3 measurements and the (B) last 3 measurements.

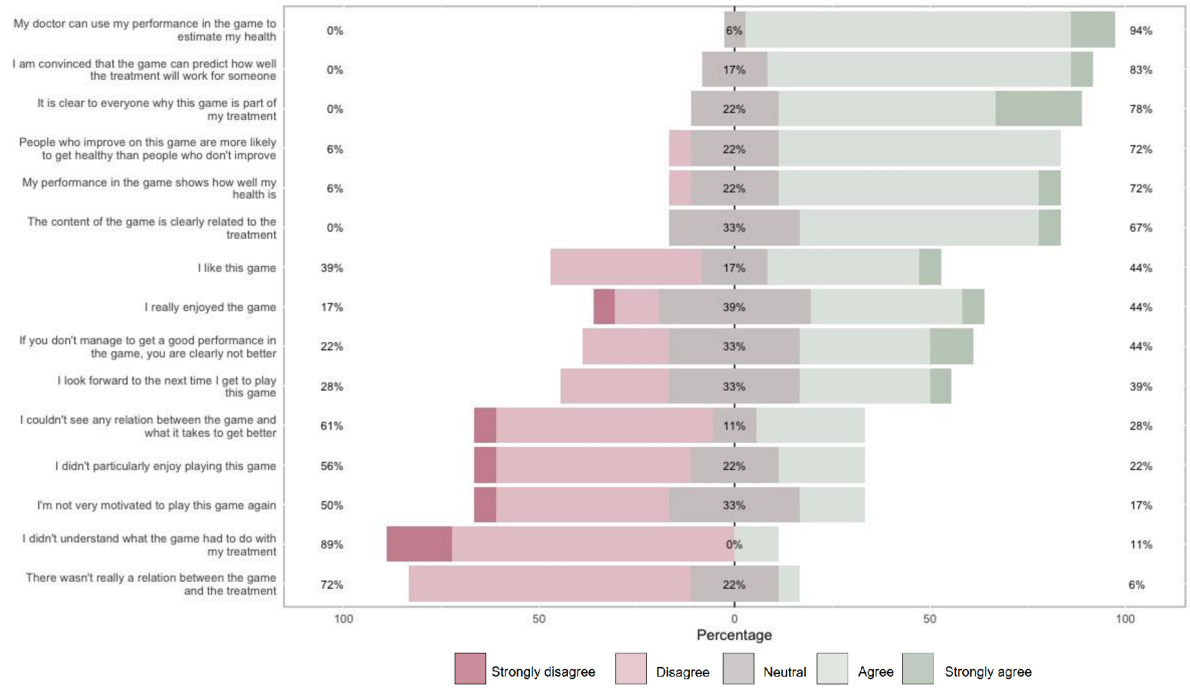


Patient Satisfaction and Motivation

Our findings demonstrate that patients recognize how important swallowing training is during ICU stay. The patients understood the importance of the game in relation to their swallowing problems (16/18, 89%), which kept them motivated to participate in the training sessions (9/18, 50%). The findings demonstrate that patients think that the ICU doctor can use the performance in the game to estimate the severity of illness (19/18, 95%) and predict how the treatment will affect their dysphagia (15/18, 83%). The patients were also convinced that the game can predict the effect of the training and in which way this reflects the patients' health status. The results of the Likert scale questions are summarized in Figure 3.

XSL•FO
RenderX

Figure 3. Summary response of the patient questionnaire.



Discussion

Principal Findings

In this study, we showed that the Rephagia biofeedback system using an interactive game for the stimulation of swallowing in dysphagic ICU patients is feasible in an ICU environment. Individual changes in obtained sEMG signals during swallowing actions were not useful to evaluate improvements in dysphagia.

Playing interactive games and engaging patients in an ICU setting are feasible [19,20]. However, the use of a surface electrode using detected EMG signals while performing muscle tasks in the ICU represents a novel approach in this field. Previous studies in healthy participants demonstrated a relation between obtained sEMG signals from the masseter muscle and effortful swallow maneuvers [21]. Moreover, using sEMG signals from masseter and hyoid muscles, age could be identified as a risk factor for dysphagia [22]. Decreased sEMG signals were found to be associated with swallowing disorders in older adult patients with sarcopenia [23]. However, it may also prove difficult to use sEMG signals as derivatives of swallowing function, since different swallow-specific tasks in healthy adults may cause increased or decreased submental muscle activities [24].

All participating patients showed significant improvements in their swallowing function over the course of the study. However, it was disappointing that these clinical improvements did not coincide with observable changes in surface EMG signals during swallows. In previous studies, the benefits of training with biofeedback sEMG were tested over multiple sessions in patients diagnosed with dysphagia [1], whereas a relatively low number of practice sessions was attainable in this study caused by a delayed availability of FEES; no availability of trained ICU nurses; fatigue of patients due to, for example, dialysis, fever, or sickness; other procedures such as computed tomography or magnetic resonance imaging; and family visits. These factors require attention in future studies with the Rephagia system.

Participants did not display a significant increase in surface EMG intensity if the first and last swallow actions were evaluated. However, all patients did clinically improve in terms of swallowing function, resulting eventually in decannulation in all cases. This confirms previous observations in patients with clinical dysphagia compared with those with no dysphagia where no differences in sEMG signals were found between both groups [1].

These results are not so surprising because the sEMG signal is a derivative of isotonic muscle contractions and is characterized by a nonlinear relationship between force and EMG signal. In addition, the coordination and duration of contractions of muscle groups during dynamic muscle contractions are complex [11]. Moreover, other factors are in play as well including improved sensibility and coordination, a compensatory mechanism consisting of smaller sips or bites to swallow or changing the consistency, or a combination of those factors [6].

In the ICU environment, patients may also experience additional muscle atrophy and weakness, which could impact their ability to swallow effectively. Notably, the obtained sEMG signals in this study (around 55 µV) were considerably lower than values obtained in non-ICU patients with swallowing problems (around 105 µV) [1]. However, raw differences in EMG amplitudes across different configurations and electrode placements may not be a valid means to verify muscle activity or intensity, and this could be a contributing factor to the observed differences with non-ICU patients. While older age may also be associated with lower sEMG levels [25], the mean age in both studies was comparable, suggesting that age alone is unlikely to explain the present results. Despite this, clinical improvements in

XSL•FO
RenderX

swallowing function can still occur due to adaptive and compensatory strategies, which might not be fully captured by sEMG alone. These voluntary compensatory strategies might involve adjustments in head and neck posture or the recruitment of postural muscles to aid in swallowing. Other approaches using accelerometers as a sensor for biofeedback games may be a useful alternative as was demonstrated in a small study describing the positive effects on swallowing in poststroke dysphagia using a "frog swallows mosquito game" [26]. Moreover, involuntary adaptive movements and displacements in adjacent or related structures may also compensate for swallowing deficiencies in various anatomical areas [27]. In addition to compensatory strategies, improved sensitivity and coordination can contribute to clinical improvements in swallowing. Patients may adopt smaller sips or bites to swallow or modify the consistency of food, and a combination of these factors might lead to enhanced swallowing function [6]. Indeed, all participants in this study learned to modulate their swallowing musculature in response to the biofeedback to play the game. In addition, despite the severe swallowing disorder of the impaired participant, the patients were able to play the game and understood the importance of the game in relation to their swallowing problems which kept them motivated to participate in the training sessions. This confirms our previous experience [28] and is important to engage patients in their personal recovery and rehabilitation activities [29].

Strengths and Limitations

This study has several strengths. First, it is the first study concerning the feasibility of the Rephagia system in tracheotomized ICU patients in an ICU setting. Second, swallowing characteristics were evaluated in a relatively new manner, using sEMG values by comparing swallowing actions of patients with each other as well as the change in swallow actions in 1 exercise within a group of patients. Third, we quantified the interest of patients in swallowing training, which had not been done before. The positive feedback from participants, together with the encouraging physiological effects, strengthens the evidence of the feasibility of this swallowing training program.

This study has several limitations that should be mentioned. First, it was a small case series performed in 1 center in the Netherlands, limiting the generalizability of the results. Second, we did not track participant adherence with other exercises (or swallowing of food or liquid) given by an SLP or an ICU nurse and therefore do not know how many repetitions of exercises each participant completed. Third, the interventions applied in this study did not specifically target each individual's underlying deficits in their swallowing physiology, rather the intervention protocol was a predetermined set of swallowing exercise that is adopted in clinical practice. Fourth, the specific EMG parameters corresponding to swallowing time and intensity were not recorded or analyzed in detail. Instead, only the mean EMG values before and after the patient's training were considered. This approach may not fully capture the nuanced changes in swallowing mechanics and muscle activation over time. As a result, it is difficult to definitively conclude that the observed EMG changes were directly related to improvements in dysphagia, as other factors influencing muscle activity during swallowing may have contributed to the results. Fifth, the questionnaire was administered after the first session of sEMG biofeedback, meaning participants had limited experience with the swallowing exercise. However, this approach was chosen to maximize the number of responses , despite the potential risk of response bias, as participants might have felt inclined to provide more positive feedback to please the researcher. To mitigate this, they were informed that the purpose of the study was to evaluate a new swallowing training, in which the researcher had no vested interest, to determine its potential recommendation for other ICU patients and clinicians. They were also reassured that the questionnaire forms remained anonymous and were analyzed at a later date. Therefore, we think that the risk of response bias was minimized as far as possible.

Conclusions

The Rephagia biofeedback system for stimulating swallowing actions in dysphagic tracheotomized ICU patients is feasible. Patients understood the importance of the game in relation to their swallowing problems and understood the relationship with their treatment. No relation was found between clinical improvement in swallowing function and sEMG signals obtained by the system. A prospective intervention study is required to assess the potential additional role of the Rephagia system in swallowing rehabilitation in the ICU.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Patient questionnaires
[DOCX File, 14 KB - games_v13i1e60685_app1.docx]

Multimedia Appendix 2
Baseline Penetration Aspiration Scale scores and Murray Secretion Scale
[DOCX File, 14 KB - games_v13i1e60685_app2.docx]

References

XSL•FO
RenderX

1. Spronk PE, Spronk LEJ, Lut J, et al. Prevalence and characterization of dysphagia in hospitalized patients. Neurogastroenterol Motil 2020 Mar;32(3):e13763. [doi: 10.1111/nmo.13763] [Medline: 31742866]
2. Logemann JA. The evaluation and treatment of swallowing disorders. Curr Opin Otolaryngol Head Neck Surg 1998 Dec;6(6):395-400. [doi: 10.1097/00020840-199812000-00008]
3. Skoretz SA, Flowers HL, Martino R. The incidence of dysphagia following endotracheal intubation: a systematic review. Chest 2010 Mar;137(3):665-673. [doi: 10.1378/chest.09-1823] [Medline: 20202948]
4. Macht M, Wimbish T, Clark BJ, et al. Postextubation dysphagia is persistent and associated with poor outcomes in survivors of critical illness. Crit Care 2011;15(5):R231. [doi: 10.1186/cc10472] [Medline: 21958475]
5. Altman KW, Yu GP, Schaefer SD. Consequence of dysphagia in the hospitalized patient: impact on prognosis and hospital resources. Arch Otolaryngol Head Neck Surg 2010 Aug;136(8):784-789. [doi: 10.1001/archoto.2010.129] [Medline: 20713754]
6. Macht M, Wimbish T, Bodine C, Moss M. ICU-acquired swallowing disorders. Crit Care Med 2013 Oct;41(10):2396-2405. [doi: 10.1097/CCM.0b013e31829caf33] [Medline: 23939361]
7. van Snippenburg W, Kröner A, Flim M, et al. Awareness and management of dysphagia in Dutch intensive care units: a nationwide survey. Dysphagia 2019 Apr;34(2):220-228. [doi: 10.1007/s00455-018-9930-7] [Medline: 30069599]
8. Brodsky MB, Nollet JL, Spronk PE, González-Fernández M. Prevalence, pathophysiology, diagnostic modalities, and treatment options for dysphagia in critically ill patients. Am J Phys Med Rehabil 2020;99(12):1164-1170. [doi: 10.1097/PHM.0000000000001440]
9. Spronk PE, Spronk LEJ, Egerod I, et al. Dysphagia in Intensive Care Evaluation (DICE): an international cross-sectional survey. Dysphagia 2022 Dec;37(6):1451-1460. [doi: 10.1007/s00455-021-10389-y] [Medline: 35092486]
10. Dziewas R, Stellato R, van der Tweel I, et al. Pharyngeal electrical stimulation for early decannulation in tracheotomised patients with neurogenic dysphagia after stroke (PHAST-TRAC): a prospective, single-blinded, randomised trial. Lancet Neurol 2018 Oct;17(10):849-859. [doi: 10.1016/S1474-4422(18)30255-2] [Medline: 30170898]
11. Disselhorst-Klug C, Schmitz-Rode T, Rau G. Surface electromyography and muscle force: Limits in SEMG–force relationship and new approaches for applications. Clin Biomech (Bristol, Avon) 2009 Mar;24(3):225-235. [doi: 10.1016/j.clinbiomech.2008.08.003]
12. Benfield JK, Hedstrom A, Everton LF, Bath PM, England TJ. Randomized controlled feasibility trial of swallow strength and skill training with surface electromyographic biofeedback in acute stroke patients with dysphagia. J Oral Rehabil 2023 Jun;50(6):440-451. [doi: 10.1111/joor.13437] [Medline: 36810785]
13. Battel I, Walshe M. An intensive neurorehabilitation programme with SEMG biofeedback to improve swallowing in idiopathic Parkinson's disease (IPD): a feasibility study. Int J Lang Commun Disord 2023 May;58(3):813-825. [doi: 10.1111/1460-6984.12824] [Medline: 36478035]
14. Albuquerque LCA, Pernambuco L, da Silva CM, Chateaubriand MM, da Silva HJ. Effects of electromyographic biofeedback as an adjunctive therapy in the treatment of swallowing disorders: a systematic review of the literature. Eur Arch Otorhinolaryngol 2019 Apr;276(4):927-938. [doi: 10.1007/s00405-019-05336-5] [Medline: 30771061]
15. Murray J, Langmore SE, Ginsberg S, Dostie A. The significance of accumulated oropharyngeal secretions and swallowing frequency in predicting aspiration. Dysphagia 1996;11(2):99-103. [doi: 10.1007/BF00417898] [Medline: 8721067]
16. Rosenbek JC, Robbins JA, Roecker EB, Coyle JL, Wood JL. A penetration-aspiration scale. Dysphagia 1996;11(2):93-98. [doi: 10.1007/BF00417897] [Medline: 8721066]
17. Steele CM, Peladeau-Pigeon M, Barrett E, Wolkin TS. The risk of penetration-aspiration related to residue in the pharynx. Am J Speech Lang Pathol 2020 Aug 4;29(3):1608-1617. [doi: 10.1044/2020_AJSLP-20-00042] [Medline: 32598168]
18. Gnacke E, Mijnes D. SEMG-normaalwaarden van slikkracht en sliktiming in de ziekenhuispopulatie [SEMG normal data for swallow force and swallow timing in the hospitalized population]. [Unpublished Bachelor's thesis].: Zuyd University of Applied Sciences; 2018.
19. Gomes TT, Schujmann DS, Fu C. Rehabilitation through virtual reality: physical activity of patients admitted to the intensive care unit. Rev Bras Ter Intensiva 2019;31(4):456-463. [doi: 10.5935/0103-507X.20190078] [Medline: 31967219]
20. Kho ME, Damluji A, Zanni JM, Needham DM. Feasibility and observed safety of interactive video games for physical rehabilitation in the intensive care unit: a case series. J Crit Care 2012 Apr;27(2):219. [doi: 10.1016/j.jcrc.2011.08.017] [Medline: 21944880]
21. Bahia MM, Lowell SY. Surface electromyographic activity of the masseter muscle during regular and effortful saliva swallows: a preliminary study. Dysphagia 2024 Apr;39(2):231-240. [doi: 10.1007/s00455-023-10605-x] [Medline: 37477753]
22. Chang WH, Chen MH, Liu JF, Chung WL, Chiu LL, Huang YF. Surface electromyography for evaluating the effect of aging on the coordination of swallowing muscles. Dysphagia 2023 Oct;38(5):1430-1439. [doi: 10.1007/s00455-023-10572-3] [Medline: 37106228]
23. Sakai K, Nakayama E, Rogus-Pulia N, et al. Submental muscle activity and its role in diagnosing sarcopenic dysphagia. Clin Interv Aging 2020;15:1991-1999. [doi: 10.2147/CIA.S278793] [Medline: 33116453]
24. Söyler AK, Serel Arslan S, Demir N, Kıylıoğlu N. Surface electromyographic activity of submental muscles during different types of swallow-specific tasks: a comparative study. J Oral Rehabil 2023 Jul;50(7):580-586. [doi: 10.1111/joor.13448] [Medline: 36932464]

XSL•FO
RenderX

25. Barikroo A, Im I, Crary M, Carnaby G. Effects of age and lingual-palatal pressure levels on submental surface electromyography measures. Dysphagia 2023 Oct;38(5):1277-1285. [doi: 10.1007/s00455-023-10553-6] [Medline: 36656382]
26. Li CM, Wang TG, Lee HY, et al. Swallowing training combined with game-based biofeedback in poststroke dysphagia. PM R 2016 Aug;8(8):773-779. [doi: 10.1016/j.pmrj.2016.01.003] [Medline: 26791426]
27. Buchholz DW, Bosma JF, Donner MW. Adaptation, compensation, and decompensation of the pharyngeal swallow. Gastrointest Radiol 1985;10(3):235-239. [doi: 10.1007/BF01893106] [Medline: 4029539]
28. Spronk P, Iperen I, Flim M, Ruijter A, Vries R. Interactive Gaming Potentially Increases ICU Patients' Motivation to Be Engaged in Rehabilitation Programs. : De Intensivist; 2023.
29. Söderberg A, Karlsson V, Fagevik Olsén M, Thelandersson A, Johansson A. Patient as active partner—clue to successful early mobilization in intensive care. Physiother Theory Pract 2024 Oct;40(10):2298-2308. [doi: 10.1080/09593985.2023.2239891] [Medline: 37489585]

Abbreviations

FEES: flexible endoscopic evaluation of swallowing
ICU: intensive care unit
PAS: Penetration Aspiration Scale
sEMG: surface electromyography
SLP: speech and language pathologist

Edited by A Coristine, M Schmid; submitted 18.05.24; peer-reviewed by S Hamzic, Y Xie; revised version received 25.11.24; accepted 27.11.24; published 05.03.25.

Please cite as:
Jansen M, van Iperen ID, Kroner A, Hemler R, Dekker-Holverda E, Spronk PE
Kangaroo Stimulation Game in Tracheostomized Intensive Care–Related Dysphagia: Interventional Feasibility Study
JMIR Serious Games 2025;13:e60685
URL: https://games.jmir.org/2025/1/e60685
doi:10.2196/60685

© Marjolein Jansen, Ingrid D van Iperen, Anke Kroner, Raphael Hemler, E Dekker-Holverda, Peter E Spronk. Originally published in JMIR Serious Games (https://games.jmir.org), 5.3.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Improvement of Game Users' Depressive Symptoms via Behavioral Activation in a Massive Multiplayer Online Game: Randomized Controlled Trial

Kenji Yokotani[1,2], PhD; Masanori Takano[3], PhD; Nobuhito Abe[4], PhD; Takahiro A Kato[5], MD, PhD

[1]Graduate School of Technology, Industrial and Social Sciences, Tokushima University, Minamijosanjimacho 1-1, Tokushima, Japan

[2]Department of Data Science, PsychoBit, Inc., Kobe, Japan

[3]Multidisciplinary Information Science Center, CyberAgent (Japan), Tokyo, Japan

[4]Institute for the Future of Human Society, Kyoto University, Kyoto, Japan

[5]Department of Psychiatry, Hokkaido University Graduate School of Medicine, Sapporo, Japan

Corresponding Author:
Kenji Yokotani, PhD
Graduate School of Technology, Industrial and Social Sciences, Tokushima University, Minamijosanjimacho 1-1, Tokushima, Japan

Abstract

Background: Online games developed to improve mental health symptoms are reportedly effective among game users. However, it has not been verified whether massive multiplayer online games (MMOGs) developed for leisure purposes are effective in improving users' mental health symptoms.

Objective: Based on 2 theoretical frameworks, this study examined whether MMOGs improve depression and social anxiety. First, behavioral activation theory posits that depressive symptoms improve through the repetition of reward-linked behaviors. Second, inhibitory learning theory suggests that exposure to social stimuli (eg, being the center of attention) previously perceived as threatening reduces fear responses over time.

Methods: Participants were Pigg Party users with at least 3 months of previous experience. Overall, 1105 participants were randomly assigned to either the experimental (n=548) or waitlist groups (n=557). Participants in the experimental group were instructed to ring a friend's room bell (an action that clearly draws attention) on weekdays and to customize their avatars on weekends. Those completing ≥60% of the tasks received additional monthly rewards. The waitlist group received no interventions but was given random additional rewards. Both groups completed questionnaires on depression (Quick Inventory of Depressive Symptomatology) and social anxiety (Brief Liebowitz Social Anxiety Scale) at baseline, and again at 1, 2, and 3 months.

Results: The experimental group showed a significantly higher frequency of bell ringing compared to that of the waitlist group (standardized mean difference [SMD]=0.13), whereas no significant difference was observed between the groups in avatar customization frequency. As predicted, the experimental group showed a significantly greater reduction in depressive symptoms, with a small effect size observed (SMD=–0.12). However, no significant difference was determined between groups in social anxiety symptoms.

Conclusions: This study demonstrated that MMOGs, when combined with administrator-led interventions, can reduce users' depressive symptoms, albeit with a small effect size. Further studies are needed to test the intervention effects on social anxiety symptoms in MMOGs, with improved exposure scenarios.

Trial Registration: OSF Registries osf.io/wdb9u; https://osf.io/wdb9u

(JMIR Serious Games 2025;13:e73734) doi:10.2196/73734

KEYWORDS

massive multiplayer online games; behavioral activation; exposure therapy; depression; social anxiety

Introduction

Interventions using online games to improve the mental health symptoms of game users have been increasingly implemented in recent years [1]. These online games reportedly improve depression [2] and anxiety [3] symptoms. Although most of these online games have been developed to improve mental health symptoms, massive multiplayer online games (MMOGs) developed for leisure purposes [4] might have similar effects on depressive [5] and anxiety symptoms [6]. In fact, user networks within MMOGs represent sources of social support

XSL•FO
RenderX

[7], and gameplay experiences in MMOGs are associated with subjective well-being [8].

We especially focused on the Pigg Party (Figure 1), a well-known MMOG in Japan, which has been linked to extensive mental health research. For instance, previous studies have shown that bipolar disorder [9], depression [10], and social anxiety symptoms [11] can be estimated from user interactions in Pigg Party. These studies suggested that interaction logs in the Pigg Party contained indicators of mood and anxiety disorders. Moreover, online support within the Pigg Party had been reported to represent users' social resources [12], and these resources were suggested to reduce users' risk of mental illness [6]. Briefly, while user interactions in the Pigg Party may serve as indicators of mental illness, the support provided within the Pigg Party has the potential to mitigate those risks. Therefore, one randomized control trial in Pigg Party suggested that user interactions in the game may alleviate users' depressive symptoms [5]. However, the sample size for this trial was small, and the completion rate of the experimental task was not high. Therefore, this study will examine whether completing tasks that are easy to perform within the Pigg Party, with a sample size of over 1000 participants, contributes to the improvement of users' depressive and anxiety symptoms.

Figure 1. Players' behavior in the Pigg Party. Each player has an avatar and a private room, both of which can be customized to visit the private rooms of other players. Permission must be obtained from the owner in advance to visit a private room. Translation of the figure. (a) "I'm going to enter the room." (b) "Hi." (c) "What anime do you watch? Tell me what you are into." (d) "It's a bumper crop this season." (e) "I want to know that too."



This study used 2 theoretical frameworks. First is the behavioral activation of depressive symptoms [13]. Individuals with depressive symptoms have difficulty learning the association between their behaviors and the rewards linked to them [14], which makes it challenging for them to experience positive emotional states and facilitates the maintenance of these symptoms [15]. By increasing behaviors associated with rewards, they can learn reward-associated behaviors, experience positive emotional states, and improve depressive symptoms [16]. Meta-analysis of behavioral activation shows that activation of reward-associated behaviors improves depressive symptoms [17]. Activation of reward-associated behaviors—implemented through online games, such as webpage applications [18], short message services [19], and smartphone apps [20]—has demonstrated efficacy in improving depressive symptoms. Activation of reward-associated behaviors within online games also improved users' depressive symptoms [21]. Indeed, reciprocal reward relationships—such as mutually exchanging "likes" in an MMOG—can reduce depressive symptoms [5]. These studies suggest that the activation of reward-associated behaviors in MMOGs improves the depressive symptoms of game users [5].

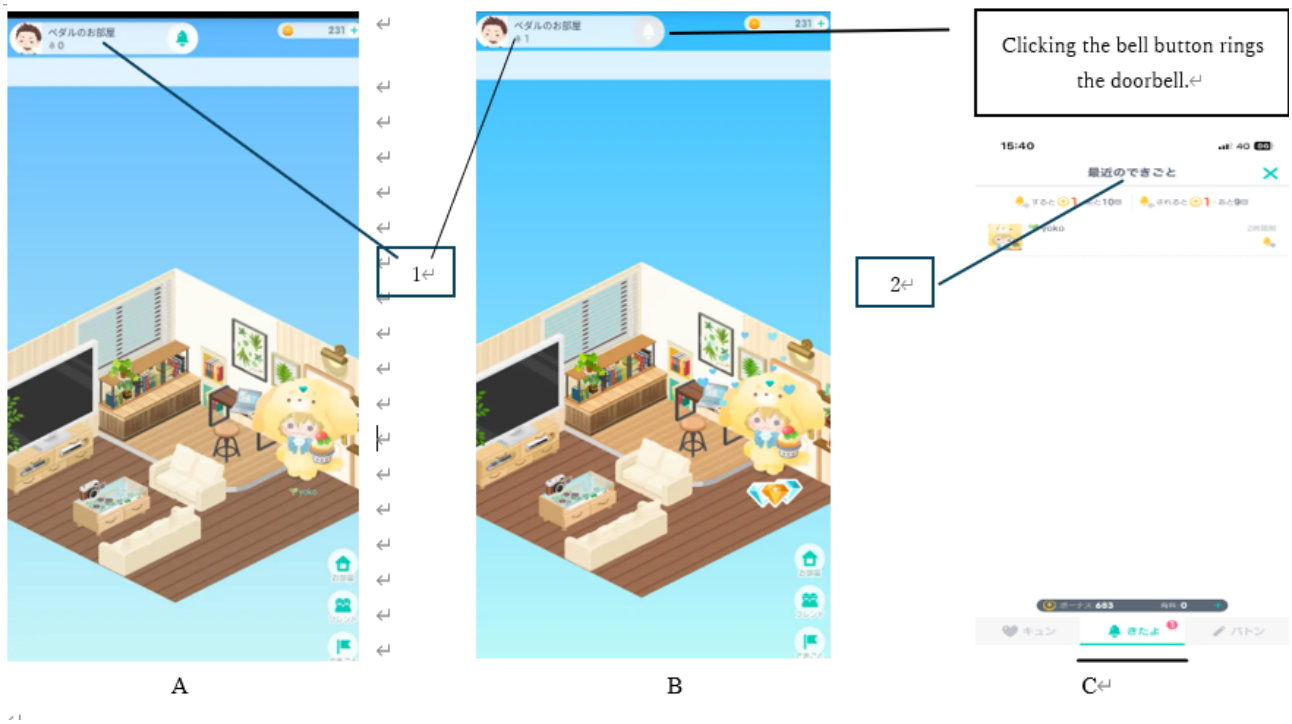
Online games also improve the anxiety symptoms of game users [3].

Inhibitory learning theory for anxiety symptoms serves as our second theoretical framework [22]. According to this theory, repeated relearning—where individuals who perceive specific stimuli (eg, being the focus of attention) as threatening reinterpret them as safe—reduces the perceived danger of those stimuli [23]. This theory is empirically supported by exposure therapy. Individuals with social anxiety symptoms tend to become anxious because they have learned to perceive social situations as dangerous stimuli [24]. By repeatedly experiencing social situations with little danger, they learn to perceive such situations as safe stimuli, which in turn improves their anxiety symptoms [25]. Their relearning of social situations has been confirmed in neuropsychological experiments [26], and the effectiveness of exposure therapy in improving social anxiety symptoms has been widely demonstrated [27]. Virtual reality exposure therapy, which constructs a virtual social community within online games and enables users to repeatedly experience social situations in the virtual community, has shown high intervention efficacy for social anxiety [28]. Virtual reality exposure therapy has been demonstrated to be effective for social anxiety symptoms in children [29], adolescents [30], and adults [31]; therefore, various virtual communities have been developed [32]. These findings suggest that repetitive experience of virtual social situations improves the social anxiety of game users [6].

XSL•FO
RenderX

In exposure therapy, it is essential to provide exposure scenarios in which participants experience fear [33]. Individuals with social anxiety disorder fear situations in which they are the focus of attention by others [34]; therefore, it is necessary to provide such attention-focused situations as the exposure scenarios. In a Pigg Party task where the player rings a bell in a friend's room, the player must first enter the room before ringing the bell. Upon entry, they become the focus of attention from people already in the room (Figure 1). Furthermore, when the bell is rung, the ringer is highlighted with an animation, drawing even more attention (Figure 2B). Thus, the bell-ringing task includes situations where the individual is noticed both upon entry and at ringing the bell, making it an appropriate exposure situation for individuals with social anxiety disorder. In fact, individuals with social anxiety disorder avoid attention even in the MMOG [11], and the situation where they receive attention from others is set as the exposure scenario [35].

Figure 2. Process of entering another user's room and ringing the doorbell on the Pigg Party. (A) One user visits another user's room, (B) One user rings the doorbell of another user's room. (C) The owner of the room can see who rang the doorbell. 1: "Pedal Room" (where "Pedal" refers to a person's name); 2: "Recent Events."



Based on these studies, we created an experimental group in which game users repetitively experienced social situations in an MMOG. Specifically, in Pigg Party, users in the experimental group were prompted to engage in the behavior of ringing a doorbell in another user's private room as the exposure scenarios (Figure 2). In addition, users were prompted to regularly customize their own avatars (Multimedia Appendix 1), enhancing self-identification with their avatars [6,36] and making social behavior in MMOG more analogous to social behavior in a physical community [37]. These prompts enabled the experimental group to receive virtual reality exposure therapy in the Pigg Party. Furthermore, by providing monetary rewards for ringing the doorbell, the behavior of ringing the doorbell becomes a reward-associated behavior in an MMOG [38]. In contrast, the waitlist group did not receive any intervention but did receive a random reward. As a result, the experimental group learned reward-associated behaviors, but the waitlist group did not. Thereby, only the experimental group received behavioral activation for reward-associated behaviors in the MMOG.

Our hypotheses are as follows: (1) according to behavioral activation on depressive symptoms [13,21], the experimental group would exhibit improvements in depressive symptoms compared to the waitlist group. (2) According to inhibitory learning theory on social anxiety symptoms [28,39], the experimental group would exhibit improvements in social anxiety, both within an MMOG and in the physical community, compared to that of the waitlist group. In addition, we examined whether Hypotheses 1 and 2 would be supported among individuals with depressive disorder and those with social anxiety disorder.

Methods

Design

This study involved a single-site, randomized, parallel-group comparative trial (Trial Registration: wdb9u). The study protocol was approved by the ethics committee of the National University A in Japan. The detailed protocol for this experiment has been disclosed previously [40].

Participants

The participants of this study were users who had been using the Pigg Party before November 15, 2023. Pigg Party is a mobile game that can be played on Android and iPhone devices, but its development environment has not been disclosed. On December 5, 2023, the administrator of the Pigg Party emailed

XSL•FO
RenderX

the users requesting their participation in the study. By December 15, 2023, 1331 users had expressed their willingness to participate, and then their eligibility for participation was assessed. As a result, 226 users were removed from the participants; among these, 175 users had withdrawn their consent to participate in the experiment, and after the recruitment period, 51 users requested to participate. Eligibility criteria required users to have played Pigg Party by November 15, 2023, and to understand Japanese. All users who completed the first questionnaire met both criteria.

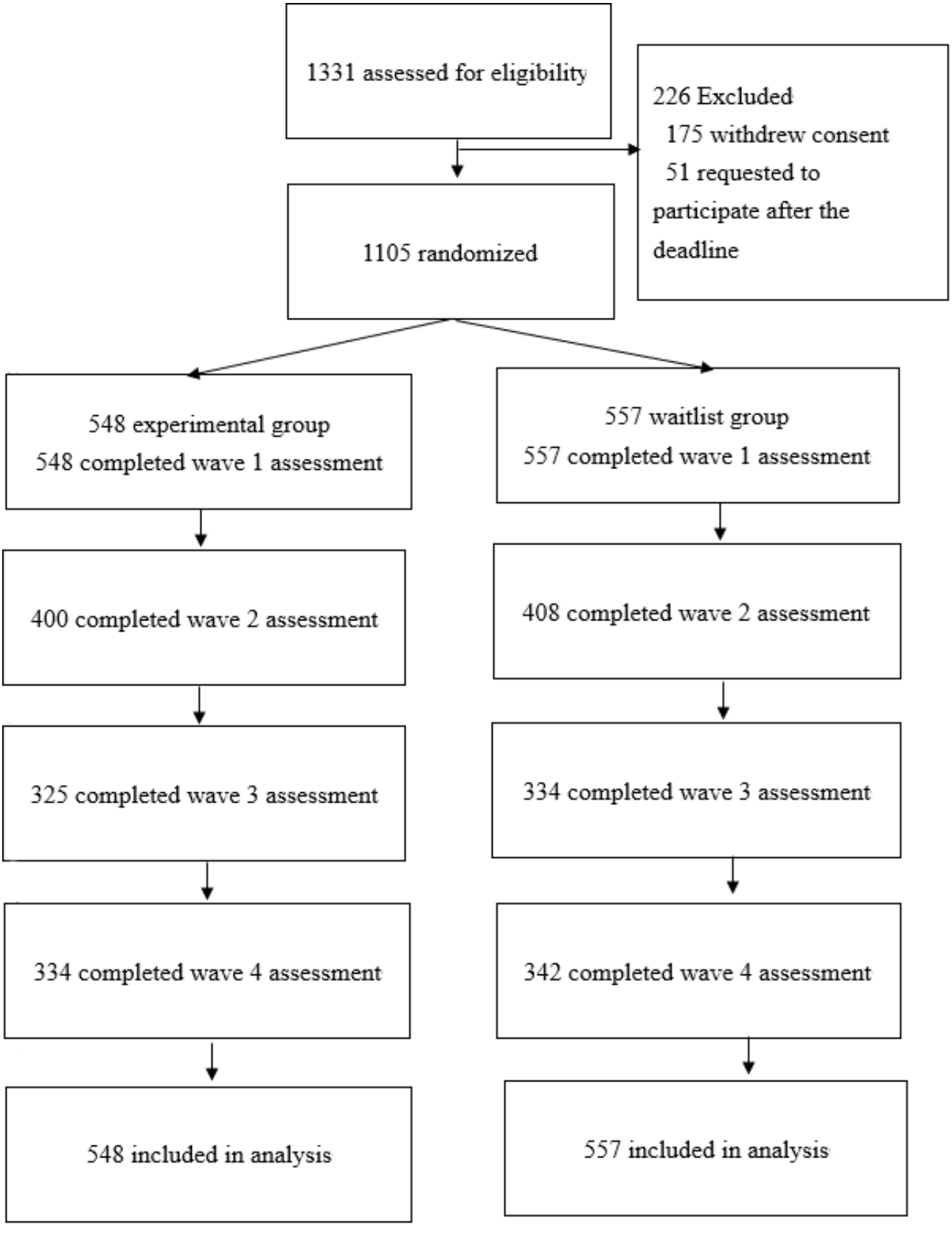
The final sample consisted of 1105 Pigg Party users. Participants' mean age was 23.68 (SD 10.44) years, ranging from 9 to 82 years. No exclusions were made based on participants' age or severity of psychiatric symptoms. Baseline social anxiety symptoms were assessed using the Japanese version of the Brief Liebowitz Social Anxiety Scale (Brief LSAS-J) [6,41]. Because the Brief LSAS-J includes 14 items, whereas the original LSAS contains 24, raw scores were multiplied by 24/14 to ensure comparability with the full LSAS metric [6,41]. Applying the established clinical cut-off of 30 points [42,43] to the adjusted scores indicated that 944 individuals (944/1105, 85.4 %) met the threshold for social anxiety; thus, the majority of the sample can be classified within the social anxiety disorder spectrum (Multimedia Appendix 2). To assess the depressive symptoms of participants, the Japanese version of the Quick Inventory of Depressive Symptomatology (QIDS-J) was used [44,45]. The QIDS-J is a 16-item questionnaire that uses a 4-point Likert scale, with higher scores indicating greater severity of depression. Applying the established clinical cut-off of 6 points [46,47] revealed that 677 individuals (677/1105, 61.2%) met the criterion for depression. Consequently, more than half of the present sample can be classified within the depressive disorder spectrum (Multimedia Appendix 3).

Randomization and Masking

A total of 1105 participants were randomly assigned to the experimental and waitlist groups in a 1:1 ratio (Figure 3). The Mersenne Twister method [44], which is implemented by default in R, was used for the random assignment. As the participants' Pigg Party tasks differed depending on the assigned group, assigned group masking was not implemented for the participants.

XSL•FO
RenderX

Figure 3. Participant flow diagram.



Procedures

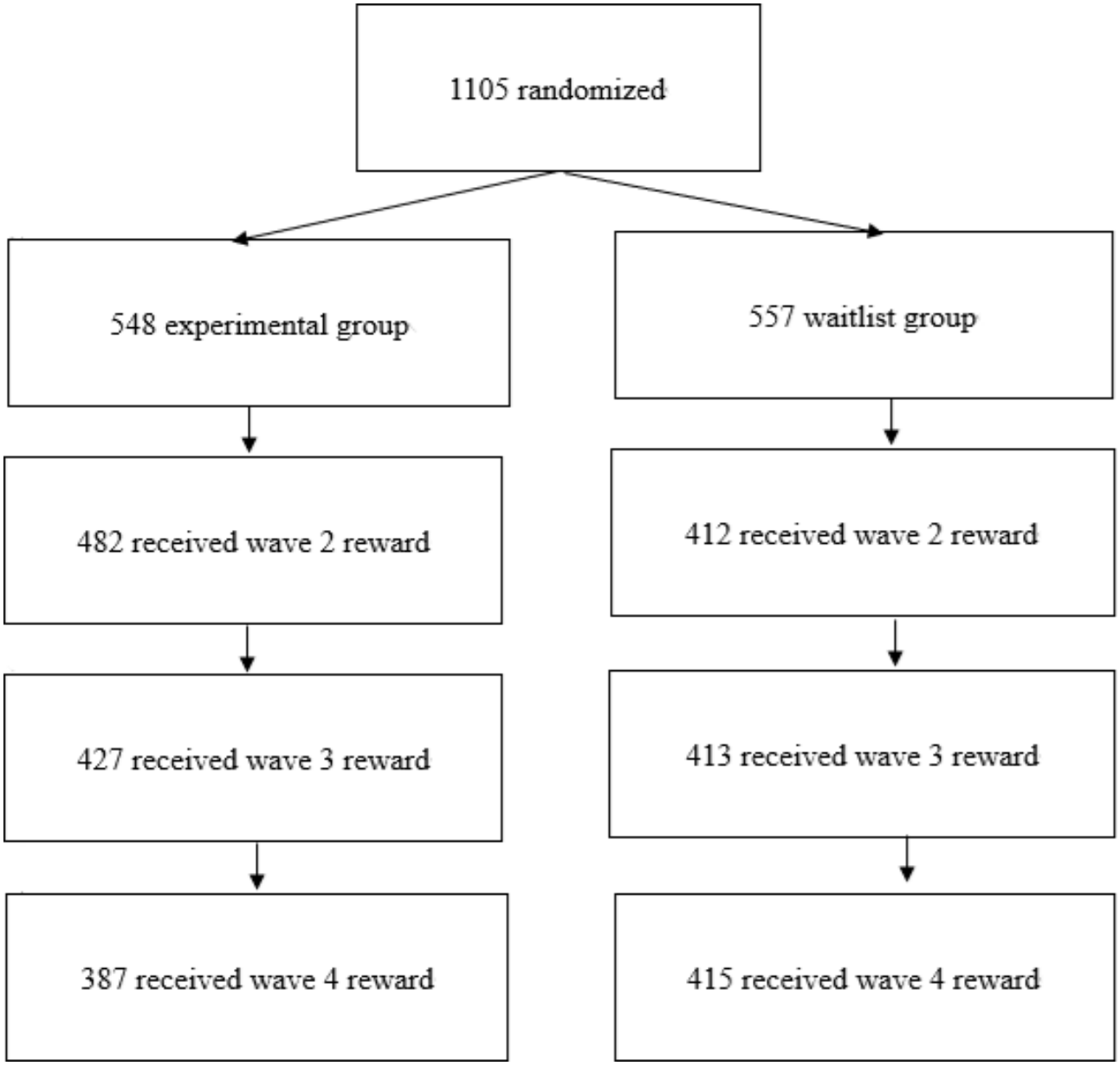
Participants completed questionnaires on demographic characteristics and outcome measures of mental health symptoms in the first wave (baseline, December 5 - 15, 2023). In the second wave (January 16 - 26, 2024), third wave (February 16 - 26, 2025), and fourth wave (March 16 - 26, 2025), they completed questionnaires solely on outcome measures of mental health symptoms. Participants who completed the questionnaires for the first and fourth waves

XSL•FO
RenderX

within the designated period were provided with an Amazon gift voucher worth JPY 500 (US $3.39) or a Google Play gift voucher, whereas those who completed the questionnaires for the second and third waves within the designated period were provided with an Amazon gift voucher worth JPY 200 (US $1.36) or a Google Play gift voucher. Each questionnaire was available for 10 days, and the intervention lasted 30 days, with these periods alternating (Multimedia Appendix 4).

The experimental group was prompted to visit another user's room once per day and ring the bell on weekdays (Monday through Friday, excluding public holidays; refer to Figure 2 for the method of ringing the bell) during the 3-month experimental period (from December 16, 2023, to March 15, 2024). In addition, on weekends (Saturday and Sunday) or public holidays, they were prompted to customize their avatar once per day (Multimedia Appendix 1 for the method of customization). Participants who accomplished these tasks on at least 18 days (18/30, 60%) out of 30 days were awarded an additional bonus in the form of an Amazon or Google Play gift voucher worth JPY 500 (US $3.39). As the experimental period spanned three months, gift vouchers worth a maximum of JPY 1500 (US $10.24) were provided as additional bonuses per user. Figure 4 shows the number of participants in the experimental group who received additional rewards.

Figure 4. The number of participants who received additional rewards in the experimental and waitlist groups. The experimental group received additional rewards for completing 60% or more of the tasks, whereas the waitlist group received additional rewards through a lottery.



In contrast, the waitlist group received no prompts during the experimental period. Within the waitlist group, 73% (412-415/557) of the users were randomly selected each month and were awarded an additional bonus of either Amazon or Google Play gift vouchers worth JPY 500 (US $3.39; Figure 4). One previous study on Pigg Party [5] reported that 73% of participants in the experimental group continued the intervention (sending "likes" to other users on at least 18 of 30 d). Therefore, we assumed a similar proportion would receive rewards in the experimental group in the present study and offered the same

XSL•FO
RenderX

proportion to the waitlist group. Since the experimental period lasted 3 months, gift vouchers worth a maximum of JPY 1500 (US $10.24) were provided as additional bonuses per user. Figure 4 illustrates the number of participants in the waitlist group who received additional rewards.

Outcome Measures

To assess depressive symptoms, the QIDS-J was used [45,46]. The QIDS-J is a 16-item questionnaire that uses a 4-point Likert scale, with higher scores indicating greater severity of depression. The Cronbach α coefficient for the QIDS-J at baseline was 0.86.

To assess social anxiety symptoms, the Brief LSAS-J was used [6,41]. The Brief LSAS-J can measure social anxiety symptoms in both virtual and physical communities. The term “virtual community” refers to the Pigg Party, whereas “physical community” denotes face-to-face interactions. Each version of the questionnaire comprises 14 items that assess anxiety symptoms and avoidance behaviors within the respective community. Details of the questionnaire are provided in Multimedia Appendix 5. Higher scores on the Brief LSAS-J indicate greater severity of social anxiety symptoms in both virtual and physical communities. The Cronbach α coefficients for the Brief LSAS-J at baseline in virtual and physical communities were 0.95 and 0.97, respectively.

Game Behavioral Logs for Manipulation Checks

To verify the frequency with which users rang the bell, it was measured monthly. Since the frequency of bell ringing is proportional to the number of friends and follows a power-law distribution (Multimedia Appendix 6), the value was calculated by adding one and then applying a base 2 logarithmic transformation.

Similarly, to verify the frequency with which users changed their avatars’ clothing, it was measured monthly. As with the frequency of bell ringing, avatar customization frequency followed a power-law distribution (Multimedia Appendix 7). The resulting frequencies were transformed by adding one and applying a base 2 logarithmic transformation.

Statistical Analysis

The final number of participants required was 1043 (based on an estimate for the latent growth model with Cohen *d*=0.10, α=.05, β=.90) [47,48]. In latent growth model sample size estimation, priority is given to the number of participants in each group at the final time point, rather than to the total number of measurement points; thus, the number of time points does not influence the required sample size or statistical power [49]. After recruiting 1105 participants, it was deemed that a sufficient sample size had been reached, and further recruitment was terminated.

Latent growth curve modeling (LGCM) was used in this study. The current structural equations of the LGCM accounted for the effect of gender, following a previous study [50]. The equations are as follows:

$$Y[it]=B0[i]+B1[i]F[t]+E2[it] \quad (1)$$

$$B0[i]=B00+B0exp*Gexp[i]+B0female*Gfemale[i]+E0[i] \quad (2)$$

$$B1[i]=B10+B1exp*Gexp[i]+B1female*Gfemale[i]+E1[i] \quad (3)$$

Where Y[it] represents the outcome measures (depressive symptoms, social anxiety symptoms) at time t for participant i. B0[i] and B1[i] are the intercept and slope contributing to Y[it] for participant i. E2[it] is the error contributing to Y[it] at time t for participant i . In line with previous studies [51], F[t] is obtained by logarithmically transforming the time of 4t weeks after adding 1. Specifically, for 4*t*=0, 4, 8, and 12 weeks, F[t] corresponds to 0.000, 1.000, 1.585, and 2.000, respectively. Improvement in psychiatric symptoms tends to follow a pattern in which there is a rapid improvement at the initial stage, followed by a gradually diminishing but sustained effect over time [52]. This pattern approximates the graph of a logarithmic function with a base>1 (ie, large initial change followed by a decelerated rate of change). Accordingly, many LGCMs for psychiatric symptoms use log-transformed time variables [53], and this study follows that convention.

The intercept B0[i] is influenced by B00, B0exp, B0female, and E0, which represent the intercept for the intercept, the intercept effect of the experimental group, the intercept effect of the female group, and the intercept effects of individuals, respectively. Gexp[i] is a group variable indicating whether participant i belongs to the experimental group (1 if yes, 0 if no). Similarly, Gfemale[i] is another group variable indicating the sex of the participant (1 for females, 0 for no females).

The slope B1[i] is influenced by B10, B1exp, B1female, and E1, which represent the intercept for the slope, slope effect of the experimental group, slope effect of the female group, and slope effect of individuals, respectively. B1exp represents the time-variant effects of the experimental group on the outcome measure Y[it]. Hence, this effect represents the intervention effects of the experimental group on the outcome measures. The intervention effects were analyzed using 95% CI and Wald tests. The Widely Applicable Bayesian Information Criterion (WAIC) and Leave-One-Out cross-validation (LOO) were used as indicators of the model’s overall fit.

To estimate the intervention’s effect size, the standardized mean difference (SMD) was calculated using the following formula, which is identical to Cohen *d* and widely used [54].

$$\frac{(Wave4exp-Wave1exp)-(Wave4wait-Wave1wait)}{\sqrt{\frac{(nexp-1)\sigma exp^2+(nwait-1)\sigma wait^2}{nexp+nwait-2}}} \quad (4)$$

Where Wave4exp and Wave1exp represent average outcome measure at waves 4 and 1 in the experimental group. Wave4wait and Wave1wait represent average outcome measure at waves 4 and 1 in the waitlist group. The nexp and nwait represent the number of participants in the experiment and waitlist groups, respectively. The σexp is the standard deviation of the difference between outcome measure waves 4 and 1 in the experimental group. Similarly, σwait is the SD of the difference between outcome measure waves 4 and 1 in the waitlist group.

In accordance with clinical trial guidelines [55], intention-to-treat analyses were performed. Full Information Maximum Likelihood (FIML) was used to handle missing data [56], as it is commonly used for estimating missing data in clinical trials [50]. During the FIML procedure, 0.0001 was added to the diagonal matrix to ensure that the covariance matrix remained positive. In addition, instead of an inverse matrix, a

XSL•FO
RenderX

pseudo-inverse matrix was used to manage errors when the inverse matrix did not exist. The initial values were set as the mean of each variable, and the optimization process was accelerated using the Limited-memory Broyden-Fletcher-Goldfarb-Shanno algorithm-bound optimization approach [57].

An independent-samples *t* test was conducted to compare the experimental and waitlist groups at baseline, with Cohen *d* reported as the effect size. Pearson's correlation coefficient was used for correlation analyses. A *P* value of <.05 was considered statistically significant.

Ethical Considerations

This study was approved by the ethics committee of the National University A in Japan as of December 1, 2023 (registration number 301). Informed consent was obtained from all the participants in this study. All procedures were conducted in accordance with the revised 1964 Helsinki Declaration and its later amendments, or comparable ethical standards. The participants' data were anonymized. Participants received either an Amazon or Google Play gift card as compensation for completing the questionnaires. The amounts were as follows in JPY (where JPY 150 is equivalent to around US $1): JPY 500 (US $3.39) for the first round, JPY 200 (US $1.36) for the second, JPY 200 (US $1.36) for the third, and JPY 500 (US $3.39) for the fourth. In addition, during the intervention phase, participants received a JPY 500 (US $3.39) bonus for each of the first 3 rounds. Thus, participants received a minimum of JPY 500 (US $3.39) and a maximum of JPY 2900 (US $19.80) for their participation in the experiment. No personally identifiable information appears in the data, figures, or tables of this study.

Results

Comparison of Demographic and Outcome Variables at Baseline Between the Experimental and Waitlist Groups

Before hypothesis testing, the demographic and outcome variables at the baseline of the experimental and waitlist groups were compared. In the experimental group, the proportion of females was significantly lower than that in the waitlist group (Table 1). The phi coefficient was 0.06, indicating a smaller-than-small effect (small=0.10); however, no significant differences were observed in other demographic variables, such as age, educational status, and employment status (Table 1). Further, no significant differences were found when comparing depressive, virtual social anxiety, and real-life social anxiety symptoms at baseline survey (Table 1). These results indicate that, aside from the proportion of females, the experimental and waitlist groups were evenly distributed.

XSL•FO
RenderX

Table . Comparison of demographic and outcome variables at baseline between the experimental and waitlist groups. .

	Experimental group (n=548)	Waitlist group (n=557)	Statistics				
			t test (*df*)	Chi-square (*df*)	*P* value	Cohen d	ϕ
Age (years), mean (SD)	24.151 (10.528)	23.226 (10.347)	–1.473 (1101.7)	—[a]	.141	0.089	—
Female, n (%)	394 (0.719)	432 (0.776)	—	4.394 (1)	.036[b]	—	0.063
Retired, n (%)	1 (0.002)	0 (0)	—	0 (1)	.993	—	0.000
Homemaker, n (%)	53 (0.097)	49 (0.088)	—	0.159 (1)	.691	—	0.012
Employed, n (%)	203 (0.37)	186 (0.334)	—	1.458 (1)	.227	—	0.036
On leave from studies or work, n (%)	22 (0.04)	26 (0.047)	—	0.148 (1)	.700	—	0.012
Student, n (%)	244 (0.445)	277 (0.497)	—	2.798 (1)	.094	—	0.050
Unpaid family worker, n (%)	29 (0.053)	24 (0.043)	—	0.389 (1)	.533	—	0.019
Preparing for university entrance or employment, n (%)	8 (0.015)	13 (0.023)	—	0.712 (1)	.399	—	0.025
Unemployed, n (%)	46 (0.084)	46 (0.083)	—	0 (1)	>.99	—	0.000
On long-term vacation, n (%)	15 (0.027)	17 (0.031)	—	0.018 (1)	.894	—	0.004
Not falling under any of the above categories, n (%)	12 (0.022)	22 (0.039)	—	2.309 (1)	.129	—	0.046
Outcome variables at baseline							
Depressive symptoms, mean (SD)	8.78 (6.06)	8.1 (5.99)	–1.903 (1103)	—	.057	0.114	—
Social anxiety symptoms in virtual, mean (SD)	28.95 (18.97)	28.39 (19.07)	–0.486 (1103)	—	.627	0.029	—
Social anxiety symptoms in physical, mean (SD)	46.81 (23.49)	45.66 (23.22)	–0.817 (1103)	—	.414	0.049	—

[a]Not applicable.

[b]$P<.05$.

Verification of the Experimental Manipulation

In addition, we confirmed that the manipulation applied to the experimental group was conducted appropriately. The experimental group exhibited a significantly higher number of bell-ringing behaviors than the waitlist group (Table 2). The effect size was 0.13, confirming a small effect (Table 2). Figure 5 also shows that although there was almost no difference in the number of bell-ringing behaviors during the first wave (baseline), the experimental group exhibited a higher number of bell-ringing behaviors during the second, third, and fourth waves. In contrast, no significant difference was observed in the frequency of avatar customization between the experimental and waitlist groups (Table 2). Because the prompt for ringing the bell occurred on weekdays (5/7 d) and the prompt for customization occurred on weekends (2/7 d), the frequency of prompts may have influenced the experimental manipulation. These results indicate that the prompt for ringing the bell on the MMOG in the experimental group was sufficiently implemented, whereas the prompt for customizing the avatar was not.

XSL•FO
RenderX

Table . Intention-to-treat analysis of latent growth curve modeling of online game behaviors in the overall sample. The Gelman-Rubin statistic (rhat) for all coefficients was 1.0.

	Intervention effect					Baseline (Wave1)	Wave 2	Wave 3	Wave 4	SMD[a]
	β	95% LL[b]	95% UL[c]	Wald	*P* value	mean (SD)	mean (SD)	mean (SD)	mean (SD)	
The number of bells rang (LOG2)[d]										
Experiment	0.195	0.031	0.36	2.321	0.020	8.07 (2.56)	7.93 (2.4)	7.29 (2.89)	6.65 (2.97)	0.13
Waitlist	__[e]	—	—	—	—	8.15 (2.42)	7.8 (2.4)	6.99 (3.01)	6.35 (3.17)	—
The number of avatar customizations (LOG2)[f]										
Experiment	0.069	-0.065	0.197	1.03	0.303	7.71 (1.67)	7.56 (1.92)	6.5 (2.12)	6.13 (2.29)	0.07
Waitlist	—	—	—	—	—	7.65 (1.7)	7.48 (1.9)	6.39 (2.05)	5.9 (2.37)	—

[a]SMD: standardized mean difference.

[b]LL: lower limit.

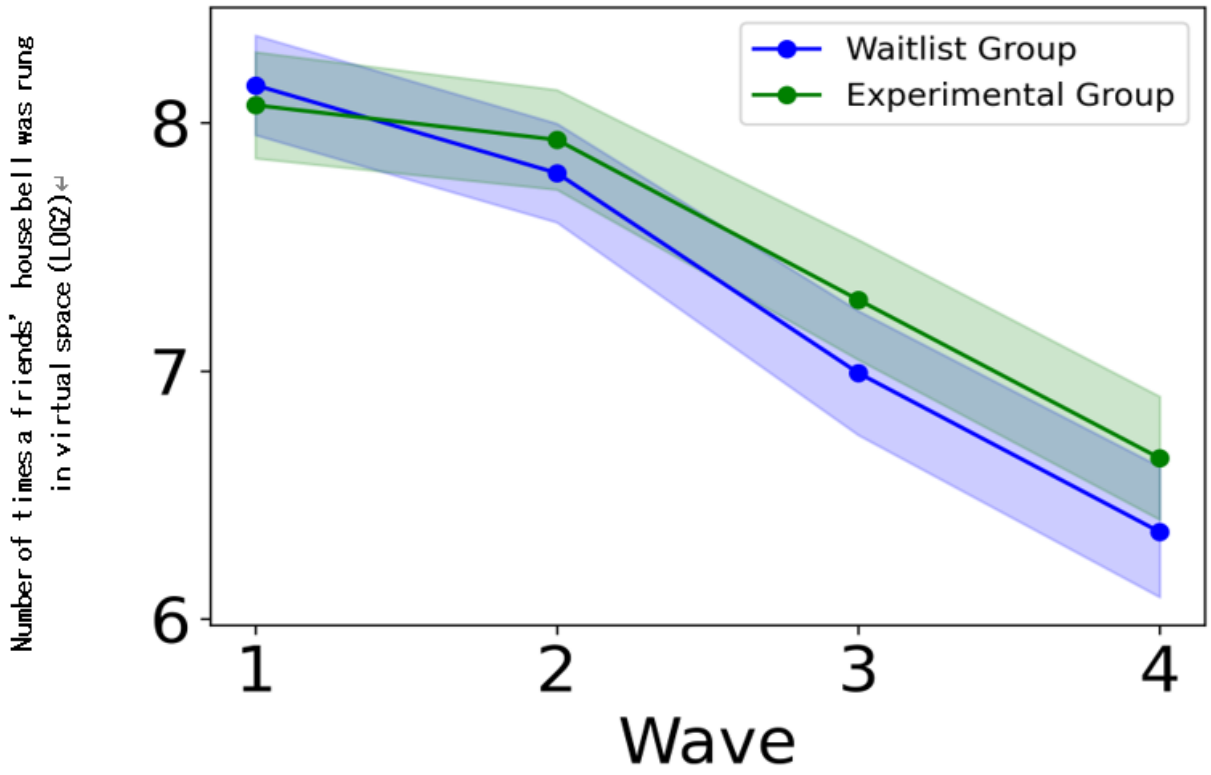
[c]UL: upper limit.

[d]Widely Applicable Bayesian Information Criterion (WAIC) -8376.6 leave-one-out (LOO) -8544.8.

[e]Not applicable.

[f]WAIC -7392.2, LOO -7523.4.

Figure 5. Comparison of the number of times a friend's house bell was rung in a virtual space between the experimental and waitlist groups (LOG2). The shaded area around the line represents the 95% confidence interval. As the y-axis values of 8, 7, and 6 are plotted on a base-2 logarithmic scale, they correspond to actual bell ring counts of 256, 128, and 64, respectively.



XSL•FO
RenderX

Examination of the Intervention Effects on Depressive Symptoms

We compared depressive symptoms between the experimental and waitlist groups. The experimental group exhibited a significantly greater improvement in depressive symptoms compared to the waitlist group (Table 3). The effect size was −0.12, indicating a small effect (Table 3). Figure 6 further illustrates that the experimental group demonstrated a consistent improvement in depressive symptoms. These findings support hypothesis 1.

Figure 6. Comparison of depressive symptoms between experimental and waitlist groups. The shaded area around the line represents the 95% CI.

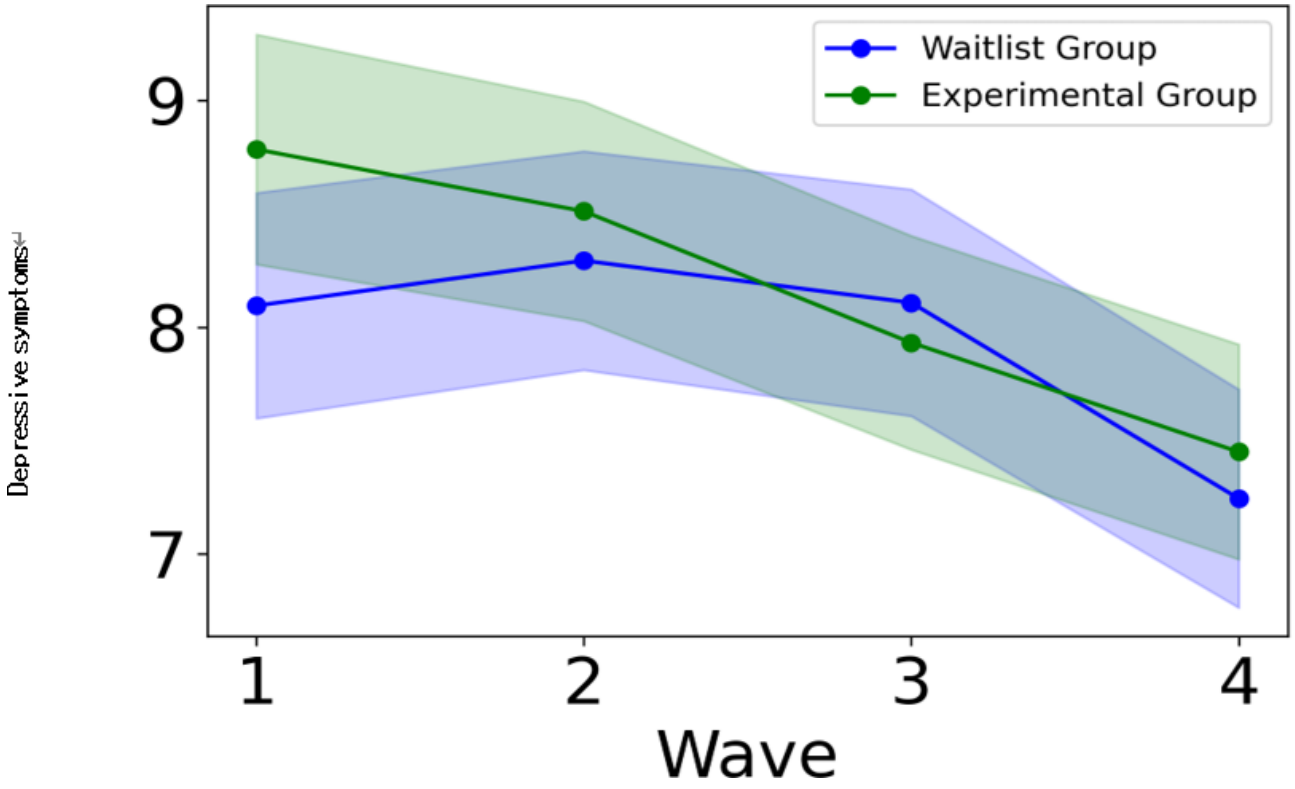


Table . Intention-to-treat analysis of latent growth curve modeling of depressive symptoms in the overall sample[a]. The Gelman–Rubin statistic (rhat) for all coefficients was 1.0.

Depressive symptoms	Intervention effect					Baseline (Wave 1), mean (SD)	Wave 2, mean (SD)	Wave 3, mean (SD)	Wave 4, mean (SD)	SMD[b]
	B	95% LL[c]	95% UL[d]	Wald	*P* value					
Experiment	–0.319	–0.543	–0.08	–2.703	.007[e]	8.78 (6.06)	8.51 (5.78)	7.93 (5.63)	7.45 (5.66)	–0.12
Waitlist	—[f]	—	—	—	—	8.1 (5.99)	8.29 (5.81)	8.11 (6.02)	7.24 (5.8)	—

[a]Widely Applicable Bayesian Information Criterion (WAIC) -10658.7, leave-one-out (LOO) -10801.2.

[b]SMD: Standardized Mean Difference.

[c]LL: lower limit.

[d]UL: upper limit.

[e]*P*<.01.

[f]Not applicable.

In addition, we tested hypothesis 1 among participants with depressive disorder only (n=677; QIDS-J ≥6). As presented in Table 4, when restricting the sample to individuals with depressive disorder, no significant difference in depressive symptoms was observed between the experimental and waitlist control groups. This finding suggests that, although behavioral activation delivered via an online game may reduce depressive symptoms in the overall sample, its disorder-specific effectiveness among individuals with depressive disorder may be limited.

XSL•FO
RenderX

Table . Intention-to-treat analysis of latent growth curve modeling of depressive symptoms in individuals with depressive disorder[a]. The Gelman–Rubin statistic (rhat) for all coefficients was 1.0. The number of participants in the experiment and waitlist groups was 346 and 331, respectively.

Depressive symptoms[a]	Intervention effect					Baseline (Wave 1), mean (SD)	Wave 2, mean (SD)	Wave 3, mean (SD)	Wave 4, mean (SD)	SMD[b]
	B	95% LL[c]	95% UL[d]	Wald	*P* value					
Experiment	–0.28	–0.61	0.04	–1.64	.10	12.27 (4.86)	11.31 (5.29)	10.53 (5.25)	9.86 (5.53)	–0.08
Waitlist	__[e]	—	—	—	—	11.90 (4.75)	11.09 (5.37)	10.96 (5.74)	9.83 (5.75)	—

[a]Widely Applicable Bayesian Information Criterion (WAIC) -6716.6, leave-one-out (LOO) -6799.8.

[b]SMD: standardized mean difference.

[c]LL: lower limit.

[d]UL: upper limit.

[e]Not applicable.

Examination of the Intervention Effects on Social Anxiety Symptoms

We compared social anxiety symptoms between the experimental and waitlist groups. No significant difference was observed between the experimental and waitlist groups regarding social anxiety in either virtual or physical communities (Table 5). Thus, these results did not support hypothesis 2.

Table . Intention-to-treat analysis of latent growth curve modeling of social anxiety symptoms in the overall sample. The Gelman–Rubin statistics (rhat) for all coefficients was 1.0.

	Intervention effect					Baseline (Wave 1), mean (SD)	Wave 2, mean (SD)	Wave 3, mean (SD)	Wave 4, mean (SD)	SMD[a]
	β	95% LL[b]	95% UL[c]	Wald	*P* value					
Social anxiety symptoms in virtual communities[d,e]										
Experiment	0.666	–0.109	1.48	1.649	.099	28.95 (18.97)	31.89 (17.85)	31.87 (17.64)	31.96 (18.4)	0.11
Waitlist	__[f]	—	—	—	—	28.39 (19.07)	31.12 (18.04)	30.79 (17.87)	29.79 (18.92)	—
Social anxiety symptoms in physical communities[g]										
Experiment	0.319	–0.534	1.095	0.771	.441	46.81 (23.49)	48.93 (21.21)	47.79 (20.41)	47.61 (20.53)	0.06
Waitlist	—	—	—	—	—	45.66 (23.22)	46.24 (21.3)	46.07 (20.51)	45.54 (21.16)	—

[a]SMD: Standardized Mean Difference.

[b]LL: lower limit.

[c]UL: upper limit.

[d]Widely Applicable Bayesian Information Criterion (WAIC) -15995.2, leave-one-out (LOO) -16147.0.

[e]LOO -16467.5.

[f]Not applicable.

[g]WAIC -16332.1.

We likewise tested hypothesis 2 among participants with social anxiety disorder only (n=944; Modified LSAS-J≥30). As presented in Table 6, even when restricting the sample to individuals with social anxiety disorder, no statistically significant difference in social anxiety symptoms was observed between the experimental and waitlist control groups. This result suggests that the effectiveness of online game–based exposure therapy may be limited, irrespective of the severity level of social anxiety symptoms.

XSL•FO
RenderX

Table . Intention-to-treat analysis of latent growth curve modeling of social anxiety symptoms in individuals with social anxiety disorders[a,b]. The Gelman–Rubin statistics (rhat) for all coefficients was 1.0. The number of participants in the experiment and waitlist groups was 470 and 474, respectively.

	Intervention effect					Baseline (Wave 1), mean (SD)	Wave 2, mean (SD)	Wave 3, mean (SD)	Wave 4, mean (SD)	SMD[c]
	B	95% LL[d]	95% UL[e]	Wald	*P* value					
Social anxiety symptoms in virtual communities[a]										
Experiment	0.82	–0.05	1.68	1.83	.07	32.17 (18.02)	34.85 (17.05)	34.80 (16.80)	34.93 (17.59)	0.13
Waitlist	—[f]	—	—	—	—	32.07 (17.87)	33.97 (16.92)	33.68 (16.75)	32.97 (18.00)	—
Social anxiety symptoms in physical communities[b]										
Experiment	0.27	–0.61	1.12	0.61	.54	53.37 (18.31)	53.75 (18.12)	52.58 (17.22)	52.35 (17.50)	0.04
Waitlist	—	—	—	—	—	52.38 (18.00)	51.36 (17.83)	50.84 (17.43)	50.73 (17.92)	—

[a]Widely Applicable Bayesian Information Criterion (WAIC) -13643.1, leave-one-out (LOO) -13778.3.

[b]WAIC -13775.6, LOO -13895.1.

[c]SMD: standardized mean difference.

[d] LL: lower limit.

[e]UL: upper limit.

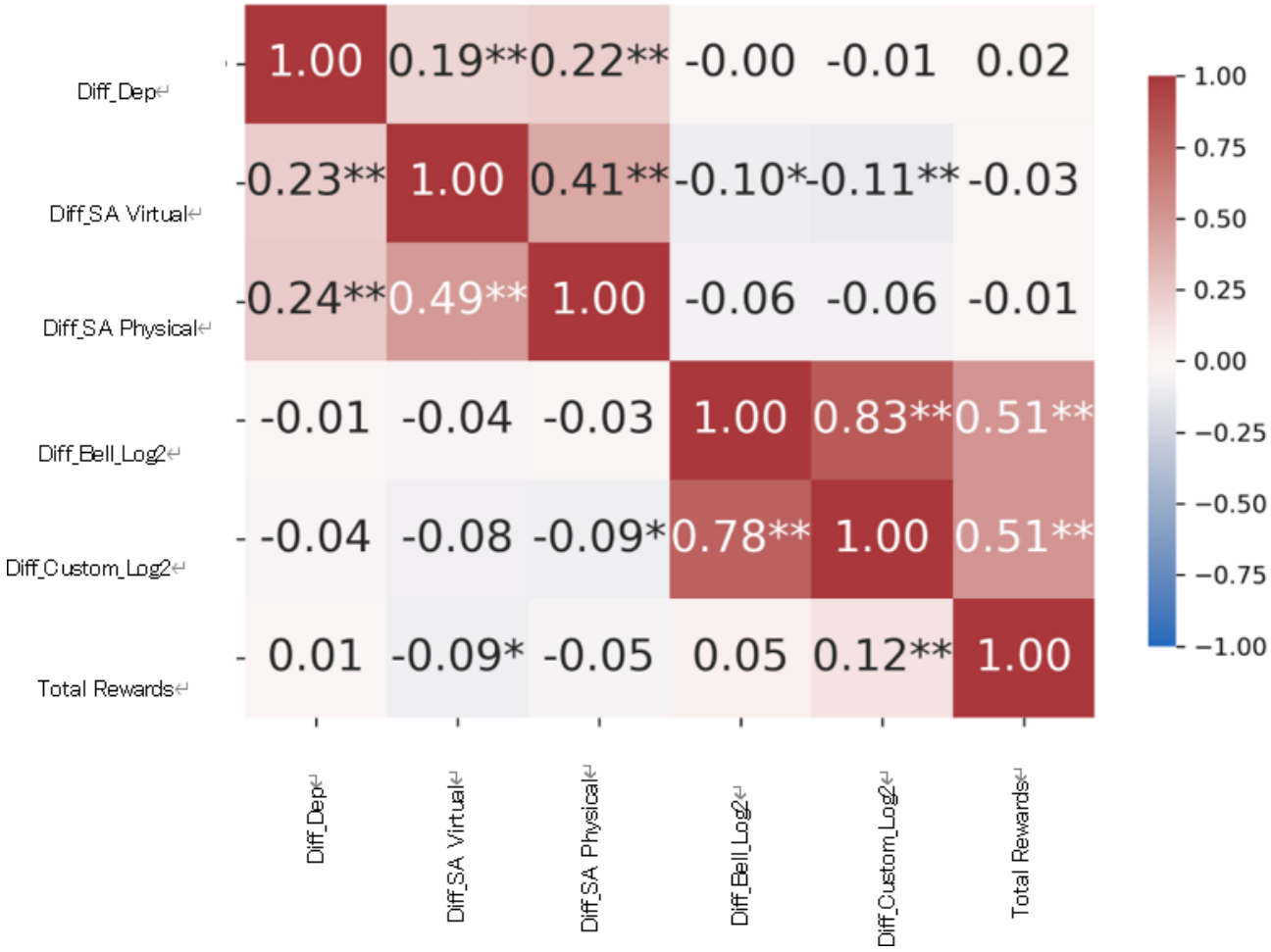
[f]Not applicable.

Association Between Online Game Behaviors and Depressive Symptoms

The association between online game behaviors and depressive symptoms was examined through correlation analysis. In Figure 7, Diff_Dep represents the value obtained by subtracting depressive symptoms at wave 1 from those at wave 4. That is, a more negative value indicates improvement during the intervention period. Similarly, Diff_bell_countLOG2 is the difference in bell-ringing frequency (log2-transformed) between wave 4 and wave 1, and Diff_customl_countLOG2 is the difference in avatar customization frequency (log2-transformed) between the same waves. In addition, total reward indicates the number of bonus rewards received, ranging from 0 to 3.

XSL•FO
RenderX

Figure 7. Correlation matrix among depressive symptoms, social anxiety, and online game behaviors. The upper half is the correlation coefficient for the experimental group, and the lower half is the correlation coefficient for the waitlist group. Diff indicates the change score calculated by subtracting the baseline value from the follow-up value; Dep denotes depression; SA denotes social anxiety; Bell indicates the number of times the bell was rung; Custom refers to the number of times the avatar's clothing was customized; and Log2 denotes that the data were log-transformed using base 2.*P<.05, **P<.01.



The upper and lower halves of Figure 7 showed correlation matrices for the experimental and waitlist groups, respectively. No significant correlations were observed between online game behaviors and depressive symptoms in either group.

Association Between Online Game Behaviors and Social Anxiety Symptoms

The association between online game behaviors and social anxiety symptoms was examined through correlation analysis. In Figure 7, Diff_SA_V represents the difference in virtual-community social anxiety symptoms in the Pigg Party between waves 4 and 1. Similarly, Diff_SA_P represents the difference in physical-community social anxiety symptoms between those waves. A more negative value indicates greater improvement during the intervention.

In the experimental group, a higher frequency of bell ringing was associated with a reduction in social anxiety symptoms (r=–0.10). Likewise, a higher frequency of avatar customization was associated with a reduction in virtual social anxiety symptoms (r=–0.11). Even in the waitlist group, a higher frequency of avatar customization was associated with a reduction in physical-community social anxiety symptoms (r=–0.09). Higher frequencies of receiving additional bonuses were also associated with greater improvement in virtual social anxiety symptoms (r=–0.09). However, the effect sizes of these correlation coefficients were all below the threshold for a small effect (0.2), and consistent results were not observed between the experimental and waitlist groups; therefore, these findings should be interpreted with caution.

Discussion

Principal Results

This study investigated whether mental health symptoms can be improved through interventions on MMOGs by conducting a clinical trial with over 1000 participants. Consistent with previous findings [2,5], our results confirm that depressive symptoms can be alleviated through interventions on MMOGs. Given the large user base of MMOGs, such interventions may significantly contribute to reducing depressive symptoms on a broad scale [1]. Additionally, since the intervention was integrated as part of the in-game tasks, users could potentially experience an improvement in depressive symptoms while enjoying the game. Previous studies have used specific applications to improve depressive symptoms, which imposed a certain burden on users in terms of the intervention [18-21]. In contrast, since the participants in this study engaged with an

XSL•FO
RenderX

MMOG interface they were already familiar with, the burden on users was minimal. This suggests that MMOG-based interventions could serve as an effective and accessible approach to alleviate depressive symptoms without imposing additional burden [11].

Although depressive symptoms in the experimental group improved significantly, no significant correlation was observed between bell-ringing behaviors and depressive symptoms, unlike the results reported in face-to-face behavioral activation studies [13-16]. This finding implies that bell-ringing behaviors may act merely as triggers, whereas other in-game activities—such as mutual visits or increased time spent chatting—could have contributed to the improvement of depressive symptoms [5].

Furthermore, contrary to previous studies [24-27,39], this study did not find evidence supporting an improvement in social anxiety symptoms. One possible explanation is the insufficient frequency of avatar customization in this study. Research suggests that if users do not customize their avatars frequently enough, they may struggle to identify with their avatars [36]. When users fail to fully customize their avatars, the discrepancy between in-game behaviors and real-life behaviors may increase [58], potentially explaining why the experimental group did not experience significant reductions in social anxiety symptoms [37]. Therefore, the participants in the experimental group may also have found it difficult to identify with their avatars [6], which consequently made it challenging to achieve the full effects of exposure therapy [28-30]. Moreover, in the experimental group, some participants may have intentionally rung bells when no one was present to avoid being noticed by others. In such cases, exposure to attention-focused situations may not have been effectively implemented. It is necessary to provide exposure situations that are difficult to avoid.

Future Work

Although this study did not identify specific in-game behaviors related to depressive symptoms, future studies could let users self-select enjoyable in-game activities, followed by an examination of whether such activities contribute to improvement in depressive symptoms [5]. Identifying these specific in-game behaviors is crucial for implementing behavioral activation in MMOGs [17]. Although the effect size was small, greater avatar customization was associated with reduced social anxiety in both MMOGs and physical communities, suggesting that promoting avatar customization may still offer benefits. To uniformly increase avatar customization frequency, it may be effective to limit interventions to new users, for whom previous customization frequency is zero. This condition may enhance the observable intervention effects. Future studies should explicitly test whether increasing avatar customization frequency reduces social anxiety symptoms.

Limitations

This study had 7 limitations. First, despite the random assignment of participants, the experimental group had a significantly lower proportion of females compared to the waitlist group. The gender imbalance may have influenced the results. Second, the platform was limited to Pigg Party [59]. Given that immersion is a key factor in exposure therapy [32], further research is needed to determine whether the finding can be replicated in MMOGs with a virtual reality headset. Third, it was suggested that randomly distributed rewards contributed to a reduction in virtual social anxiety symptoms in the waitlist group only. Moreover, this effect was insufficiently controlled. Fourth, it is also possible that the generally low baseline scores for depressive symptoms led to insufficient observable intervention effects. Fifth, because this study did not include indicators beyond symptom severity—such as quality of life—clinical implications cannot be fully inferred. Sixth, associations between specific gaming behaviors and depressive symptoms were not confirmed, and the mechanism for improvement of depressive symptoms in MMOGs could not be explained. Seventh, this study included participants across a broad age range, which precluded classification into a single age cohort. Future studies should be conducted with narrower and clearly defined age groups.

Conclusions

This study, with a randomized controlled trial with over 1000 participants, highlights the potential of MMOG-based interventions to alleviate depressive symptoms. Notably, the MMOG used in this study was developed for leisure purposes rather than for therapeutic purposes [59]. This suggests that users experienced improvements in depressive symptoms while enjoying a leisure activity [4]. Using MMOGs to improve mental health may help prevent depressive symptoms in various individuals by providing an accessible interface through online games [3,11].

Acknowledgments

This study was funded by a grant from CyberAgent, Inc. (akblab-0005) and KAKENHI (23K22365). Funding statement Kenji Yokotani received funding from CyberAgent, Inc (grant akblab-0005). CyberAgent, Inc, primarily provided financial support for the present study and imposed no restrictions on the publication of its academic findings. Furthermore, CyberAgent, Inc, does not impose constraints on the experimental design unless such constraints would result in substantial disadvantages for Pigg Party users. ChatGPT-4o was used for translating Japanese text into English. However, the translated text has been substantially revised and supplemented by the authors and an English editor after thorough content verification.

XSL•FO
RenderX

Data Availability

The data cannot be shared publicly, even in anonymized form, due to the privacy policy of CyberAgent, Inc., which prohibits the disclosure of users' personal information. CyberAgent, Inc, provides access to the data only to researchers and only within the scope of the informed consent obtained from participants.

Authors' Contributions

KY: Conceptualization, Methodology, Formal analysis, Writing – original draft preparation, Writing – review and editing, Funding acquisition, Project administration
MT: Conceptualization, Data Curation, Investigation, Funding acquisition
NA: Conceptualization, Writing – review and editing
TK: Conceptualization
All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

KY was funded by CyberAgent, Inc. MT was an employee of CyberAgent, Inc. CyberAgent, Inc, developed and operates Pigg Party.

Multimedia Appendix 1
Movie showing example of an avatar changing clothes.
[MP4 File, 51 KB - games_v13i1e73734_app1.mp4]

Multimedia Appendix 2
Percentage of participants with social anxiety disorder at baseline.
[DOCX File, 45 KB - games_v13i1e73734_app2.docx]

Multimedia Appendix 3
Percentage of participants with depressive disorder at baseline.
[DOCX File, 37 KB - games_v13i1e73734_app3.docx]

Multimedia Appendix 4
Timetable of Questionnaire Response Period and Intervention Period.
[DOCX File, 22 KB - games_v13i1e73734_app4.docx]

Multimedia Appendix 5
Brief Liebowitz Social Anxiety Scale in the Japanese version.
[DOCX File, 21 KB - games_v13i1e73734_app5.docx]

Multimedia Appendix 6
Frequency distribution of bell ringing events at baseline: comparison between original and log-transformed data.
[DOCX File, 67 KB - games_v13i1e73734_app6.docx]

Multimedia Appendix 7
Frequency distribution of avatar customization at baseline: comparison between original and log-transformed data.
[DOCX File, 66 KB - games_v13i1e73734_app7.docx]

Checklist 1
CONSORT checklist.
[PDF File, 1295 KB - games_v13i1e73734_app8.pdf]

References

1. Lau CKY, Saad A, Camara B, Rahman D, Bolea-Alamanac B. Acceptability of digital mental health interventions for depression and anxiety: Systematic review. J Med Internet Res 2024 Oct 28;26(1):e52609. [doi: 10.2196/52609] [Medline: 39466300]
2. Six SG, Byrne KA, Tibbett TP, Pericot-Valverde I. Examining the effectiveness of gamification in mental health apps for depression: Systematic review and meta-analysis. JMIR Ment Health 2021 Nov 29;8(11):e32199. [doi: 10.2196/32199] [Medline: 34847058]

XSL•FO
RenderX

3. Cheng C, Ebrahimi OV. A meta-analytic review of gamified interventions in mental health enhancement. Comput Human Behav 2023 Apr;141:107621. [doi: 10.1016/j.chb.2022.107621]
4. Raith L, Bignill J, Stavropoulos V, et al. Massively multiplayer online games and well-being: A systematic literature review. Front Psychol 2021 Jun 30;12. [doi: 10.3389/fpsyg.2021.698799]
5. Yokotani K, Takano M, Abe N. Can likes returned by peers within a day improve users' depressive/manic levels in a massive multiplayer online game? A randomized controlled trial. J Comput Soc Sc 2024 Dec;7(3):2333-2357. [doi: 10.1007/s42001-024-00312-4]
6. Yokotani K, Takano M, Abe N, Kato TA. Improving social anxiety in lesbian, gay, bisexual, transgender, questioning, intersex and asexual individuals through avatar customization and communication. Asian J of Social Psycho 2025 Mar;28(1):e12666. [doi: 10.1111/ajsp.12666]
7. Prochnow T, Patterson MS. It's not just a game: social networks, isolation and mental health in online gamers. JPMH 2024 Jun 5;23(2):140-153. [doi: 10.1108/JPMH-11-2023-0099]
8. Anto A, Basu A, Selim R, Foscht T, Eisingerich AB. Open-world games' affordance of cognitive escapism, relaxation, and mental well-being among postgraduate students: Mixed methods study. J Med Internet Res 2024 Dec 17;26(1):e63760. [doi: 10.2196/63760] [Medline: 39689301]
9. Yokotani K, Takano M, Abe N. Abnormal behavior of following peers in an online game indicates bipolar disorder and manic/hypomanic episodes. 2023 Nov 6 Presented at: Proceedings of the 2023 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining; Kusadasi, Turkiye p. 464-469. [doi: 10.1145/3625007.3627319]
10. Yokotani K, Takano M. Avatars' social rhythms in online games indicate their players' depression. Cyberpsychol Behav Soc Netw 2022 Nov 1;25(11):718-732. [doi: 10.1089/cyber.2022.0058]
11. Yokotani K, Takano M, Abe N, Kato TA. Predicting social anxiety disorder based on communication logs and social network data from a massively multiplayer online game: Using a graph neural network. Psychiatry Clin Neurosci 2025 May;79(5):274-281. [doi: 10.1111/pcn.13804] [Medline: 40007197]
12. Takano M, Yokotani K. Online social support via avatar communication buffers harmful effects of offline bullying victimization. 2022 Presented at: Proceedings of the International AAAI Conference on Web and Social Media p. 980-992. [doi: 10.1609/icwsm.v16i1.19351]
13. Dimidjian S, Barrera M, Martell C, Muñoz RF, Lewinsohn PM. The origins and current status of behavioral activation treatments for depression. Annu Rev Clin Psychol 2011;7(1):1-38. [doi: 10.1146/annurev-clinpsy-032210-104535] [Medline: 21275642]
14. Kumar P, Goer F, Murray L, et al. Impaired reward prediction error encoding and striatal-midbrain connectivity in depression. Neuropsychopharmacol 2018 Jun;43(7):1581-1588. [doi: 10.1038/s41386-018-0032-x]
15. Jacobson NS, Martell CR, Dimidjian S. Behavioral activation treatment for depression: Returning to contextual roots. Clinical Psychology: Science and Practice 2001;8(3):255-270. [doi: 10.1093/clipsy.8.3.255]
16. McCauley E, Gudmundsen G, Schloredt K, et al. The Adolescent Behavioral Activation Program: Adapting Behavioral Activation as a Treatment for Depression in Adolescence. J Clin Child Adolesc Psychol 2016;45(3):291-304. [doi: 10.1080/15374416.2014.979933] [Medline: 25602170]
17. Cuijpers P, van Straten A, Warmerdam L. Behavioral activation treatments of depression: a meta-analysis. Clin Psychol Rev 2007 Apr;27(3):318-326. [doi: 10.1016/j.cpr.2006.11.001] [Medline: 17184887]
18. Weitzel EC, Pabst A, Luppa M, et al. Are self-managed online interventions for depression effective in improving behavioral activation? A secondary analysis of a cluster-randomized controlled trial. J Affect Disord 2022 Jul 1;308:413-420. [doi: 10.1016/j.jad.2022.04.090] [Medline: 35460734]
19. Santopetro N, Jones D, Garron A, Meyer A, Joyner K, Hajcak G. Examining a fully automated mobile-based behavioral activation intervention in depression: Randomized controlled trial. JMIR Ment Health 2024 Aug 30;11(1):e54252. [doi: 10.2196/54252] [Medline: 39212598]
20. Vanderkruik RC, Ferguson C, Kobylski LA, et al. Testing a behavioral activation gaming app for depression during pregnancy: Multimethod pilot study. JMIR Form Res 2024 Jan 26;8(1):e44029. [doi: 10.2196/44029] [Medline: 38277191]
21. Paul M, Bullock K, Bailenson J, Burns D. Examining the efficacy of extended reality-enhanced behavioral activation for adults with major depressive disorder: Randomized controlled trial. JMIR Ment Health 2024 Apr 15;11(1):e52326. [doi: 10.2196/52326] [Medline: 38437873]
22. Craske MG, Kircanski K, Zelikowsky M, Mystkowski J, Chowdhury N, Baker A. Optimizing inhibitory learning during exposure therapy. Behav Res Ther 2008 Jan;46(1):5-27. [doi: 10.1016/j.brat.2007.10.003] [Medline: 18005936]
23. Weisman JS, Rodebaugh TL. Exposure therapy augmentation: A review and extension of techniques informed by an inhibitory learning approach. Clin Psychol Rev 2018 Feb;59:41-51. [doi: 10.1016/j.cpr.2017.10.010] [Medline: 29128146]
24. Bishop SJ. Neurocognitive mechanisms of anxiety: an integrative account. Trends Cogn Sci (Regul Ed) 2007 Jul;11(7):307-316. [doi: 10.1016/j.tics.2007.05.008]
25. McNally RJ. Mechanisms of exposure therapy: how neuroscience can improve psychological treatments for anxiety disorders. Clin Psychol Rev 2007 Jul;27(6):750-759. [doi: 10.1016/j.cpr.2007.01.003] [Medline: 17292521]
26. Hur JW, Shin H, Jung D, et al. Virtual reality-based psychotherapy in social anxiety disorder: fMRI study using a self-referential task. JMIR Ment Health 2021 Apr 14;8(4):e25731. [doi: 10.2196/25731] [Medline: 33851931]

XSL•FO
RenderX

27. Powers MB, Emmelkamp PMG. Virtual reality exposure therapy for anxiety disorders: A meta-analysis. J Anxiety Disord 2008;22(3):561-569. [doi: 10.1016/j.janxdis.2007.04.006] [Medline: 17544252]
28. Horigome T, Kurokawa S, Sawada K, et al. Virtual reality exposure therapy for social anxiety disorder: a systematic review and meta-analysis. Psychol Med 2020 Nov;50(15):2487-2497. [doi: 10.1017/S0033291720003785] [Medline: 33070784]
29. Beidel DC, Tuerk PW, Spitalnick J, Bowers CA, Morrison K. Treating childhood social anxiety disorder with virtual environments and serious games: A randomized trial. Behav Ther 2021 Nov;52(6):1351-1363. [doi: 10.1016/j.beth.2021.03.003] [Medline: 34656191]
30. Kahlon S, Lindner P, Nordgreen T. Gamified virtual reality exposure therapy for adolescents with public speaking anxiety: a four-armed randomized controlled trial. Front Virtual Real 2023 Aug 29;4. [doi: 10.3389/frvir.2023.1240778]
31. Reeves R, Elliott A, Curran D, Dyer K, Hanna D. 360° Video virtual reality exposure therapy for public speaking anxiety: A randomized controlled trial. J Anxiety Disord 2021 Oct;83:102451. [doi: 10.1016/j.janxdis.2021.102451] [Medline: 34304028]
32. Chard I, van Zalk N. Virtual reality exposure therapy for treating social anxiety: A scoping review of treatment designs and adaptation to stuttering. Front Digit Health 2022;4. [doi: 10.3389/fdgth.2022.842460]
33. Foa EB, Kozak MJ. Emotional processing of fear: Exposure to corrective information. Psychol Bull 1986;99(1):20-35. [doi: 10.1037/0033-2909.99.1.20]
34. American Psychiatric Association. American Psychiatric Association Diagnostic and Statistical Manual of Mental Disorders: DSM-5-TR, 5th edition: Amer Psychiatric Pub Inc; 2022.
35. Bouchard S, Dumoulin S, Robillard G, et al. Virtual reality compared with in vivo exposure in the treatment of social anxiety disorder: a three-arm randomised controlled trial. Br J Psychiatry 2017 Apr;210(4):276-283. [doi: 10.1192/bjp.bp.116.184234] [Medline: 27979818]
36. Takano M, Taka F. Fancy avatar identification and behaviors in the virtual world: Preceding avatar customization and succeeding communication. Computers in Human Behavior Reports 2022 May;6:100176. [doi: 10.1016/j.chbr.2022.100176]
37. Dechant MJ, Birk MV, Shiban Y, Schnell K, Mandryk RL. How avatar customization affects fear in a game-based digital exposure task for social anxiety. Proc ACM Hum-Comput Interact 2021 Oct 5;5(CHI PLAY):1-27. [doi: 10.1145/3474675]
38. Hopko DR, Lejuez CW, LePage JP, Hopko SD, McNeil DW. A brief behavioral activation treatment for depression. A randomized pilot trial within an inpatient psychiatric hospital. Behav Modif 2003 Sep;27(4):458-469. [doi: 10.1177/0145445503255489] [Medline: 12971122]
39. Craske MG, Treanor M, Conway CC, Zbozinek T, Vervliet B. Maximizing exposure therapy: an inhibitory learning approach. Behav Res Ther 2014 Jul;58:10-23. [doi: 10.1016/j.brat.2014.04.006] [Medline: 24864005]
40. Yokotani K, Takano M, Abe N, Kato T. Effects of avatar interaction in massively multiplayer online games on users' social anxiety symptoms: A double-blind, randomized controlled trial. OSF. 2023 Dec 4. [doi: 10.17605/OSF.IO/WDB9U]
41. Olivares J, Sánchez-García R, López-Pina JA. The Liebowitz social anxiety scale for children and adolescents. Psicothema Colegio Oficial De Psicologos Del Principado De Asturias 2009;21(3):486-491.
42. Mennin DS, Fresco DM, Heimberg RG, Schneier FR, Davies SO, Liebowitz MR. Screening for social anxiety disorder in the clinical setting: using the Liebowitz Social Anxiety Scale. J Anxiety Disord 2002;16(6):661-673. [doi: 10.1016/s0887-6185(02)00134-2] [Medline: 12405524]
43. Nanamori M, Tomita N, Kametani C, Matsuda N, Kumano H. Triggers of self-focused attention: an ecological momentary assessment study. Biopsychosoc Med 2023 Apr 23;17(1):16. [doi: 10.1186/s13030-023-00273-6] [Medline: 37087445]
44. Matsumoto M, Nishimura T. Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. ACM Trans Model Comput Simul 1998 Jan 1;8:3-30. [doi: 10.1145/272991.272995]
45. Fujisawa. Reliability and validity of Quick Inventory of Depressive Symptomatology in Japanese version. 2008 May. URL: https://cir.nii.ac.jp/crid/1570854176020420992 [accessed 2021-02-21]
46. Reilly TJ, MacGillivray SA, Reid IC, Cameron IM. Psychometric properties of the 16-item Quick Inventory of Depressive Symptomatology: a systematic review and meta-analysis. J Psychiatr Res 2015 Jan;60:132-140. [doi: 10.1016/j.jpsychires.2014.09.008] [Medline: 25300442]
47. Cohen J. Statistical Power Analysis for the Behavioral Sciences, 2nd edition: Routledge; 1988. [doi: 10.4324/9780203771587ISBN:978-0-203-77158-7]
48. Faul F, Erdfelder E, Buchner A, Lang AG. Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. Behav Res Methods 2009 Nov;41(4):1149-1160. [doi: 10.3758/BRM.41.4.1149] [Medline: 19897823]
49. Feingold A. Effect sizes for growth-modeling analysis for controlled clinical trials in the same metric as for classical analysis. Psychol Methods 2009;14(1):43-53. [doi: 10.1037/a0014699]
50. Champion KE, Newton NC, Gardner LA, et al. Health4Life eHealth intervention to modify multiple lifestyle risk behaviours among adolescent students in Australia: a cluster-randomised controlled trial. Lancet Digit Health 2023 May;5(5):e276-e287. [doi: 10.1016/S2589-7500(23)00028-6] [Medline: 37032200]
51. Hancock GR, Mueller RO. Structural Equation Modeling: A Second Course: IAP; 2006.
52. Robinson L, Delgadillo J, Kellett S. The dose-response effect in routinely delivered psychological therapies: A systematic review. Psychother Res 2020 Jan 2;30(1):79-96. [doi: 10.1080/10503307.2019.1566676]

XSL•FO
RenderX

53. Drake RJ, Husain N, Marshall M, et al. Effect of delaying treatment of first-episode psychosis on symptoms and social outcomes: a longitudinal analysis and modelling study. Lancet Psychiatry 2020 Jul;7(7):602-610. [doi: 10.1016/S2215-0366(20)30147-4]
54. Luo Y, Funada S, Yoshida K, Noma H, Sahker E, Furukawa TA. Large variation existed in standardized mean difference estimates using different calculation methods in clinical trials. J Clin Epidemiol 2022 Sep;149:89-97. [doi: 10.1016/j.jclinepi.2022.05.023] [Medline: 35654267]
55. Gupta SK. Intention-to-treat concept: A review. Perspect Clin Res 2011 Jul;2(3):109-112. [doi: 10.4103/2229-3485.83221] [Medline: 21897887]
56. Enders CK, Bandalos DL. The relative performance of full information maximum likelihood estimation for missing data in structural equation models. Structural Equation Modeling: A Multidisciplinary J 2001 Jul 1;8(3):430-457. [doi: 10.1207/S15328007SEM0803_5]
57. Zhu C, Byrd RH, Lu P, Nocedal J. Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization. ACM Trans Math Softw 1997 Dec;23(4):550-560. [doi: 10.1145/279232.279236]
58. Johannes Dechant M, Frommel J, Mandryk R. Assessing social anxiety through digital biomarkers embedded in a gaming task. 2021 May 6 Presented at: CHI '21; Yokohama Japan. [doi: 10.1145/3411764.3445238]
59. Takano M, Tsunoda T. Self-disclosure of bullying experiences and social support in avatar communication: analysis of verbal and nonverbal communications. 2019 Presented at: Proceedings of the International AAAI Conference on Web and Social Media p. 473-481. [doi: 10.1609/icwsm.v13i01.3353]

Abbreviations

Brief LSAS-J: Japanese version of the Brief Liebowitz Social Anxiety Scale
FIML: Full Information Maximum Likelihood
LGCM: Latent Growth Curve Modeling
LOO: Leave-One-Out
MMOG: massive multiplayer online game
QIDS-J: Quick Inventory of Depressive Symptomatology
SMD: standardized mean difference
WAIC: Widely Applicable Bayesian Information Criterion

Edited by A Coristine; submitted 11.03.25; peer-reviewed by BA Hives, H Shin; revised version received 05.08.25; accepted 06.08.25; published 24.09.25.

Please cite as:
Yokotani K, Takano M, Abe N, Kato TA
Improvement of Game Users' Depressive Symptoms via Behavioral Activation in a Massive Multiplayer Online Game: Randomized Controlled Trial
JMIR Serious Games 2025;13:e73734
URL: https://games.jmir.org/2025/1/e73734
doi:10.2196/73734

© Kenji Yokotani, Masanori Takano, Nobuhito Abe, Takahiro A Kato. Originally published in JMIR Serious Games (https://games.jmir.org), 24.9.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Perceptions of Patients With Stroke Regarding an Immersive Virtual Reality–Based Exercise System for Upper Limb Rehabilitation: Questionnaire and Interview Study

Jiayin Chen[1], PhD; Calvin Kalun Or[2], PhD; Zhixian Li[3], BM; Eric Hiu Kwong Yeung[4], MSc; Tianrong Chen[2], PhD

[1]Department of Ergonomics and Healthcare, College of Furniture and Industrial Design, Nanjing Forestry University, Nanjing, China

[2]Department of Data and Systems Engineering, University of Hong Kong, Hong Kong, China (Hong Kong)

[3]Department of Rehabilitation Medicine, Dingzhou People's Hospital, Dingzhou, China

[4]The University of Hong Kong–Shenzhen Hospital Department of Physiotherapy, The University of Hong Kong–Shenzhen Hospital, Shenzhen, China

Corresponding Author:
Calvin Kalun Or, PhD
Department of Data and Systems Engineering
University of Hong Kong
Room 8-7, Haking Wong Building, The University of Hong Kong
Pokfulam Road
Hong Kong
China (Hong Kong)
Phone: 852 39172587
Fax: 852 28586535
Email: klor@hku.hk

Abstract

Background: With substantial resources allocated to develop virtual reality (VR)–based rehabilitation exercise programs for poststroke motor rehabilitation, it is important to understand how patients with stroke perceive these technology-driven approaches, as their perceptions can determine acceptance and adherence.

Objective: This study aimed to examine the perceptions of patients with stroke regarding an immersive VR-based exercise system developed to deliver shoulder, elbow, forearm, wrist, and reaching exercises.

Methods: A questionnaire was used to assess the perceptions of 21 inpatients who had experienced stroke (mean time from stroke onset: 37.2, SD 25.9 days; Brunnstrom stage of stroke recovery for the arm: 3-5) regarding the perceived usefulness of, ease of use of, attitude toward, intrinsic motivation for, and intention to use the exercise system. The measurement items were rated on a 7-point Likert scale ranging from 1 (very strongly disagree) to 7 (very strongly agree), with higher values indicating more positive perceptions. Descriptive statistics were used to summarize the responses. Moreover, we conducted semistructured interviews that were audio recorded, transcribed, and subjected to content analysis to identify thematic patterns.

Results: The questionnaire results revealed that the patients' perceptions of the exercise system were positive (mean ratings >6). The content analysis revealed 6 positive themes from 73 statements about the exercise system: ease of use, usefulness, enjoyment, motivation, accessibility, and game design. Conversely, 15 statements reflected negative perceptions, which were clustered into 3 themes: difficulty in handling VR devices, uncomfortable experiences when using VR devices, and monotony.

Conclusions: Integrating VR technology into poststroke functional exercises holds significant promise based on patient interests. However, patient preferences and adaptability must be considered to promote the technology's success. VR-guided exercises should be user-friendly, health-promoting, engaging, and well-designed. Furthermore, addressing challenges, such as bulkiness, motion sickness, discomfort, and exercise monotony, is crucial for the widespread adoption and diffusion of this technology.

(JMIR Serious Games 2025;13:e49847) doi:10.2196/49847

KEYWORDS

virtual reality; stroke; perception; rehabilitation; questionnaire; interview

XSL•FO
RenderX

Introduction

Persistent upper limb impairment following stroke has a substantial impact on patients' daily activities and quality of life [1-5]. Therapeutic exercises are necessary to restore motor function and independence. Patients commonly attend face-to-face therapy sessions at clinics and perform exercises under the supervision of therapists. However, this traditional approach has limitations, including high costs, scheduling conflicts, limited access to therapy services, and the need for patients to travel to clinics [6,7].

Technologies, such as immersive virtual reality (VR), have been applied to promote the accessibility and affordability of poststroke therapeutic exercises [8-10]. The effectiveness of immersive VR-based rehabilitation programs on upper limb motor recovery has been examined in previous studies [11-13]. Besides effectiveness, it is equally important to consider patients' perceptions and acceptance of such programs because negative perceptions and nonacceptance of these technologies, which is a common challenge in practice, can lead to implementation failure, losses to stakeholders, and undesirable health care outcomes [14,15].

Several studies have examined the experiences of stroke survivors with immersive VR-based motor rehabilitation programs, evaluating various aspects such as perceived usefulness, discomfort, motivation, and intention to use these programs. Results have shown that the application of immersive VR is both feasible and acceptable among patients [13,16-18]. However, these studies primarily relied on quantitative methods, specifically rating scales, to capture the experiences of patients. This approach, while valuable, fails to uncover the underlying factors that influence these experiences, leaving critical elements unexplored [19]. To address this gap, incorporating qualitative methods, such as in-depth interviews, is essential. Qualitative research can provide rich, detailed data that reveals the complexities of patients' interactions with VR technology, their emotional responses, and the contextual factors that shape their experiences. By exploring these dimensions, researchers can identify barriers and facilitators to the formation of positive perceptions and subsequent acceptance and implementation. This will allow them to tailor interventions to meet the specific needs of patients with stroke and enhance the overall effectiveness of VR-based rehabilitation programs.

Currently, qualitative studies evaluating the experiences of stroke survivors with immersive VR-based rehabilitation programs are limited. Previous studies with small sample sizes have provided some insights but are insufficient to understand broader experiences and perceptions of such patients [20,21]. Therefore, this study aimed to conduct both quantitative and qualitative analyses of the perception of patients with stroke regarding an immersive VR-based exercise system for poststroke upper limb exercises. Through this mixed methods approach, we aim to obtain a better understanding of the factors influencing perceptions and identify areas for improvement in the design and implementation of VR-based rehabilitation programs.

Methods

Data Source

The data source was a questionnaire survey and semistructured interviews conducted as part of a proof-of-concept randomized controlled trial (RCT) [11]. The trial examined the effectiveness and safety of an immersive VR-based exercise system for poststroke upper limb exercises. The exercise system comprised 5 games (Figures 1-5): dumbbell-lifting game for shoulder flexion and abduction, fishing game for elbow flexion, sheep-whacking game for forearm pronation and supination, apple-picking game for wrist flexion and extension, and balloon-popping game for reaching exercises. The details of the 5 games are presented in Multimedia Appendix 1. A VR headset and a wireless handheld controller (HTC VIVE Pro, HTC Corporation) were used to display and interact with the virtual environments (Figure 6).

Study participants were inpatients at a public hospital in China who have had a stroke. Patients were considered eligible for the study if they were aged 19 to 75 years; had their first ever unilateral stroke, as confirmed by their computed tomography or magnetic resonance imaging records; were experiencing motor impairments in one of the upper limbs, with Brunnstrom stages of stroke recovery of the arm of 3 to 5; had an active range of motion of at least 10° in the shoulder and elbow of the affected arm; could maintain autonomous upright seating for at least 45 minutes; and had normal or corrected-to-normal vision and hearing. Patients with injuries or other health conditions restricting their upper limb mobility, unilateral spatial neglect, unstable medical conditions, a history of seizures or epilepsy, communication difficulty, mood instability, or concurrent participation in any other ongoing investigational drug studies were excluded. Physicians and physical therapists in the rehabilitation medicine department initially introduced the study to their inpatients. Patients expressing interest in participation were invited to a screening session, during which they tried the study's VR devices, and their eligibility was assessed. The researcher (JC) provided detailed explanations of the study to the patients initially determined eligible, confirmed their eligibility, and obtained their written informed consent for participation. In the RCT, participants were randomly assigned to one of the 2 study groups. The intervention group received the aforementioned exercise system and was instructed to use the system to perform upper limb exercises for 35 minutes each day, 6 days a week, for 2 weeks. The control group received commercial games downloaded from Steam (Valve Corporation) and was instructed to play them for the same duration and frequency [22]. These commercial games served as a sham VR program to conceal group assignments from participants. They were not directly aligned with the exercises provided in the exercise system and were used solely for entertainment. Multimedia Appendix 2 provides details of these commercial games. The control group used the same headset and handheld controller as the intervention group. All participants continued their prescribed stroke rehabilitation therapy in the hospital throughout the study period.

XSL•FO
RenderX

Figure 1. Dumbbell-lifting game for shoulder exercise.

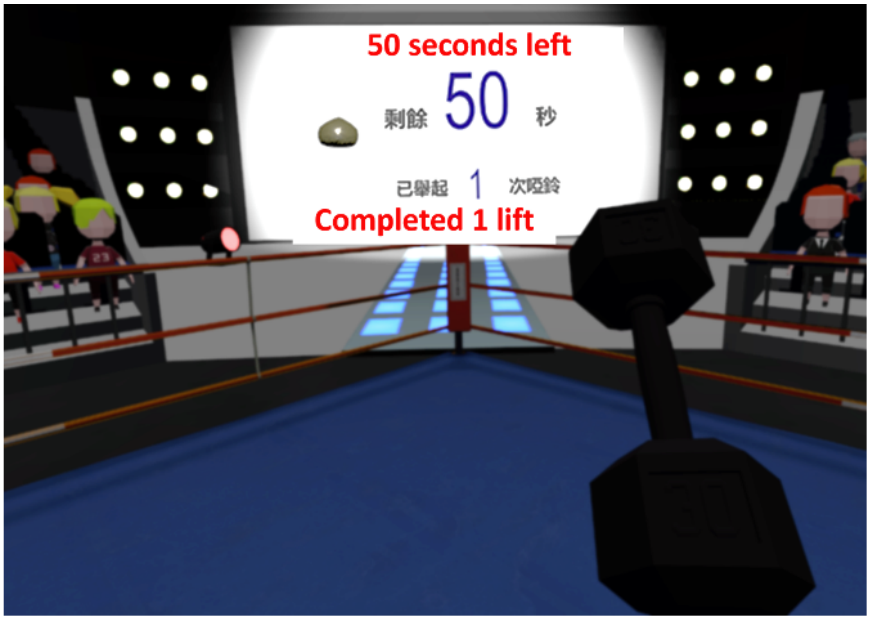


Figure 2. Fishing game for elbow exercise.



XSL•FO
RenderX

Figure 3. Sheep-whacking game for forearm exercise.



Figure 4. Apple-picking game for wrist exercise.



XSL•FO
RenderX

Figure 5. Balloon-popping game for reaching exercise.



Figure 6. A participant performing exercises using the virtual reality device.



The participants' demographic and clinical characteristics, including age, sex, education level, stroke type, time from stroke onset, Brunnstrom stage of the affected arm, and previous experience with VR, were collected at baseline. The effectiveness and safety of the immersive VR-based exercise system were measured at baseline and the 1-week and 2-week follow-up assessments. Owing to the exclusive use of the exercise system by the intervention group, perceptions of the exercise system were only evaluated in this group. A questionnaire survey and semistructured interviews were conducted during the participants' last follow-up assessments; the details are provided in the Questionnaire and Interview sections.

Ethics Approval

The RCT was approved by The University of Hong Kong–Shenzhen Hospital Human Research Ethics Committee (approval number [2021]137) and Dingzhou People's Hospital (approval number hx-2021-04) in China. It has been registered in the Chinese Clinical Trial Registry (registration number ChiCTR2100047150). All the participants provided written informed consent before participating in the RCT. No cash compensation was provided. All data used were deidentified.

Questionnaire

A questionnaire was adapted from technology acceptance models and relevant studies examining individuals' and patients' perceptions of various technologies [23-28]. It was used to assess participants' perceptions of the exercise system, including perceived usefulness, perceived ease of use, attitude, intrinsic motivation, and behavioral intention. Textbox 1 presents the measurement items. The items were rated on a 7-point Likert scale ranging from 1 (very strongly disagree) to 7 (very strongly agree), with higher values indicating more positive perceptions.

XSL•FO
RenderX

Textbox 1. Perception outcomes and measurement items.

Perceived usefulness (PU)

- PU 1: using the immersive virtual reality (VR)–based exercise system to perform upper limb exercises would improve upper limb motor function.
- PU 2: using the immersive VR-based exercise system to perform upper limb exercises would increase exercise efficiency.
- PU 3: using the immersive VR-based exercise system to perform upper limb exercises would enhance exercise effectiveness.
- PU 4: you found the immersive VR-based exercise system useful for performing upper limb exercises.

Perceived ease of use (PEOU)

- PEOU 1: learning to use the immersive VR-based exercise system was easy for you.
- PEOU 2: you found it easy to get the immersive VR-based exercise system to do what you wanted to do.
- PEOU 3: it was easy to become skillful at using the immersive VR-based exercise system.
- PEOU 4: you found the immersive VR-based exercise system easy to use.

Attitude (ATT)

- ATT 1: using the immersive VR-based exercise system to perform upper limb exercises is a good idea.
- ATT 2: using the immersive VR-based exercise system to perform upper limb exercises is a wise idea.
- ATT 3: you like the idea of using the immersive VR-based exercise system to perform upper limb exercises.
- ATT 4: using the immersive VR-based exercise system to perform upper limb exercises was a pleasant experience.

Intrinsic motivation (IM)

- IM 1: you found using the immersive VR-based exercise system to perform upper limb exercises enjoyable.
- IM 2: the actual process of using the immersive VR-based exercise system to perform upper limb exercises was pleasant.
- IM 3: you had fun using the immersive VR-based exercise system to perform upper limb exercises.

Behavioral intention (BI)

- BI 1: you intend to continue using the immersive VR-based exercise system to perform upper limb exercises.
- BI 2: You plan to continue using the immersive VR-based exercise system to perform upper limb exercises.

Interview

As the RCT was a proof-of-concept trial, individual, face-to-face, and semistructured interviews were conducted to learn more about the participants' experiences with the exercise system, allowing us to improve it further. Open-ended questions were asked regarding the exercise system's comfort, comprehensibility, usefulness, and the participants' preferences and intention to use the system. Textbox 2 presents the details of the interview questions. During the interview, a researcher (JC) reformulated and clarified questions to ensure the participants' complete comprehension. Probing questions, such as "What do you mean by that?" and "Could you be more specific?" were also asked to obtain deeper insights into the participants' responses when necessary.

XSL•FO
RenderX

Textbox 2. Interview questions.

Comfort

- Do you feel any discomfort when using the immersive virtual reality (VR)–based exercise system?
- Do virtual environments make you feel uncomfortable?
- Does the headset or the handheld controller make you feel uncomfortable?
- (Ask if the participants reported discomfort) What makes you feel uncomfortable?

Comprehensibility

- Is it easy for you to understand the exercise tasks?
- Is it easy for you to use the headset or the handheld controller?
- (Ask if the participants reported difficulties) What do you find difficult to understand or use?

Usefulness

- Do you think that the immersive VR-based exercise system is useful for your upper limb rehabilitation?
- What makes you think the immersive VR-based exercise system is useful for your rehabilitation?
- What makes you think that the immersive VR-based exercise system is useless for your rehabilitation?

Preference

- Do you like the immersive VR-based exercise system?
- What aspects of the immersive VR-based exercise system do you like?
- What aspects of the immersive VR-based exercise system do you dislike?

Intention to use

- Do you intend to use the immersive VR-based exercise system for rehabilitation in the future?
- What motivates you to use the immersive VR-based exercise system for rehabilitation?
- What fails to make you intend to use the immersive VR-based exercise system for rehabilitation?

Procedures

Written informed consent was obtained from each participant before enrollment into the RCT. Before starting, the researcher (JC) explained the objectives of the questionnaire and interview. The participants then completed the questionnaire to indicate their perceptions of the exercise system. Next, the researcher interviewed the participants. Verbal consent for audio recording was verified again before the start of the interviews. During the interviews, the participants were encouraged to express their experience with the exercise system as much as possible. The interview was audio recorded for subsequent transcription. The interviews were conducted in Chinese. The administration of the questionnaire and interview lasted approximately 20 minutes for each participant.

Data Analysis

Questionnaire

Descriptive statistics, including the mean, SD, median, and rating distribution of the measurement items, were calculated.

Interview

The interview audiotapes were transcribed verbatim by the researcher (JC). Two researchers (JC and TC) with backgrounds in human factors and health care informatics were involved in the analysis. The interview transcripts were analyzed using a qualitative content analysis approach with 3 steps [29]. Step 1: the interview transcripts were read several times before being independently analyzed by the 2 researchers. During the initial reading, the 2 researchers read through the interview transcripts while listening to the interview audiotapes to verify the accuracy of the transcripts and to gain an overall sense of the interview content. During the subsequent reading, the 2 researchers read through the interview transcripts in detail to gain a thorough understanding of the interview content. Step 2: an inductive approach was used to analyze the data because this approach allows the identification of codes and themes that answer the research question [30]. During the data analysis process, the first interview transcript was independently analyzed by the 2 researchers. Relevant statements from the interview transcript were identified, and codes were generated, which were then refined and developed into themes. The same process was repeated for each interview transcript, with the codes and themes added, revised, or developed each time. In addition, the frequencies of the statements expressed by the participants were recorded. Step 3: after all the interview transcripts were analyzed, the codes and themes were compared between the 2 researchers. In case of discrepancies, the original interview transcripts were checked, and the discrepancies were resolved through discussion between the 2 researchers until a consensus was reached.

XSL•FO
RenderX

NVivo (version 12; Lumivero) was used for content analysis, which was based on a checklist of consolidated criteria for reporting qualitative research [31].

Results

Participant Characteristics

Of the 25 intervention group participants, 3 were discharged from the hospital in advance and 1 withdrew from the study, leaving 21 participants who completed the questionnaire and interviews. The demographic and clinical characteristics of the 21 participants are presented in Table 1. The mean age was 55.3 (SD 12.9) years, and the participants were predominantly male (15/21, 71%). Two-thirds of the participants had completed their secondary school education (14/21, 67%). Two-thirds of the participants had an ischemic stroke (14/21, 67%), with a mean time from stroke onset until study enrollment of 37.2 (SD 25.9) days. More than half of the participants (13/21, 62%) were at Brunnstrom stage 3. Most participants had no experience using a VR device (20/21, 95%).

Table 1. Characteristics of the participants (N=21).

Characteristics	Values
Age (y)	
Mean (SD)	55.3 (12.9)
Median (range; IQR)	54 (28-75; 46-66.5)
Sex, n (%)	
Male	15 (71)
Female	6 (29)
Education level, n (%)	
No schooling completed	0 (0)
Some level of primary schooling	0 (0)
Primary school completed	6 (29)
Some level of secondary schooling	0 (0)
Secondary school completed	14 (67)
Diploma, advanced diploma, associate degree, or equivalent	1 (5)
Bachelor’s degree	0 (0)
Stroke type, n (%)	
Ischemic	14 (67)
Hemorrhagic	7 (33)
Time from stroke onset at enrollment (d)	
Mean (SD)	37.2 (25.9)
Median (range; IQR)	34 (7-107; 18-56)
Brunnstrom stage of the affected arm, n (%)	
3	13 (62)
4	4 (19)
5	4 (19)
Experience using a VR[a] device, n (%)	
Never	20 (95)
Rarely	1 (5)
Sometimes	0 (0)
Often	0 (0)
Always	0 (0)

[a]VR: virtual reality.

XSL•FO
RenderX

Questionnaire

Table 2 presents the distribution, mean (SD), and median of the participants' ratings for each measurement item on the questionnaire. The mean values of the overall ratings for perceived usefulness, perceived ease of use, attitude, intrinsic motivation, and behavioral intention were 6.20 (SD 0.69), 6.33 (SD 0.80), 6.25 (SD 0.74), 6.37 (SD 0.75), and 6.31 (SD 0.84), respectively. The median ratings for each measurement item were 6 or 7.

Table 2. Distribution, mean (SD), and median of the participant's responses to the measurement items (N=21).

Outcomes	Rating distribution, n (%)							Mean (SD)	Median
	1	2	3	4	5	6	7		
PU[a]									
PU 1	0 (0)	0 (0)	0 (0)	0 (0)	5 (24)	7 (33)	9 (43)	6.19 (0.81)	6
PU 2	0 (0)	0 (0)	0 (0)	0 (0)	5 (24)	8 (38)	8 (38)	6.14 (0.79)	6
PU 3	0 (0)	0 (0)	0 (0)	0 (0)	4 (19)	10 (48)	7 (33)	6.14 (0.73)	6
PU 4	0 (0)	0 (0)	0 (0)	0 (0)	2 (10)	10 (48)	9 (43)	6.33 (0.66)	6
PEOU[b]									
PEOU 1	0 (0)	0 (0)	0 (0)	0 (0)	3 (14)	6 (29)	12 (57)	6.43 (0.75)	7
PEOU 2	0 (0)	0 (0)	0 (0)	1 (5)	3 (14)	7 (33)	10 (48)	6.24 (0.89)	6
PEOU 3	0 (0)	0 (0)	0 (0)	1 (5)	2 (10)	7 (33)	11 (52)	6.33 (0.86)	7
PEOU 4	0 (0)	0 (0)	0 (0)	1 (5)	3 (14)	5 (24)	12 (57)	6.33 (0.91)	7
ATT[c]									
ATT 1	0 (0)	0 (0)	0 (0)	0 (0)	5 (24)	6 (29)	10 (48)	6.24 (0.83)	6
ATT 2	0 (0)	0 (0)	0 (0)	1 (5)	3 (14)	7 (33)	10 (48)	6.24 (0.89)	6
ATT 3	0 (0)	0 (0)	0 (0)	0 (0)	5 (24)	8 (38)	8 (38)	6.14 (0.79)	6
ATT 4	0 (0)	0 (0)	0 (0)	0 (0)	3 (14)	7 (33)	11 (52)	6.38 (0.74)	7
IM[d]									
IM 1	0 (0)	0 (0)	0 (0)	0 (0)	3 (14)	7 (33)	11 (52)	6.38 (0.74)	7
IM 2	0 (0)	0 (0)	0 (0)	0 (0)	4 (19)	6 (29)	11 (52)	6.33 (0.80)	7
IM 3	0 (0)	0 (0)	0 (0)	0 (0)	3 (14)	7 (33)	11 (52)	6.38 (0.74)	7
BI[e]									
BI 1	0 (0)	0 (0)	0 (0)	1 (5)	2 (10)	7 (33)	11 (52)	6.33 (0.86)	7
BI 2	0 (0)	0 (0)	0 (0)	1 (5)	2 (10)	8 (38)	10 (48)	6.29 (0.85)	6

[a]PU: perceived usefulness.

[b]PEOU: perceived ease of use.

[c]ATT: attitude.

[d]IM: intrinsic motivation.

[e]BI: behavioral intention.

Interviews

Overview

A total of 87.23 minutes of interview audiotapes were recorded. The interviews lasted between 2.43 and 8.30 (mean 4.15; median 3.88) minutes. The results of the content analysis were categorized into positive and negative perceptions of the exercise system. The details of the qualitative data are presented in the following sections.

Positive Perceptions of the Exercise System

Overview

A total of 73 statements derived from 21 participants were coded as positive perceptions. Of these, 25 (34%) statements were related to ease of use, 20 (27%) to benefits, 9 (12%) to enjoyment, 7 (10%) to assistance, 7 (10%) to accessibility, and 5 (7%) to game design. The findings are summarized in Table 3.

XSL•FO
RenderX

Table 3. Positive perceptions of the exercise system (N=21).

Themes, subthemes, and statements	Participants who made the statement, n
Ease of use	
Easy to understand games	
Easy to understand how to play the games	20
Easy to use VR[a] devices	
Easy to use the head-mounted display and handheld controller	5
Benefits	
Improved upper limb motor function	
Improved muscle strength	5
Improved range of motion	3
Improved stability	3
Improved overall movement performance	2
Improved flexibility	1
Improved daily living skills	
Improved self-care skills in daily life	1
Improved cognitive ability	
Improved reaction speed	2
Promoted brain thinking	1
Improved mood	
Was in a good mood	2
Enjoyment	
Had fun exercising	
Had fun during the exercise	9
Assistance	
Assisted in motor learning	
Assisted the affected upper limb in developing normal movement patterns	7
Accessibility	
Promoted access to exercise	
Provided more opportunities for exercises	7
Game design	
Virtual environments	
Well-created virtual environments	2
Feeling of immersion	
High immersion level	2
Novelty	
Novel games	1

[a]VR: virtual reality.

Ease of Use

Of the 21 participants, 20 expressed that the games were easy to understand. One of them felt that it was difficult to play the games the first time because of unfamiliarity, although it became easy after a few trials. A total of 5 participants reported that the headset and handheld controller were easy to use. Some participants stated the following:

> *[I] know how to play each [game].* [Participant 20]
>
> *It is difficult [to play the games] the first time when you are not familiar, but after you practice a few*

XSL•FO
RenderX

times, it is not difficult, because you get to know what is going on. [Participant 1]

Picking up apples [the apple-picking game] is used to extend [the wrist]. I know it. [Participant 18]

No difficulties [in using the headset or the handheld controller]. [Participant 3]

Benefits

Some participants reported that their upper limb motor function had improved. Specifically, 5 out of 21 participants expressed that their muscle strength had increased, 3 reported that their range of motion had improved, 3 expressed that their affected upper limb became more stable, 2 expressed that their overall upper limb movement performance had improved, and 1 said that their hand became more flexible. Some examples of the participants' statements are as follows:

Muscle strength is a little better. There is no obvious [feeling] when putting the weight [on the wrist]. [It is] not tiring. [Participant 17]

[I] can get the elbow straight. [Participant 13]

Wasn't it [the arm] wobbly before? The balance does not seem good. Now it [the arm] is not wobbly like that. [Participant 18]

It [the immersive VR-based exercise system] makes your hands more flexible and softer. [Participant 1]

The joints can move to the correct location. [Participant 7]

It [the immersive VR-based exercise system] is good for arm movements, meaning your movements are more standard. [Participant 3]

One participant observed improvements in their ability to perform self-care eating activities:

I can easily pick up and hold the food now. [I needed to] break it [into small pieces] before. I was afraid that [I] could not hold it and dropped it. Now I do not need to break it [into small pieces], I can hold it and eat it directly. I can also hold the cup when brushing [my] teeth and cleaning [my] mouth. [Participant 18]

Two participants mentioned that playing games improved their reaction speed to the computer, and 1 participant expressed that playing games could promote thinking:

To the computer...I feel that [my] reaction is faster. [Participant 9]

[The immersive VR-based exercise system] can facilitate your thinking and make the brain work hard. [Participant 1]

Two participants expressed that they were in a good mood when immersed in the virtual environments.

I have been kind of stressed since I got this disease, but once I listen to that music, [my] mood turns good....I feel like I am alive again, walking into that wood and picking apples like [in] real [life], and being immersed into that environment. I feel like ... my mind broadens. [Participant 1]

[I] feel like myself...anyway, I feel as if I am still useful, I will be able to do housework like before, to do this and that. It seems like the feeling of pessimism has been reduced. [Participant 1]

Enjoyment

Of the 21 participants, 9 described their feelings of enjoyment and fun when performing VR exercises. For example, some participants stated as follows:

At least performing [exercises] with it [the immersive VR-based exercise system] is not that annoying, not that boring. [Participant 5]

You are free to play this [game]. It is also exciting. [Participant 8]

It [the immersive VR-based exercise system] is quite fun and can also train the function of the limbs. [Participant 19]

Assistance

Of the 21 participants, 7 expressed that they might not know how to perform some movements in the real world; however, the virtual objects in the games guided and helped them to learn and encouraged them to try to perform the movements. Some participants stated as follows:

You know [I] could not do that movement [elbow extension] before. I feel it [the immersive VR-based exercise system] is useful. [The elbow] could extend when [I] put the fishing rod down. [Participant 5]

[When] you would attempt to perform [the movements], it [the arm] could not move, but you would attempt to do it. [Participant 6]

For example, that apple-picking [game], together with that bird, could guide my hand to perform the movements [wrist flexion and extension]. [Participant 7]

Accessibility

Of the 21 participants, 7 reported that the exercise system provided more opportunities for exercise. Some examples of the participants' statements are as follows:

[I like using the immersive VR-based exercise system because I] can exercise. I can always exercise. [Participant 4]

[I like using the immersive VR-based exercise system because it] provides arm exercises. [Participant 13]

Game Design

Several participants expressed interest in the game design. Two out of 21 participants stated that the virtual environments were attractive, 2 stated that the games provided a sense of immersion, and 1 stated that the games were innovative. Some participants stated as follows:

[It is] just like being in a wonderland, right? It is also very clear, just like the 3D movies. I like looking around and performing the exercises. [Participant 21]

XSL•FO
RenderX

There are images, music, and sounds of the birds. It makes you feel like you are in the woods. I feel very excited. [Participant 1]

The games are novel. [Participant 17]

Negative Perceptions of the Exercise System

Overview

A total of 15 statements derived from 8 participants were identified as negative perceptions. Of those, 8 (53%) statements were related to difficulty in handling VR devices, 6 (40%) statements to uncomfortable experiences when using VR devices, and 1 (7%) statement to monotony. The findings are summarized in Table 4.

Table 4. Negative perceptions of the exercise system (N=21).

Themes, subthemes, and statements	Participants who made the statement, n
Difficulty in handling VR[a] devices	
Bulky headset	
Heavy headset	3
Sweating and stuffiness when wearing the headset	3
Cable constraint	1
Heavy handheld controller	
Heavy to hold handheld controller	1
Uncomfortable experiences when using VR devices	
Dizziness	
Feeling dizzy when using the headset	2
Fatigue	
Feeling tired during the exercise	2
Eyestrain	
Eyestrain when using the headset	1
Pain	
Pain during exercise	1
Monotony	
Boring exercise	
Repetitive movements leading to monotony	1

[a]VR: virtual reality.

Difficulty in Handling VR Devices

Of the 21 participants, 3 indicated that the headset was heavy, 3 mentioned that wearing the headset made them feel sweaty and stuffy, 1 stated that the cable connecting the headset and laptop constrained their head movement, and 1 mentioned that the handheld controller was heavy to hold. Some participants stated the following:

I feel that the headset is a little heavy. [Participant 5]

It is a little sweaty [when wearing the headset]. It is not tiring; it is stuffy. [Participant 8]

The head is being pressed. It is uncomfortable. [I] feel that it would be good if it were wireless [headset]. [Participant 5]

[If the headset is wireless], turning it [the head] around would be convenient and free. [Participant 5]

It [the handheld controller] is just heavy. The total weight is heavy. [Participant 1]

Uncomfortable Experiences When Using VR Devices

Two out of the 21 participants mentioned that wearing the headset led to dizziness, 2 reported that they experienced fatigue during the VR exercises, 1 highlighted that wearing the headset caused eyestrain, and 1 reported experiencing pain in the upper limb during the VR exercises. Some of the participants stated as follows:

[I] felt a little dizzy after taking off [the headset]. [Participant 5]

Watching [the virtual environments], the eyes are not very comfortable. [Participant 6]

[The] part [of the upper limb that] worked hard...was sore. [Participant 7]

It was a little painful when stretching. [Participant 18]

XSL•FO
RenderX

Monotony

One participant reported that repeating the same exercises every day was boring, but the participant kept doing the exercises to recover faster.

> *For my own health, [I] mean to recover faster, the exercises must be performed, right? ...It is, of course, boring, always repeating the same movement.* [Participant 18]

Discussion

Principal Findings

Overview

This study aimed to investigate the perceptions of stroke survivors regarding the use of an immersive VR-based exercise system for poststroke upper limb exercises. Quantitative and qualitative analyses were used to explore these perceptions. The questionnaire results revealed that the participants recognized the exercise system's usefulness for rehabilitation, found the exercise system easy to use, maintained a positive attitude toward the exercise system, enjoyed using the exercise system for exercises, and expressed their intention to continue using the exercise system. The interviews further uncovered the following 6 themes related to positive perceptions of the exercise system: (1) ease of use, (2) benefits, (3) enjoyment, (4) assistance, (5) accessibility, and (6) game design. Conversely, 3 themes were associated with negative perceptions: (1) difficulty in handling VR devices, (2) uncomfortable experiences when using VR devices, and (3) monotony. The interview findings may help interpret the results obtained from the questionnaire and provide further understanding of the factors influencing the patients' acceptance and use of VR-based rehabilitation programs. For example, some patients expressed interest in the highly immersive game scenarios, which might have attracted them to use the exercise system to perform exercises. In contrast, some patients expressed undesirable feelings about wearing the headset, which might have led to a negative perception regarding the exercise system and impeded its use. Moreover, our findings serve as a concrete illustration of established technology acceptance models, such as the Technology Acceptance Model and Unified Theory of Acceptance and Use of Technology model. For example, previous research has underscored perceived usefulness as a key determinant of technology acceptance by users [23,24,32,33]. Our study further elucidated that motor function, daily living skills, cognitive ability, and mood were the main aspects of usefulness valued by the participants. In the following sections, we delve into the questionnaire and interview findings in greater detail.

Perception Ratings

In contrast to previous studies reporting that individuals with limited prior technology experience are less inclined to accept the technology [8,32,34-37], our findings indicated that a lack of experience with VR may not impede patients' positive feelings about using the exercise system. The interview responses, which are discussed in the following sections, may help explain this. However, the highly positive perceptions might be a result of selection bias, meaning that participants who were satisfied with the exercise system were more likely to adhere to the VR exercises, whereas those who were unsatisfied with it could have withdrawn from the study. Therefore, the patients' perceptions of the exercise system might have been overestimated.

Ease of Use

Ease of use is a critical perceptual aspect affecting technology acceptance [23,24,32,33]. Most patients who have experienced stroke are older adults who often have limited experience with technology and may face declining physical or cognitive functions, as well as technology anxiety [38-41]. Therefore, ensuring that VR-based rehabilitation exercises are easy to use is particularly important for this demographic to promote acceptance and adoption. In this study, almost all participants reported that it was easy to understand how to play the rehabilitation exercise games and use the VR devices. Although 1 participant initially found the games difficult due to unfamiliarity, the participant quickly adapted and found the games easier after practicing several times. This may be because the games were related to daily life (eg, fishing), and there were no complicated tasks or complex operations with the handheld controller. Consequently, participants could easily become familiar with the rules of the games and the use of the VR devices without much training or practice. Another possible explanation may be that most participants were able to perform the exercises independently. Necessary assistance was provided only to ensure their safety and prevent the development of abnormal movements (eg, shrugging the shoulders) by the researcher (JC), with the guidance of physical therapists [11]. These findings suggest several specific implications for VR-based rehabilitation programs. First, the simplicity of game design, mimicking daily life activities, can significantly enhance usability for patients with stroke. Second, minimal training and practice requirements make it feasible for older adults with limited technology experience to quickly adapt to the exercises. Third, providing tailored support to prevent incorrect movements can further improve patient engagement and safety. Consequently, designing technology-based rehabilitation exercises with these considerations can foster greater acceptance, enhance patient adherence, and ultimately improve rehabilitation outcomes for stroke survivors.

Benefits

Participants perceived improved upper limb motor function and daily living skills after using the exercise system. Although this exercise system was initially designed for motor rehabilitation, participants perceived improvements in reaction speed, thinking, and mood. Such gains have rarely been reported in previous studies. One possible explanation may be that the repetitive practice led to familiarity with the exercises, thereby helping the participants increase their reaction speed. Furthermore, the setup of the games in the exercise system, such as the 1-minute countdown and real-time feedback on task performance (refer to details of the games in Figures 1-5 and Multimedia Appendix 1), might provide participants with a sense of tension and competition. This, in turn, might "facilitate your thinking and make the brain work hard," as 1 participant expressed. Another

XSL•FO
RenderX

reason may be the advantages of immersive VR; that is, wearing the headset isolated the participants from the physical world, so they could fully focus on the tasks in the games. Moreover, being able to complete the tasks in the games provided the participants with a sense of achievement, which might have elicited positive emotions. As 1 participant described, being able to play games made the participants feel "useful," and thus, "the feeling of pessimism" has been alleviated. Furthermore, the natural scenes and relaxing background music in the games might have helped relieve the participants' negative emotions due to the stroke. A participant stated as follows:

> *I have been kind of stressed since I got this disease, but once I listen to that music, [my] mood turns good... walking into that wood and picking apples like [in] real [life], and being immersed into that environment. I feel like... my mind broadens.*

The results indicate that the positive experiences associated with these games can help reduce the psychological burden of rehabilitation, making the process feel less like a chore and more like a recreational activity. These findings also suggest that the exercise system possesses the potential to influence multiple dimensions of stroke rehabilitation, which should be considered when designing and implementing such VR-based interventions.

Enjoyment

Consistent with previous studies that used video games for rehabilitation [42-46], performing exercises by playing the games in the exercise system was described as "exciting" and "fun" in this study. These findings suggest that the exercise system offers an enjoyable possibility for performing upper limb exercises. The enjoyment factor has specific implications for VR-based rehabilitation programs. First, enjoyable exercises can increase patient motivation and adherence and can consequently lead to more consistent participation in the rehabilitation regimen. This is particularly important for stroke survivors, as sustained engagement is crucial for recovery. Second, incorporating enjoyable elements into the exercise design can help alleviate technology anxiety, making patients more willing to try and continue using the system. Motivational and enjoyable elements can include reward systems that provide a sense of accomplishment, features that facilitate social interaction, and real-time positive feedback and encouragement.

Assistance

Some participants expressed that the virtual objects in the games guide and encourage them to perform movements that they might not be able to perform in the real world. For example, 1 participant said, "That apple-picking [game], together with that bird, could guide my hand to perform the movements." The rationale behind this effect may be that learning a motor task is more effective when learners' attention is focused on achieving specific goals (ie, focusing on external cues) rather than focusing on performing specific movements (ie, focusing internally on the movements themselves) [47-49]. For example, the apple-picking game was designed for wrist flexion and extension exercises. It required the participants to extend their wrists to pick a red apple and flex their wrists to drop it. Picking up and dropping the apple (ie, external focus) was relatively easy for the participants to understand because the instructions were specific and the goal was clear. However, if the participants were told to contract and relax muscles in the wrist and forearm (ie, internal focus), it would have been too abstract to understand and execute the movements. These findings suggest that the objects in virtual environments should be carefully designed because they play a vital role in helping patients learn movements and perform therapeutic tasks.

Accessibility

The participants reported that the exercise system provided extra opportunities to perform rehabilitation exercises, reflecting their need for exercises in addition to conventional rehabilitation therapy. A possible explanation may be that the needs for rehabilitation in patients who have experienced stroke, may not always be fulfilled because of constraints on professional and institutional resources in public hospitals [50,51]. For example, the patients in this study usually had to wait several days before getting an appointment with a physical therapist. The patients could only receive 30 minutes of physical therapy with the guidance of therapists every day during hospitalization [11], which was less than the treatment duration suggested in previous research [52,53].

Game Design

The games in the exercise system were characterized as "novel"; the virtual scenes and auditory components within the games offered the participants a highly immersive experience, described as "being in a wonderland." These findings suggest that the virtual environment design and the VR device type might significantly impact user experience, which should be carefully considered when designing VR-based interventions. The implications of these findings are multifaceted. First, the novelty and immersive nature of the virtual environment can greatly enhance user engagement and enjoyment, making the rehabilitation process more appealing to patients. This implies that designers should focus on creating visually rich and interactive environments that capture patients' attention and interest. Second, the type of VR device used can affect the level of immersion and comfort. High-quality VR devices that offer clear visuals, realistic audio, and ergonomic design can prevent discomfort and enhance the overall experience. This suggests that investment in advanced VR hardware could be beneficial. Third, the positive descriptions of the experience indicate that carefully crafted virtual environments can provide a significant psychological boost to patients. Feeling as though they are "in a wonderland" can distract patients from the monotony of repetitive exercises, making the rehabilitation process feel less burdensome and more enjoyable. Fourth, these findings suggest that regular updates and new content should be incorporated to maintain novelty and engagement levels. Stale or repetitive content could undermine the benefits of the initial immersive experience, so ongoing development and refreshing content are crucial.

Difficulty in Handling VR Devices

Negative perceptions were mainly related to using the headset and handheld controller. Wearing the headset was heavy, making participants feel sweaty and stuffy, the cable used to connect

XSL•FO
RenderX

the headset to the laptop sometimes limited the free movement of the participants' heads, and the handheld controllers were heavy for some participants. This is due to the technical limitations of current VR technology. A large headset is necessary to support a high-resolution display, and a physical cable is needed to support the transmission of massive amounts of real-time data at a high speed and low latency. A better user experience with an immersive VR device relies on advances in hardware and software. For example, the development of retina displays can reduce the size of a headset and improve its wearable comfort, and advances in wireless transmission technology can facilitate the development of wireless VR devices to promote users' freedom of movement in virtual environments [54,55].

Uncomfortable Experiences When Using VR Devices

A few participants experienced mild and short-lived discomfort with the VR exercise system, such as dizziness and eyestrain, as observed by the researcher. These findings indicate that the exercise system overall was well tolerated by the participants. Nevertheless, factors that may be associated with the side effects of immersive VR should be further identified and minimized in the design and development of immersive VR-based interventions for patients. The implications of these side effects are crucial for the safe implementation of VR-based rehabilitation programs. First, identifying the root causes of discomfort, such as prolonged use, visual strain from screen resolution, or motion sickness due to discrepancies in visual and vestibular input, is essential. Addressing these issues could involve optimizing session durations, improving screen technology, and ensuring smooth and realistic motion tracking. Second, providing clear guidelines and training for patients on how to use the VR system correctly can help minimize discomfort. Educating users on taking regular breaks, adjusting headset fit, and recognizing early signs of discomfort can enhance their overall experience. Third, personalizing the VR experience based on individual tolerance levels can further mitigate adverse effects. Implementing customizable settings that allow patients to adjust brightness, contrast, and motion sensitivity can help tailor the experience to their comfort. Fourth, continuous monitoring and feedback mechanisms can be integrated into the VR system to track patient responses in real time. This would allow for immediate adjustments and support, ensuring that the therapy remains both effective and comfortable.

Monotony

Although the upper limb exercises were designed as games, repetitive practice can become boring and monotonous for some individuals, potentially lowering their motivation to adhere to the exercise system [56]. This finding underscores the need for further exploration of strategies to sustain patients' interest in VR-based interventions in the long term. Several strategies can be considered. First, incorporating a variety of game scenarios and levels can keep the exercises fresh and engaging, thereby preventing boredom. For instance, rotating through different themes, such as sports, adventure, and daily activities, can maintain patients' interest. Second, designing exercises with progressively challenging tasks can provide a sense of achievement and keep patients motivated. Third, personalizing exercise routines to match the individual's progress and preferences can make the sessions more enjoyable and relevant. Fourth, integrating social elements, such as multiplayer modes or community challenges, can enhance the fun aspect and encourage patients to stay committed to their rehabilitation.

Implications for Practice

This study offers the following implications for practice. First, the immersive VR-based exercise system was initially accepted by patients because it offered increased exercise opportunities and was easy to use, enjoyable, and beneficial for motor rehabilitation. Thus, health care technology developers, public health decision makers, and health care providers can consider the possibility of integrating such VR-based interventions to complement conventional therapeutic exercises. The aforementioned characteristics should also be addressed to increase the likelihood of system acceptance. Second, therapeutic games can function as an appealing approach to encourage patients to engage in exercises. The preferences demonstrated by patients toward the games within the exercise system, including well-designed virtual environments and a high level of immersion, offer valuable insights for future refinement of VR-based exercise systems. Consequently, this can enhance the user experience and foster the acceptance of such systems. Third, considering patients may have limited prior technology experience, exercise systems must prioritize ease of use. To achieve this, integrating usability inspections and tests throughout the developmental phase is recommended. This approach ensures that the newly developed systems align with users' needs and expectations, ultimately making them usable and useful [57-62]. Fourth, it is important to recognize that patients may encounter discomfort during VR exercises. Although this aspect was highlighted by patients only a few times, health care professionals must remain vigilant about the potential risks of VR technology in the use of VR-based exercise systems. During the design phase, the types of VR devices to be used should be carefully chosen or evaluated, considering the patients' abilities. In addition, elements within virtual scenes that might induce discomfort, such as poor environmental illuminance and highly realistic graphics, should be avoided. During the implementation phase, health care providers should meticulously assess patients' susceptibility to side effects related to VR-based therapy. Education can play a vital role in minimizing the risk of discomfort (eg, advising patients to avoid rapid head movements in virtual environments), and individualized therapy plans should be tailored to each patient's unique needs and circumstances.

Implications for Research

This study's findings provide implications for future research. First, possible reasons for the participants' withdrawal from the study should be further examined, as these may shed light on factors that impede patients' acceptance of and adherence to the intervention. Such information would help improve the design and implementation of VR-based interventions, improve patients' experiences with and acceptance of such interventions, and increase their benefits for patients. Further studies are warranted to collect information using approaches such as interviews with patients and heuristic evaluation by experts

XSL•FO
RenderX

[63]. Second, fewer statements related to negative perceptions of the exercise system were identified than those related to positive perceptions. One possibility for this finding may be that the respondents were inclined to be agreeable and positive to avoid the disapprobation of the interviewers [64,65]. Future studies should consider the impact of response bias on the validity of the results, detect such biases when administering questionnaires and conducting interviews, and deal with the bias during data analysis. Third, the participants used the immersive VR-based exercise system for up to 2 weeks. However, poststroke motor rehabilitation is often a long-term process, and patients' experiences with the exercise system may change with time. Further research is needed to examine patients' perceptions of such VR-based interventions in the long term. For example, it has been suggested that >15 hours of total intervention are needed to produce improvements in upper limb motor function [66]. Further studies should accumulate patients' experiences with such VR-based interventions at different time points. Fourth, the implementation of VR-based interventions requires the involvement of health care providers and caregivers, whose attitude toward such interventions can also influence patients' acceptance of and adherence to VR-based therapy [24,32,33]. Therefore, research examining the opinions of health care providers and caregivers regarding VR-based interventions is warranted.

Limitations

This study has several limitations. First, the participants included in the study were predominantly male individuals (15/21, 71%), were in the subacute stage of stroke (≤6 months after stroke onset), and had limited experience using VR devices. Therefore, further research is needed to assess the generalizability of the findings to other user groups. Second, some withdrawals from the study may have been caused by dissatisfaction with the VR-based exercise system, which might have led to an overestimation of the participants' perceptions. In addition, the participants' reasons for withdrawal were not recorded, resulting in their barriers to accepting and adhering to VR exercises and the usability issues of the exercise system being unknown. Third, participants' perceptions of the exercise system were based on a relatively short period of use. Long-term perceptions of the system should be investigated in future studies. Fourth, some interviews were brief. Consistent with challenges documented in relevant studies [67,68], some of our participants exhibited limited communication and articulation. This limitation could be attributed to reduced sensitivity to perceptions and increased susceptibility to fatigue following brain injuries. For future studies, it may be beneficial to explore other strategies that could improve overall understanding, such as engaging caregivers and family members who can represent patients' experiences to provide complementary information; using group discussions to enhance participation, expression, and recall; and establishing user-friendly communication methods (eg, avoiding highly technical terms) [67-69].

Conclusions

Integrating VR to support poststroke functioning improvement exercises appears promising and acceptable based on patient interests. However, individual preferences and adaptability must be considered to achieve optimal outcomes, as some patients may find VR challenging. VR-guided exercise approaches of this nature should be simple to use, promote improvement in health conditions, foster a sense of fun, facilitate exercise access, and be well-designed in terms of VR exercise content. In addition, these approaches must address challenges such as bulky VR equipment, motion sickness, discomfort, and exercise monotony. Overall, assessing and addressing both favorable and unfavorable patient concerns should be done early in the development process, as this is critical for diffusion and adoption.

Acknowledgments

The authors would like to acknowledge the support from Dingzhou People's Hospital and The University of Hong Kong–Shenzhen Hospital. This study was funded by Seed Funding for Basic Research, The University of Hong Kong, Hong Kong, China (project number 201811159012; principal investigator: CKO). The funder has played no role in the research.

Data Availability

The datasets generated during and analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

JC and CKO designed the study. ZL and EHKY supported participant recruitment and experiment implementation. JC conducted data collection. JC and TC performed data analysis. JC drafted the manuscript. CKO reviewed and significantly revised the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Games developed and used by the intervention group in the randomized controlled trial.
[DOCX File , 23 KB - games_v13i1e49847_app1.docx]

XSL•FO
RenderX

Multimedia Appendix 2
Commercial games used by the control group in the randomized controlled trial.
[DOCX File , 19 KB - games_v13i1e49847_app2.docx]

References

1. Langhorne P, Bernhardt J, Kwakkel G. Stroke rehabilitation. Lancet 2011 May 14;377(9778):1693-1702. [doi: 10.1016/S0140-6736(11)60325-5] [Medline: 21571152]
2. Langhorne P, Coupar F, Pollock A. Motor recovery after stroke: a systematic review. Lancet Neurol 2009 Aug;8(8):741-754. [doi: 10.1016/S1474-4422(09)70150-4] [Medline: 19608100]
3. Mayo NE, Wood-Dauphinee S, Côté R, Durcan L, Carlton J. Activity, participation, and quality of life 6 months poststroke. Arch Phys Med Rehabil 2002 Aug;83(8):1035-1042. [doi: 10.1053/apmr.2002.33984] [Medline: 12161823]
4. Campbell BC, Khatri P. Stroke. Lancet 2020 Jul 11;396(10244):129-142. [doi: 10.1016/S0140-6736(20)31179-X] [Medline: 32653056]
5. Johnson W, Onuma O, Owolabi M, Sachdev S. Stroke: a global response is needed. Bull World Health Organ 2016 Sep 01;94(9):634-634A [FREE Full text] [doi: 10.2471/BLT.16.181636] [Medline: 27708464]
6. Sibley LM, Glazier RH. Reasons for self-reported unmet healthcare needs in Canada: a population-based provincial comparison. Healthc Policy 2009 Aug;5(1):87-101 [FREE Full text] [doi: 10.12927/hcpol.2009.20934] [Medline: 20676253]
7. Billinger SA, Arena R, Bernhardt J, Eng JJ, Franklin BA, Johnson CM, et al. Physical activity and exercise recommendations for stroke survivors: a statement for healthcare professionals from the American Heart Association/American Stroke Association. Stroke 2014 Aug;45(8):2532-2553. [doi: 10.1161/STR.0000000000000022] [Medline: 24846875]
8. Chen J, Or CK, Chen T. Effectiveness of using virtual reality-supported exercise therapy for upper extremity motor rehabilitation in patients with stroke: systematic review and meta-analysis of randomized controlled trials. J Med Internet Res 2022 Jun 20;24(6):e24111 [FREE Full text] [doi: 10.2196/24111] [Medline: 35723907]
9. Domínguez-Téllez P, Moral-Muñoz JA, Salazar A, Casado-Fernández E, Lucena-Antón D. Game-based virtual reality interventions to improve upper limb motor function and quality of life after stroke: systematic review and meta-analysis. Games Health J 2020 Feb 01;9(1):1-10 [FREE Full text] [doi: 10.1089/g4h.2019.0043] [Medline: 32027185]
10. Chen J, Xie Z, Or C. Effectiveness of immersive virtual reality-supported interventions for patients with disorders or impairments: a systematic review and meta-analysis. Health Technol 2021 Jul 16;11(4):811-833. [doi: 10.1007/s12553-021-00561-7]
11. Chen J, Or CK, Li Z, Yeung EH, Zhou Y, Hao T. Effectiveness, safety and patients' perceptions of an immersive virtual reality-based exercise system for poststroke upper limb motor rehabilitation: a proof-of-concept and feasibility randomized controlled trial. Digit Health 2023 Sep 26;9:20552076231203599 [FREE Full text] [doi: 10.1177/20552076231203599] [Medline: 37766909]
12. Crosbie JH, Lennon S, McGoldrick MC, McNeill MD, McDonough SM. Virtual reality in the rehabilitation of the arm after hemiplegic stroke: a randomized controlled pilot study. Clin Rehabil 2012 Sep;26(9):798-806. [doi: 10.1177/0269215511434575] [Medline: 22275463]
13. Lee SH, Jung HY, Yun SJ, Oh BM, Seo HG. Upper extremity rehabilitation using fully immersive virtual reality games with a head mount display: a feasibility study. PM R 2020 Mar 03;12(3):257-262. [doi: 10.1002/pmrj.12206] [Medline: 31218794]
14. Greenhalgh T, Hinder S, Stramer K, Bratan T, Russell J. Adoption, non-adoption, and abandonment of a personal electronic health record: case study of HealthSpace. BMJ 2010 Nov 16;341(nov16 1):c5814 [FREE Full text] [doi: 10.1136/bmj.c5814] [Medline: 21081595]
15. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017 Nov 01;19(11):e367 [FREE Full text] [doi: 10.2196/jmir.8775] [Medline: 29092808]
16. Lim I, Cha B, Cho DR, Park E, Lee KS, Kim M. Safety and potential usability of immersive virtual reality for brain rehabilitation: a pilot study. Games Health J 2023 Feb 01;12(1):34-41. [doi: 10.1089/g4h.2022.0048] [Medline: 36206236]
17. Specht J, Schroeder H, Krakow K, Meinhardt G, Stegmann B, Meinhardt-Injac B. Acceptance of immersive head-mounted display virtual reality in stroke patients. Comput Hum Behav Reports 2021 Aug;4:100141. [doi: 10.1016/j.chbr.2021.100141]
18. Heinrich C, Morkisch N, Langlotz T, Regenbrecht H, Dohle C. Feasibility and psychophysical effects of immersive virtual reality-based mirror therapy. J Neuroeng Rehabil 2022 Oct 07;19(1):107 [FREE Full text] [doi: 10.1186/s12984-022-01086-4] [Medline: 36207720]
19. Kelle U. Combining qualitative and quantitative methods in research practice: purposes and advantages. Qual Res Psychol 2006;3(4):293-311. [doi: 10.1177/1478088706070839]
20. Gustavsson M, Kjörk EK, Erhardsson M, Alt Murphy M. Virtual reality gaming in rehabilitation after stroke - user experiences and perceptions. Disabil Rehabil 2022 Nov 31;44(22):6759-6765 [FREE Full text] [doi: 10.1080/09638288.2021.1972351] [Medline: 34465269]

XSL•FO
RenderX

21. Moan ME, Vonstad EK, Su X, Vereijken B, Solbjør M, Skjæret-Maroni N. Experiences of stroke survivors and clinicians with a fully immersive virtual reality treadmill exergame for stroke rehabilitation: a qualitative pilot study. Front Aging Neurosci 2021 Nov 2;13:735251 [FREE Full text] [doi: 10.3389/fnagi.2021.735251] [Medline: 34795576]
22. Steam. URL: https://store.steampowered.com/ [accessed 2024-11-28]
23. Davis FD, Bagozzi RP, Warshaw PR. User acceptance of computer technology: a comparison of two theoretical models. Manag Sci 1989 Aug;35(8):982-1003. [doi: 10.1287/mnsc.35.8.982]
24. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003;27(3):425-478. [doi: 10.2307/30036540]
25. Taylor S, Todd PA. Understanding information technology usage: a test of competing models. Inf Syst Res 1995 Jun;6(2):144-176. [doi: 10.1287/isre.6.2.144]
26. Yan M, Or C. Factors in the 4-week acceptance of a computer-based, chronic disease self-monitoring system in patients with type 2 diabetes mellitus and/or hypertension. Telemed J E Health 2018 Feb;24(2):121-129. [doi: 10.1089/tmj.2017.0064] [Medline: 28737995]
27. Liu K, Or CK, So M, Cheung B, Chan B, Tiwari A, et al. A longitudinal examination of tablet self-management technology acceptance by patients with chronic diseases: integrating perceived hand function, perceived visual function, and perceived home space adequacy with the TAM and TPB. Appl Ergon 2022 Apr;100:103667. [doi: 10.1016/j.apergo.2021.103667] [Medline: 34920356]
28. Xie Z, Or CK. Acceptance of mHealth by elderly adults: a path analysis. Proc Hum Factors Ergon Soc Annu Meet 2021 Feb 09;64(1):755-759. [doi: 10.1177/1071181320641174]
29. Graneheim UH, Lundman B. Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. Nurse Educ Today 2004 Feb;24(2):105-112. [doi: 10.1016/j.nedt.2003.10.001] [Medline: 14769454]
30. Bengtsson M. How to plan and perform a qualitative study using content analysis. NursingPlus Open 2016;2:8-14. [doi: 10.1016/j.npls.2016.01.001]
31. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
32. Or CK, Karsh BT, Severtson DJ, Burke LJ, Brown RL, Brennan PF. Factors affecting home care patients' acceptance of a web-based interactive self-management technology. J Am Med Inform Assoc 2011;18(1):51-59 [FREE Full text] [doi: 10.1136/jamia.2010.007336] [Medline: 21131605]
33. Yan M, Or C. A 12-week pilot study of acceptance of a computer-based chronic disease self-monitoring system among patients with type 2 diabetes mellitus and/or hypertension. Health Informatics J 2019 Sep 18;25(3):828-843 [FREE Full text] [doi: 10.1177/1460458217724580] [Medline: 28820007]
34. Chen T, Or CK. Perceptions of a machine learning-based lower-limb exercise training system among older adults with knee pain. Digit Health 2023 Jul 04;9:20552076231186069 [FREE Full text] [doi: 10.1177/20552076231186069] [Medline: 37426581]
35. Yan M, Or CK, Xie Z, Liu H. Facilitators of and barriers to the use of a computer-based self-monitoring system by type 2 diabetic and/or hypertensive patients. Int J Ind Ergon 2023 Nov;98:103509. [doi: 10.1016/j.ergon.2023.103509]
36. Or CK, Karsh BT. A systematic review of patient acceptance of consumer health information technology. J Am Med Inform Assoc 2009;16(4):550-560 [FREE Full text] [doi: 10.1197/jamia.M2888] [Medline: 19390112]
37. Chen T, Chen J, Or CK, Lo FP. Path analysis of the roles of age, self-efficacy, and TAM constructs in the acceptance of performing upper limb exercises through immersive virtual reality games. Int J Ind Ergon 2022 Sep;91:103360. [doi: 10.1016/j.ergon.2022.103360]
38. Fonarow GC, Reeves MJ, Zhao X, Olson DM, Smith EE, Saver JL, et al. Age-related differences in characteristics, performance measures, treatment trends, and outcomes in patients with ischemic stroke. Circulation 2010 Feb 23;121(7):879-891. [doi: 10.1161/CIRCULATIONAHA.109.892497] [Medline: 20142445]
39. Chen J, Or C. Assessing the use of immersive virtual reality, mouse and touchscreen in pointing and dragging-and-dropping tasks among young, middle-aged and older adults. Appl Ergon 2017 Nov;65:437-448. [doi: 10.1016/j.apergo.2017.03.013] [Medline: 28395855]
40. Atkinson HH, Cesari M, Kritchevsky SB, Penninx BW, Fried LP, Guralnik JM, et al. Predictors of combined cognitive and physical decline. J Am Geriatr Soc 2005 Jul 11;53(7):1197-1202 [FREE Full text] [doi: 10.1111/j.1532-5415.2005.53362.x] [Medline: 16108938]
41. Tsai TH, Lin WY, Chang YS, Chang PC, Lee MY. Technology anxiety and resistance to change behavioral study of a wearable cardiac warming system using an extended TAM for older adults. PLoS One 2020;15(1):e0227270 [FREE Full text] [doi: 10.1371/journal.pone.0227270] [Medline: 31929560]
42. Forsberg A, Nilsagård Y, Boström K. Perceptions of using videogames in rehabilitation: a dual perspective of people with multiple sclerosis and physiotherapists. Disabil Rehabil 2015;37(4):338-344 [FREE Full text] [doi: 10.3109/09638288.2014.918196] [Medline: 24833535]

XSL•FO
RenderX

43. Tatla SK, Shirzad N, Lohse KR, Virji-Babul N, Hoens AM, Holsti L, et al. Therapists' perceptions of social media and video game technologies in upper limb rehabilitation. JMIR Serious Games 2015 Mar 10;3(1):e2 [FREE Full text] [doi: 10.2196/games.3401] [Medline: 25759148]
44. Celinder D, Peoples H. Stroke patients' experiences with Wii Sports® during inpatient rehabilitation. Scand J Occup Ther 2012 Sep 20;19(5):457-463 [FREE Full text] [doi: 10.3109/11038128.2012.655307] [Medline: 22339207]
45. Lam MY, Tatla SK, Lohse KR, Shirzad N, Hoens AM, Miller KJ, et al. Perceptions of technology and its use for therapeutic application for individuals with hemiparesis: findings from adult and pediatric focus groups. JMIR Rehabil Assist Technol 2015 Feb 10;2(1):e1 [FREE Full text] [doi: 10.2196/rehab.3484] [Medline: 28582236]
46. Hale LA, Satherley JA, McMillan NJ, Milosavljevic S, Hijmans JM, King MJ. Participant perceptions of use of CyWee Z as adjunct to rehabilitation of upper-limb function following stroke. J Rehabil Res Dev 2012;49(4):623-634 [FREE Full text] [doi: 10.1682/jrrd.2001.04.0070] [Medline: 22773264]
47. Wulf G, Landers M, Lewthwaite R, Töllner T. External focus instructions reduce postural instability in individuals with Parkinson disease. Phys Ther 2009 Feb;89(2):162-168. [doi: 10.2522/ptj.20080045] [Medline: 19074619]
48. Wulf G, McNevin NH, Fuchs T, Ritter F, Toole T. Attentional focus in complex skill learning. Res Q Exerc Sport 2000 Sep;71(3):229-239. [doi: 10.1080/02701367.2000.10608903] [Medline: 10999260]
49. Merians AS, Fluet GG, Qiu Q, Saleh S, Lafond I, Davidow A, et al. Robotically facilitated virtual rehabilitation of arm transport integrated with finger movement in persons with hemiparesis. J Neuroeng Rehabil 2011 May 16;8(1):27 [FREE Full text] [doi: 10.1186/1743-0003-8-27] [Medline: 21575185]
50. Jutai JW, Teasell RW. The necessity and limitations of evidence-based practice in stroke rehabilitation. Top Stroke Rehabil 2015 Feb 02;10(1):71-78. [doi: 10.1310/crda-pgfw-khel-20e1]
51. Teasell R, Meyer MJ, McClure A, Pan C, Murie-Fernandez M, Foley N, et al. Stroke rehabilitation: an international perspective. Top Stroke Rehabil 2009 Aug 04;16(1):44-56. [doi: 10.1310/tsr1601-44] [Medline: 19443347]
52. Dobkin BH. Strategies for stroke rehabilitation. Lancet Neurol 2004 Sep;3(9):528-536 [FREE Full text] [doi: 10.1016/S1474-4422(04)00851-8] [Medline: 15324721]
53. Hebert D, Lindsay MP, McIntyre A, Kirton A, Rumney PG, Bagg S, et al. Canadian stroke best practice recommendations: stroke rehabilitation practice guidelines, update 2015. Int J Stroke 2016 Jun;11(4):459-484. [doi: 10.1177/1747493016643553] [Medline: 27079654]
54. Xiong J, Hsiang EL, He Z, Zhan T, Wu ST. Augmented reality and virtual reality displays: emerging technologies and future perspectives. Light Sci Appl 2021 Oct 25;10(1):216 [FREE Full text] [doi: 10.1038/s41377-021-00658-8] [Medline: 34697292]
55. Guerra V, Rabadan J, Perez-Jimenez R. Suitability of optical wireless communication receivers for virtual reality applications. In: Proceedings of the 15th International Conference on Telecommunications. 2019 Presented at: ConTEL 2019; July 3-5, 2019; Graz, Austria. [doi: 10.1109/contel.2019.8848557]
56. Miller KK, Porter RE, DeBaun-Sprague E, Van Puymbroeck M, Schmid AA. Exercise after stroke: patient adherence and beliefs after discharge from rehabilitation. Top Stroke Rehabil 2017 Mar 23;24(2):142-148. [doi: 10.1080/10749357.2016.1200292] [Medline: 27334684]
57. Maramba I, Chatterjee A, Newman C. Methods of usability testing in the development of eHealth applications: A scoping review. Int J Med Inform 2019 Jun;126:95-104. [doi: 10.1016/j.ijmedinf.2019.03.018] [Medline: 31029270]
58. Cheung DS, Or CK, So MK, Tiwari A. Usability testing of a smartphone application for delivering Qigong training. J Med Syst 2018 Sep 05;42(10):191. [doi: 10.1007/s10916-018-1048-9] [Medline: 30187139]
59. Or C, Tao D. Usability study of a computer-based self-management system for older adults with chronic diseases. JMIR Res Protoc 2012 Nov 08;1(2):e13 [FREE Full text] [doi: 10.2196/resprot.2184] [Medline: 23612015]
60. Karsh BT, Holden RJ, Or CK. Human factors and ergonomics of health information technology implementation. In: Carayon P, editor. Handbook of Human Factors and Ergonomics in Health Care and Patient Safety. Boca Raton, FL: CRC Press; 2012.
61. Tao D, Or C. A paper prototype usability study of a chronic disease self-management system for older adults. In: Proceedings of the IEEE International Conference on Industrial Engineering and Engineering Management. 2012 Presented at: IEEM 2012; December 10-13, 2012; Hong Kong, China. [doi: 10.1109/ieem.2012.6837946]
62. Or CK, Chan AH. Inspection methods for usability evaluation. In: Stephanidis C, Salvendy G, editors. User Experience Methods and Tools in Human-Computer Interaction. Boca Raton, FL: CRC Press; 2024.
63. Nielsen J. Usability Engineering. Cambridge, MA: Academic Press; 1993.
64. Knowles ES, Nathan KT. Acquiescent responding in self-reports: cognitive style or social concern? J Res Pers 1997 Jun;31(2):293-301. [doi: 10.1006/jrpe.1997.2180]
65. van de Mortel TF. Faking it: social desirability response bias in self-report research. Aus J Adv Nurs 2008;25(4):40-48. [doi: 10.3316/ielapa.210155003844269]
66. Laver KE, Lange B, George S, Deutsch JE, Saposnik G, Crotty M. Virtual reality for stroke rehabilitation. Cochrane Database Syst Rev 2017 Nov 20;11(11):CD008349 [FREE Full text] [doi: 10.1002/14651858.CD008349.pub4] [Medline: 29156493]

XSL•FO
RenderX

67. Carlsson E, Paterson BL, Scott-Findlay S, Ehnfors M, Ehrenberg A. Methodological issues in interviews involving people with communication impairments after acquired brain damage. Qual Health Res 2007 Dec 01;17(10):1361-1371. [doi: 10.1177/1049732307306926] [Medline: 18000075]
68. Paterson B, Scott-Findlay S. Critical issues in interviewing people with traumatic brain injury. Qual Health Res 2002 Mar 01;12(3):399-409. [doi: 10.1177/104973202129119973] [Medline: 11918104]
69. Patient experience in adult NHS services: improving the experience of care for people using adult NHS services. National Institute for Health and Care Excellence. 2012 Feb 24. URL: https://www.nice.org.uk/guidance/cg138 [accessed 2024-11-28]

Abbreviations

RCT: randomized controlled trial
VR: virtual reality

Edited by T Leung, G Eysenbach; submitted 12.06.23; peer-reviewed by L Sheehy, H Wang, S Crowe; comments to author 05.12.23; revised version received 29.04.24; accepted 07.11.24; published 01.01.25.

Please cite as:
Chen J, Or CK, Li Z, Yeung EHK, Chen T
Perceptions of Patients With Stroke Regarding an Immersive Virtual Reality–Based Exercise System for Upper Limb Rehabilitation: Questionnaire and Interview Study
JMIR Serious Games 2025;13:e49847
URL: https://games.jmir.org/2025/1/e49847
doi:10.2196/49847
PMID:

©Jiayin Chen, Calvin Kalun Or, Zhixian Li, Eric Hiu Kwong Yeung, Tianrong Chen. Originally published in JMIR Serious Games (https://games.jmir.org), 01.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Improving Pediatric Patients' Magnetic Resonance Imaging Experience With an In-Bore Solution: Design and Usability Study

Annerieke Heuvelink[1], PhD; Privender Saini[2], PDEng; Özgür Taşar[3], MFA; Sanne Nauts[1], PhD

[1]Imaging Innovation, Philips Medical Systems, Eindhoven, Netherlands

[2]Data Science and AI, Royal Philips, Eindhoven, Netherlands

[3]Ambient Experience, Philips Medical Systems, Best, Netherlands

Corresponding Author:
Annerieke Heuvelink, PhD
Imaging Innovation
Philips Medical Systems
High Tech Campus 34
Eindhoven, 5656 AE
Netherlands
Phone: 31 639833471
Email: Annerieke.Heuvelink@philips.com

Abstract

Background: Annually, millions of children undergo a magnetic resonance imaging (MRI) examination. Hospitals increasingly aim to scan young children awake, as doing so benefits both patients and health care systems. To help hospitals reduce the need for anesthesia, we have developed solutions to prepare pediatric patients at home and in the hospital.

Objective: The goal of our project was to design, develop, and test a solution that extends our preparation solutions by guiding and engaging children during their MRI examination.

Methods: Pediatric In-bore was designed to deliver a familiar experience by reusing design elements from our preparation solutions. It offers child-friendly movies and auditory and visual guidance about examination progress and breath holding. To evaluate children's liking and understanding of the solution, we conducted a usability study. Ten healthy children participated in a mock MRI examination featuring pediatric In-bore. We observed task compliance (ability to lie still and hold one's breath) and conducted guided interviews to assess their experience and understanding of the guidance offered.

Results: Participants (aged 5 to 10 years) were generally positive about pediatric In-bore. They liked the main character (Ollie the elephant) and her movie. Auditory and visual guidance were generally liked and understood. All but one participant successfully managed to lie still during the mock examination, and 6 (60%) out of 10 participants successfully held their breath.

Conclusions: Pediatric In-bore appears promising for engaging and guiding young children during awake MRI. It completes the Pediatric Coaching solution that now offers guidance throughout the MRI journey. Future research can expand on this work by evaluating the clinical impact of the Pediatric Coaching solution in a larger and more diverse sample of pediatric patients.

(JMIR Serious Games 2025;13:e55720) doi:10.2196/55720

KEYWORDS

MRI; magnetic resonance imaging; imaging; radiology; pediatrics; children; patient guidance; patient experience; design; usability; breath hold

Introduction

Background

Using magnetic resonance imaging (MRI), physicians can diagnose a host of different pediatric conditions without exposing children to harmful ionizing radiation. Every year, millions of children get an MRI examination. To be able to have a successful examination, children need to enter a room with a large machine, lie down on a table that slides into this machine, and keep very still for an extended period of time (20-40 min, with some examinations taking up to an hour [1]). For some examinations, children need to hold their breath during various scans to enable physicians to create sharp images of the child's heart or abdomen. During the examination, the MRI scanner makes very loud noises. Children report that this noise and boredom are the most stressful factors of having an MRI [2].

XSL•FO
RenderX

Given these challenges, young children are often scanned using general anesthesia or sedation, which enables high-quality scans along with a smooth and predictable workflow. Unfortunately, using anesthesia or sedation is costly, and waiting lists are typically longer for examinations with anesthesia or sedation than for awake examinations [3]. More importantly, there are concerns about the potential effect of repeated anesthesia exposure on children's neurodevelopment (for an overview, see [4-7]). For these reasons, an increasing number of hospitals are trying to scan children awake.

Existing Solutions

There are many initiatives to support hospitals that aim to scan children awake (see [7] for a review). Some initiatives focus on preparing children at home, for example, using a gamified app or virtual reality (eg, [8-13]). Other initiatives focus on preparing children in the hospital, for example, using face-to-face training programs with child life specialists [14,15] or educational (mock) scanners (eg, [16], for a review see [17]). During the MRI examination, children can sometimes watch a movie in the bore of the MRI scanner (eg, using movie goggles or a screen), play a digital game [18,19], or enjoy a child-friendly imaging room (for a review of different approaches, see [20]).

Next to initiatives that focus on one specific touchpoint in a child's diagnostic pathway (such as interventions at home or in the hospital or the MRI room), there are multifaceted concepts that cover the entire pathway from home to examination. An example of such a holistic approach is Children Centered Care [21], which prepares children at home (with an app) and in the hospital (using an educational scanner) and supports them during their MRI examination with a child-friendly imaging room. The Pingunauten program [7,8,18] prepares children at home but also provides an in-bore game featuring the same child-friendly characters and theme (penguins in space) as the Pingunauten app.

To help hospitals reduce the need for anesthesia or sedation, we have previously created a gamified mobile app to prepare children at home (Scan Buddy; [22]). We also redesigned an educational scanner for the waiting room (the Kitten Scanner; [23]) to feature the same design elements, voice, and terminology as Scan Buddy. The main character featured in our solutions is a child-friendly elephant that is well-liked by children: Ollie the Elephant [24]. Consistently using design elements like characters, visuals, voices, and terminology throughout multiple touchpoints in the care pathway can provide children with a sense of familiarity. Providing such a sense of familiarity in an otherwise unfamiliar, stressful medical setting can give children and their parents a sense of comfort, continuity, and safety [25]. Using recognizable elements like characters can also provide staff with a starting point for conversation, helping them build rapport with the child.

Objective

To complete the full Pediatric Coaching solution (Figure 1), we aimed to design, develop, and test a solution that guides and engages children during their MRI examination. The goal was to develop a pediatric In-bore Experience solution (pediatric In-bore) as a variant of an existing In-bore Experience solution using design elements of our preparation solutions. By means of a usability study, we wanted to test whether children would like pediatric In-bore and understand the guidance it offered.

Figure 1. The complete Pediatric Coaching solution consists of the Scan Buddy app (left), enabling preparation at home days prior to the examination; the Kitten Scanner (middle), enabling preparation on-site moments before the examination; and pediatric In-bore (right), providing guidance during the examination and reusing familiar elements from the preparation.



Methods

Design of Pediatric In-bore

Background

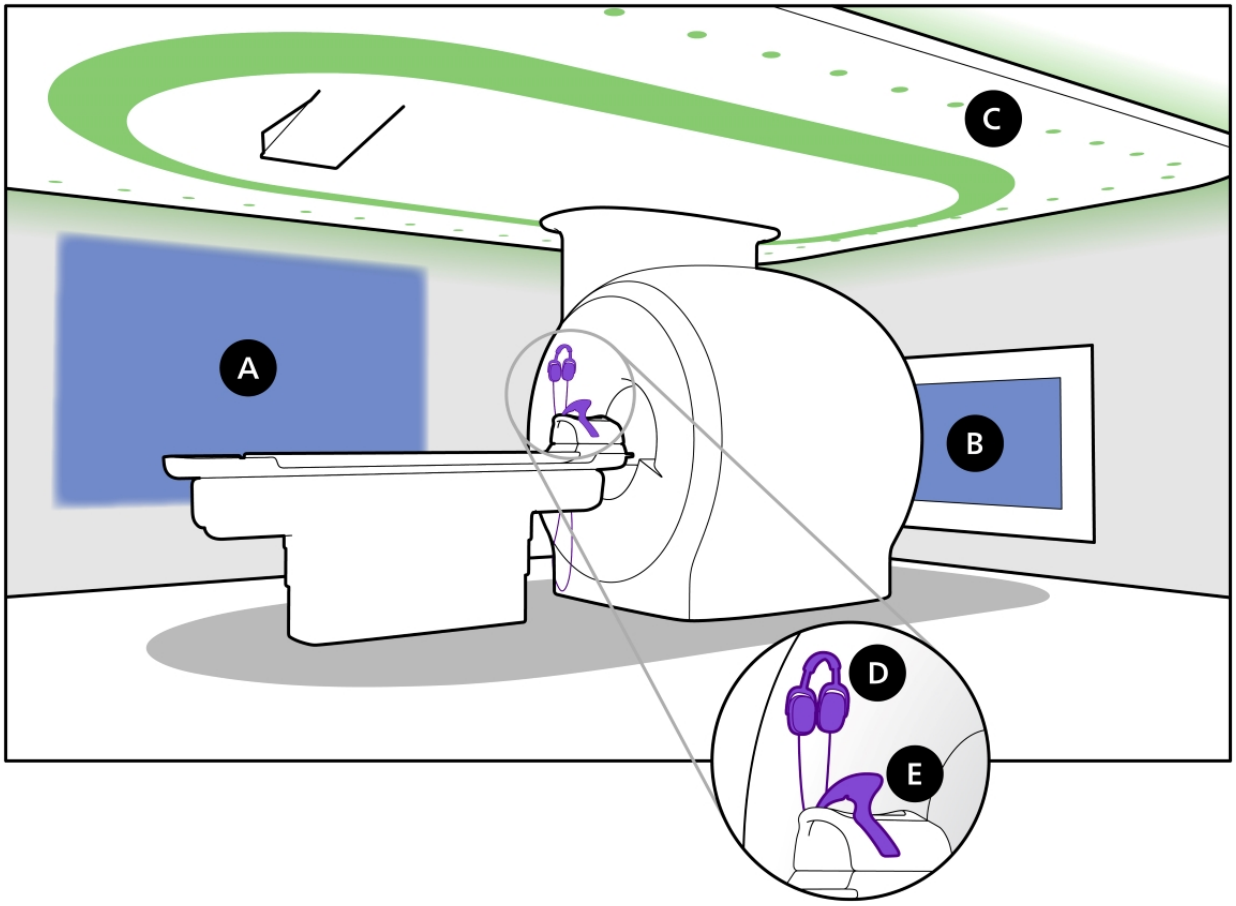
Pediatric In-bore is a child-friendly adaptation of an existing In-bore Experience solution that engages and guides patients during their examination (see Figure 2 for an overview of its hardware elements). Supporting patients in the MRI room is important, as patients' physiological anxiety levels are relatively high at the beginning of an MRI examination, with a peak when patients enter the bore [26]. Providing patients with information during the examination (eg, to inform them about how long the examination lasts), is associated with lower cortisol levels (an important indicator of physiological stress; [27]).

XSL•FO
RenderX

With In-bore Experience, patients can look outside the bore during the MRI examination using a mirror. In this mirror, they can see a screen showing audiovisual content, such as a movie. Patients wear headphones to hear the music or audio accompanying the content displayed on the screen. They can also use these headphones to communicate with staff and to hear automated voice messages. These automated messages ("AutoVoice") instruct patients about when to hold still because the table moves and how to breathe during a breath-hold scan. They also inform patients about the duration of the upcoming scan (eg, "The next scan will last for three minutes"). Through a connection with the MRI scanner ("In-bore Connect"), information about the progress of the examination and breath holds can be shown as overlays on the audiovisual content.

For pediatric In-bore, we adapted three distinct elements of the In-bore Experience solution: the audiovisual themes, AutoVoice, and In-bore Connect. Below, we describe the design rationale for the pediatric version of each of these elements.

Figure 2. The hardware elements enabling In-bore Experience are (A) wall projection; (B) In-bore screen; (C) dynamic lighting that matches the colors of the visual content displayed on the wall projection or screen; (D) headphones; and (E) mirror.



Themes

Themes consist of a video with sounds and music for the wall projection and In-bore screen, and associated lighting patterns for the dynamic lightning (Figure 2). There are themes for adults, as well as children, and patients can choose which theme they want to watch.

To complement the Pediatric Coaching solution, new videos were created featuring three characters: Ollie the Elephant, Chris the Crocodile, and Doris the Chicken. Children have been familiarized with these characters with the Kitten Scanner and (for Ollie) with the Scan Buddy app. Each character has his or her own story (Multimedia Appendix 1). The goal of themes is to engage children and prevent boredom and as such support them with lying still. The themes were designed with a storyline (eg, "Ollie's space journey") that consists of several substories (eg, "Ollie visits an underwater planet" and "Ollie visits a mushroom planet") separated by a repetitive scene (eg, "Ollie in her rocket"). This setup was chosen because the length of an MRI examination can differ considerably; the modular setup ensures that children do not miss an exciting part of a story, because the examination finished midway.

Moreover, given that for some head and brain scans, even minor head motion can negatively affect image quality [28], the themes were designed with one central focus. Peripheral cues can draw attention, involuntarily directing people's gaze toward the cue [29]. Some research states that such reflexive eye movements can cause motion artifacts [30]. As such, themes that avoid inducing many eye movements are preferable to themes that may induce a lot of eye movement.

XSL•FO
RenderX

AutoVoice

Background

Pediatric AutoVoice is a version of AutoVoice redesigned to provide a consistent experience throughout the Pediatric Coaching solution. The existing "adult" AutoVoice was transformed into a pediatric version tailored to the needs of young children by making several changes elaborated below.

Voice

For all 15 languages in which pediatric AutoVoice was created, we worked with the same voice actors as in the Scan Buddy app and Kitten Scanner. This was done to ensure that children hear the same voice at home, in the hospital, and during the MRI scanner. While adult AutoVoice has a practical, professional tone of voice, pediatric AutoVoice has an encouraging, lively tone of voice, using more inflection.

Language Use

The language used in pediatric AutoVoice was tailored to young children by:

- Using informal language for those languages that have a distinction between the formal and informal language (eg, "Sie" or "du" in German; "u" or "jij" in Dutch).
- Using the terminology used throughout the Pediatric Coaching solution, which is a child-friendly version of the terminology used for adults. For example, where adult AutoVoice uses terminology such as "table" and "scan," pediatric AutoVoice uses terminology such as "bed" and "picture."
- Using generic time indicators to indicate the duration of the next scan. For example, when adult AutoVoice says "The next scan will last for half a minute," pediatric AutoVoice says "The next picture will be really quick."

The announcements of scan duration were adapted because research shows that children younger than 8 years of age typically do not have an explicit sense of time that can be measured in concrete time units such as minutes or hours [31]. This is in line with Piaget's [32] classic theory, which posits that children start to understand abstract concepts and symbolic representations such as time around age 7-8 years, when they move into the concrete-operational stage of development.

Timing

Providing automated breathing instructions is an important feature of adult AutoVoice. During a so-called breath-hold scan, the patient is instructed how to breathe and when to hold, most commonly with the instructions "Breathe in ... breathe out ... and hold your breath," after which the scanner starts acquiring data. When the acquisition is finalized, the patient is told "You may breathe again."

On average, children have higher breathing rates than adults [31]; as such, the AutoVoice timing was adjusted to better match children's natural breathing patterns. In adult AutoVoice, on average, patients are instructed to breathe in and breathe out in approximately 2.5 seconds each (the exact length can differ across languages). This resembles a breathing frequency of 12 breaths per minute (BPM), which is at the low end of the natural breathing range spectrum for adults at rest.

For pediatric AutoVoice, we ensured that instructions had the same timing for all languages and shortened them to accommodate children's natural breathing rates. In particular, the "Breathe in ..." and "breathe out ..." instructions were timed 2 seconds apart, in accordance with a breathing frequency of 15 BPM. With these instructions, we enable children to deeply breathe in a rhythm that matches their age, as 15 BPM is at the low end of the natural breathing range spectrum for young children at rest [33].

In-bore Connect

Background

In-bore Connect enables real-time visual patient guidance through a link with the MRI scanner. From the information received in real-time from the scanner about breath-hold scans and scan progress, overlays are constructed that are superimposed on the content displayed on the screen.

Breath-Hold Guidance (Visual)

When a breath-hold scan is coming up, AutoVoice informs the patient that the next scan is a breath-hold scan. To capture the patient's attention, the content is dimmed, the progress bar hidden, and a "breathe" icon is displayed in the center of the screen. The moment the patient is instructed to hold their breathing, the icon changes and the circle that surrounds it is progressively filled to indicate the remaining time (Multimedia Appendix 2).

For pediatric In-bore Connect, the visuals of the breath-hold guidance were redesigned to have a more familiar and playful component. Upon announcing the breath-hold scan, the pediatric "breathe" icon is accompanied by Ollie, who encouragingly points to it to support the child to pay attention to what is coming next. The moment the child is instructed to hold their breath, the icon changes and the circle is animated to indicate the remaining time. Upon completing the breath hold, the icon changes back to the "breathe" icon and stars appear as a small reward for completing the breath hold (Multimedia Appendix 3).

Progress Bar

During the MRI examination, In-bore Connect enables the overlay of a progress bar at the bottom of the screen on top of the content shown. The progress bar indicates the total number of scans, their length, and the examination progress (Multimedia Appendix 4).

For pediatric In-bore, the progress bar was redesigned in such a way that it appeals to children and has the same familiar look and feel as the progress bar introduced in the Scan Buddy app and the Kitten Scanner (Multimedia Appendix 5).

Animated Ollie

Next to redesigning the breath-hold guidance and progress bar overlays, pediatric In-bore Connect introduces an animated Ollie as the third type of overlay. This animated Ollie is visible at the start and the end of the examination and displays various gestures like waving to the patient (Multimedia Appendix 6).

XSL•FO
RenderX

The goal of showing Ollie at the start of the examination is to actively engage and distract young children when they lie down on the table and the table slides into the MRI scanner, as this is a moment of high anxiety for patients [26]. As soon as the first scan starts, Ollie disappears and the progress bar starts filling, informing the child about the progress of the examination. When all the scans are finished and the progress bar has completely filled, Ollie pops up again. In this way, Ollie creates a small celebratory moment when the examination has been completed. She also bridges the time between finishing the last scan, and the operator coming into the room and moving the child out of the scanner.

Pediatric In-bore Usability Study

Study Design

To investigate whether children would like and understand pediatric In-bore, we conducted a usability study with healthy participants. During a mock examination, participants were exposed to the different elements of pediatric In-bore: a pediatric theme, pediatric AutoVoice (in their native language), and the three overlays of pediatric In-bore Connect.

The mock examination was conducted in a usability laboratory featuring a mock scanner that looks exactly like a functional MRI scanner (Figure 3). To simulate the full pediatric In-bore Experience during a real MRI examination, all the visuals (theme, as well as overlays) and sounds (MRI scanner sounds, as well as AutoVoice and theme sounds) were prerecorded and assembled into one video. The video was displayed on a screen behind the bore, which children could see and hear using an MRI mirror and headphones. After the mock examination, the children participated in a guided interview.

Figure 3. Picture of the mock MRI scanner in the usability laboratory, where the ~10-minute mock examination with the healthy participants (N=10; aged 5-10 years) of the pediatric In-bore usability study took place. MRI: magnetic resonance imaging.



Each study session was facilitated by two researchers, one led the session and one observed and logged participants' responses. Sessions were recorded to help code participants' behavior inside the bore.

The study was conducted over the course of 4 weeks from mid-September to mid-October 2022 and lasted a maximum of 1 hour per participant.

Participants

Ten participants were recruited by an external recruitment agency that specializes in recruiting participants for qualitative research studies. They recruited the participants from their existing panel. Recruitment criteria were healthy children between the ages of 4 and 10 years, and speaking native English or Dutch. Participation was only allowed if the child was

XSL•FO
RenderX

accompanied by a parent or legal guardian. Parents received a small compensation from the recruitment agency in the form of voucher points, while we gave the children a small gift package related to the Pediatric Coaching solution (small elephant snuggle, stickers, magnets, and a coloring book) after participation.

Ethical Considerations

Ethical approval for the study and all required documents, including the information letter, informed consent (with a section on privacy and confidentiality), and interview guide, were obtained from the Philips Internal Committee for Biomedical Ethics (study ICBE-S-000879), ensuring the study was conducted in accordance with the Declaration of Helsinki. The recruitment agency sent the information letter and informed consent to parents as part of recruitment, and consent was obtained prior to the study. During the introduction of the study on-site, child assent was also obtained by the researchers. Raw data such as video files and paper forms with identifying information were deleted or otherwise destroyed after extracting the relevant information and digitizing the data without identifying information other than gender, age, and participation number.

Outcomes

The main goal of the study was to investigate to what extent pediatric In-bore was liked and the guidance it provides understood by young children. As a secondary outcome, we gauged the ability of the child to lie still during the examination and to follow the breath-hold instructions.

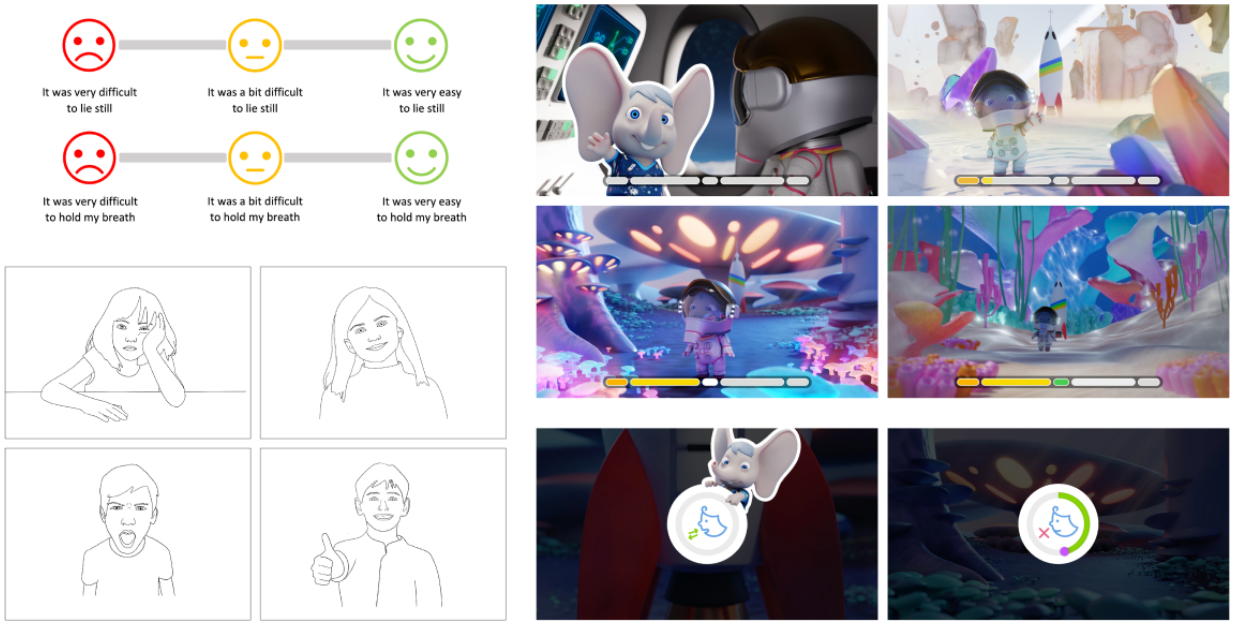
Study Procedure and Material

Study sessions all started with a general introduction of what an MRI machine is and what it is used for, followed by an explanation of the study procedure. After that, the informed consent was signed by the parent, and a walk-through of child assent was done.

The main part of the study consisted of a mock examination of approximately 10 minutes. First, the child was equipped with a small camera on their head to enable the observers to inspect the behavior of the child when lying in the MRI scanner. Next, they were asked to climb onto the MRI bed. They were asked to lie as still as they could, like children undergoing a real MRI, and to watch the video and follow any instructions that were provided in the video. Then, they were donned with headphones so they could hear the audio embedded in the study video. The mirror was positioned, after which the bed was moved into the bore. While lying in the bore, the child watched the study video mimicking an MRI examination containing 5 MRI scans including 1 breath-hold scan.

Throughout the study, the Ollie theme was playing full screen, accompanied by the progress bar displaying the examination progress (Figure 4, top right). Pediatric AutoVoice announced the duration for each of the scans without breath holds and provided instructions throughout the breath-hold scan. The middle scan was the breath-hold scan containing 4 breath holds of varying lengths (4-10 s each), during which the breath-hold guidance was shown. During breath-hold guidance, the theme was dimmed and the progress bar was invisible (Figure 4, bottom right). At the start and the end of the study video, an Animated Ollie was shown (Figure 4, top middle).

Figure 4. Examples of the rating scale and probing material (emotion line drawings, screenshots of the video) used during the ~35-minute guided interview with the healthy participants (N=10; aged 5-10 years) of the pediatric In-bore usability study.



The last part of the study was a guided interview to learn about the child's experience during the mock examination and their opinion about the various elements of pediatric In-bore. This part lasted for approximately 35 minutes. To facilitate a smooth interview, screenshots of the study video were used to jog participants' memory (Figure 4, right), and printed probes—line

XSL•FO
RenderX

drawings of children with various emotions—were used to encourage talking and expression of feelings (Figure 4, bottom left). Three-point smiley rating scales were used to get an indication of how difficult or easy tasks were (Figure 4, top left).

Measures

To estimate how well the child executed the various parts of the mock examination ("Tasks"), children's behavior was observed both directly, as well as during data analysis (using the video recording made inside the bore). For each task (see Table 1 for an overview of the tasks), it was coded if the child was able to perform the task "successfully (without any issues)," "successfully with minor issues," or "unsuccessfully (with major issues)".

Table 1. The tasks embedded in the ~10-minute mock examination of the pediatric In-bore usability study in chronological order, and the intended behavior of the healthy participants (N=10; aged 5-10 years) for each of them.

Tasks	Intended behavior
Lying down, going into the bore, and viewing animated Ollie	Going in without issues; watching the animation without movement
Lying still during the first two scans	Lying still without movement
Executing the breath holds during the middle, breath-hold scan	Correctly executing the four breath holds (after exhaling, holding their breath for the indicated duration)
Lying still during the last two scans, and viewing animated Ollie	Lying still without movement

For all tasks, small (localized) movements of the extremities (like wiggling toes or moving fingers) were coded as "minor issues," as for the large majority of scans, these movements will not lead to images that are of nondiagnostic quality. Large body movements (shifting, rolling, arm gestures, attempting to get up) that are likely to have a major impact on the diagnostic quality of the scan were coded as "major issues."

For the breath-hold task, task compliance was coded as "successful with minor issues" if the child conducted the breath holds, but (partially) did so in a wrong way (eg, after inhaling instead of exhaling). It was coded as "unsuccessful (with major issues)" if, for one or more of the breath holds, the instructions were observably not followed (no attempt to hold the breath or not able to maintain it).

The guided interview after the mock examination contained open questions and questions the child could answer by pointing to a position on a printed slider showing three emoticons (Figure 4, top left for an example). Besides general questions about the mock examination, specific questions were asked about the various elements of pediatric In-bore (see Multimedia Appendix 7 for a full overview of the self-reported outcome measures). For data analysis, the sentiment of the answers to the open questions was coded to a 3-point scale (negative, neutral, and positive). For example, if a child said "I love Ollie!" this answer would be coded as positive for likeability of Ollie. The slider position was coded as a 5-point scale.

Results

Participants

Ten healthy children (4 male and 6 female) aged 5 to 10 (mean age 7.6, SD 1.9) years, who had never undergone an MRI, participated in the study. All the participants underwent the mock examination and participated in the guided interview. Despite the supporting interview material, the two 5-year-old participants did not always seem to understand the question, nor were they always able to clearly communicate their opinions to the investigators, especially at the end of the interview. When possible, their answers were scored (this frequently resulted in a negative or neutral or did not notice score), while for some outcomes, no data are available at all (in these cases, N<10).

Appreciation of Pediatric In-bore and its Elements

Overview

Table 2 provides the overview of the opinions voiced by children during the guided interview in the open questions. Table 3 provides the data from the slider scales.

XSL•FO
RenderX

Table 2. Self-reported outcomes of the healthy participants (N=10; aged 5-10 years) interpreted from the open questions asked during the ~35-minute guided interview of the pediatric In-bore usability study.

	Negative, n	Neutral, n	Positive, n	Did not notice, n
Overall examination experience (n=10)	0	2	8	N/A[a]
Likability of Ollie (n=10)	0	2	8	N/A
Likability of the Ollie theme (n=10)	0	2	8	N/A
Likability of the voice (n=10)	0	3	7	N/A
Understanding of verbal (scan duration) information (n=10)	1	1	7	1
Understanding of verbal breath-hold guidance (n=10)	0	1	9	N/A
Understanding of visual breath-hold guidance (n=10)	0	0	10	N/A
Understanding the progress bar (n=10)	0	0	8	2
Understanding the intent of animated Ollie at the start (n=10)	1	1	8	N/A
Understanding the intent of animated Ollie at the end (n=10)	0	0	8	2

[a]Not applicable.

Table 3. Self-reported outcomes of the healthy participants (N=10; aged 5-10 years) measured with sliders during the ~35-minute guided interview of the pediatric In-bore usability study.

	No, n	Not really, n	Do not know, n	Think so, n	Yes, n
Knowing when to lie still (n=9)	2	0	0	0	7
Knowing what to do (n=10)	1	0	1	0	8
Pediatric In-bore considered helpful (n=10)	1	0	1	2	6
Knowing when the first scan started (n=9)	2	0	0	0	7
Knowing when the last scan ended (n=8)	0	0	0	0	8

Ollie

Ollie and her theme were generally liked, with the same two participants (a 5-year-old and an 8-year-old) scoring neutral for both outcomes. The animated Ollie (Figure 4, top middle) was not clearly identifiable or memorable for all participants in comparison to the Ollie shown in the theme. When asking the participants about animated Ollie and what they thought Ollie was doing, they often started to explain the theme's storyline. By explicating that we were interested in what they thought the animated Ollie was doing, the majority of participants (8/10, 80%) correctly indicated that Ollie was waving hi, or welcoming them at the start of the examination, and waving goodbye, or applauding them at the end; for example, a participant said:

> *He means like good job. I saw it because of the wink, because when you wink it often means you did a good job.*

Progress Bar

A total of 8 (80%) out of 10 participants understood the progress bar and were able to explain it in their own words:

> *That was really nice. Because, when it was a short one, it has a short bar, and the long one that took a long time, there was a long one. And every time it made a different sound with the bar…It was really handy, because you sort of knew when you'd be done.*

Both 5-year-old participants mentioned not seeing or noticing the progress bar, and they either did not answer, or answered negatively to the questions about whether they knew when the first scan started, or the last scan finished.

Task Success

Overview

As a secondary outcome, task success was coded. One participant (a 5-year-old child) executed the mock examination with major issues on all tasks, while the rest of the 9 participants were able to perform most of the tasks without any issues or with minor issues (Table 4).

XSL•FO
RenderX

Table 4. Observed task success of the healthy participants (N=10; aged 5-10 years) during the ~10-minute mock examination of the pediatric In-bore usability study.

	Unsuccessful (with major issues), n	Successful with minor issues, n	Successful (without any issues), n
Lying down, going into the bore, and viewing animated Ollie (n=10)	1	0	9
Lying still during the first two scans (n=10)	1	4	5
Executing the breath holds during the breath-hold scan (n=10)	4	4	2
Lying still during the last two scans, and viewing animated Ollie (n=10)	1	3	6

Next to observing their behavior, we also asked children about their perceived ability to lie still and hold their breath (see Table 5). The participant who was unable to complete the tasks successfully (due to large body movements) indicated that she found it very difficult to lie still, and she failed to answer the question about holding one's breath.

Table 5. Self-reported task success of the healthy participants (N=10; aged 5-10 years) during the ~10-minute mock examination of the pediatric In-bore usability study.

	Very difficult, n	Difficult, n	A bit difficult, n	Pretty easy, n	Very easy, n
Perceived ability to lie still (n=10)	1	0	0	3	6
Perceived ability to hold breath (n=9)	0	0	2	3	4

Breath-Hold Scan

Of all the tasks, correct execution of the 4 breath holds during the breath-hold scan was clearly the most difficult. Only 2 participants (a 6-year-old and a 9-year-old) perfectly held their breath after exhaling for all 4 breath holds. Four participants did successfully hold their breath during all breath holds, but not always in the exact way AutoVoice indicated, while 4 participants did not correctly hold their breath for 1 or 2 of the 4 breath holds.

The AutoVoice said "Breathe in ... breathe out and hold your breath" with the visual guidance supporting the transition from breathing to the breath hold (Multimedia Appendix 3). Objectively, we observed the children make solid attempts, but frequently they took a quick breath in before holding, instead of pausing breathing after exhaling as was instructed. We also noted that some children displayed a learning curve, taking a bigger breath when instructed to breathe in and out during the later breath holds compared to the earlier ones.

When discussing their performance, several children—including ones who paused their breathing at the correct moment—questioned the accuracy of the instructions. For example, one participant remarked:

> *Hold your breath, I did not get that completely. Because it was breathe in, breathe out, and hold your breath, but how is that possible? Because your breath is already gone!*

Emotionally, the breath-hold scan was one of the most discussed and commented parts of the mock examination, with some participants expressing feelings of pride in being able to hold their breath and others elaborating on similar breath-holding experiences, for example, during swimming underwater.

Discussion

Principal Findings

We developed a pediatric In-bore Experience solution using design elements of our preparation solutions, thereby creating a holistic coaching solution to support patients at home, in the hospital, and during their MRI. The usability study demonstrated that overall, young children were positive about pediatric In-bore and indicated that it helped them during the examination. Ollie and the theme were liked, and pediatric AutoVoice and accompanying breath-hold guidance and progress bar visual overlays were generally liked and understood by most children. All but one child successfully completed the majority of the tasks embedded in the mock examination, that is, lying down on the bed, entering the bore, and lying still during the scans, and most participants successfully held their breath.

Breath-Hold Scans

One of the most challenging aspects of the mock examination was the breath-hold scan, which is required for a small subset of all MRI examinations: namely, those of anatomies that are affected by breathing motion, such as cardiac scans and liver scans. Breath-hold scans make up a small percentage of all scans since the majority of pediatric examinations are head scans or brain scans (see eg, [34]). Nevertheless, we wanted to provide some support for this type of scan because it is known to be a challenging examination for (pediatric) patients.

In this usability study, we instructed participants to hold their breath after exhaling, as this is the common approach in clinical practice. The reason for this is that holding one's breath after exhaling is—when executed correctly—more reproducible than holding it after inhaling, thus resulting in better image quality (see eg, [35]). However, since holding one's breath is familiar to children after inhaling (this is what many children learn during swimming lessons), the counterintuitive instructions of holding one's breath after exhaling can be difficult to adhere to. This is consistent with anecdotal evidence indicating that

XSL•FO
RenderX

experienced MRI technologists often opt to conduct breath-hold scans after inhaling for young patients.

Familiarizing the child with the instructions and expected behavior during breath-hold scans and practicing breath holds before their examination is also likely to improve their breath-hold execution [36,37]. Integrating such a feature in patient preparation solutions is therefore recommended based on the results of our usability study.

Technical Advancements that Support Awake Scans

In addition to improved preparation and guidance, there are ongoing technical advancements in MRI that are making it easier to scan young children awake. For example, breath holds are becoming shorter [38], and for an increasing number of breath-hold scans, free breathing alternatives are being developed, yielding acceptable diagnostic quality that is, in many cases, comparable to the diagnostic quality of breath-hold scans [39-42]. Moreover, emerging artificial intelligence–based techniques provide accelerated acquisition and motion robustness for MRI [43]. The technical advancements that shorten MRI scans and make them more resilient to motion, coupled with effective patient preparation and guidance solutions, create a powerful combination to support the minimization of sedation or anesthesia use in pediatric MRI.

Limitations

In this study, one child found it very difficult to lie still in the scanner and to follow the instructions. This could be partially because the participants were not prepared for the mock examination. Pediatric In-bore is designed to be part of a broader solution to support children in their MRI journey, consisting of preparation at home (Scan Buddy app), in the hospital (Kitten Scanner), and guidance during the examination (pediatric In-bore), but this usability study was limited in that it only included the last part of the solution. More research is needed to investigate the impact of the full Pediatric Coaching solution on patient experience and the clinical workflow.

Another limitation of this study was that only children between the ages of 5 and 10 years were included, while pediatric In-bore may also be valuable for younger and older children. We did not find marked age differences in liking pediatric In-bore or Ollie, but this could be due to the small sample size: we only included 10 participants. The generalizability of our findings is particularly limited because there are large intraindividual differences between children in terms of, for example, their developmental level, preferences, coping styles, and ability to understand instructions. This diversity between children is even more pronounced within a hospital setting, in which children's chronological age may not necessarily align with their developmental level (eg, in the case of children with cognitive impairments [44]). Conceivably, in clinical practice, there may be some teenagers or even adults (eg, those with cognitive impairment) who prefer pediatric In-bore over the adult In-bore Experience.

In the end, any patient preparation or guidance will be most effective if a certified child life specialist, play therapist, or other staff member (such as an MRI technologist or the child's referring physician) can tailor it to patients [44]. For stronger conclusions on the liking and acceptance of pediatric In-bore in the pediatric patient population, more research is needed in a larger and more diverse sample of pediatric patients.

Conclusions

The pediatric In-bore Experience solution includes child-friendly themes, automated voice instructions tailored to children, and three types of informative and engaging overlays; all were designed to support and guide young children during their MRI. Pediatric In-bore seems a promising approach to engage and guide young children during awake MRI examinations, as this usability study showed that it was liked and understood by most of the participants aged 5-10 years.

Although patient preparation and guidance may strongly reduce the need for anesthesia or sedation in pediatric MRI, we expect that there will always be children for whom it will be challenging to keep still during the scan. This is in line with previous studies that suggest that most, but not all children are able to undergo a successful awake examination (even if they have been prepared for their examination). For example, both Geuens et al [24] and Runge et al [21] report a success rate of 95% (in these studies, 5% of children could not successfully be scanned awake). If a child is scheduled for an awake examination but ends up requiring anesthesia, this can be challenging for hospitals, as well as for children and their families.

Pediatric In-bore complements the Pediatric Coaching solution, which further consists of the Scan Buddy app (to prepare children at home [23]) and the Kitten Scanner (to prepare children in the hospital [24]). More research is needed to investigate the impact of the full Pediatric Coaching solution on patient experience and the clinical workflow. Given the ongoing debate on the safety of repeated use of anesthesia in young children, it is crucial to find solutions to scan children awake that are safe, effective, and child-friendly. Holistic solutions that can prepare children for their examinations and guide them throughout the process can provide hospitals with new tools to reduce the need for anesthesia or sedation in pediatric MRI, although more research is needed to fully understand how such solutions affect anesthesia and sedation rates in a hospital setting.

Acknowledgments

We would like to thank Alexandra Kovács, Bärbel van Zanten, Valerié Drost, Frank van Houten, and Christel de Bakker for their support in carrying out the usability study, and Elizabeth van Vorstenbosch-Lynn, Simon de Turck, and Vincent Hooijer for their help in designing and testing pediatric In-bore.

XSL•FO
RenderX

Data Availability

The data sets generated during or analyzed during this study are not publicly available due to the company's privacy policy.

Conflicts of Interest

The authors are employees of Philips.

Multimedia Appendix 1
Screenshots of the three pediatric themes created for pediatric In-bore Experience featuring the characters of the Pediatric Coaching solution: Ollie the elephant (top); Chris the crocodile (middle); and Doris the chicken (bottom).
[PNG File , 1366 KB - games_v13i1e55720_app1.png]

Multimedia Appendix 2
Screenshots of the visual breath-hold guidance provided by the regular In-bore Experience solution through In-bore Connect: alternately displaying a "breathe" icon (top) and a "hold breath" icon + animated progress circle (bottom).
[PNG File , 2539 KB - games_v13i1e55720_app2.png]

Multimedia Appendix 3
Screenshots of the visual breath-hold guidance provided by pediatric In-bore through In-bore Connect: breath-hold scan announcement displaying the "breathe" icon and Ollie (top); breath-hold progress displaying the "hold breath" icon and the progress (middle); and breath hold completed displaying the "breathe" icon and celebrative stars (bottom).
[PNG File , 711 KB - games_v13i1e55720_app3.png]

Multimedia Appendix 4
Example progress bar provided by the regular In-bore Experience solution through In-bore Connect that shows the following: the first two scans have been completed, the third scan is ongoing, and three scans are remaining.
[PNG File , 1613 KB - games_v13i1e55720_app4.png]

Multimedia Appendix 5
Illustration of the colorful examination progress bar as a recurring design element in the Pediatric Coaching solution: in the Scan Buddy app (top); the Kitten Scanner (middle); and pediatric In-bore (bottom).
[PNG File , 922 KB - games_v13i1e55720_app5.png]

Multimedia Appendix 6
Screenshots of pediatric In-bore displaying that at the start (top) and end (bottom) of the MRI examination Ollie the elephant actively welcomes and says goodbye to the child. MRI: magnetic resonance imaging.
[PNG File , 1346 KB - games_v13i1e55720_app6.png]

Multimedia Appendix 7
Structured overview per element of pediatric In-bore of the outcome measures and their question type that were embedded in the ~35-minute guided interview with the healthy participants (N=10; aged 5-10 years) of the pediatric In-bore usability study.
[DOCX File , 35 KB - games_v13i1e55720_app7.docx]

References

1. Hogan D, DiMartino T, Liu J, Mastro KA, Larson E, Carter E. Video-based education to reduce distress and improve understanding among pediatric MRI patients: a randomized controlled study. J Pediatr Nurs 2018;41:48-53. [doi: 10.1016/j.pedn.2018.01.005] [Medline: 29370960]
2. Inhestern L, Herrmann J, Schürmann J, Meister RL, Nawka MT, Mynarek M, et al. Child-centredness in paediatric magnetic resonance imaging: information needs and experiences of children requiring magnetic resonance imaging and their parents. Child Care Health Dev 2024;50(1):e13157. [doi: 10.1111/cch.13157] [Medline: 37581953]
3. Vanderby SA, Babyn PS, Carter MW, Jewell SM, McKeever PD. Effect of anesthesia and sedation on pediatric MR imaging patient flow. Radiology 2010;256(1):229-237. [doi: 10.1148/radiol.10091124] [Medline: 20505061]
4. Artunduaga M, Liu CA, Morin CE, Serai SD, Udayasankar U, Greer MLC, et al. Safety challenges related to the use of sedation and general anesthesia in pediatric patients undergoing magnetic resonance imaging examinations. Pediatr Radiol 2021;51(5):724-735 [FREE Full text] [doi: 10.1007/s00247-021-05044-5] [Medline: 33860861]

XSL•FO
RenderX

5. Ing C, Warner DO, Sun LS, Flick RP, Davidson AJ, Vutskits L, et al. Anesthesia and developing brains: unanswered questions and proposed paths forward. Anesthesiology 2022;136(3):500-512 [FREE Full text] [doi: 10.1097/ALN.0000000000004116] [Medline: 35015802]
6. Useinovic N, Jevtovic-Todorovic V. Controversies in anesthesia-induced developmental neurotoxicity. Best Pract Res Clin Anaesthesiol 2023;37(1):28-39 [FREE Full text] [doi: 10.1016/j.bpa.2023.03.004] [Medline: 37295851]
7. Thestrup J, Hybschmann J, Madsen TW, Bork NE, Sørensen JL, Afshari A, et al. Nonpharmacological interventions to reduce sedation and general anesthesia in pediatric MRI: a meta-analysis. Hosp Pediatr 2023;13(10):e301-e313. [doi: 10.1542/hpeds.2023-007289] [Medline: 37727937]
8. Liszio S, Masuch M. Virtual reality MRI: playful reduction of children's anxiety in MRI exams. 2017 Presented at: IDC '17: Proceedings of the 2017 Conference on Interaction Design and Children; June 27-30, 2017; Stanford, CA p. 127-136 URL: https://doi.org/10.1145/3078072.3079713 [doi: 10.1145/3078072.3079713]
9. Liszio S, Graf L, Basu O, Masuch M. Pengunaut trainer: a playful VR app to prepare children for MRI examinations—in-depth game design analysis. 2020 Presented at: IDC '20: Proceedings of the Interaction Design and Children Conference; June 21-24, 2020; London, United Kingdom p. 470-482. [doi: 10.1145/3392063.3394432]
10. Stunden C, Stratton K, Zakani S, Jacob J. Comparing a virtual reality-based simulation app (VR-MRI) with a standard preparatory manual and child life program for improving success and reducing anxiety during pediatric medical imaging: randomized clinical trial. J Med Internet Res 2021;23(9):e22942 [FREE Full text] [doi: 10.2196/22942] [Medline: 34550072]
11. van Spaendonck Z, Leeuwenburgh KP, Dremmen M, van Schuppen J, Starreveld D, Dierckx B, et al. Comparing smartphone virtual reality exposure preparation to care as usual in children aged 6 to 14 years undergoing magnetic resonance imaging: protocol for a multicenter, observer-blinded, randomized controlled trial. JMIR Res Protoc 2023;12:e41080 [FREE Full text] [doi: 10.2196/41080] [Medline: 36692931]
12. McGlashan HL, Dineen RA, Szeszak S, Whitehouse WP, Chow G, Love A, et al. Evaluation of an internet-based animated preparatory video for children undergoing non-sedated MRI. Br J Radiol 2018;91(1087):20170719 [FREE Full text] [doi: 10.1259/bjr.20170719] [Medline: 29688030]
13. Olloni SS, Villadsen N, Mussmann B. Pediatric MRI without anesthesia: the effect of application-supported communication to prepare the child. J Radiol Nurs 2021;40(1):56-60. [doi: 10.1016/j.jradnu.2020.09.002]
14. Szeszak S, Man R, Love A, Langmack G, Wharrad H, Dineen RA. Animated educational video to prepare children for MRI without sedation: evaluation of the appeal and value. Pediatr Radiol 2016;46(12):1744-1750. [doi: 10.1007/s00247-016-3661-4] [Medline: 27568023]
15. Bharti B, Malhi P, Khandelwal N. MRI customized play therapy in children reduces the need for sedation—a randomized controlled trial. Indian J Pediatr 2016;83(3):209-213. [doi: 10.1007/s12098-015-1917-x] [Medline: 26477350]
16. Fletcher S, Lardner D, Bagshawe M, Carsolio L, Sherriff M, Smith C, et al. Effectiveness of training before unsedated MRI scans in young children: a randomized control trial. Pediatr Radiol 2023;53(7):1476-1484. [doi: 10.1007/s00247-023-05647-0] [Medline: 37010547]
17. Suzuki A, Yamaguchi R, Kim L, Kawahara T, Ishii-Takahashi A. Effectiveness of mock scanners and preparation programs for successful magnetic resonance imaging: a systematic review and meta-analysis. Pediatr Radiol 2023;53(1):142-158. [doi: 10.1007/s00247-022-05394-8] [Medline: 35699762]
18. Liszio S, Basu O, Masuch M. A universe inside the MRI scanner: an in-bore virtual reality game for children to reduce anxiety and stress. 2020 Presented at: CHI PLAY '20: Proceedings of the Annual Symposium on Computer-Human Interaction in Play; November 2-4, 2020; Virtual event, Canada p. 46-57. [doi: 10.1145/3410404.3414263]
19. Qian K, Arichi T, Price A, Dall'Orso S, Eden J, Noh Y, et al. An eye tracking based virtual reality system for use inside magnetic resonance imaging systems. Sci Rep 2021;11(1):16301 [FREE Full text] [doi: 10.1038/s41598-021-95634-y] [Medline: 34381099]
20. Anwar I, McCabe B, Simcock C, Harvey-Lloyd J, Malamateniou C. Paediatric magnetic resonance imaging adaptations without the use of sedation or anaesthesia: a narrative review. J Med Imaging Radiat Sci 2022;53(3):505-514. [doi: 10.1016/j.jmir.2022.04.048] [Medline: 35585018]
21. Runge SB, Christensen NL, Jensen K, Jensen IE. Children centered care: minimizing the need for anesthesia with a multi-faceted concept for MRI in children aged 4-6. Eur J Radiol 2018;107:183-187. [doi: 10.1016/j.ejrad.2018.08.026] [Medline: 30292264]
22. Saini P, Koehn C, Heuvelink A, Tasar O, van Vorstenbosch-Lynn E, Nauts S, et al. Scan buddy: a gamified app to prepare children for an MRI scan. In: Human-Computer Interaction. Theoretical Approaches and Design Methods. Cham, Switzerland: Springer; 2022:594-612. [doi: 10.1007/978-3-031-05311-5_42]
23. Saini P, Heuvelink A, Taşar O, van Vorstenbosch-Lynn E, van de Vleuten-Chraibi S, Nauts S. Re-design of an educational scanner to prepare children for an MRI or CT scan. : IEEE; 2023 Presented at: 2023 IEEE 11th International Conference on Serious Games and Applications for Health (SeGAH); August 28-30, 2023; Athens, Greece. [doi: 10.1109/segah57547.2023.10253811]
24. Geuens S, Lemiere J, Nijs J, Treunen M, Aertsen M, Toelen J, et al. Testing a home solution for preparing young children for an awake MRI: a promising smartphone application. Children (Basel) 2023;10(12):1866 [FREE Full text] [doi: 10.3390/children10121866] [Medline: 38136068]

XSL•FO
RenderX

25. Haslund-Thomsen H, Bauditz SJ, Kristensen HN. Familiar in the unfamiliar—children's, parents', and healthcare professionals' experiences of procedure-related anesthesia in the pediatric oncological ward. J Pediatr Nurs 2021;61:40-46. [doi: 10.1016/j.pedn.2021.03.008] [Medline: 33744621]
26. van Minde D, Klaming L, Weda H. Pinpointing moments of high anxiety during an MRI examination. Int J Behav Med 2014;21(3):487-495. [doi: 10.1007/s12529-013-9339-5] [Medline: 24043600]
27. Tazegul G, Etcioglu E, Yildiz F, Yildiz R, Tuney D. Can MRI related patient anxiety be prevented? Magn Reson Imaging 2015;33(1):180-183. [doi: 10.1016/j.mri.2014.08.024] [Medline: 25172986]
28. Zaitsev M, Maclaren J, Herbst M. Motion artifacts in MRI: a complex problem with many partial solutions. J Magn Reson Imaging 2015;42(4):887-901 [FREE Full text] [doi: 10.1002/jmri.24850] [Medline: 25630632]
29. Schmitz A, MacQuarrie A, Julier S, Binetti N, Steed A. Directing versus attracting attention: exploring the effectiveness of central and peripheral cues in panoramic videos. : IEEE; 2020 Presented at: 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR); March 22-26, 2020; Atlanta, GA p. 63-72. [doi: 10.1109/vr46266.2020.1581102716289]
30. Budrys T, Veikutis V, Lukosevicius S, Gleizniene R, Monastyreckiene E, Kulakiene I. Artifacts in magnetic resonance imaging: how it can really affect diagnostic image quality and confuse clinical diagnosis? J Vibroeng 2018;20(2):1202-1213. [doi: 10.21595/jve.2018.19756]
31. Droit-Volet S. Time perception in children: a neurodevelopmental approach. Neuropsychologia 2013;51(2):220-234. [doi: 10.1016/j.neuropsychologia.2012.09.023] [Medline: 22999968]
32. Piaget J. The Origins of Intelligence in Children. New York, NY: W. W. Norton & Co; 1952. [doi: 10.1037/11494-000]
33. Wallis LA, Healy M, Undy MB, Maconochie I. Age related reference ranges for respiration rate and heart rate from 4 to 16 years. Arch Dis Child 2005;90(11):1117-1121 [FREE Full text] [doi: 10.1136/adc.2004.068718] [Medline: 16049061]
34. Largo-Pineda CE, Arenas-Correa ID, Ángel-González GJ, Vélez-Arango JM, Calvo-Betancur VD, Arango-Zapata AN. Adverse events in paediatric patients taken to magnetic resonance imaging under sedation or anaesthesia. Colomb J Anesthesiol 2017;45(1):8-14. [doi: 10.1097/01819236-201701000-00003]
35. Vu KN, Haldipur AG, Roh ATH, Lindholm P, Loening AM. Comparison of end-expiration versus end-inspiration breath-holds with respect to respiratory motion artifacts on t1-weighted abdominal MRI. AJR Am J Roentgenol 2019;212(5):1024-1029. [doi: 10.2214/AJR.18.20239] [Medline: 30835515]
36. Liszewski MC, Hersman FW, Altes TA, Ohno Y, Ciet P, Warfield SK, et al. Magnetic resonance imaging of pediatric lung parenchyma, airways, vasculature, ventilation, and perfusion: state of the art. Radiol Clin North Am 2013;51(4):555-582. [doi: 10.1016/j.rcl.2013.04.004] [Medline: 23830786]
37. Sodhi KS, Ciet P, Vasanawala S, Biederer J. Practical protocol for lung magnetic resonance imaging and common clinical indications. Pediatr Radiol 2022;52(2):295-311 [FREE Full text] [doi: 10.1007/s00247-021-05090-z] [Medline: 34037828]
38. Rajiah PS, François CJ, Leiner T. Cardiac MRI: state of the art. Radiology 2023;307(3):e223008. [doi: 10.1148/radiol.223008] [Medline: 37039684]
39. Jaimes C, Kirsch JE, Gee MS. Fast, free-breathing and motion-minimized techniques for pediatric body magnetic resonance imaging. Pediatr Radiol 2018;48(9):1197-1208. [doi: 10.1007/s00247-018-4116-x] [Medline: 30078042]
40. Treutlein C, Zeilinger MG, Dittrich S, Roth JP, Wetzl M, Heiss R, et al. Free-breathing and single-breath hold compressed sensing real-time MRI of right ventricular function in children with congenital heart disease. Diagnostics (Basel) 2023;13(14):2403 [FREE Full text] [doi: 10.3390/diagnostics13142403] [Medline: 37510147]
41. Koch V, Merklein D, Zangos S, Eichler K, Gruenewald LD, Mahmoudi S, et al. Free-breathing accelerated whole-body MRI using an automated workflow: comparison with conventional breath-hold sequences. NMR Biomed 2023;36(2):e4828. [doi: 10.1002/nbm.4828] [Medline: 36082477]
42. Longère B, Abassebay N, Gkizas C, Hennicaux J, Simeone A, Rodriguez Musso A, et al. A new compressed sensing cine cardiac MRI sequence with free-breathing real-time acquisition and fully automated motion-correction: a comprehensive evaluation. Diagn Interv Imaging 2023;104(11):538-546 [FREE Full text] [doi: 10.1016/j.diii.2023.06.005] [Medline: 37328394]
43. Shimron E, Perlman O. AI in MRI: computational frameworks for a faster, optimized, and automated imaging workflow. Bioengineering (Basel) 2023;10(4):492 [FREE Full text] [doi: 10.3390/bioengineering10040492] [Medline: 37106679]
44. Kinnebrew SL, Dove CG, Midwin CM, Olson TM, Guimaraes CVA. The role of child life in pediatric radiology. Pediatr Radiol 2020;50(11):1509-1513. [doi: 10.1007/s00247-020-04795-x] [Medline: 32935242]

Abbreviations

BPM: breaths per minute
MRI: magnetic resonance imaging

XSL•FO
RenderX

Edited by A Coristine; submitted 22.12.23; peer-reviewed by S Liszio, K Qian, A Grilo, R Toomey; comments to author 16.02.24; revised version received 31.05.24; accepted 20.11.24; published 13.02.25.

Please cite as:
Heuvelink A, Saini P, Taşar Ö, Nauts S
Improving Pediatric Patients' Magnetic Resonance Imaging Experience With an In-Bore Solution: Design and Usability Study
JMIR Serious Games 2025;13:e55720
URL: https://games.jmir.org/2025/1/e55720
doi:10.2196/55720
PMID:

©Annerieke Heuvelink, Privender Saini, Özgür Taşar, Sanne Nauts. Originally published in JMIR Serious Games (https://games.jmir.org), 13.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Testing the Effectiveness of a Gamified Emotional Cognitive Bias Modification Task as an Intervention for Low Mood: Randomized Controlled Trial

Rumeysa Kuruoğlu[1], MSc; Angela Attwood[1,2], Prof Dr; Ian Penton-Voak[1,2], Prof Dr

[1]School of Psychological Science, University of Bristol, 5 Priory Rd, Bristol, United Kingdom

[2]NIHR Biomedical Research Centre, University Hospitals Bristol and Weston NHS Foundation Trust, Bristol, United Kingdom

Corresponding Author:
Rumeysa Kuruoğlu, MSc
School of Psychological Science, University of Bristol, 5 Priory Rd, Bristol, United Kingdom

Abstract

Background: Emotion recognition bias in depression is well-documented and is proposed to play a causal role in depression. A cognitive bias modification (CBM) intervention targeting the bias in emotional expression perception was developed, but despite robust training effects on emotion perception, the effect on mood was unreliable and weak. We propose a new gamified cognitive bias modification (GCBM) to address potential limitations that may attenuate therapeutic effects.

Objective: This study aimed to investigate the effectiveness of a single session of GCBM on emotion perception and to assess whether the gamified version of the task would produce the same robust training effects on the interpretation of emotional expressions as the original CBM. The second aim was to compare the effectiveness of a single session of CBM training, CBM control (no training), and GCBM training on immediate mood.

Methods: We report a between-subjects fully automated and web-based experimental study that recruited participants via the web from the general population (N=916). We tested the effectiveness of GCBM in changing participants' responses to ambiguous facial expressions. The primary outcome was emotion recognition bias, measured as the increased identification of happy faces. We also compared the effects of a single session of GCBM training (n=397), CBM training (n=400), and CBM control (n=119) conditions on immediate mood, measured using the Immediate Mood Scaler.

Results: Results showed that participants in the GCBM intervention condition classified more ambiguous faces as "happy" after the training compared with controls, indicating an increased perception of happiness in ambiguous faces (B=1.57, *P*<.001). There was also evidence that GCBM training produced more positive changes in immediate mood compared with the CBM control condition (B=−3.64, *P*=.003) and compared with the CBM training condition (B=1.69, *P*=.048).

Conclusions: GCBM demonstrated promising results in changing participants' emotion recognition bias to ambiguous facial expressions and enhance the immediate mood compared with both CBM and control conditions. These results suggest that GCBM holds promise to be a better alternative to CBM for addressing mood-related cognitive biases. Further exploration of GCBM's long-term effects on mood and its clinical application is needed.

(JMIR Serious Games 2025;13:e65103) doi:10.2196/65103

KEYWORDS

low mood; mood; depression; depressive; mental health; cognitive bias modification; perceptions; gamification; emotion recognition bias; emotional facial expressions; emotions; facial; expression; biases; mental health

Introduction

Depression is associated with a negative bias in the interpretation of facial emotional expressions [1,2]. This negative bias has been proposed to play an important role in the onset and maintenance of depression, as successful pharmacological interventions have been found to be associated with the reduction of negative biases [3]. It was previously thought that treating depression would lead to improvements in emotion recognition bias. However, there is evidence from psychopharmacological studies of antidepressant action that suggests that changes in emotion recognition biases precede and predict later mood improvement [3]. This finding implies that altering negative biases through psychological interventions may be an effective way to improve the mood of individuals with depression [3,4]. Consequently, cognitive bias modification (CBM) interventions that target this negative bias may reduce depressive symptoms [4].

A CBM intervention was developed to aim at changing negative biases in the perception of emotional facial expressions [4]. This CBM technique shifts participants' responses toward more

XSL•FO
RenderX

positive interpretations of ambiguous facial expressions [4]. However, the improvements in mood that are predicted to follow a reduction in negative bias by the neurocognitive models of depression have been weak and unreliable [5-8]. We conducted a meta-analysis of 6 CBM training studies using the same intervention to assess the overall effectiveness of this CBM technique on mood outcomes [9]. The meta-analytic results showed that the effectiveness of the CBM on mood (ie, reducing overall depressive symptoms) was consistent with neurocognitive models of depression; indirect effects of training on improved mood were mediated by decreased negative emotional bias. However, the direct effects of training in the mediation model were in the opposite direction, with allocation to the active training condition leading to lower mood in comparison with the control condition. This was contrary to our predictions and indicated that some aspects of the active training condition were having a negative effect on mood. Overall, the indirect effects (in the therapeutic direction) and direct effects (in the nontherapeutic direction) largely canceled each other; resulting in a weak overall effect of CBM training on mood [9].

In the CBM studies reviewed in our meta-analysis, participants were given negative (ie, "incorrect") and positive (ie, "correct") text-based feedback to shift the perception of ambiguous faces from sad to happy (ie, reduce a negative cognitive bias [9]. In the CBM, participants in the training condition receive more negative feedback (ie, are told that they are incorrect) as the mechanism to change their perception of facial emotion from negative (sad) to positive (happy) judgments. Participants in the control conditions receive less negative feedback than those in the active training condition, as the CBM technique does not aim to change their responses from baseline. Based on the meta-analysis results, we speculated that the relatively frequent negative (ie, "incorrect") feedback in the task might elicit the negative effect of mood associated with the training condition revealed in the meta-analysis. Research suggests that continuous negative feedback decreases positive emotional experiences, and can trigger self-esteem problems and induce negative emotions [10,11]. Furthermore, receiving negative feedback elicits negative emotions, which can reduce the acceptance and use of future feedback [12]. In contrast, positive feedback reduces social anxiety [13] and prevents or attenuates depression [14].

These reported effects of negative feedback suggest that removing or changing the delivery of negative feedback may improve the effectiveness of this CBM technique on mood-related outcomes. However, in order to change emotion recognition bias, some form of feedback is necessary to drive the modification as this "corrects" the response in the intended direction. However, such feedback should be delivered in a constructive way. The education literature points to the importance of constructive feedback to increase the effectiveness of the feedback and positive emotional experience [15]. Given this information, the impact of negative feedback on cognitive interventions should be investigated in order to improve the emotional experience, and subsequent mood outcomes, of CBM.

There are different ways to give feedback to participants in cognitive tasks. In this study, we present a new version of our previous CBM task, in which negative feedback is replaced by positive reinforcement of the desired response via gamification techniques that adapt game-like features to nongame tasks [16]. Gamification methods have been used in long and repetitive cognitive tasks to improve the engagement of participants [17,18]. The gamified cognitive bias modification (GCBM) uses a scoring system, instead of the "correct and incorrect" feedback, to change participant's judgments of emotional faces, and potentially improve mood outcomes after the CBM training. The participants earn points for each correct answer but do not earn any points for the incorrect answers. They earn more points for "correct" answers to more ambiguous images that are harder to judge (ie, more ambiguous faces that are more difficult to judge).

In this study, we investigated whether replacing negative feedback with a gamified points feature delivers similar training effects to those seen in our previous studies [5-8]. The hypothesis is that a single session of GCBM would result in a positive change in the recognition of emotional facial expressions, similar to CBM training. We also, compared the effectiveness of a single session of GCBM training, CBM training, and CBM control conditions on immediate mood. Previous studies suggested that younger people benefited more from CBM training in terms of mood improvement [8]. Therefore, we recruited young adults between the ages of 18 and 30 years. We hypothesized that participants in the GCBM condition would have significantly higher immediate mood scores after training compared with the CBM group.

Methods

Overview

This study compared the effectiveness of a single session of CBM training, CBM (no training) control, and GCBM training on balance points and immediate mood. Preregistration information for this study is available on the Open Science Framework (OSF), where the experimental data can also be accessed [19]. There were deviations from the protocol as the CBM control condition was added post hoc. These deviations are explained in the Multimedia Appendix 1, but for clarity, we have presented the study as though all conditions were run concurrently ("Protocol Deviations: Adding a control condition post hoc" Multimedia Appendix 1). In addition, the development of the GCBM study, which tested its effectiveness on balance points with a smaller sample size, is provided in "Development and initial test of GCBM training" in Multimedia Appendix 1.

Participants

The meta-analysis of CBM studies revealed an indirect effect size of g=0.23 for the effect of training on mood outcomes [9]. An a priori power analysis was conducted using G*Power [20]. We calculated a sample size of a total of 788 (n=394 in GCBM and n=394 CBM training conditions) would allow us to detect d=0.2 with 80% power. A further 120 participants were recruited for the control condition (refer to "Protocol Deviations: Sample size determination for the control group" in Multimedia Appendix 1 for this sample size determination). In total, we recruited 927 participants to allow for participants failing attention checks or not completing the study through Prolific Academic (Oxford University). The study was advertised to

XSL•FO
RenderX

participants in the United Kingdom through Prolific. The study was run on Gorilla (Cauldron Science Ltd), a platform for creating web-based experimental tasks [21].

To be eligible for this study, participants had to be aged between 18 and 30 years, fluent in English, and have normal or corrected-to-normal vision. Participants were ineligible if they had consumed alcohol within the last 12 hours (self-report).

Ethical Considerations

Ethics approval was obtained from the School of Psychological Science Research Ethics Committee at the University of Bristol (approval 15,735). The study was conducted according to the revised Declaration of Helsinki (2013) and the 1996 ICH (International Council for Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use) Guidelines for Good Clinical Practice E6 (R1).

At the beginning of the session, participants provided informed consent by checking a box after reading an information page. The consent form outlined that participation was voluntary, data would be anonymized and confidential, and participants could withdraw at any time. It also informed participants that anonymized data would be made publicly available as "open access" after the study, in line with best research practices. For full details of the consent form, please refer to the Multimedia Appendix 1. Participants were reimbursed £2.25 (US $2.84) via Prolific for their time (median completion time was 11, IQR 4.33 min).

Study Design

This experimental study used a between-subjects design with 1 factor of intervention (CBM control, CBM training, and GCBM training) with immediate mood as the outcome. Participants were fully randomized to CBM and GCBM conditions with a 1:1 ratio by Gorilla randomizer. Participants were allocated to the control condition in a second phase of testing (refer to "protocol deviations" in Multimedia Appendix 1). Participants were blind to the condition to which they were assigned. Mood was self-assessed using the Immediate Mood Scaler (IMS) before and after training. The primary outcome is the difference in IMS scores between the CBM and GCBM groups.

Materials

GCBM

The GCBM training used in this study was adapted from the CBM intervention used in our earlier work. The original CBM intervention involved repeated trials where the participant was briefly shown an image of a morphed emotional face and asked to decide whether the displayed face was happy or sad (refer to Figure 1 for example faces). These facial images (15 in total) were morphed between happy and sad emotions, taken from a continuum ranging from an unambiguously happy emotional face to an unambiguously sad emotional face, with ambiguous images toward the center of the continuum (Figure 1 illustrates an example of morphed happy and sad faces). Hence, each individual stimulus differed in the proportion of each emotion presented. The facial images are computer-generated faces and do not belong to any actual individuals [4].

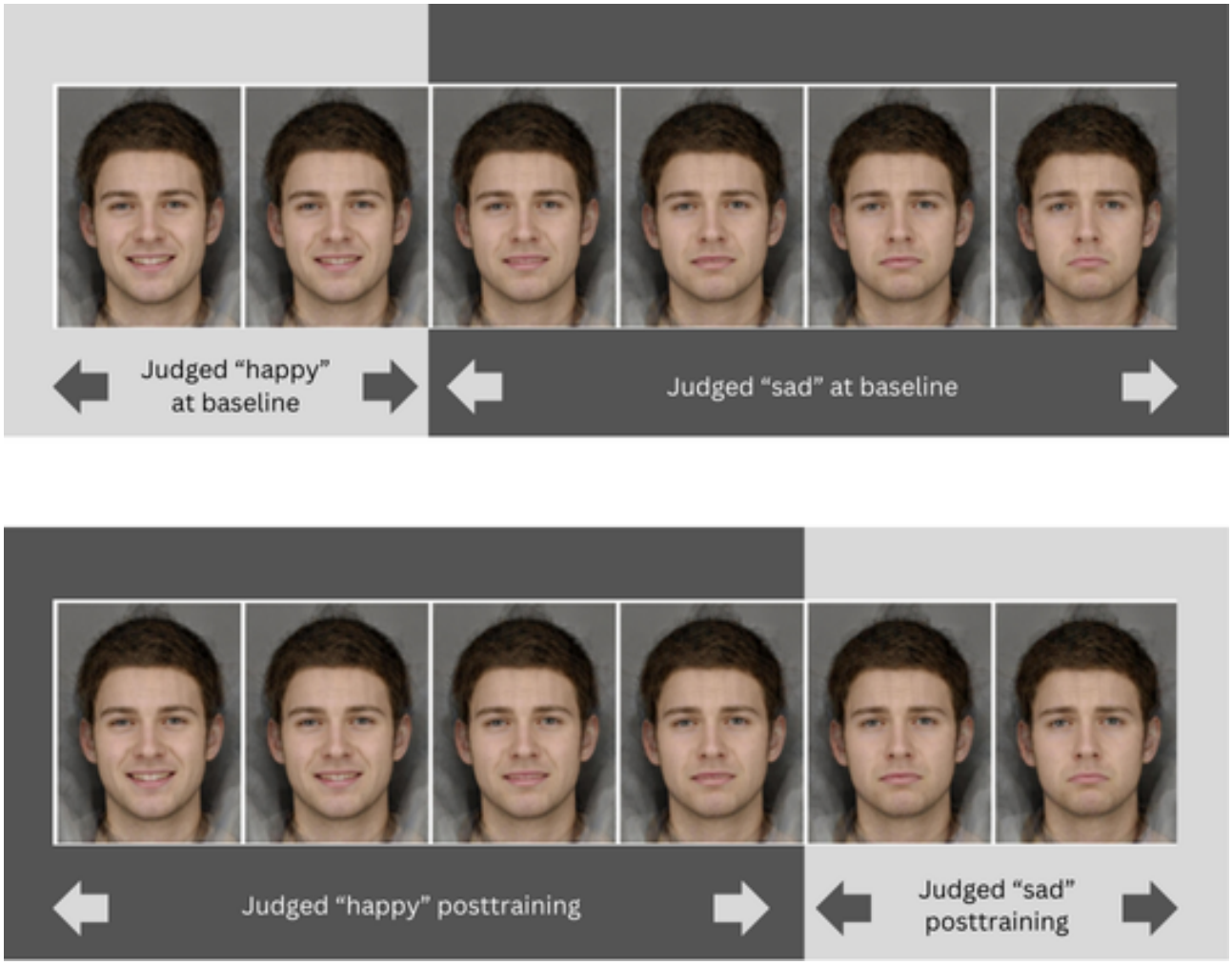
Figure 1. Examples of morphed faces used in the gamified cognitive bias modification intervention. The images depict unambiguously sad and happy expressions at the end points, with more ambiguous expressions in the center. These faces were presented to participants one at a time during the intervention.



The primary outcome was a "balance point," which indicated the image at which a participant was equally likely to categorize the face as happy or sad. A lower balance point indicates a stronger negative emotion recognition bias toward ambiguous faces, as it reflects a tendency to classify more faces from the continuum as sad. Starting balance points were calculated for participants during the baseline block. In the training block, participants receive feedback on each trial, indicating whether their categorization was "correct" or "incorrect." This feedback is tailored to each participant, based on their baseline balance point. For participants in the active training condition, feedback was designed to gradually shift their responses toward a more positive bias by categorizing more faces as happy. In the control condition, feedback aligned with their baseline balance point, meaning no shift in responses was expected. The balance point was calculated as the proportion of "happy" responses out of the total number of trials in the baseline or test block, multiplied by the total number of stimuli. For example, if a participant classified 30 out of 45 faces as happy during the baseline block, their baseline balance point would be (30÷45)×15=10. In the training condition, feedback would be given on the basis of a calculated target balance point of ([30÷45]×15+2=12; Figure 2 illustrates an example balance point change). Balance points were rounded to the nearest whole number for purposes of setting training feedback [4].

XSL•FO
RenderX

Figure 2. Example of intervention group responses to emotional facial expressions. The top image represents judgments at baseline, while the bottom image represents judgments after training.



There were 4 blocks—baseline, 2 training blocks, and a test. The (identical) baseline and test blocks determined the participant's pretraining and posttraining balance points. In a typical trial, a fixation cross was displayed for 600 ms, followed by a face stimulus for 200 ms. The baseline and test blocks consisted of 45 trials; the 15 facial stimuli were each presented 3 times in random order. All participants started with a baseline block to calculate their balance points. After that they continued to do the 2 training blocks, there were 31 trials, in which images 1-2 (unambiguously happy) and 14 - 15 (unambiguously sad) were presented once, images 3-5 and 11-13 were presented twice, and images 6-10 were presented 3 times. This design was implemented because the most ambiguous images (6-10) were harder to classify, and most participants' balance points tended to fall within this range, these images were presented more frequently to provide more training trails for the most difficult trials, at the critical level of ambiguity. In the training blocks, they received feedback based on their baseline balance point and the group they were in (training-control). After the training blocks, there was a test block in which the posttraining balance point was calculated. The experiment took roughly 10 minutes to complete.

The new GCBM intervention included the features mentioned above, but the method of feedback to participants was changed. In the CBM training, participants received positive or negative feedback as text (eg, "Correct! That face was happy" or "Incorrect! That face was Sad"). However, in the GCBM, participants earned points as a reward for correct answers but received no points for incorrect answers (Figure 3). In GCBM, more points were scored for correct answers closer to their baseline balance point so participants received a higher reward for correctly responding to more ambiguous expressions that were more difficult to classify (refer to "Scoring System for GCBM" in Multimedia Appendix 1 for scoring system details). Participants were able to see their total scores on the right upper corner of the screen throughout both training blocks. In the control condition, points were awarded based on a participant's baseline balance point without any modification (ie, participants were rewarded for responding in the same way as they responded in the baseline block), and therefore no change in their balance point was expected.

XSL•FO
RenderX

Figure 3. Example of a training block in the web-based gamified cognitive bias modification intervention.

XSL•FO
RenderX

GCBM also differs from CBM in terms of stimulus presentation duration and the use of visual noise masks. In GCBM, stimuli were presented for 200 ms (compared with 150 ms in CBM) to give participants more time to process facial expressions while minimizing the "flashing" effect. In addition, the visual noise mask used in CBM was removed, as participants in earlier trials reported it as unpleasant. These modifications were based on qualitative feedback from previous CBM studies to enhance participant experience.

IMS

IMS is a self-report questionnaire that measures well-being. It consists of 22 questions and responses are made on a 5-point scale (Multimedia Appendix 1).

Demographic Questions

Participants were asked about their ethnicity, gender, and age. The demographic questions were as follows:

- "Where are you from?" Participants were provided with a list of country options to select from.
- "What is your ethnicity?" Options included (1) Asian or Asian British; (2) Black, African, Caribbean, or Black British; (3) White; (4) Mixed or Multiple Ethnic Groups; (5) prefer not to say; and (6) other (please specify).
- "How old are you?" Participants answered using a text box.
- "What is your gender?" Options included (1) men, (2) women, (3) prefer not to say, and (4) other (please specify).

Attention Check

The participants were asked to type "seven" into a text box if they were still paying attention to the task after the IMS questionnaire.

Procedure

At the beginning of the experiment, participants were presented with an information sheet and asked to complete a consent form via the web in which they confirmed eligibility against the criteria listed above. If participants confirmed eligibility and consent, they started the experiment by answering demographic questions. Next, they filled out IMS questionnaire for mood measurement. Next, the training was presented (either CBM control, CBM training, or GCBM training per the randomization). After the training, they completed the IMS questionnaire again and were presented with the feedback text box and attention check text box. Finally, they were provided with the debrief and researcher information, and reimbursed via their Prolific account.

Statistical Analysis

We used box plots to identify and remove outliers (ie, data points that fall 1.5 times above or below the IQR). Data were assessed for normality using skewness and kurtosis statistics and were found to meet the assumptions of normality.

We conducted a linear regression analysis using SPSS 28 software (IBM Corp). This analysis compared posttraining balance points between the GCBM condition and other conditions (CBM training and CBM control), as well as all other pairwise combinations of conditions. We adjusted for baseline balance point and reported adjusted and unadjusted models.

To assess the effects of training on mood scores, we employed a mixed-design ANOVA. This analysis included a between-subjects factor with 3 levels—GCBM training, CBM training, and CBM control, and a within-subjects factor with 2 levels—pretraining and posttraining IMS scores. We looked at the interaction effect group and time, which assessed whether the change in mood scores from pretraining to posttraining varied across different groups.

We ran linear regression analysis using SPSS 28 software as a planned comparison to compare posttraining IMS scores in the CBM and GCBM conditions (the primary outcome in our protocol), and additionally all other pairwise combinations of conditions as post hoc tests. We adjusted for baseline mood, age, and gender, and reported adjusted and unadjusted models.

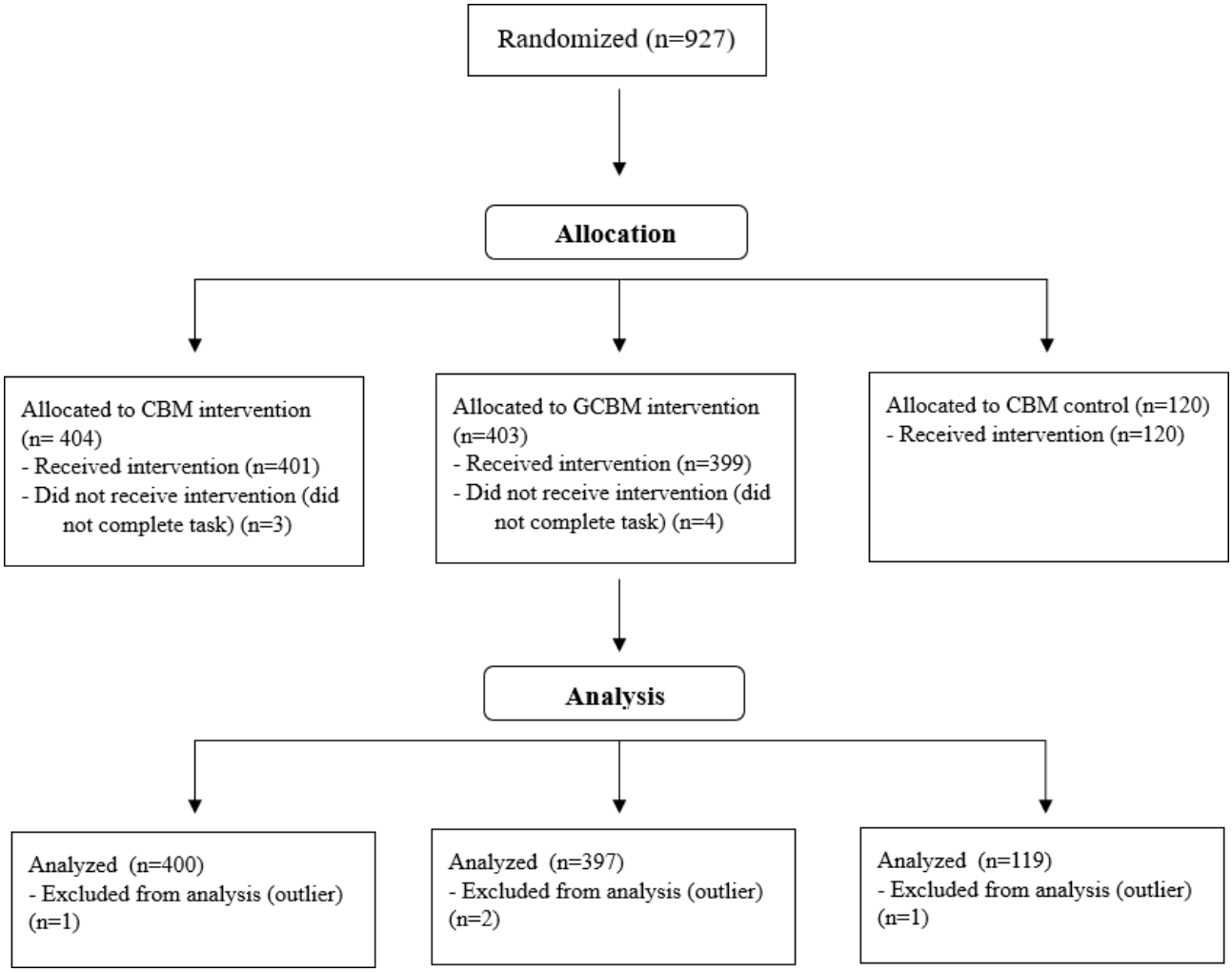
Results

Overview

We collected data from 927 participants (CBM=404, GCBM=403, and control=120) to account for potential exclusions due to failed attention checks and technical issues, and still reach our desired minimum sample size from our power calculation. Of the total, 7 participants did not complete the task (CBM=3, GCBM=4, and control=0). We removed 4 outliers that fell 1.5 times above or below the IQR (CBM=1, GCBM=2, and control=1) (refer to Figure 4 for CONSORT [Consolidated Standards of Reporting Trials] flowchart; Checklist 1). Of the remaining 916 participants (CBM=400, GCBM=397, and control=119), 55.1% (n=505) were women and 392 men; 19 participants did not state their gender. The mean age was 25.46 (SD 3.18) years. Out of 927, 76% (n=696) of the participants were White, 11.2% (n=103) were Asian, 6.6% (n=60) were Black, African, or Caribbean, 4.9% (n=45) were Mixed or from Multiple Ethnic Groups, and 1.3% (n=12) chose "other" as an option.

XSL•FO
RenderX

Figure 4. CONSORT (Consolidated Standards of Reporting Trials) flowchart. CBM: cognitive bias modification; GCBM: gamified cognitive bias modification.

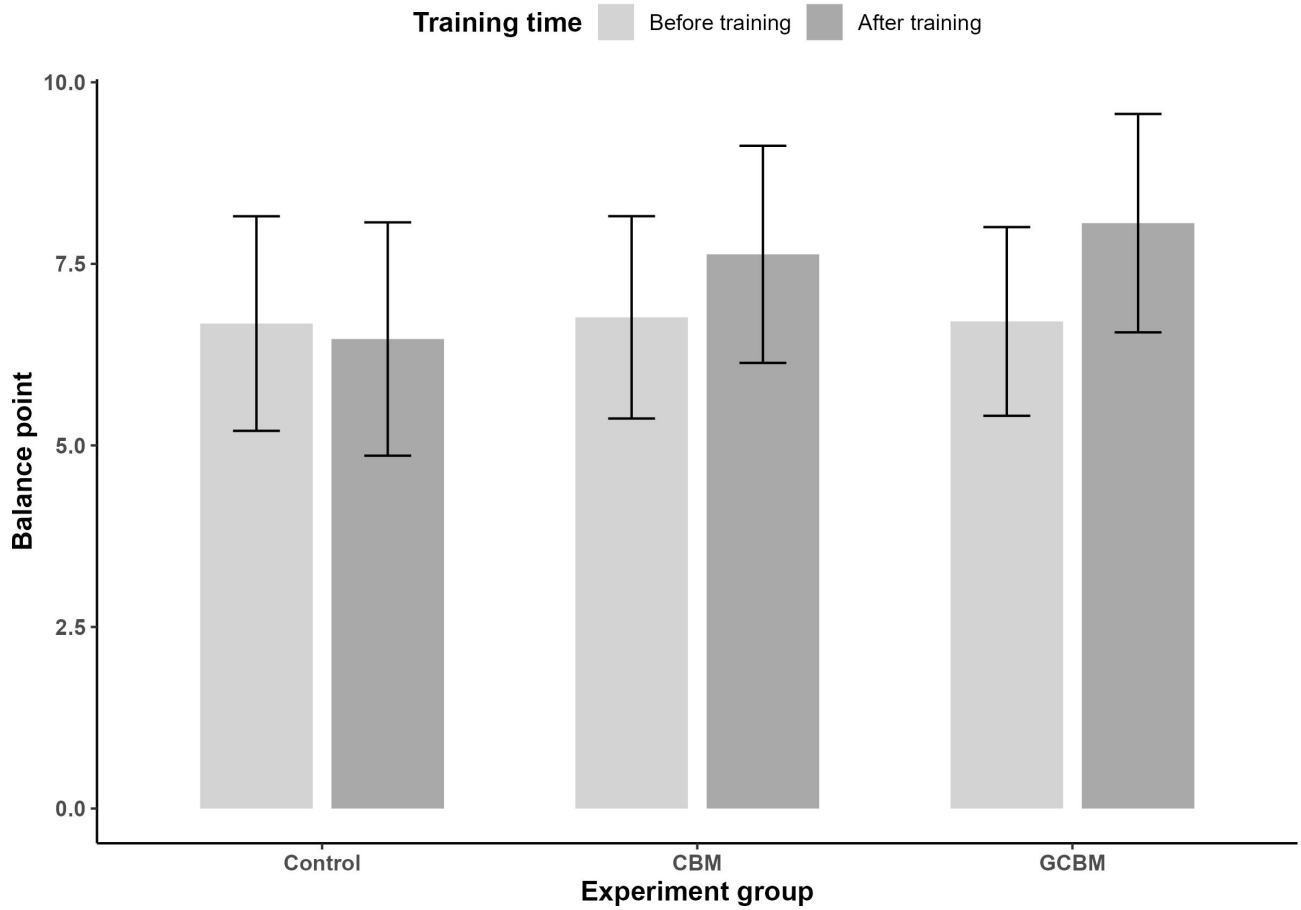


Balance Points

To assess the effect of training on emotion recognition bias, we examined participants' balance points pre- and posttraining. Both CBM and GCBM training showed positive changes after training indicating that participants in both conditions categorized more ambiguous faces as "happy" after training while the control condition showed no meaningful change (Figure 5). On average, participants in CBM training showed a 0.87-point increase, while those in GCBM training demonstrated a 1.35-point increase.

XSL•FO
RenderX

Figure 5. Graph for pre- and posttraining balance points for cognitive bias modification, gamified cognitive bias modification, and control conditions. Both cognitive bias modification and gamified cognitive bias modification training positively shifted participants' balance points (higher balance points indicate participants categorized more faces as happy, suggesting a reduction in negative bias when interpreting ambiguous emotional facial expressions, and error bars represent SDs). CBM: cognitive bias modification; GCBM: gamified cognitive bias modification.



To further evaluate the effect of cognitive bias training on balance points, we conducted a 3 (group: CBM, GCBM, and control)×2 (Time: pre-training and post-training) ANOVA with balance point as the dependent variable. The results showed evidence of an interaction effect between time (pre- and posttraining) and group (CBM, GCBM, and control), F_{2913}=127.11, P<.001, partial η^2=0.22.

We conducted exploratory, stratified linear regressions to compare posttraining balance points of CBM and GCBM training with each other and the control group. The results showed that both CBM and GCBM training groups had improved posttraining balance points compared with control group (B=−.55, 95% CI −0.64 to −0.45], P<.001; B=−1.57, 95% CI −1.76 to −1.38, P<.001, respectively) Also, there is evidence that GCBM training induced a larger training effect than CBM, (B=0.48, 95% CI 0.35-0.61, P<.001). These results are from the adjusted model for the pretraining balance point. Results observed in the unadjusted models were qualitatively the same in terms of direction and strength of evidence (Table S1 in Multimedia Appendix 1).

Mood Outcomes

To assess the effect of training on immediate mood, we examined participants' posttraining IMS scores. Participants in all conditions showed improved mood after the training. Control group had a 3.63-point improvement, CBM group had a 5.35-point improvement, and GCBM group had a 7.01-point improvement on average (Table 1). To evaluate the effect of cognitive bias training on immediate mood, a mixed method ANOVA with IMS as the outcome, time (pretraining or posttraining) as a within-subjects factor, and group (CBM, GCBM, and control) as a between-subjects factor was performed. The ANOVA results indicated that there is evidence for an interaction effect between time (pre- and posttraining IMS) and group (CBM, GCBM, and control), F_{2913}=4.36, P=.01, partial η^2=0.01.

XSL•FO
RenderX

Table . Descriptive statistics for pre- and posttraining Immediate Mood Scaler scores cognitive bias modification, gamified cognitive bias modification, and control conditions (higher Immediate Mood Scaler scores indicate better mood).

	CBM[a]		GCBM[b]		Control	
	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n
Pretraining IMS score	96.39 (25.32)	400	98.12 (27.27)	397	95.50 (26.54)	119
Posttraining IMS score	101.74 (26.43)	400	105.13 (28.11)	397	99.13 (27.91)	119

[a]CBM: cognitive bias modification.

[b]GCBM: gamified cognitive bias modification.

To investigate this interaction effect, we ran stratified linear regressions to assess the effect of CBM, GCBM, and control groups on posttraining IMS scores as specified in our preregistration. The results indicated no substantial evidence that participants in the CBM training experienced improved immediate mood compared with the control group (B=−1.00, 95% CI −2.23 to 0.23, *P*=.11). However, participants in the GCBM training group reported improved immediate mood compared with the control group (B=−3.64, 95% CI −6.02 to −1.25, *P*=.003). In addition, there is weak evidence that GCBM leads to more positive changes in participants' moods compared with CBM (B=1.69, 95% CI 0.02-3.37, *P*=.048; our primary prespecified outcome). These results are from adjusted model for pretraining IMS scores, age, and gender. Results observed in the unadjusted models were qualitatively the same in terms of direction and strength of evidence (Table S2 in Multimedia Appendix 1).

Discussion

Principal Findings

In this study, our objective was to assess the effectiveness of a new GCBM task to change judgments of emotional facial expressions and the immediate mood of participants. The results demonstrated that GCBM training positively changed the participants' responses to emotional faces compared with control condition. Furthermore, findings suggest that GCBM is more effective than the CBM task at changing participants' judgments of emotional expression. The results provided evidence that GCBM improved immediate mood more than the control and CBM groups.

Results showed that the difference between pre- and posttraining balance points was greater in both CBM and GCBM groups compared with control. This indicates that both training methods generated a positive shift in the responses to emotional facial expressions, in comparison with the control (no training) condition. Also, all 3 groups demonstrated an improvement in immediate mood after the training. However, participants in the GCBM group showed greater immediate mood improvement compared with those in the control and CBM conditions. There was no statistical evidence for a difference between the CBM and control groups in terms of immediate mood improvement. These results suggest that GCBM training is successful in both modifying responses to emotional facial expressions and enhancing participants' immediate mood, above and beyond the (unusual, in our previous experience) positive effects of financial compensation for participation. The greater immediate mood improvement in the GCBM group, compared with both the CBM and control groups, suggests that the gamified aspect of the training may provide an additional benefit for improving mood beyond our original CBM intervention. This could imply that the engaging, game-like nature of GCBM might enhance participant motivation or make the training more impactful, leading to better mood outcomes.

Comparison With the Previous Work

Participants in the GCBM group showed a greater positive shift in their responses to emotional faces compared with the CBM group. This result suggests that gamification of the task increased engagement with the cognitive target of the intervention. This outcome aligns with the rationale behind the CBM training which suggests that changing people's perspective toward ambiguous facial expressions can improve their mood and potentially improve the symptoms of depression [4]. The results might suggest that greater improvement in emotion perception was associated with better improvement in immediate mood. On the other hand, greater improvement in immediate mood in the GCBM group might be solely the result of the absence of the negative effect of feedback in CBM as we predicted.

While some previous studies have shown mood improvement after multiple sessions of CBM training using the outlined method, most studies have not demonstrated significant evidence of improvement [5-8,22]. Results from a meta-analysis revealed that CBM has no overall positive effect on mood. However, the relationship between CBM and mood is mediated by the reduction in emotion recognition bias after training [9]. These results suggest that some therapeutic benefits of CBM may be masked or attenuated by the negative feedback in the training condition. GCBM was designed to address this issue by changing the delivery of feedback. Given that GCBM has been shown to reduce negative emotion recognition bias more effectively than CBM, it may be more effective when administered through multiple training sessions for longer-term mood improvement.

Strengths and Limitations

One of the limitations of this study is that we only tested participants between the ages of 18 and 30 years old. This decision was based on previous research suggesting that younger participants tend to benefit more from this emotion recognition CBM task [8]. In this study, our primary objective was to gain

XSL•FO
RenderX

a clear understanding of the effectiveness of gamifying this CBM task.

Second limitation of this study is that we did not examine the improvement of depressive symptoms; we only assessed the immediate mood of the participants. The IMS asks participants how they are feeling at the moment. However, previous CBM studies have used questionnaires that assess depressive symptoms. Although immediate mood can be an indicator of depressive symptoms [23], the effect of GCBM on depressive symptoms still needs further investigation.

Third limitation of the study is the absence of a control group where participants played a game unrelated to emotion recognition training. The only control group in this study was the CBM control condition, which is similar to CBM training but without the training component. This means we did not compare the impact of simply playing a game versus the impact of GCBM training on immediate mood. It is possible that GCBM influences immediate mood through the enjoyable aspects of playing a game.

Fourth limitation is that this study only assessed the effectiveness of a single session of GCBM on immediate mood. This approach was chosen to test whether GCBM effectively changes balance points similarly to CBM and to compare its mood-related effects with CBM. Economic considerations also influenced this decision, as detecting small effect sizes requires large samples, making multisession studies expensive. Before committing to a multisession study, we wanted to evaluate GCBM's effectiveness on mood relative to CBM in a single-session format to determine its potential.

Future Directions

Future studies should examine whether the greater improvement in immediate mood observed in GCBM is specifically driven by the absence of negative feedback or whether other factors, such as increased engagement or improved task design, contribute to its effectiveness. Investigating the specific mechanisms underlying these differences will help clarify the role of feedback in mood enhancement.

In addition, future research should assess the long-term effects of multiple GCBM training sessions on mood, as mood changes related to depression typically develop over an extended period [3].

Furthermore, future studies should compare GCBM with a similar game-based control condition to determine whether improvements in immediate mood are a result of the cognitive mechanisms targeted by the intervention or simply the enjoyment of playing a game.

Finally, further research is necessary to explore potential differences in how younger and older individuals respond to this CBM intervention, as age-related factors may influence its effectiveness.

Conclusion

This study provided evidence that GCBM led to a more pronounced shift in participants' responses to emotional faces after training compared with the CBM control and training conditions. Importantly, the immediate mood of participants in the GCBM group improved to a greater extent than in the CBM control and training conditions. These findings suggest that GCBM may be a promising and acceptable intervention for individuals experiencing mood disorders. Given the importance of developing remote and cost-effective interventions for depression, future studies should investigate the long-term effects of GCBM on depression symptoms and recruit participants diagnosed with depression, as they constitute the actual target group for GCBM.

Acknowledgments

We would like to thank Geoff Gallagher for their help with coding.

This study was funded by the National Institute for Health and Care Research Bristol Biomedical Research Centre (NIHR). The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

Data Availability

The data supporting the findings of this study are available on the Open Science Framework (OSF).

Authors' Contributions

IP-V, AA, and RK contributed to the conceptualization of this study. RK was responsible for data collection, analysis, interpretation, and writing the original draft of the paper. IP-V and AA reviewed and provided feedback on the paper.

Conflicts of Interest

IPV is a director of Jericoe Ltd, which develops software for assessing and modifying emotion perception. Authors RK and AA declare no conflicts of interest.

Multimedia Appendix 1
Supplementary materials.
[DOCX File, 155 KB - games_v13i1e65103_app1.docx]

XSL•FO
RenderX

Checklist 1
CONSORT-eHEALTH checklist.
[PDF File, 1179 KB - games_v13i1e65103_app2.pdf]

References

1. Bourke C, Douglas K, Porter R. Processing of facial emotion expression in major depression: a review. Aust N Z J Psychiatry 2010 Aug;44(8):681-696. [doi: 10.3109/00048674.2010.496359] [Medline: 20636189]
2. Persad SM, Polivy J. Differences between depressed and nondepressed individuals in the recognition of and response to facial emotional cues. J Abnorm Psychol 1993 Aug;102(3):358-368. [doi: 10.1037//0021-843x.102.3.358] [Medline: 8408947]
3. Harmer CJ, Goodwin GM, Cowen PJ. Why do antidepressants take so long to work? A cognitive neuropsychological model of antidepressant drug action. Br J Psychiatry 2009 Aug;195(2):102-108. [doi: 10.1192/bjp.bp.108.051193] [Medline: 19648538]
4. Penton-Voak IS, Bate H, Lewis G, Munafò MR. Effects of emotion perception training on mood in undergraduate students: randomised controlled trial. Br J Psychiatry 2012 Jul;201(1):71-72. [doi: 10.1192/bjp.bp.111.107086] [Medline: 22539781]
5. Rawdon C, Murphy D, Motyer G, Munafò MR, Penton-Voak I, Fitzgerald A. An investigation of emotion recognition training to reduce symptoms of social anxiety in adolescence. Psychiatry Res 2018 May;263:257-267. [doi: 10.1016/j.psychres.2018.02.023] [Medline: 29602534]
6. Penton-Voak IS, Adams S, Button KS, et al. Emotional recognition training modifies neural response to emotional faces but does not improve mood in healthy volunteers with high levels of depressive symptoms. Psychol Med 2021 May;51(7):1211-1219. [doi: 10.1017/S0033291719004124] [Medline: 32063231]
7. Van Meter A, Stoddard J, Penton-Voak I, Munafò MR. Interpretation bias training for bipolar disorder: a randomized controlled trial. J Affect Disord 2021 Mar 1;282:876-884. [doi: 10.1016/j.jad.2020.12.162] [Medline: 33601731]
8. Suddell S, Müller-Glodde M, Lumsden J, et al. Emotional bias training as a treatment for anxiety and depression: evidence from experimental medicine studies in healthy and medicated samples. Psychol Med 2023 Feb;53(3):696-705. [doi: 10.1017/S0033291721002014] [Medline: 34057058]
9. Kuruoglu R, Suddell S, Penton-Voak IS. Meta-analysis of studies testing cognitive bias modification (CBM) intervention for emotion recognition bias. PsyArXiv. . [doi: 10.31234/osf.io/xk75t]
10. Ming X, Lou Y, Zou L, Lei Y, Li H, Li Y. The cumulative effect of positive and negative feedback on emotional experience. Psychophysiology 2021 Dec;58(12):e13935. [doi: 10.1111/psyp.13935] [Medline: 34459511]
11. Ilies R, De Pater IE, Judge T. Differential affective reactions to negative and positive feedback, and the role of self‐esteem. Journal of Managerial Psychology 2007 Aug 21;22(6):590-609. [doi: 10.1108/02683940710778459]
12. Sargeant J, Mann K, Sinclair D, Van der Vleuten C, Metsemakers J. Understanding the influence of emotions and reflection upon multi-source feedback acceptance and use. Adv in Health Sci Educ 2008 Aug;13(3):275-288. [doi: 10.1007/s10459-006-9039-x]
13. Cohen L, Huppert JD. Positive emotions and social anxiety: the unique role of pride. Cogn Ther Res 2018 Aug;42(4):524-538. [doi: 10.1007/s10608-018-9900-2]
14. Santos V, Paes F, Pereira V, et al. The role of positive emotion and contributions of positive psychology in depression treatment: systematic review. Clin Pract Epidemiol Ment Health 2013;9(1):221-237. [doi: 10.2174/1745017901309010221] [Medline: 24358052]
15. Fong CJ, Williams KM, Williamson ZH, Lin S, Kim YW, Schallert DL. “Inside out”: appraisals for achievement emotions from constructive, positive, and negative feedback on writing. Motiv Emot 2018 Apr;42(2):236-257. [doi: 10.1007/s11031-017-9658-y]
16. Khaleghi A, Aghaei Z, Mahdavi MA. A gamification framework for cognitive assessment and cognitive training: qualitative study. JMIR Serious Games 2021 May 18;9(2):e21900. [doi: 10.2196/21900] [Medline: 33819164]
17. Vermeir JF, White MJ, Johnson D, Crombez G, Van Ryckeghem DML. The effects of gamification on computerized cognitive training: systematic review and meta-analysis. JMIR Serious Games 2020 Aug 10;8(3):e18644. [doi: 10.2196/18644] [Medline: 32773374]
18. Lumsden J, Edwards EA, Lawrence NS, Coyle D, Munafò MR. Gamification of cognitive assessment and cognitive training: a systematic review of applications and efficacy. JMIR Serious Games 2016 Jul 15;4(2):e11. [doi: 10.2196/games.5888] [Medline: 27421244]
19. Kuruoglu R. GCBM study 2 data set. URL: https://osf.io/mv6sk [accessed 2025-04-17]
20. Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007 May;39(2):175-191. [doi: 10.3758/bf03193146] [Medline: 17695343]
21. Anwyl-Irvine AL, Massonnié J, Flitton A, Kirkham N, Evershed JK. Gorilla in our midst: an online behavioral experiment builder. Behav Res Methods 2020 Feb;52(1):388-407. [doi: 10.3758/s13428-019-01237-x] [Medline: 31016684]
22. Peters SE, Lumsden J, Peh OH, Penton-Voak IS, Munafò MR, Robinson OJ. Cognitive bias modification for facial interpretation: a randomized controlled trial of transfer to self-report and cognitive measures in a healthy sample. R Soc Open Sci 2017 Dec;4(12):170681. [doi: 10.1098/rsos.170681] [Medline: 29308221]

XSL•FO
RenderX

23. Nahum M, Van Vleet TM, Sohal VS, et al. Immediate Mood Scaler: tracking symptoms of depression and anxiety using a novel Mobile Mood Scale. JMIR Mhealth Uhealth 2017 Apr 12;5(4):e44. [doi: 10.2196/mhealth.6544] [Medline: 28404542]

Abbreviations

CBM: cognitive bias modification
CONSORT: Consolidated Standards of Reporting Trials
GCBM: gamified cognitive bias modification
ICH: International Council for Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use
IMS: Immediate Mood Scaler
OSF: Open Science Framework

Edited by A Coristine; submitted 05.08.24; peer-reviewed by A Poulose, H Kou; revised version received 06.03.25; accepted 10.03.25; published 01.05.25.

Please cite as:
Kuruoğlu R, Attwood A, Penton-Voak I
Testing the Effectiveness of a Gamified Emotional Cognitive Bias Modification Task as an Intervention for Low Mood: Randomized Controlled Trial
JMIR Serious Games 2025;13:e65103
URL: https://games.jmir.org/2025/1/e65103
doi:10.2196/65103

© Rumeysa Kuruoğlu, Angela Attwood, Ian Penton-Voak. Originally published in JMIR Serious Games (https://games.jmir.org), 1.5.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Long-Term Usage of Breeze, a Gamified Breathing Training App, and Its Effect on Momentary Relaxation in People With Cancer: Cohort Study

Anja Yvonne Bischof[1*], PhD; Tobias Budig[2,3*], BSc; Sonja Schläpfer[4], PhD; Yanick Xavier Lukic[1,5,6], PhD; Fabian Schneider[7], BSc; Prabhakaran Santhanam[8], MSc; Jürgen Barth[4], PhD; Claudia M Witt[4], Prof Dr Med, MBA; Tobias Kowatsch[1,2,3], Prof Dr

[1]School of Medicine, University of St. Gallen, St.Gallen, Switzerland

[2]Institute for Implementation Science in Health Care, University of Zurich, Zurich, Switzerland

[3]Centre for Digital Health Interventions, Department of Management, Technology, and Economics, ETH Zurich, Weinbergstr. 56/58, Zurich, Switzerland

[4]Institute for Complementary and Integrative Medicine, University Hospital Zurich and University of Zurich, Zurich, Switzerland

[5]School of Engineering, ZHAW Zurich University of Applied Sciences, Winterthur, Switzerland

[6]Codeklang GmbH, Zurich, Switzerland

[7]Byldr GmbH, Zurich, Switzerland

[8]Datahouse AG, Zurich, Switzerland

[*]these authors contributed equally

Corresponding Author:
Tobias Kowatsch, Prof Dr
School of Medicine, University of St. Gallen, St.Gallen, Switzerland

Abstract

Background: People with cancer often experience stress. Digital health interventions can help individuals increase momentary relaxation. Breeze is a gamified breathing training that can be embedded into digital health interventions. Its effectiveness in controlled cross-sectional studies has been demonstrated. However, adherence to Breeze and its effect on momentary relaxation in longitudinal interventional studies has yet to be investigated.

Objective: This study aimed to assess Breeze across 3 key dimensions: (1) its usage over time compared to other intervention components; (2) the accuracy with which users adhered to the guiding breathing pattern provided by Breeze; and (3) its effect on momentary relaxation, including the impact of the duration of the breathing session delivered by Breeze.

Methods: Breeze was 1 of 7 relaxation techniques included in the CanRelax 2.0 app, designed specifically for individuals diagnosed with cancer. Participants could select any of the 7 techniques to promote momentary relaxation. The intervention was designed to last 10 weeks. However, participants were allowed to use the CanRelax 2.0 app after that period. Participants were adults diagnosed with cancer in the last 5 years recruited across Switzerland, Germany, and Austria. Adherence to the intended breathing pattern was measured using a Pearson correlation coefficient. Momentary relaxation was measured pre- and post-exercise using an 11-point visual analog scale. Statistical analyses included linear mixed effects models and calculations of effect sizes. We further analyzed the relationship between session duration and the magnitude of momentary relaxation and compared Breeze's efficacy and time efficiency to the other available techniques.

Results: Of 279 participants, 118 (42.3%) used Breeze at least once. The 118 participants engaged in 754 breathing sessions with Breeze. Momentary relaxation was assessed and calculated for 249 (33.0%) Breeze sessions. The use of Breeze declined initially but remained stable even after 10 weeks. Participants followed the predefined breathing rates of Breeze (r=0.9). On average, a small effect (−0.42; P<.001; d=0.19) on momentary relaxation was observed, with 2-minute breathing sessions (−0.30; P=.02; d=0.13) showing a small effect, whereas a large effect (−1.53; P=.03; d=0.74) was observed for breathing sessions of 4 minutes or longer.

Conclusions: This study demonstrates the potential of Breeze to alleviate acute stress in individuals with chronic conditions, such as cancer. By combining gamification with evidence-based breathing techniques, Breeze fosters sustained user engagement and momentary relaxation. Participants adhered well to the guided breathing. While even short breathing sessions of 2 minutes provided modest momentary relaxation, longer durations (≥4 min) were considerably more effective. Future research aims to assess the impact of Breeze on other populations and chronic conditions.

(JMIR Serious Games 2025;13:e70297) doi:10.2196/70297

XSL•FO
RenderX

KEYWORDS

stress management; digital health; gamification; people with cancer; breathing; mindfulness; self-care; self-guided; momentary relaxation

Introduction

Managing chronic conditions is complex, as it often involves high medication adherence, coping with disease progression, managing decreased quality of life, and addressing uncertainty [1-3], all of which can increase perceived stress. Therefore, incorporating regular exercises into disease management is crucial, as they help strengthen personal resilience and develop strategies to manage the chronic condition more effectively by reducing stress. Long-term elevated stress levels might emerge if mental health is not regarded in people with a chronic condition. Chronic stress characterizes itself as a prolonged period of challenging or threatening life conditions that interfere with daily life and last for a longer period, that is, for a minimum of 1 month [4]. People dealing with chronic stress face a higher risk of developing further chronic conditions, a higher risk of mortality, and faster biological aging [5-7].

Approaches for stress reduction in people with chronic conditions, such as cancer, are manifold. Some of these are meditation (eg, short meditation, mindfulness meditation, walking meditation, or body scan), formal relaxation techniques (eg, guided imagery and progressive muscle relaxation), or breathing exercises [1,8-12]. Different types of techniques and exercises may have a positive impact on reduced worry and psychological stress [13,14]. To reduce acute periods of stress, in particular, slow-paced breathing training shows positive effects on physiological markers and psychological well-being [15]. It further reduces chronic pain [16] and improves dealing with acute and chronic stress [15,17]. Reducing the impact of daily acute stressors may have a positive effect on the reduction of chronic stress as well [18].

Digital health interventions (DHIs) can support individuals with chronic conditions in consistently engaging in stress-reducing exercises, helping to lower stress levels and enhance resilience [19,20]. One approach is *Breeze* (research prototype co-developed by YXL, Helen Galliker, and TK), a smartphone-based gamified breathing training that uses breathing sounds as biofeedback to guide users through a slow-paced breathing training. In experimental cross-sectional studies and online experiments, Breeze has shown its potential for stress reduction [21-23]. Slow-paced, controlled breathing has the potential to achieve momentary relaxation [12].

As promising as DHIs are, one major challenge is low adherence [24-26]. Studies show dropout rates of 50% to 83% in DHIs for various chronic conditions [25,27-29]. Adherence is a prerequisite to achieving intended health outcomes [30]. Gamification can enhance the user experience with a DHI and, thus, may address low adherence rates [31,32].

However, long-term use and effectiveness in a longitudinal study assessing a gamified breathing training with people with chronic conditions are still open. To this end, we formulate the following 3 research questions (RQs): (RQ1) How often is Breeze used over time? (RQ2) How accurately do participants adhere to the guided breathing rate of Breeze? and (RQ3) What is the effect of Breeze on momentary relaxation?

Methods

Overview

To answer the research questions, this study leverages usage data of the smartphone-based CanRelax 2.0 app (research prototype developed by the study team) [33,34], which was recently evaluated in a randomized controlled trial (RCT; German Clinical Trials Register DRKS00027546; registered on February 23, 2022). The RCT with a third randomized arm examined the effectiveness of CanRelax 2.0 in reducing stress among people with cancer over 10 weeks. The participants were allowed to use the CanRelax 2.0 app even after the end of the study intervention period. The app contained 7 different relaxation techniques, including Breeze. Participants were able to self-select exercises, and no exercise was highlighted. Where applicable, the study adhered to STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines. Detailed information on the RCT procedures and the CanRelax 2.0 app has been published [33-35].

Recruitment

Participants were recruited between July 2022 and February 2023 in Switzerland, Germany, and Austria. Recruitment efforts included social media posts (via Facebook, LinkedIn, and X [formerly Twitter]) and traditional outreach methods, such as consultations with health care providers, printed flyers, newsletters, and a press release issued by the University Hospital Zurich. A dedicated project website provided essential study information and app store download links to support recruitment. Eligibility criteria included being at least 18 years old, fluent in German, owning a smartphone (Android or iOS), and having received a cancer diagnosis within the past 5 years. Exclusion criteria included self-reported suicidal ideation and pregnancy at baseline. Further information on the recruitment strategy was published previously [33,34].

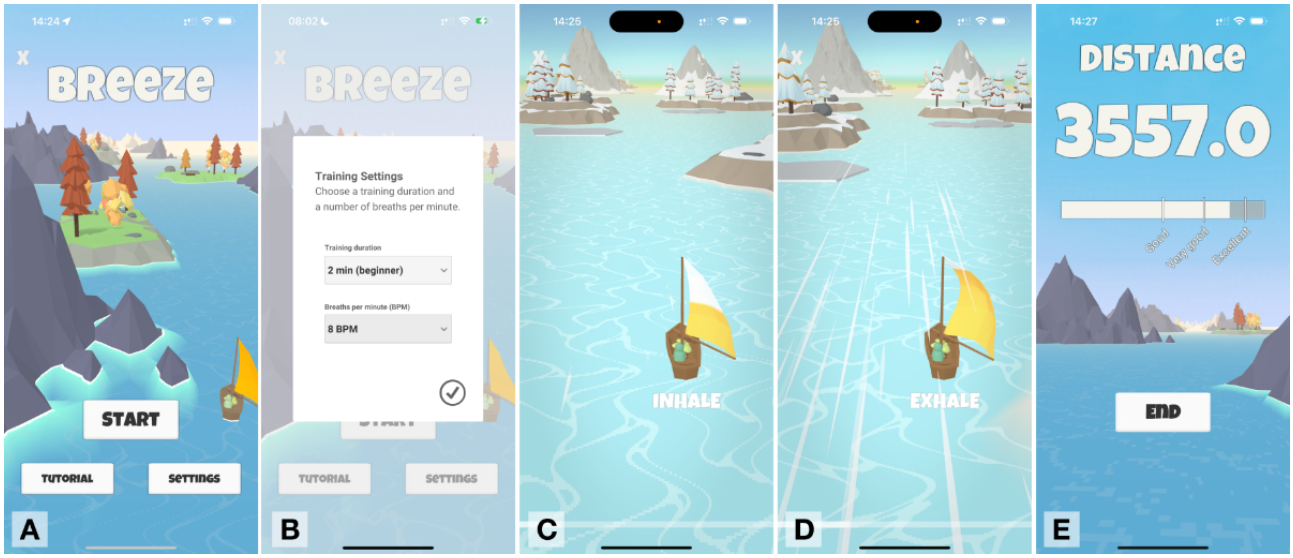
Intervention

The breathing training Breeze was 1 of 7 distinct relaxation exercises integrated into the CanRelax 2.0 app [33,34]. Breeze supports slow-paced breathing in a gamified setting, where the smartphone's microphone is used to continuously detect breathing phases in quasi-real time to accelerate a virtual sailboat if an individual is exhaling through the mouth [21,36]. A visual animation in the sail provides guidance according to a slow-paced breathing pattern (8 breaths per minute, the default setting; the breathing rhythm consists of inhale, pause, exhale). Breath detection has been proven to be highly accurate [23]. Figure 1 provides an overview of a slow-paced breathing session of Breeze. The intended usage pattern was 3 relaxation exercises per week, which participants could adjust by self-set goals, meaning that some weeks may have been completed without using Breeze.

XSL•FO
RenderX

Consistent with prior work [34], momentary relaxation was assessed using an 11-point visual analog scale (0=very relaxed and 10=very tense) before and after every third time a participant used any of the relaxation exercises included in the CanRelax 2.0 app. Momentary relaxation change was then calculated as the difference between the pre- and post-exercise values of perceived momentary relaxation. The default duration of the breathing exercise was set to 2 minutes, and the sailboat animation guided individuals to follow 8 breaths per minute. Although slow-paced breathing training is characterized as being especially effective with 6 breaths per minute [17], many people are not able to apply this rate and need to train toward it. Participants were able to change the duration and the guided breathing rate according to personal preferences. Breeze incorporates a concise, slide-based tutorial designed to familiarize users with the benefits of slow-paced breathing training while providing a brief overview of its guidance and interaction features. On initial access, the start button on the home screen was inactive. Completion of the tutorial activated the start button, allowing users to initiate their first training session [23].

Figure 1. One entire session of Breeze. (A) The home screen offers 3 primary options: initiating the Breeze session directly, accessing a tutorial on how to use the Breeze application, or adjusting the breathing session settings; (**B**) In the settings section, participants can set the training duration and the breaths per minute; (**C, D**) Textual and visual feedback is provided to guide participants through the breathing training; (**E**) On completing the session, the distance covered by the sailing boat is displayed as a performance metric, serving as an indicator of the participant's adherence to the guided breathing instructions. The study was conducted using a German-language interface.



Data Preparation

The CanRelax 2.0 app collected demographic information, momentary relaxation levels, app usage metrics, and additional information required for the RCT. For this work, we focused exclusively on data related to the intervention component Breeze. We extracted the timestamp of the exercise, its duration, and perceived momentary relaxation before and after the exercises. Moreover, we used the breathing detection classifier of Breeze 2 to measure the accuracy of the breathing rate [23].

The initial dataset comprised 787 Breeze sessions and 265 pairs of pre- and post-exercise values of participants' momentary relaxation, as only every third session was evaluated to minimize survey fatigue in the whole CanRelax study. We applied a series of exclusion criteria to ensure data quality and reliability. First, 2 pre- and post-exercise values of participants' momentary relaxation were removed, as they lacked either a pre- or post-exercise value (eg, Breeze session not completed). Additionally, 11 pairs of pre- and post-exercise values were excluded due to incomplete datasets, which made it impossible to retrieve essential exercise details (eg, duration). Moreover, 7 exercise sessions were excluded where user IDs were missing, as these were crucial for linking data across sessions. Furthermore, a duration-based outlier criterion was applied to remove exercise sessions that were either shorter than 30 seconds or longer than 1000 seconds, resulting in the exclusion of 13 exercise sessions. After these exclusions, the dataset consisted of 754 (754/787, 95.8%) Breeze sessions and 249 (249/265, 93.9%) pairs of pre- and post-exercise values.

Statistical Analysis

Descriptive Statistics

During participant registration, demographic data on sex, age, and education were collected. The use of CanRelax 2.0 relaxation exercises was tracked, enabling the calculation of each user's activity. In general, a participant is a person who was part of the study; an active user refers to an individual who performed at least one Breeze session within the corresponding time frame (eg, the entire study, a specific week).

Usage of Breeze Over Time

To assess potential differences between participants who engaged with the Breeze application (Breeze users) and those who did not (nonusers), a series of post hoc comparative analyses were conducted. Demographic characteristics, including sex, education level (nontertiary, tertiary), and age group (18 - 44, 45 - 64, and ≥65 years), were compared between Breeze users and nonusers using chi-squared tests. Furthermore, overall app engagement, quantified by the average weekly number of all exercises performed, was compared between

XSL•FO
RenderX

Breeze users and nonusers using a Welch *t* test. A *P* value of <.05 was considered statistically significant for all analyses.

For Breeze usage over time, we evaluated the recorded sessions over time. Each participant had a timestamp of the first session, which was used to calculate the relative day of each session. Moreover, the self-adjusted target for exercises was used to calculate the weekly share of Breeze. To compare the user engagement distribution before and after the intervention period (weeks >10), 2 metrics were used. First, the weekly number of distinct participants was tracked over time. Second, patient retention was assessed by comparing the number of participants from a given week (week 1) who reappeared in a subsequent week (week 2). This comparison generated a matrix illustrating the return rate of users from a source week (week 1) in a future week (week 2).

Breathing Adherence

Participant adherence to the guided breathing rate was assessed for each Breeze session. The application's validated sound detection algorithm [23] provided an estimated performed breathing rate. Adherence was then analyzed by (1) visually comparing distributions of measured breathing rates against the intended (user-selected) breathing rates via box plots, categorized by the intended rate; and (2) quantitatively assessing the relationship through a linear regression fit of measured rates on intended rates, supplemented by the Pearson correlation coefficient between the two.

Effect on Momentary Relaxation

To quantify the impact of the breathing exercise on momentary relaxation, a linear mixed effects fit was performed. This approach was chosen because the design involved repeated measures, with each participant providing momentary relaxation ratings at 2 time points (pre- and post-exercise). Mixed effects modeling allows us to account for within-session variability by including random intercepts for each session, thereby controlling for baseline differences in stress levels [37].

Each measured exercise in the cleaned dataset includes a pair of pre- and post-exercise values for the participant's momentary relaxation. To estimate the average change in momentary relaxation following the intervention, we fit a linear mixed effects model. The model included a dummy-coded fixed effect for the measurement time point (is_post: 0=preexercise and 1=postexercise). To account for repeated measures within the same exercise session, we included a random intercept for *experiment_id*. This resulted in the Model 1: *Perceived_stress_measure ~ is_post+(1 | experiment_id)*.

For the implementation [38], the Python (version 3.12.7) package statsmodels (version 0.14.2) [39] was used. The linear fixed effect model calculations in the package are based on Lindstrom et al [40].

Moreover, a second model was used to investigate the effect of other covariates. The outcome (*momentary relaxation change*) was defined as the difference between the pre- and post-exercise values of participants' perceived momentary relaxation. Lower values on this scale indicate a greater increase in relaxation, reflecting the intended effect of the intervention. It included the number of Breeze exercises already practiced by the person (*usage_count*), the duration of Breeze (*duration_minutes*), the average use of Breeze per week (*avg_weekly_earned*), the measured breathing rate of Breeze (*breathing_rate)*, and the number of days the participant had participated in the study up to that point (*relative_day*). The usage count, average weekly use, and number of relative days in the study were considered potential indicators of a learning effect during the observed period. Additionally, duration and breathing rate of the exercise were hypothesized to influence the momentary relaxation change directly. To this end, model 2 is formulated as follows: *momentary relaxation change ~ usage_count+duration_minutes + avg_weekly_earned+ breathing_rate+relative_day+(1 | experiment_id)*.

The results were evaluated based on the *P* value of each variable during a test for significance. Here, a 5% threshold was applied [41]. To assess the effect size of Breeze to promote momentary relaxation (see RQ3), we calculated Cohen *d* by the method proposed by Westfall et al [42] as Brysbaert and Stevens [43] recommended for linear mixed effect models. To this end, it was computed by dividing the fixed effect coefficient by the square root of the sum of the random intercept variance and residual variance, accounting for both between-subject and within-subject variability in the mixed effects framework.

Ethical Considerations

The trial was conducted in accordance with the principles of the Declaration of Helsinki. The Ethics Committee of Zurich, Switzerland, reviewed the study synopsis and concluded that the project does not fall under the Swiss Human Research Act (BASEC 2021-01071); therefore, a full ethics review was not required. All participants were adults (aged ≥18 y) and provided informed electronic consent within the app. Participants were informed about the study purpose, the voluntary nature of participation, the data collected through the app, and their right to withdraw at any time without negative consequences. Neither the study team nor the CanRelax 2.0 app requested any personally identifiable information at any stage of the study. Participants could provide a (fictional) nickname to be used by the conversational agent of the CanRelax 2.0 app. App-derived and survey data were stored on secure, access-restricted servers. All data handling complied with applicable Swiss data protection regulations. Participants did not receive any financial compensation for their participation, but the app was free of charge. The study was registered in the German Clinical Trials Register (DRKS00027546) on February 23, 2022. The first participant was enrolled on July 20, 2022, and the last app session took place on September 30, 2023.

Results

Descriptive Statistics

Assessments of momentary relaxation were collected both pre- and post-exercise for 249 (33.0%) of the sessions. Table 1 summarizes the demographic characteristics of the participants, whereas Table 2 provides an overview of engagement with the intervention components.

XSL•FO
RenderX

Analyses comparing Breeze users and nonusers revealed no statistically significant differences in terms of sex (χ^2_1=0.04; *P*=.85), age group (χ^2_2=0.2; *P*=.90), education level (χ^2_1=2.8; *P*=.10), and distress before study start (t_1=0.66; *P*=.51). Moreover, Breeze users performed a significantly higher average weekly number of all exercises (mean 3.85, SD 3.32) compared to nonusers (mean 1.99, SD 2.04; t_1=5.07; *P*<.001), indicating that Breeze use co-occurred with higher general app activity.

Table . Descriptive statistics for the 118 participants who used Breeze at least once.

Characteristics	Total number, n (%)
Sex assigned at birth (female)	96 (81.4)
Age (y)	
18 - 44	15 (12.7)
45 - 64	84 (71.2)
>64	19 (16.1)
Educational attainment	
Nontertiary	28 (23.7)
Tertiary	56 (47.5)
Missing	34 (28.8)
Number of completed Breeze sessions	754 (100)
Number of Breeze sessions with pre- and post-exercise relaxation values	249 (33.0)

Table . Descriptive statistics of Breeze usage.

Characteristic	Values, mean (SD)	Minimum	Maximum
Total number of Breeze sessions per Breeze users	6.39 (10.61)	1	92
Duration of a Breeze session (sec)	150 (41)	31	376
Preexercise value of momentary relaxation	4.59 (2.15)	1	10
Postexercise value of momentary relaxation	4.17 (2.35)	0	10
Momentary relaxation after having used Breeze[a]	−0.42 (1.67)	–6	7

[a]Calculated as the difference between the post- and preexercise value of momentary relaxation.

How Often Is Breeze Used Over Time?

Breeze usage is shown in Figures 2-5. Figures 2 and 3 illustrate the relationship between the number of Breeze sessions per week and the number of distinct Breeze users per week. While Figure 2 shows a decline in absolute numbers over time, Figure 3 demonstrates that the number of Breeze sessions per user remained relatively stable, fluctuating between 1.5 and 2 sessions per user throughout the study period. Figures 4 and 5 compare the usage of Breeze with the other 6 relaxation techniques available in the CanRelax 2.0 app. During weeks 1 to 3, participants exceeded the target number of exercises; however, this trend did not persist for the remainder of the intervention period. After the intervention ended, participants continued to use the CanRelax 2.0 app, but the average number of only Breeze exercises performed declined to approximately 2 per week. Notably, the proportion of Breeze sessions relative to all relaxation techniques in the app decreased over time.

XSL•FO
RenderX

Figure 2. Number of Breeze sessions per week over time (dashed line) and number of distinct users per week who used Breeze in that week (solid line). The dotted line at week 10 indicates the end of the study intervention period; participants were allowed to use the CanRelax 2.0 app afterward.

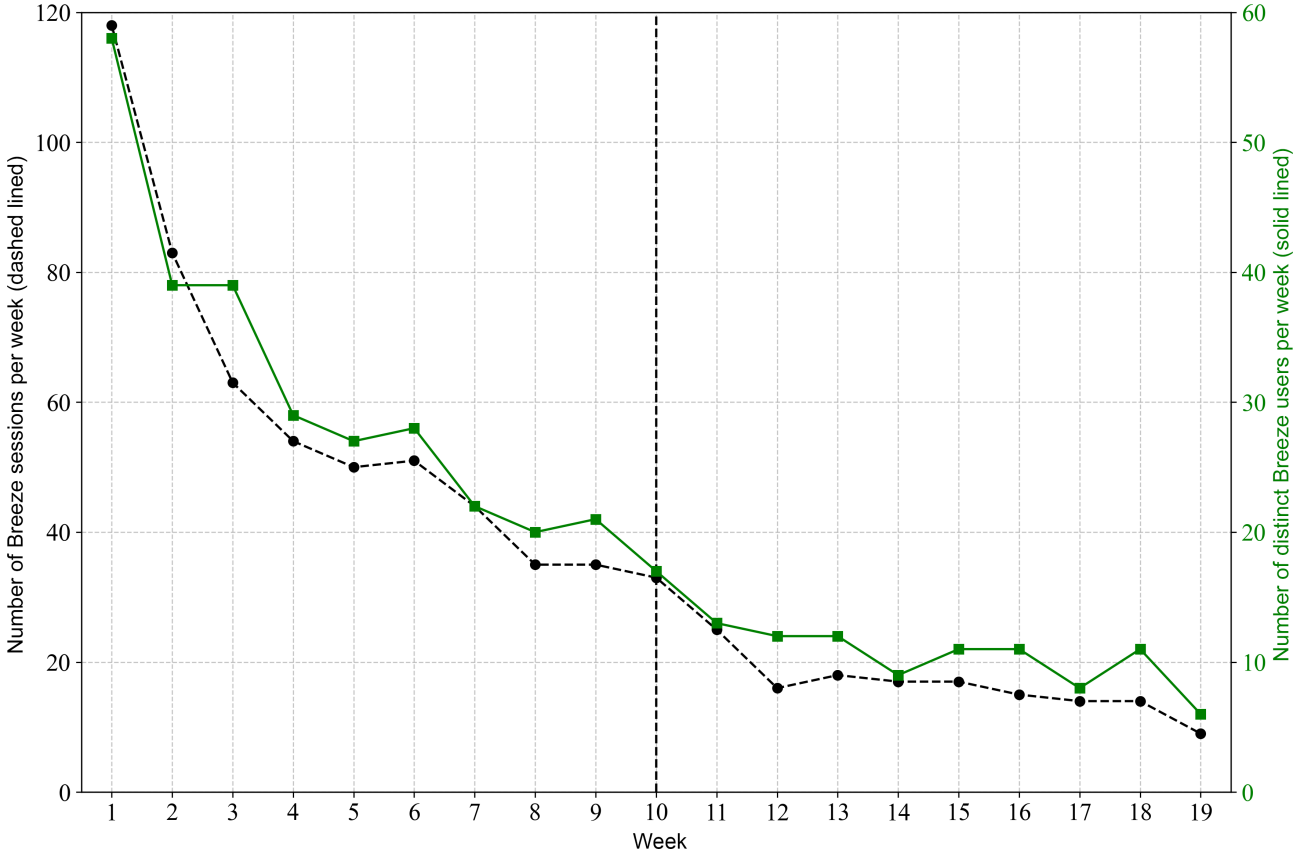
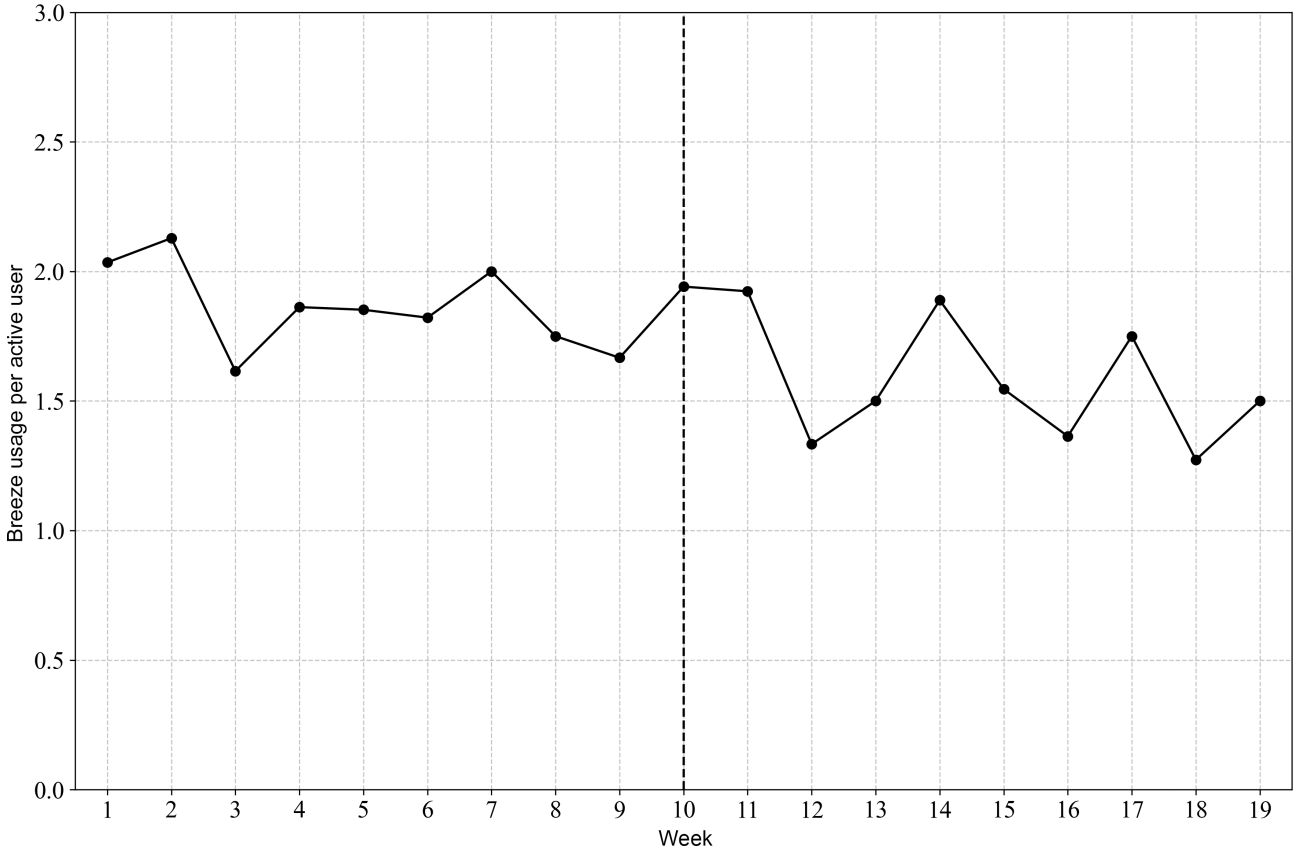


Figure 3. Average weekly Breeze sessions by active Breeze users.



XSL•FO
RenderX

Figure 4. Average weekly relaxation goal (black) and average number of relaxation exercises practiced per participant per week (green; all types of relaxation techniques). Note: The shadow represents the SDs. The dotted line at week 10 indicates the end of the study intervention period; participants were allowed to use the CanRelax 2.0 app afterward. However, from week 10 onward, the weekly relaxation goals could no longer be adjusted.

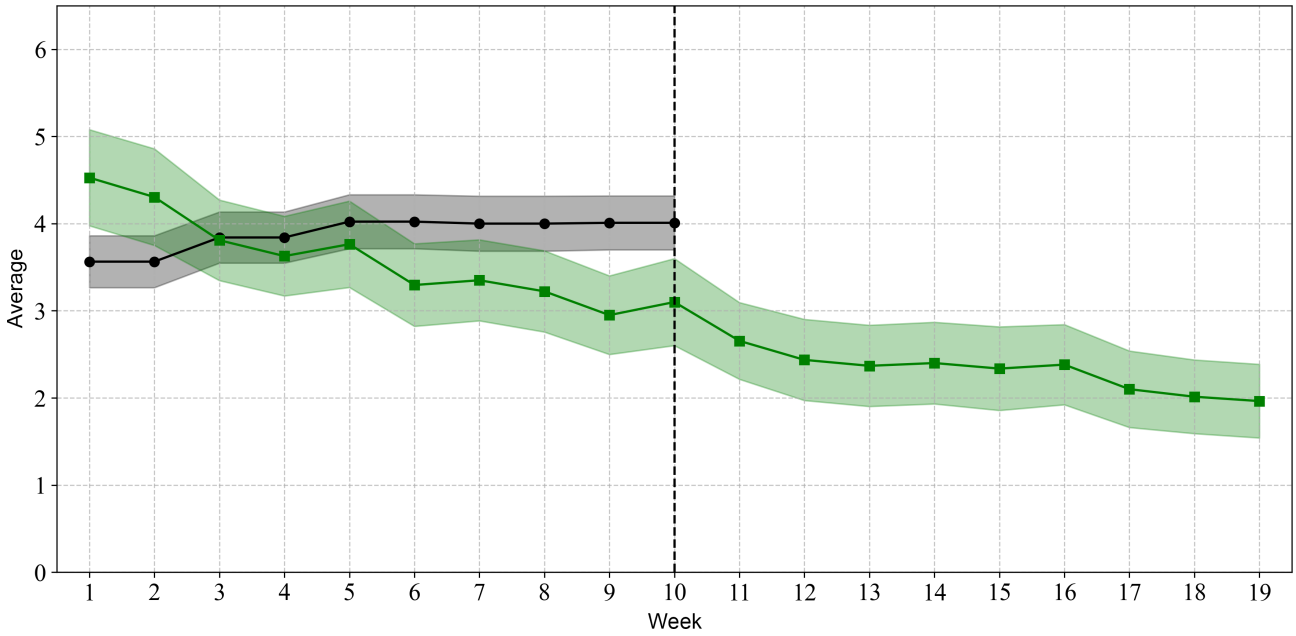
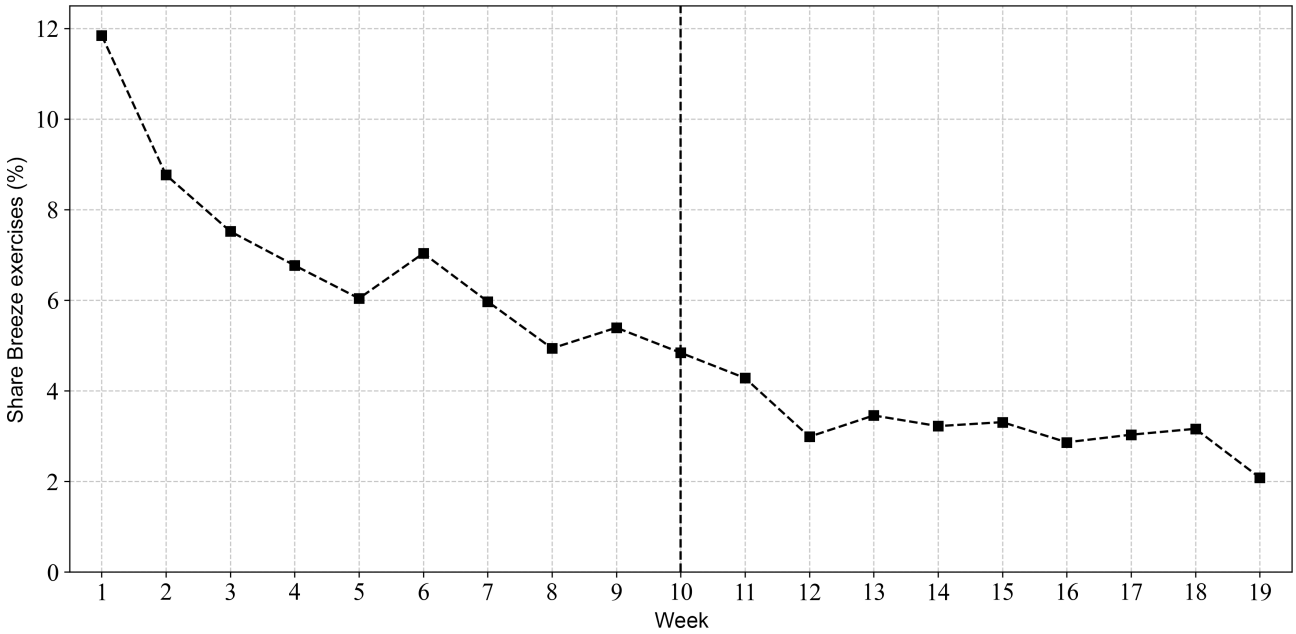


Figure 5. Share of practiced Breeze exercises relative to all 7 relaxation techniques.



For Breeze usage after the intervention period (weeks >10), user engagement stabilized at approximately 10.3 (SD 1.5) distinct users per week (Figure 2). A cohort analysis (Multimedia Appendix 1) further revealed that approximately 4.2 (SD 1.1) of these were returning users. This suggests a user base comprising a stable core alongside a broader, more sporadic group without a dominant trend toward either.

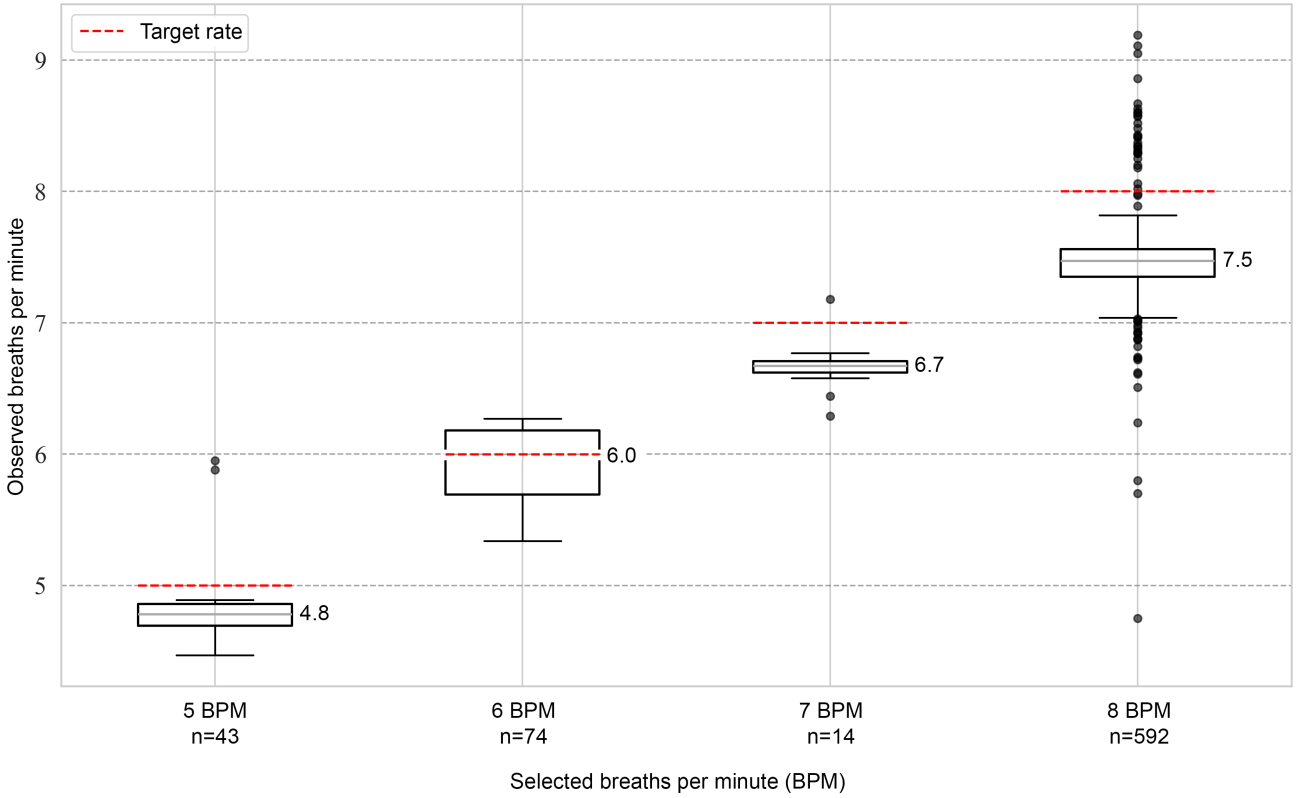
How Accurately Do Participants Adhere to the Guided Breathing Rate of Breeze?

Figure 6 shows the relationship between the measured breathing rate and the visually guided breathing rate selected by the Breeze users. In most sessions (592/754, 78.5%), users chose the default breathing rate of 8 breaths per minute.

A linear fit of the measured versus selected breathing rates indicates an R^2 of 0.81 and a root mean squared error of 0.38 breaths per minute. Moreover, the Pearson coefficient for the correlation is 0.9. These results indicate that Breeze guided participants well with visual cues and acoustic biofeedback.

XSL•FO
RenderX

Figure 6. Boxplot of measured breathing rate according to the selected breathing rate. Horizontal dashed lines indicate the selected (target) breathing rate for each group. The number next to the box plot shows the median of that group. Note: only 723 (95.9%) of 754 sessions are shown; the remaining 4.1% either had more or fewer breaths per minute selected and are thus not displayed.



What Is the Effect of Breeze on Momentary Relaxation?

Figure 7 illustrates the distribution of momentary relaxation changes for all measurements given the preexercise value. One finding is that higher preexercise distress values are linked to a greater change in momentary relaxation (Pearson correlation −0.26). In a first linear fixed-effects model with the experiment treated as a random effect, performing a Breeze exercise was associated with a significant decrease in the outcome (estimate −0.42; SE 0.11; P<.001), corresponding to a small effect size (Cohen d=0.19). In the covariate analysis (model 2), the duration of Breeze sessions was the only significant predictor of the outcome (Table 3). Longer Breeze sessions were associated with a reduction in the outcome (−0.56; SE 0.17; P=.002), corresponding to a small-to-medium effect (Cohen d=0.31). None of the other covariates, including the number of previously practiced Breeze exercises, average Breeze use per week, measured breathing rate, or number of days participated in the study, were significantly associated with the outcome (P values >.05). Plotting the momentary relaxation change against the exercise duration (Figure 8) suggests a favorable effect of longer durations as illustrated by Table 4, which lists effect sizes for 2, 3, and ≥4 minutes of Breeze's exercise durations. The results indicate that longer exercise durations increase the changes.

XSL•FO
RenderX

Figure 7. Scatter plot illustrating the distribution of momentary relaxation changes from before to after the breathing exercise. Note: The x-axis represents participants' momentary relaxation before using Breeze. The y-axis shows the change in momentary relaxation between post- and pre-exercise ratings. A negative change indicates an increase in participants' relaxation after using Breeze.

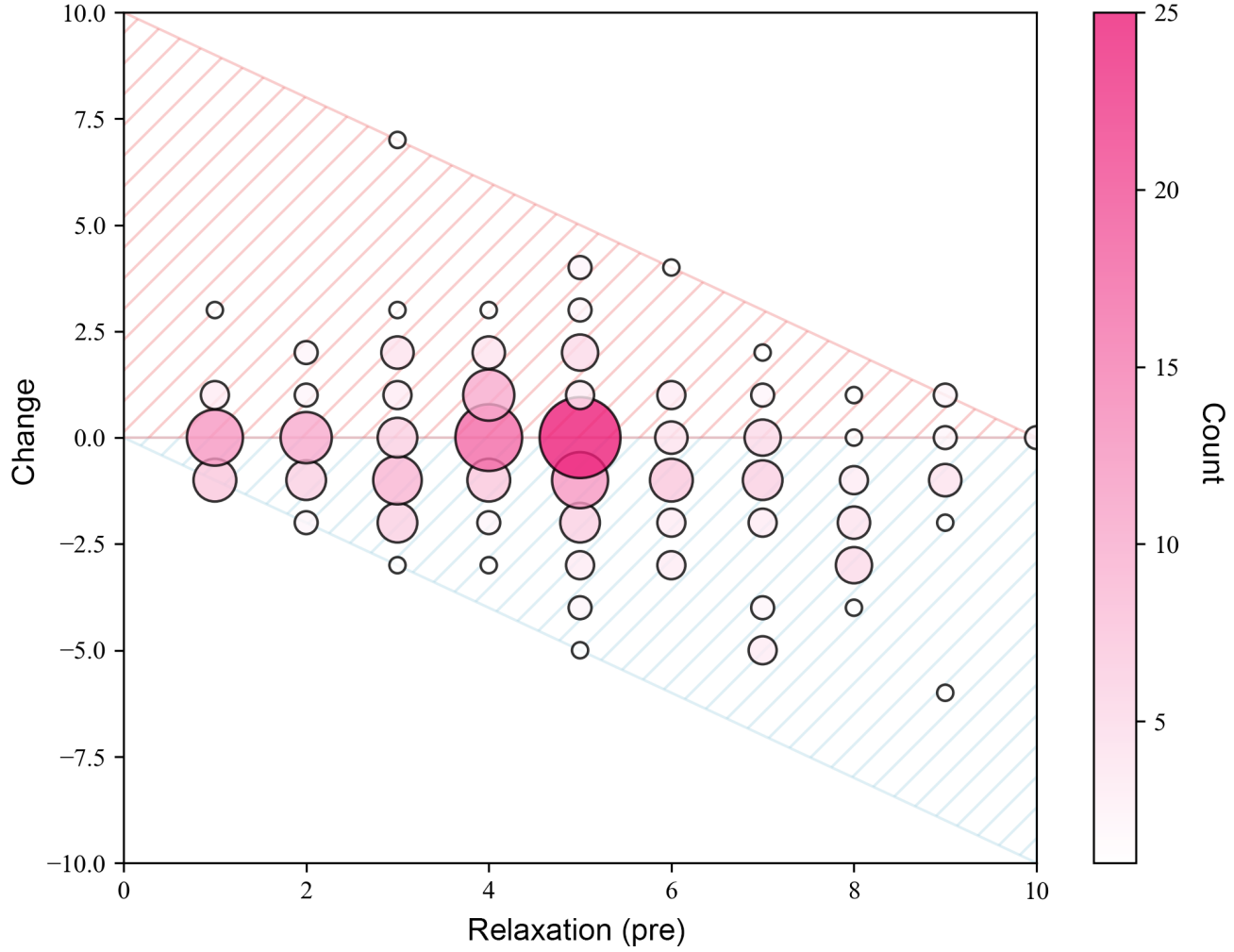


Table . Results of the investigation of covariates with model 2.

Variable	Estimate	SE	*P* value	Cohen *d*[a]
Number of Breeze exercises already practiced by the person	−0.007	0.008	.41	N/A[b]
Duration of Breeze	−0.557	0.172	.002	0.31
Average use of Breeze per week	−0.029	0.026	.27	N/A
Measured breathing rate guided by Breeze	0.023	0.135	.87	N/A
Number of days participated in the study	−0.001	0.002	.52	N/A

[a]Cohen *d* is only reported if the level of significance is ≤.05.

[b]N/A: not applicable.

XSL•FO
RenderX

Figure 8. Scatter plot of relaxation change (post-pre) against breathing duration together with linear fit. Note: The x-axis represents the duration of the Breeze session. The y-axis represents the change between the post- and pre-exercise value of participants' momentary relaxation. A negative change indicates an increase in participants' relaxation after using Breeze.

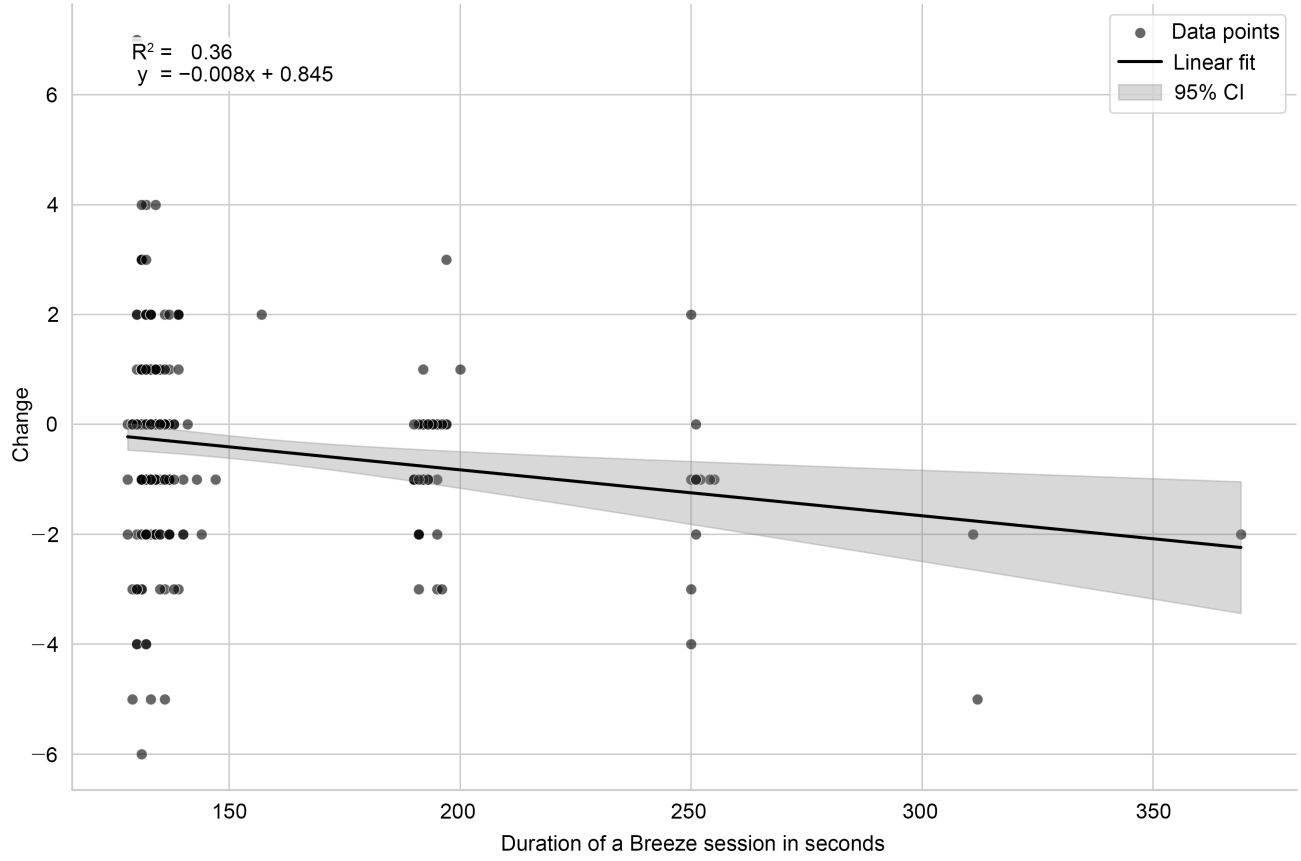


Table . Summary of stress reduction effects by breathing exercise duration.

Duration	Effect estimate	SE	*P* value	Cohen *d*	Observations	Unique users
2 min	−0.30	0.13	.017	0.13	382	68
3 min	−0.56	0.18	.035	0.26	86	19
4+ min	−1.53	0.42	.028	0.74	30	8

Discussion

Principal Findings

In this study, 3 dimensions of Breeze as a tool for helping participants with chronic conditions manage stress were assessed: its usage over time, the accuracy with which participants adhered to the visually guided, slow-paced breathing rate, and its effect on momentary relaxation. The analysis of Breeze usage over time demonstrated that the number of Breeze evaluations was comparable to that of the other 6 relaxation techniques included in the CanRelax 2.0 app. Although the intervention ended after 10 weeks, the participants were allowed to use the CanRelax 2.0 beyond the end of the study intervention period, and some users continued engaging with Breeze. This sustained usage suggests that the intervention effectively promotes long-term user engagement.

Accuracy in adhering to the preset breathing rate was also a key focus, as precise use is critical for achieving optimal relaxation. The default breathing rate was set to 8 breaths per minute. Participants could adjust this rate based on personal preference. As shown in Figure 6, the 8-breaths-per-minute rate was selected in most Breeze sessions (592/754, 78.5%). Regardless of the breathing rate, the Pearson correlation coefficient for participants' accuracy in following the breathing rate guided by Breeze is 0.9. This high adherence to the target rate suggests that the exercise is intuitively designed and easy to follow.

Regarding momentary relaxation, findings indicated larger effects in participants with higher preexercise values of perceived stress. Although relaxation effects were also observed in participants with lower stress levels, the effect size was smaller. This pattern suggests that Breeze effectively reduces momentary stress in people with cancer, underscoring the intervention's relevance in high-stress scenarios.

Finally, the relationship between Breeze session duration and its impact on momentary relaxation was examined. The results indicated that relaxation effects were associated with session duration, with participants who chose longer sessions reporting larger stress reduction. While the default 2-minute duration was associated with a stress reduction effect of −0.30, sessions of 4 minutes or more were associated with an effect of −1.53. These

XSL•FO
RenderX

findings support Breeze as a low-burden intervention capable of delivering immediate momentary relaxation.

When comparing the results of Breeze with the 6 other relaxation techniques included in the CanRelax 2.0 app [34], the momentary relaxation effect of a 2- or 3-minute Breeze exercise initially appears modest (a change of −0.4). It ranked seventh in absolute effect, with walking meditation (−0.6) being the second least effective and progressive muscle relaxation the most effective (−1.45). However, this comparison must be contextualized by the significantly different time investments required. Breeze sessions are exceptionally brief, with an average duration of 2.5 minutes, whereas other techniques require substantially more time, ranging from 5 to nearly 40 minutes (Multimedia Appendix 2).

When the relaxation effect was analyzed per minute of engagement, Breeze demonstrated a high level of efficiency. It was associated with a relaxation effect of 0.16 units per minute, a rate comparable to longer exercises, such as walking meditation (0.12 units per min) and progressive muscle relaxation (0.12 units per min).

Furthermore, if users chose Breeze sessions of 4 minutes or longer, the associated effect of −1.53 was the highest observed among all techniques in the study. This finding is particularly important because a 4-minute session remains shorter than any other relaxation technique offered by the CanRelax 2 app. Collectively, these results indicate that Breeze is not only a highly time-efficient intervention but also has the potential to deliver superior relaxation effects with only a modest increase in duration.

Limitations

This study presents 4 main limitations. First, Breeze was embedded within the CanRelax 2.0 app, offering 6 additional relaxation techniques [33,34]. This design allowed participants to select techniques based on their individual preferences, meaning some participants may never have used Breeze, resulting in a lack of data on these individuals. However, participants who chose to use Breeze did so voluntarily, without being specifically encouraged to select it. Despite this, the study demonstrated consistent engagement with Breeze among its users, maintaining a stable user base throughout the 10-week intervention period and beyond.

Second, momentary relaxation was assessed using a single-item scale, whereas an internationally validated, standardized evaluation scale, such as the Relaxation State Questionnaire [44], could have provided more detailed insights into the relaxation effects. However, given the study's goal to minimize participant burden, the single-item scale was appropriate and sufficient for capturing the momentary relaxation effects.

Third, the observed pattern of larger effects in participants with higher preexercise stress levels may be partially explained by regression to the mean. Participants are likely to use Breeze when experiencing elevated stress, and subsequent measurements naturally tend to return closer to their average baseline state, independent of the intervention's actual effect. While the data are consistent with Breeze providing stress reduction, this statistical phenomenon represents an alternative explanation that should be considered when interpreting the relationship between preexercise stress levels and relaxation outcomes.

Finally, findings indicated that longer Breeze sessions were associated with larger momentary relaxation effects, with sessions lasting 4 minutes or longer showing a significant decrease in stress. Hypothetically, participants who extended their session duration may have been more motivated, potentially leading to larger effects. However, this study did not assess whether elevated motivation directly contributes to increased momentary relaxation, nor did it examine the isolated effect of exercise duration on momentary relaxation. Future studies should further investigate the relationship between session duration and the momentary relaxation effects to confirm whether an extended session time consistently enhances relaxation outcomes across a broader sample.

Comparison With Prior Work

To the best of our knowledge, this is the first study that applies a gamified, biofeedback-guided breathing exercise in a real-world setting, where the exercise is available in a study conducted over a 10-week period [33,34]. Until now, Breeze has only been tested in a controlled setting, where its effectiveness has been shown [23].

The usage of Breeze over time was comparable to the other 6 relaxation techniques in the CanRelax 2.0 app. Additionally, the high Pearson correlation for participants' accuracy in following the breathing rate guided by Breeze highlights its intuitive use and the easy-to-understand approach – both key factors for adopting a DHI [45,46]. Furthermore, the gamified approach may help sustain participants' motivation over several weeks. Gamification offers a unique way for users to practice and refine skills in a safe, interactive environment, as seen in Breeze. It provides both extrinsic motivation (eg, by measuring the distance traveled with the sailing boat) and fosters intrinsic motivation through visible progress (eg, by seeing the wind fill the boat's sail when exhaling at the right time) [47]. By incorporating user control elements (ie, setting an individual duration and breathing rate), gamified interventions can enhance a sense of autonomy [48], while facilitating flow or immersion—key factors for enjoyment and sustained attention [47] to achieve higher and longer-lasting use.

Momentary relaxation is especially important for people dealing with chronic conditions, such as cancer. Cancer and its treatment may lead to increased levels of anxiety, fatigue, and physical discomfort, which can worsen overall quality of life [2,3]. Momentary relaxation provides temporary but effective relief from these stressors, helping to improve emotional well-being [49,50]. Moments of relaxation can also support building resilience and promote a more adaptive response to the challenges of chronic conditions [50]. Thus, Breeze might be an effective intervention component for various DHIs aiming to reduce acute stress.

Breathing exercises are an effective intervention for reducing stress [12]. Deep, controlled breathing slows the heart rate, lowers blood pressure, and reduces levels of stress hormones like cortisol [12,51,52]–leading to a perceived reduction in acute

XSL•FO
RenderX

stress. Additionally, breathing exercises calm the mind, which can positively impact interrupting cycles of anxiety or negative thinking, leading to a calmer mental state and improved emotional regulation [51-53]. Practicing regular breathing techniques with Breeze might, therefore, have the potential to increase resilience to stress and improve overall well-being. Breeze may therefore also offer benefits to individuals coping with other conditions, such as depression, anxiety, hypertension, or type 2 diabetes.

Conclusions

The study demonstrates that Breeze has the potential to effectively increase momentary relaxation among people with cancer. This finding supports the suitability of DHIs as a tool to assist individuals with chronic illnesses, providing them with various relaxation techniques to help manage daily challenges. Additionally, participants used the option of extending session duration and adjusting the breathing rate according to personal preferences. This suggests that future relaxation interventions may benefit from allowing users to customize settings, thereby meeting diverse needs within a single application.

Acknowledgments

The authors thank and acknowledge all study participants for their time and responses and all usability test participants for contributing throughout the app development process. The authors thank all individuals who contributed to the design and testing of the app, as well as those who facilitated or conducted recruitment, including the organizations involved. During the preparation of this manuscript, the authors used generative AI tools (ChatGPT 4o, Claude 3, and Gemini 2.5 Pro) for assistance with copy editing and for creating and reviewing the scripts for data analyses.

Funding

This work was supported by the Swiss Cancer Research foundation (KFS 4556-08-2018).

Data Availability

The data sets generated and analyzed during this study, as well as the Python analysis scripts, are openly available in the Open Science Framework (OSF) repository [54].

Authors' Contributions

Conceptualization: TK
Data curation: SS
Formal analysis: TB
Funding acquisition: JB, CW, TK
Investigation: TB, SS, JB, CW
Methodology: TB, YL, FS, PS, TK
Project administration: JB, CW, TK
Resources: JB, CW, TK
Supervision: TK
Visualization: TB
Writing – original draft: AB
Writing – review & editing: AB, TB, SS, YL, FS, PS, JB, CW, TK

Conflicts of Interest

AYB, TB, YL, and TK are affiliated with the Centre for Digital Health Interventions, a joint initiative of the Institute for Implementation Science in Health Care, University of Zurich; the Department of Management, Technology, and Economics at ETH Zurich; and the Institute of Technology Management and the School of Medicine at the University of St. Gallen. The Centre for Digital Health Interventions is funded in part by CSS, a Swiss health insurer; Mavie Next, an Austrian health insurer; and MTIP, a Swiss digital health investor. TK was also a co-founder of Pathmate Technologies, a university spin-off company that creates and delivers digital clinical pathways. However, neither CSS, Mavie Next, MTIP, nor Pathmate Technologies was involved in this study. JB received honoraria for workshops on digital health. CMW has active research grants for digital health projects awarded to the University of Zurich and Charité—Universitätsmedizin Berlin. These grants are funded by the German Health Care Innovation Fund, Krebsforschung Schweiz, Digitalisierungsinitiative der Zürcher Hochschulen, and Newsense Lab GmbH. The remaining authors have no conflicts of interest to declare.

Multimedia Appendix 1
Cohort analysis.
[DOCX File, 120 KB - games_v13i1e70297_app1.docx]

XSL•FO
RenderX

Multimedia Appendix 2
Momentary relaxation effect of non-Breeze exercises.
[DOCX File, 16 KB - games_v13i1e70297_app2.docx]

References

1. Weis J. Psychosocial care for cancer patients. Breast Care (Basel) 2015 Apr;10(2):84-86. [doi: 10.1159/000381969] [Medline: 26195935]
2. Antoni MH, Dhabhar FS. The impact of psychosocial stress and stress management on immune responses in patients with cancer. Cancer 2019 May 1;125(9):1417-1431. [doi: 10.1002/cncr.31943] [Medline: 30768779]
3. Gil F, Costa G, Hilker I, Benito L. First anxiety, afterwards depression: psychological distress in cancer patients at diagnosis and after medical treatment. Stress Health 2012 Dec;28(5):362-367. [doi: 10.1002/smi.2445] [Medline: 22972767]
4. Crosswell AD, Lockwood KG. Best practices for stress measurement: how to measure psychological stress in health research. Health Psychol Open 2020;7(2):2055102920933072. [doi: 10.1177/2055102920933072] [Medline: 32704379]
5. Nyberg ST, Fransson EI, Heikkilä K, et al. Job strain and cardiovascular disease risk factors: meta-analysis of individual-participant data from 47,000 men and women. PLoS ONE 2013;8(6):e67323. [doi: 10.1371/journal.pone.0067323] [Medline: 23840664]
6. Holt-Lunstad J, Smith TB, Baker M, Harris T, Stephenson D. Loneliness and social isolation as risk factors for mortality: a meta-analytic review. Perspect Psychol Sci 2015 Mar;10(2):227-237. [doi: 10.1177/1745691614568352] [Medline: 25910392]
7. Epel ES, Crosswell AD, Mayer SE, et al. More than a feeling: a unified view of stress measurement for population science. Front Neuroendocrinol 2018 Apr;49:146-169. [doi: 10.1016/j.yfrne.2018.03.001] [Medline: 29551356]
8. Carlson LE. Psychosocial and integrative oncology: interventions across the disease trajectory. Annu Rev Psychol 2023 Jan 18;74(1):457-487. [doi: 10.1146/annurev-psych-032620-031757] [Medline: 36104001]
9. Psychoonkologische diagnostik, beratung und behandlung von erwachsenen krebspatient innen. Leitlinienprogramm Onkologie. URL: https://www.leitlinienprogramm-onkologie.de/fileadmin/user_upload/Downloads/Leitlinien/Psychoonkologie/Version_2/LL_Psychoonkologie_Langversion_2.1.pdf [accessed 2024-10-29]
10. Riba MB, Donovan KA, Andersen B, et al. Distress Management, Version 3.2019, NCCN Clinical Practice Guidelines in Oncology. J Natl Compr Canc Netw 2019 Oct 1;17(10):1229-1249. [doi: 10.6004/jnccn.2019.0048] [Medline: 31590149]
11. Dehghan M, Jazinizade M, Malakoutikhah A, et al. Stress and quality of life of patients with cancer: the mediating role of mindfulness. J Oncol 2020;2020:3289521. [doi: 10.1155/2020/3289521] [Medline: 33381176]
12. Balban MY, Neri E, Kogon MM, et al. Brief structured respiration practices enhance mood and reduce physiological arousal. Cell Rep Med 2023 Jan 17;4(1):100895. [doi: 10.1016/j.xcrm.2022.100895] [Medline: 36630953]
13. Santos FD, Lacerda SS, Coelhoso CC, Barrichello CR, Tobo PR, Kozasa EH. The integration of meditation and positive psychology practices to relieve stress in women workers (flourish): effects in two pilot studies. Behav Sci (Basel) 2021 Mar 26;11(4):43. [doi: 10.3390/bs11040043]
14. Gao L, Curtiss J, Liu X, Hofmann SG. Differential treatment mechanisms in mindfulness meditation and progressive muscle relaxation. Mindfulness (N Y) 2018 Aug;9(4):1268-1279. [doi: 10.1007/s12671-017-0869-9]
15. Goessl VC, Curtiss JE, Hofmann SG. The effect of heart rate variability biofeedback training on stress and anxiety: a meta-analysis. Psychol Med 2017 Nov;47(15):2578-2586. [doi: 10.1017/S0033291717001003] [Medline: 28478782]
16. Chittaro L, Sioni R. Evaluating mobile apps for breathing training: the effectiveness of visualization. Comput Human Behav 2014 Nov;40:56-63. [doi: 10.1016/j.chb.2014.07.049]
17. Russell MEB, Scott AB, Boggero IA, Carlson CR. Inclusion of a rest period in diaphragmatic breathing increases high frequency heart rate variability: implications for behavioral therapy. Psychophysiology 2017 Mar;54(3):358-365. [doi: 10.1111/psyp.12791] [Medline: 27925652]
18. O'Connor DB, Thayer JF, Vedhara K. Stress and health: a review of psychobiological processes. Annu Rev Psychol 2021 Jan 4;72(1):663-688. [doi: 10.1146/annurev-psych-062520-122331] [Medline: 32886587]
19. Ang WHD, Chew HSJ, Dong J, Yi H, Mahendren R, Lau Y. Digital training for building resilience: systematic review, meta-analysis, and meta-regression. Stress Health 2022 Dec;38(5):848-869. [doi: 10.1002/smi.3154] [Medline: 35460533]
20. Schäfer SK, von Boros L, Schaubruch LM, et al. Digital interventions to promote psychological resilience: a systematic review and meta-analysis. NPJ Digit Med 2024 Feb 8;7(1):30. [doi: 10.1038/s41746-024-01017-8] [Medline: 38332030]
21. Shih CH, Tomita N, Lukic YX, Reguera Á, Fleisch E, Breeze KT. Smartphone-based acoustic real-time detection of breathing phases for a gamified biofeedback breathing training. Proc ACM Interact Mob Wearable Ubiquitous Technol 2019 Dec 11;3(4):1-30. [doi: 10.1145/3369835]
22. Lukic YX, Klein SS, Brügger V, Keller OC, Fleisch E, Kowatsch T. The impact of a gameful breathing training visualization on intrinsic experiential value, perceived effectiveness, and engagement intentions: between-subject online experiment. JMIR Serious Games 2021 Sep 14;9(3):e22803. [doi: 10.2196/22803] [Medline: 34519662]
23. Lukic YX, Teepe GW, Fleisch E, Kowatsch T. Breathing as an input modality in a gameful breathing training app. JMIR Serious Games 2022;10(3):e39186. [doi: 10.2196/39186]

XSL•FO
RenderX

24. Meyerowitz-Katz G, Ravi S, Arnolda L, Feng X, Maberly G, Astell-Burt T. Rates of attrition and dropout in app-based interventions for chronic disease: systematic review and meta-analysis. J Med Internet Res 2020 Sep 29;22(9):e20283. [doi: 10.2196/20283] [Medline: 32990635]
25. Torous J, Lipschitz J, Ng M, Firth J. Dropout rates in clinical trials of smartphone apps for depressive symptoms: A systematic review and meta-analysis. J Affect Disord 2020 Feb 15;263:413-419. [doi: 10.1016/j.jad.2019.11.167] [Medline: 31969272]
26. Pratap A, Neto EC, Snyder P, et al. Indicators of retention in remote digital health studies: a cross-study evaluation of 100,000 participants. NPJ Digit Med 2020;3(1):21. [doi: 10.1038/s41746-020-0224-8] [Medline: 32128451]
27. Lie SS, Karlsen B, Oord ER, Graue M, Oftedal B. Dropout from an eHealth intervention for adults with type 2 diabetes: a qualitative study. J Med Internet Res 2017 May 30;19(5):e187. [doi: 10.2196/jmir.7479] [Medline: 28559223]
28. Melville KM, Casey LM, Kavanagh DJ. Dropout from internet-based treatment for psychological disorders. Br J Clin Psychol 2010 Nov;49(Pt 4):455-471. [doi: 10.1348/014466509X472138] [Medline: 19799804]
29. Christensen H, Griffiths KM, Farrer L. Adherence in internet interventions for anxiety and depression. J Med Internet Res 2009 Apr 24;11(2):e13. [doi: 10.2196/jmir.1194] [Medline: 19403466]
30. Yardley L, Spring BJ, Riper H, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
31. Jakob R, Harperink S, Rudolf AM, et al. Factors influencing adherence to mHealth apps for prevention or management of noncommunicable diseases: systematic review. J Med Internet Res 2022 May 25;24(5):e35371. [doi: 10.2196/35371] [Medline: 35612886]
32. Castro O, Mair JL, Salamanca-Sanabria A, et al. Development of “LvL UP 1.0”: a smartphone-based, conversational agent-delivered holistic lifestyle intervention for the prevention of non-communicable diseases and common mental disorders. Front Digit Health 2023;5:1039171. [doi: 10.3389/fdgth.2023.1039171] [Medline: 37234382]
33. Schläpfer S, Schneider F, Santhanam P, et al. Engagement with a relaxation and mindfulness mobile app among people with cancer: exploratory analysis of use data and self-reports from a randomized controlled trial. JMIR Cancer 2024 May 31;10:e52386. [doi: 10.2196/52386] [Medline: 38819907]
34. Schläpfer S, Astakhov G, Pawel S, et al. Effects of app-based relaxation techniques on perceived momentary relaxation: observational data analysis in people with cancer. J Psychosom Res 2024 Sep;184:111864. [doi: 10.1016/j.jpsychores.2024.111864] [Medline: 39067182]
35. Barth J, Schläpfer S, Schneider F, et al. Mobile health intervention CanRelax reduces distress in people with cancer in a randomized controlled trial. NPJ Digit Med 2025 May 10;8(1):269. [doi: 10.1038/s41746-025-01688-x] [Medline: 40348854]
36. Lin IM, Tai LY, Fan SY. Breathing at a rate of 5.5 breaths per minute with equal inhalation-to-exhalation ratio increases heart rate variability. Int J Psychophysiol 2014 Mar;91(3):206-211. [doi: 10.1016/j.ijpsycho.2013.12.006] [Medline: 24380741]
37. Hilbert S, Stadler M, Lindl A, Naumann F, Bühner M. Analyzing longitudinal intervention studies with linear mixed models. TPM Test Psychom Methodol Appl Psychol 2019. [doi: 10.4473/TPM26.1.6]
38. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Softw 2015;67(1). [doi: 10.18637/jss.v067.i01]
39. Seabold S, Perktold J. Statsmodels: econometric and statistical modeling with python. Presented at: Python in Science Conference; Jun 28, 2010; Austin, Texas p. 92-96. [doi: 10.25080/Majora-92bf1922-011]
40. Lindstrom MJ, Bates DM. Newton—Raphson and EM algorithms for linear mixed-effects models for repeated-measures data. J Am Stat Assoc 1988 Dec;83(404):1014-1022. [doi: 10.1080/01621459.1988.10478693]
41. Kuznetsova A, Brockhoff PB, Christensen RHB. lmeTest package: tests in linear mixed effects models. J Stat Softw 2017;82(13). [doi: 10.18637/jss.v082.i13]
42. Westfall J, Kenny DA, Judd CM. Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. J Exp Psychol Gen 2014 Oct;143(5):2020-2045. [doi: 10.1037/xge0000014] [Medline: 25111580]
43. Brysbaert M, Stevens M. Power analysis and effect size in mixed effects models: a tutorial. J Cogn 2018 Jan 12;1(1):9. [doi: 10.5334/joc.10] [Medline: 31517183]
44. Steghaus S, Poth CH. Assessing momentary relaxation using the Relaxation State Questionnaire (RSQ). Sci Rep 2022 Sep 29;12(1):16341. [doi: 10.1038/s41598-022-20524-w] [Medline: 36175459]
45. Davis FD, Granić A. The Technology Acceptance Model - 30 Years of TAM: Springer International Publishing; 2024. [doi: 10.1007/978-3-030-45274-2]
46. Davis FD, Bagozzi RP, Warshaw PR. Technology acceptance model. J Manag Sci 1989;35(8):982-1003. [doi: 10.1287/mnsc.35.8.982]
47. Fleming T, Sutcliffe K, Lucassen M, Pine R, Donkin L. Serious Games and Gamification in Clinical Psychology: Comprehensive Clinical Psychology Elsevier; 2022:77-90. [doi: 10.1016/B978-0-12-818697-8.00011-X]
48. Fish MT, Saul AD. The gamification of meditation: a randomized-controlled study of a prescribed mobile mindfulness meditation application in reducing college students’ depression. Simul Gaming 2019 Aug;50(4):419-435. [doi: 10.1177/1046878119851821]

XSL•FO
RenderX

49. Victorson D, Kentor M, Maletich C, et al. Mindfulness meditation to promote wellness and manage chronic disease. Am J Lifestyle Med 2015 May;9(3):185-211. [doi: 10.1177/1559827614537789]
50. Asselmann E, Zenker M, Rückert F, Kische H, Pieper L, Beesdo-Baum K. Ecological momentary assessment and applied relaxation: results of a randomized indicated preventive trial in individuals at increased risk for mental disorders. PLoS ONE 2023;18(6):e0286750. [doi: 10.1371/journal.pone.0286750] [Medline: 37289760]
51. Perciavalle V, Blandini M, Fecarotta P, et al. The role of deep breathing on stress. Neurol Sci 2017 Mar;38(3):451-458. [doi: 10.1007/s10072-016-2790-8] [Medline: 27995346]
52. Ma X, Yue ZQ, Gong ZQ, et al. The effect of diaphragmatic breathing on attention, negative affect and stress in healthy adults. Front Psychol 2017;8:874. [doi: 10.3389/fpsyg.2017.00874] [Medline: 28626434]
53. Epe J, Stark R, Ott U. Different effects of four yogic breathing techniques on mindfulness, stress, and well-being. OBM Integr Complement Med 2021;06(3):1-21 [FREE Full text] [doi: 10.21926/obm.icm.2103031]
54. CanRelax: effectiveness of a mindfulness and relaxation self-care app on distress of people with cancer. OSF. URL: https://osf.io/9nqyb/overview [accessed 2025-11-26]

Abbreviations

DHI: digital health intervention
RCT: randomized controlled trial
RQ: research question
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

Edited by A Coristine; submitted 19.12.24; peer-reviewed by A Gazis, C Nnajiofor; revised version received 26.10.25; accepted 27.10.25; published 16.12.25.

Please cite as:
Bischof AY, Budig T, Schläpfer S, Lukic YX, Schneider F, Santhanam P, Barth J, Witt CM, Kowatsch T
Long-Term Usage of Breeze, a Gamified Breathing Training App, and Its Effect on Momentary Relaxation in People With Cancer: Cohort Study
JMIR Serious Games 2025;13:e70297
URL: https://games.jmir.org/2025/1/e70297
doi:10.2196/70297

© Anja Yvonne Bischof, Tobias Budig, Sonja Schläpfer, Yanick Xavier Lukic, Fabian Schneider, Prabhakaran Santhanam, Jürgen Barth, Claudia M Witt, Tobias Kowatsch. Originally published in JMIR Serious Games (https://games.jmir.org), 16.12.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Utilizing Gamification, Artificial Intelligence, and mHealth for the Professional Development of Maternal Care Providers: Exploratory Pilot Cross-Sectional Study Assessing Providers' Satisfaction in Primary Health Care Centers in Lebanon

Mohamad Alameddine[1], MPH, PhD; Nadine Sabra[2], PharmD; Nour El Arnaout[2], MPH; Asmaa El Dakdouki[2], MSc; Mahmoud El Jaouni[3], MSc; Randa Hamadeh[4], MPH, DPH; Abed Shanaa[5], MD; Shadi Saleh[2,6], MPH, PhD

[1]College of Health Sciences, University of Sharjah, Sharjah, United Arab Emirates

[2]Global Health Institute, American University of Beirut, PO Box 11-0236, Riad El Solh, Beirut, Lebanon

[3]Faculty of Health Sciences, American University of Beirut, Beirut, Lebanon

[4]Ministry of Public Health, Beirut, Lebanon

[5]The United Nations Relief and Works Agency for Palestine Refugees in the Near East, Beirut, Lebanon

[6]Department of Health Management and Policy, Faculty of Health Sciences, American University of Beirut, Beirut, Lebanon

Corresponding Author:
Shadi Saleh, MPH, PhD
Global Health Institute, American University of Beirut, PO Box 11-0236, Riad El Solh, Beirut, Lebanon

Abstract

Background: High maternal morbidity and mortality rates globally, especially in low-income and lower-middle–income countries, highlight the critical role of skilled health care providers (HCPs) in preventing pregnancy-related complications among disadvantaged populations. Lebanon, hosting over 1.5 million refugees, is no exception. HCPs face significant challenges, including resource constraints and limited professional development opportunities, underscoring the need for continuous learning and innovative educational interventions. Artificial intelligence (AI) and gamification show promise in enhancing clinical performance and evidence-based practice.

Objective: Considering the limited evidence on the effectiveness of integrating gamification and AI in a mobile app for professional development of HCPs providing maternal health services, this pilot study aims to assess the satisfaction and acceptability of HCPs with a novel mLearning tool, titled the “GAIN MHI” app (gamification, artificial intelligence, and mHealth network for maternal health improvement), at selected primary health care centers in Lebanon.

Methods: This is a cross-sectional study that presents data collected from 12 participating HCPs, primarily obstetricians and midwives who have been using the GAIN MHI mobile app for professional development and learning. The survey used included Likert scale questions to assess HCPs’ satisfaction, engagement, and evaluation of the gamification and AI components of the app. Open-ended questions gathered qualitative feedback on app preferences and potential improvements. Statistical analysis was performed to derive insights from the quantitative data collected. Subsequently, a descriptive analysis was performed, presenting the frequencies and percentages of various participant characteristics, as well as responses to the survey across all sections.

Results: A total of 85% (n=10) of the HCPs, including midwives and doctors, were satisfied with the GAIN MHI mobile app, the user interface, and various content features. Engagement levels were robust (64.6%, SD 6.2%), notably impacting clinical routines and theoretical knowledge. The gamification and AI components garnered strong positive feedback, enhancing learning enjoyment (11/12, 92%). From a qualitative perspective, users expressed appreciation for the app’s diverse content, user-friendliness, and motivation for continuous learning. Suggestions for expanding the content included a wide range of health topics, highlighting the app’s potential applicability in various health care fields.

Conclusions: HCPs, especially those practicing in underserved areas, face challenges in accessing professional development opportunities, highlighting the need for innovative pedagogical approaches using mobile technologies. This pilot study underlines the potential of using AI-based digital solutions for professional development with the aim of improving the quality of health services—in this case, maternal health services—through continuous learning and updates on the most recent evidence-based clinical guidelines. Future research should investigate the feasibility of applying similar solutions on a larger scale to reach a wider range of HCPs and cover other health topics. The applicability of such solutions in different contexts and low-resource settings should also be explored.

XSL•FO
RenderX

(JMIR Serious Games 2025;13:e53735) doi:10.2196/53735

KEYWORDS

maternal health; gamification; artificial intelligence; professional development; mHealth; health provider; medical knowledge

Introduction

Overview of Maternal Health Status on Global and National Levels

Globally, around 800 women die on a daily basis from preventable causes related to pregnancy and childbirth, with 95% of the deaths occurring in low and lower-middle-income countries [1]. In the context of Lebanon, where women of childbearing age constitute the majority of the refugee population, increased maternal mortality rates and pregnancy-related complications are documented, particularly among socioeconomically disadvantaged populations [2-5]. In terms of maternal health services utilization, the administration of care by proficient health care providers (HCPs) prior to, during, and postpartum holds the potential to mitigate mortality rates among women and neonates [1].

Health Care Providers: Roles, Challenges, and Pathways to Improvement in Maternal Health

HCPs play a vital role within the health care system, addressing a wide range of matters such as maternal health, family planning, neonatal care, child nutrition, and vaccination [6]. It is essential to acknowledge that HCPs encounter various challenges in fulfilling their roles [7,8]. Resource constraints, including limited continuing education programs and career development, a shortage in health care professionals, and disparities in geographic distribution of providers' care, can hinder their efforts, particularly in Lebanon [9-11]. A promising approach to enhance the clinical performance and the evidence-based practice of HCPs is advocating for continuous learning and professional development [12,13]. The foregoing could be attained by developing innovative approaches and educational interventions that ensure their enduring involvement [13,14].

Existing Innovative Approaches and Educational Interventions for Maternal Health Targeting Health Care Providers

Clinical competence among HCPs plays a pivotal role in ensuring the quality of care [15]. Lifelong learning serves as an indispensable component of practice-based learning and improvement for HCPs, as it directly contributes to maintaining and enhancing their clinical competence [16]. The dynamic nature of health care requires providers to stay informed about the latest medical advancements, evidence-based practices, and technological innovations [17,18]. Considering the expenses and practical difficulties associated with delivering in-person continuing education training programs for HCPs, the use of mobile phones and comparable portable electronic devices emerges as a promising approach [19,20], commonly referred to as mobile learning or mLearning [21]. This approach effectively extends the reach of continuing education programs to distant health care workers, providing them with current and relevant information [19]. Only a limited number of research endeavors have specifically addressed the use of mobile phones to enhance the knowledge and competencies of HCPs in maternal health in low- and middle-income countries [19-22]. One example of the mLearning interventions is the Safe Delivery App (SDA), which was developed to enhance the proficiency and skills of HCPs through visually animated instruction videos and guiding steps for effectively addressing fundamental obstetric and neonatal emergencies, particularly in low-resource settings. The SDA app was successfully implemented in the Democratic Republic of the Congo [21] and was subsequently implemented [21,23] in some countries of the Middle East and North Africa region [24]. In order to maximize the advantages of the SDA app, an upgraded version was developed in India, incorporating gamification elements to enhance the motivation of the users while transforming the learning process into an interactive game [25].

Gamification and Artificial Intelligence for Health Care Providers in the Field of Maternal Health

Over the past few years, there has been a notable increase in global recognition regarding the application of gamification across various domains [26]. Gamification, defined as "the practice of introducing 'game-like' dynamics into routine activities to engage users" [22], holds the promise of enabling HCPs to engage in dynamic learning, tackle clinical challenges, and garner experience within a secure and controlled environment [22,23]. Gamification is characterized by the incorporation of game design elements into non-game-related processes, with inclusion of badges, points, and scoring systems [26-28]. Given the relatively new nature of the field, research is essential to establish evidence-based practices, guidelines, and standards for gamification and its integration in mobile health (mhealth) apps [28]. Exploring gamification methods through research can assist in identifying the best game design elements for diverse contexts and populations and integrating innovative approaches into existing systems [28]. Besides gamification, the use of artificial intelligence (AI) in mHealth apps is being investigated, with the goal of delivering health care that is more personalized [29,30] and improving overall health care provision [31,32]. AI refers to the capacity of systems to analyze data and uses computers and machines to amplify human decision-making, problem-solving skills, and the innovation driven by technology [33-35]. Within the realm of clinical practice, AI aids the HCPs in making diagnoses and treatment choices [36,37], by providing them with specialized knowledge and skills [38]. As for maternal health, studies showed that AI may reduce the time needed to train HCPs in evaluating preterm deliveries [39], and help in predicting different maternal complications such as gestational diabetes [40,41] and postpartum depression [42,43]. There is clear evidence of AI's impact extending to nearly every facet of the health care industry [44], yet the integration of AI with gamification in mHealth apps is still a novel and evolving concept.

XSL•FO
RenderX

Study Aim

Considering the limited evidence on integrating gamification and AI in a mobile app targeting the HCPs in maternal health, this pilot study aims to assess the satisfaction and acceptability of HCPs with a novel mLearning tool, titled the "GAIN MHI" app (gamification, artificial intelligence, and mHealth network for maternal health improvement), at selected primary health care centers (PHCs) in Lebanon.

Methods

The GAIN MHI App

The GAIN MHI app (Figure 1) is a mobile app developed by the Global Health Institute at the American University of Beirut as part of the GAIN MHI project. Through this mobile app, the project aims to improve the quality of care provided to pregnant women by their specialized HCPs (mainly obstetrics and gynecology [OBGYN] specialists and midwives). The GAIN MHI mobile app was characterized by 2 components: the gamification of learning and the AI components, representing a novel intervention for the professional development of the HCPs. Gamification of learning is defined within the context of this project as the app of game principles and design elements for the purpose of professional development. The gamification component in the GAIN MHI mobile app included "trivia-like" concept and monthly recognition as most valuable player and most improved player (MIP), coupled with monetary incentives equivalent to US $150 each. The questions in the app fell under 5 [5] categories related to maternal health: Prevention, Diagnostic, Management, Miscellaneous, and COVID-19 related questions. The category selection process relies on a randomized spin of a wheel, and each user is allocated a maximum of 30 attempts per month. Consequently, users are required to respond to 30 questions within the given monthly timeframe.

The project, through the GAIN MHI mobile app, used AI algorithms to identify areas with knowledge gaps based on the incorrect answers of health providers. The question generation formula encompassed both personalization and diversity. This allowed the tailoring of knowledge shared in subsequent questions to make the latter individualized to each provider based on their performance on the multiple-choice questions (MCQs). To this end, a repository of questions was built. Topic modeling was used on the repository. This allowed the identification of the topic of wrongly answered questions. Explanations were provided after answering each MCQ. Further, the application was coupled with automated WhatsApp messages sent to the users regarding their score, right/wrong answers, remaining spins for the month, monthly performance, and any gained rewards. The content of the app was available in both English and Arabic languages.

XSL•FO
RenderX

Figure 1. The GAIN MHI mobile app user interface. GAIN MHI: gamification, artificial intelligence, and mHealth network for maternal health improvement.



Project Implementation, Population, and Setting

The GAIN MHI app database was created by a medical consultant specializing in OBGYN. The database incorporated questions focused on antenatal and postnatal care practice guidelines targeting OBGYN physicians, questions focused on antenatal and postnatal care practice guidelines targeting midwives, corresponding options, correct answers, explanations, and resources for references. To cater to the Arabic-speaking HCPs, an Arabic version of the app was generated. Subsequently, a validation process was performed in which questions dedicated to OBGYNs and midwives were validated by specialists from the same field. The aim was to ensure that the database content was tailored to all intended users. After the validation, the intervention's components were eventually adjusted based on the participants' inputs and suggestions. Additionally, a testing phase was conducted to guarantee the satisfaction of users with the look, feel, and usability of the mobile app. Based on that, some features of the app were upgraded as per the shared feedback.

This study took place over a period of 21 months (from September 2021 to May 2023), in 6 PHCs located in different governorates of Lebanon, including South, Beirut, and Mount Lebanon. Among these PHCs, 3 were affiliated with the United

XSL•FO
RenderX

Nations Relief and Works Agency and the others with the Ministry of Public Health. A total of 15 HCPs (OBGYN specialists and midwives) in maternal health were enrolled in the intervention, of which 9 were midwives and 6 were OBGYN specialists. At the termination of the study, 12 participants (8 midwives, 67% and 4 OBGYN specialists, 33%) maintained their participation, whereas 3 participants dropped out. Among those who continued, 67% (n=8) were female and 33% (n=4) were male. Regarding language preferences, 10 participants used the app in English while 2 participants used it in Arabic (Table 1).

Table . General characteristics of participants (N=12): frequency and percentage distribution.

Variables	Values, n (%)
PHC[a] affiliation	
MOPH[b]	6 (50)
UNRWA[c]	6 (50)
Title	
Midwife	8 (67)
Doctor	4 (33)
Sex	
Female	8 (67)
Male	4 (33)
Language of app use	
English	10 (83)
Arabic	2 (17)

[a]PHC: primary health care center.

[b]MOPH: Ministry of Public Health.

[c]UNRWA: United Nations Relief and Works Agency.

Data Collection

In this study, both quantitative and qualitative research methods were used to comprehensively assess the GAIN MHI app, providing numerical insights and in-depth feedback concerning this app. A cross-sectional survey was designed considering the features of the GAIN MHI app. It included Likert scale and open-ended questions to gather both quantitative rates and qualitative perceptions of the app users, respectively. The 5-point Likert scale was used to assess the level of satisfaction of HCPs with the characteristics of the app, their level of engagement, and their evaluation of the Gamification and the AI components of the app. The scale consists of the following points: highly dissatisfied [1], dissatisfied [2], neither satisfied nor dissatisfied [3], satisfied [4], and highly satisfied [5]. Complementary to this, the open-ended questions explored feedback from the HCPs on the favorite aspect of the mobile app, preferred types of reward, maternal health topics that should be added to the app, health topics other than obstetrics/gynecology where this app could be applied, as well as suggestions for future improvement. Together, the quantitative and qualitative methods reveal the satisfaction with the app and identify potential areas for enhancement. The survey was filled out in person with the 12 participants after signing a written consent form.

To ensure the reliability of the questionnaire, Cronbach α (α=0.82) was calculated, indicating high internal consistency. Additionally, reliability was assessed for each subscale, with results of 0.74, 0.64, and 0.8 reflecting acceptable, moderate, and high reliability among the 3 sections, respectively. These results demonstrate that while the overall survey is reliable in this pilot study, it can be refined for improvement in future research to ensure even greater reliability. The validity of the questionnaire was also assessed. To achieve this, a thorough review of several established questionnaires in the literature was conducted. Drawing from these sources, the questionnaire was designed in collaboration with domain experts to align with our specific aims and objectives. This process ensured that our instrument accurately captures the necessary data while being tailored to the unique aspects of our study.

Ethical Considerations

Ethical approval was sought and obtained from the Institutional Review Board at the American University of Beirut (protocol number SBS-2020 - 0317) and the respective ethical committees at the Ministry of Public Health and the United Nations Relief and Works Agency. Before participation, HCPs from the recruited centers received an explanation of the study's objectives, procedures, potential risks (there is no more than the "minimal risk" in participating in this project), and benefits. Consent forms were signed in person, ensuring participants understood the study details and their rights, including their right to withdraw at any point without consequences.

In order to maintain confidentiality and privacy, the gathered data was stored on a password-protected computer only accessible to the research team members. The data was not kept in public archives. All data will be properly destroyed after the required retention period, which is typically 3 years.

XSL•FO
RenderX

Participants' privacy will be secured in all written and published data arising from this study. Names of participants or other identifying information will not be used in reports or published papers.

While no general compensation was provided to study participants for their overall participation, HCPs could receive incentives within the GAIN MHI app's gamification component based on specific performance metrics. Recognitions, such as the most valuable player or the MIP, were associated with monthly monetary rewards equivalent to US $150.

Statistical Analysis

Statistical analysis was conducted to derive insights from the quantitative data collected from the cohort of 12 participants who maintained their participation till the end of the study. The initial phase encompassed data management and recoding using Stata MP 17. Subsequently, a descriptive analysis was performed, presenting the frequencies and percentages of various participant characteristics, as well as responses to the survey across all sections. This includes assessment of satisfaction and engagement levels and evaluation of gamification and AI components in the GAIN MHI mobile app. To summarize these sections, mean percentages were calculated, accompanied by SDs and 95% CIs. Concurrently, the qualitative aspect of the study used a coding reliability thematic analysis approach, following a small-q paradigm [45]. This method emphasizes a systematic and transparent methodology to ensure the dependability of the research process, data collection, and analysis. Purposive sampling was used, where qualitative data were collected from all 12 HCPs who used the GAIN MHI app and maintained their participation till the end of the study, regardless of data saturation [46]. Thematic analysis was conducted manually for each of the 5 open-ended questions separately, beginning with an in-depth immersion in the data to gain familiarity with its content. Coding was then performed, wherein responses were systematically labeled with codes, which were subsequently organized into themes. These themes were reviewed iteratively to ensure they accurately represented the data and captured both frequently mentioned ideas and unique perspectives. Finally, a narrative synthesis was developed to integrate all relevant themes. To enhance the confirmability of the study's findings and interpretations, participants' direct quotes were incorporated into the analysis. Credibility of the study's findings was established through prolonged engagement with the data, allowing for a deep understanding of its nuances and depth. Transferability of the study's findings was addressed through the use of purposeful sampling.

Results

Quantitative Analysis

Table 2 provides an overview of participant satisfaction levels with various aspects of the GAIN MHI mobile app. The overall satisfaction percentage calculated is 85% (SD 7%). In terms of the mobile app in general, a majority of participants (10/12, 83%) reported satisfaction, with 17% (n=2) of the participants indicating high satisfaction and no dissatisfaction reported. The user interface of the mobile app showed 50% (n=6) of the participants expressing high satisfaction, 42% (n=5) of the participants expressing satisfaction, and 8% (n=1) of the participants indicating dissatisfaction. All of the surveyed app users were satisfied with the number of MCQs per month, followed by 92% satisfaction with the explanation after each question, the relevance of the links to additional resources, the scoring of the MCQs, and the user interface of the app. The area with comparatively lowest satisfaction is related to the topics covered by the MCQs.

XSL•FO
RenderX

Table . Satisfaction with gamification, artificial intelligence, and mHealth network for maternal health improvement (GAIN MHI) mobile app.

Variables	Values
Mobile app in general, n (%)	
Highly dissatisfied	0 (0)
Dissatisfied	0 (0)
Neither satisfied nor dissatisfied	0 (0)
Satisfied	10 (83)
Highly satisfied	2 (17)
Number of multiple-choice questions (MCQs) per month (30 questions per month), n (%)	
Highly dissatisfied	0 (0)
Dissatisfied	0 (0)
Neither satisfied nor dissatisfied	0 (0)
Satisfied	9 (75)
Highly satisfied	3 (25)
Topics that the MCQs covered, n (%)	
Highly dissatisfied	0 (0)
Dissatisfied	0 (0)
Neither satisfied nor dissatisfied	2 (17)
Satisfied	4 (33)
Highly satisfied	6 (50)
Explanations after each MCQs encounter, n (%)	
Highly dissatisfied	1 (8)
Dissatisfied	0 (0)
Neither satisfied nor dissatisfied	0 (0)
Satisfied	7 (58)
Highly satisfied	4 (33)
Relevance of the links mentioned in the additional resources, n (%)	
Highly dissatisfied	0 (0)
Dissatisfied	0 (0)
Neither satisfied nor dissatisfied	1 (8)
Satisfied	6 (50)
Highly satisfied	5 (42)
Point/scoring system, n (%)	
Highly dissatisfied	0 (0)
Dissatisfied	0 (0)
Neither satisfied nor dissatisfied	1 (8)
Satisfied	7 (58)
Highly satisfied	4 (33)
User interface of the mobile app, n (%)	
Highly dissatisfied	0 (0)
Dissatisfied	1 (8)
Neither satisfied nor dissatisfied	0 (0)
Satisfied	5 (42)
Highly satisfied	6 (50)

XSL•FO
RenderX

Variables	Values
Total satisfaction (%)	
Mean (SD)	85 (7.0)
95% CI	80.5-89.5

Table 3 provides insights into the participants' levels of engagement with the GAIN MHI mobile app. Participants' engagement with links from the knowledge resource center was notable, as 50% (n=6) of the participants indicated engaging "Sometimes," and 33% (n=4) of the participants reported "Often". Regarding changing clinical routines, 42% (n=5) of the participants reported that they "Often" make adjustments, and 33% (n=4) of the participants did so "Sometimes." Perceptions of theoretical knowledge improvement due to app usage were generally positive among participants. Specifically, 33% (n=4) of health care providers indicated that their theoretical knowledge "always" increased because of using the GAIN MHI app, 58% (n=7) of health care providers reported that it "often" increased, and 8% (n=1) of health care providers said it "sometimes" did. The total engagement percentage calculated is 65% (SD 6.2%).

Table . Level of engagement (N=12).

Variables	Values
I accessed the links shared below the MCQs[a] explanations, n (%)	
Never	0 (0)
Rarely	3 (25)
Sometimes	7 (58)
Often	1 (8)
Always	1 (8)
I visited and accessed links from the knowledge resource center, n (%)	
Never	2 (17)
Rarely	4 (33)
Sometimes	6 (50)
Often	0 (0)
Always	0 (0)
I changed some of my clinical routines and practice because of the overall information captured from the app, n (%)	
Never	1 (8)
Rarely	1 (8)
Sometimes	4 (33)
Often	5 (42)
Always	1 (8)
My theoretical knowledge increased because of using the app, n (%)	
Never	0 (0)
Rarely	0 (0)
Sometimes	1 (8)
Often	7 (58)
Always	4 (33)
Total engagement (%)	
Mean (SD)	64.6 (6.2)
95% CI	60.6-68.5

[a]MCQ: multiple-choice question.

Table 4 evaluates the gamification and AI components of the GAIN MHI mobile app, revealing strong agreement and satisfaction. Gamification made learning engaging and enjoyable for 92% (n=11) of participants, motivating 92% (n=11) and boosting interest in targeted areas for 100% (n=12). The monetary reward type satisfied all the respondents. Remarkably,

XSL•FO
RenderX

the incorporation of gamification components substantially elevated participants' overall satisfaction with the learning experience (100%). Collectively, participants' positive responses highlight the success of gamification, with a total average score of 88.5% (SD 7.7%). Moving to the AI components, the findings revealed positive responses (Table 5). The AI adapted to participants' needs with 83% (n=10) either agreeing or strongly agreeing. Personalization of the learning experience yielded a combined agreement of 75% (n=9). The AI component effectively guided participants to identify their weak areas with an agreement rate of 83%. They expressed full agreement with the enhancement of the overall satisfaction of the learning experience. The cumulative AI score was 82.9% (SD 15.9%). Recommendations were robust, with 100% (n=12) of respondents agreeing on suggesting the app to other colleagues. The total evaluation reached 87.2% (SD 8.9%), solidifying the app's effectiveness, particularly through its gamification and AI components.

XSL•FO
RenderX

Table . Evaluation of the gamification, artificial intelligence, and mHealth network for maternal health improvement (GAIN MHI) gamification and artificial intelligence components.

Variables	Values
The gamification component of the GAIN MHI app made learning/professional development enjoyable, fun, and more engaging, n (%)	
Strongly disagree	0 (0)
Disagree	0 (0)
Neither disagree nor agree	1 (8)
Agree	7 (58)
Strongly agree	4 (33)
The gamification component of the GAIN MHI app motivated me to use the app, n (%)	
Strongly disagree	0 (0)
Disagree	0 (0)
Neither disagree nor agree	1 (8)
Agree	4 (33)
Strongly agree	7 (58)
The gamification component of the GAIN MHI app increased my interest in the subject/knowledge areas targeted by the app, n (%)	
Strongly disagree	0 (0)
Disagree	0 (0)
Neither disagree nor agree	0 (0)
Agree	7 (58)
Strongly agree	5 (42)
I was satisfied with the most valuable player (MVP) recognition concept, n (%)	
Strongly disagree	0 (0)
Disagree	0 (0)
Neither disagree nor agree	2 (17)
Agree	5 (42)
Strongly agree	5 (42)
I was satisfied with the most improved player (MIP) recognition concept, n (%)	
Strongly disagree	0 (0)
Disagree	0 (0)
Neither disagree nor agree	2 (17)
Agree	4 (33)
Strongly agree	6 (50)
I felt challenged to showcase my best performance because of the recognition system incorporated in the GAIN MHI mobile app (MVP and MIP), n (%)	
Strongly disagree	1 (8)
Disagree	0 (0)
Neither disagree nor agree	0 (0)
Agree	2 (17)
Strongly agree	9 (75)
I was satisfied with the type of reward offered (cash US $/monetary), n (%)	
Strongly disagree	0 (0)
Disagree	0 (0)
Neither disagree nor agree	0 (0)
Agree	5 (42)

XSL•FO
RenderX

Variables	Values
Strongly agree	7 (58)
The gamification component of the GAIN MHI app improved my overall satisfaction with the learning experience, n (%)	
Strongly disagree	0 (0)
Disagree	0 (0)
Neither disagree nor agree	0 (0)
Agree	5 (42)
Strongly agree	7 (58)
Gamification components (%)	
Mean (SD)	88.5 (7.7)
95% CI	83.7-93.4
I have noticed that the GAIN MHI app was focusing on the knowledge areas that I was answering incorrectly the most, n (%)	
Strongly disagree	1 (8)
Disagree	1 (8)
Neither disagree nor agree	0 (0)
Agree	5 (42)
Strongly agree	5 (42)
The AI[a] component of the GAIN MHI app personalized my learning experience, n (%)	
Strongly disagree	1 (8)
Disagree	1 (8)
Neither disagree nor agree	1 (8)
Agree	4 (33)
Strongly agree	5 (42)
The AI component of the GAIN MHI app helped me better learn about the areas that I needed to learn about, n (%)	
Strongly disagree	0 (0)
Disagree	0 (0)
Neither disagree nor agree	2 (17)
Agree	4 (33)
Strongly agree	6 (50)
The AI component of the GAIN MHI app improved my overall satisfaction with the learning experience, n (%)	
Strongly disagree	0 (0)
Disagree	0 (0)
Neither disagree nor agree	0 (0)
Agree	8 (67)
Strongly agree	4 (33)
Artificial intelligence components (%)	
Mean (SD)	82.9 (15.9)
95% CI	72.83-93
I would recommend the GAIN MHI app to other colleagues, n (%)	
Strongly disagree	0 (0)
Disagree	0 (0)
Neither disagree nor agree	0 (0)
Agree	4 (33)

XSL•FO
RenderX

Variables	Values
Strongly agree	8 (67)
Total evaluation (%)	
Mean (SD)	87.2 (8.9)
95% CI	81.6-92.8

[a]AI: artificial intelligence.

Table . Mean percentage (%) of health providers' perception on the role of integrating artificial intelligence in the GAIN MHI mobile app.

Perceived role of artificial intelligence integration	Mean percentage, %
Targeted knowledge areas (yes)	80
Personalized learning experience (yes)	78
Exploring knowledge gaps (yes)	87
Enhanced learning satisfaction (yes)	87

Qualitative Analysis

The qualitative analysis of users' favorite aspects of the mobile app reveals several key themes. One respondent mentioned: "I liked the diversity of questions, the focus on wrong answers in the upcoming questions, and the way how the app refreshes our memory." Another participant noted: "The application was easy to use. The app motivated us to learn more." Most users appreciate the app's ease of use and intuitive interface, as well as its ability to facilitate learning and refresh memory on a variety of topics. Furthermore, participants highly value the diversity of content, encompassing a wide range of questions and subject areas. Notably, the app's motivational elements have been recognized for their capacity to stimulate users in pursuing continuous learning.

Conversely, the analysis has identified potential areas for improvement and enhancement. A total of 4 out of 12 (33%) users reported technical issues that affect the app's speed and the display of scores. Additionally, suggestions for new features stated, such as the integration of a French language option and enhanced navigation through the Knowledge Resource Center section.

Exploring rewards and incentives, the analysis reveals a general satisfaction with the current monetary rewards. However, suggestions for diversification have emerged, including proposals for increased prizes or trips. A recurring theme is the desire for professional development opportunities, with participants recommending the inclusion of workshops or conferences within the reward system. This sentiment to "organize workshops to improve our work and enhance knowledge sharing" was also expressed.

Delving into content expansion, while some participants expressed contentment with the current topics, others identified gaps related to maternal and reproductive health. Suggestions encompass a range of subjects, spanning from breastfeeding, pregnancy nutrition, family planning, and gynecological diseases, to matters of mental health during pregnancy and emergency case management. Considering the app's potential extension to other health topics, various priorities have been highlighted. Respondents emphasized chronic conditions, such as diabetes and hypertension. Pediatrics and public health themes like hygiene, food safety, and vaccinations were also suggested. Mental health emerged as a vital area, including discussions on societal issues like gender-based violence and psychological disorders. Moreover, specialized topics like geriatric medicine, emergency medicine, cancer, learning disabilities, autism spectrum disorder, and immunology were proposed. This underscores the breadth of interest and possibilities for future gamified learning programs across diverse health domains.

Discussion

Overview

Analysis from this study underscores participants' overall encouraging experiences with the app, highlighting high levels of satisfaction in various components.

mLearning Experience With the GAIN MHI App

The findings from this pilot study underscore the keen interest of HCPs in learning and pursuing continuing education. Particularly noteworthy was their engagement with the knowledge resource center. Recent pilot studies have yielded findings that propose a correlation between certain eLearning strategies and the potential equivalence in effectiveness to conventional training methodologies [19,47]. This implies that digital learning approaches, when tailored with specific parallels to traditional training methods, could potentially generate comparable outcomes in terms of skill and knowledge acquisition. The incorporation of the GAIN MHI mLearning app proved instrumental in not only transforming clinical routines but also fostering the enrichment of theoretical knowledge among the HCPs. These findings further validate outcomes from prior research, indicating that using electronic tools is recognized as a potential avenue for enhancing the quality of care provided by HCPs, by increasing their self-efficacy [21,48]. Consequently, such intervention may suggest a possible solution to tackle the issue of insufficient availability of current evidence-based training and reference resources for HCPs working in remote regions.

XSL•FO
RenderX

Gamification Component of the GAIN MHI App

To our best knowledge, the GAIN MHI app is the first in the Middle East and North Africa region to combine both gamification and AI components in one mobile app to enhance the knowledge and practices of health care workers providing maternal health services. Preliminary evaluations from this study reveal a high level of agreement among HCPs that the gamification component of the GAIN MHI app has made their learning experience more joyful, engaging, and has enhanced their interest in targeted topics. In addition, the users revealed that the gamification component motivated them to use the app with high agreement on the recognition concepts as the most valuable (eg, MIP and monetary prizes). These occurrences might be clarified by the notion that gamification amplifies the spirit of competition through the incorporation of elements like badges, rewards, etc, actively motivating and involving individuals [49-51]. The idea of providing rewards adds to the pleasure of players as they complete levels, enhance skills, and experience progress [52]. This effectively motivates them to sustain their efforts [52]. In line with this strategy, the "Mylearning" platform in the SDA in India was structured to gamify learning and push users to always pursue education and development, seeking the SDA champion title and the provided certification upon successful completion [25]. In the future, gamification will transcend its role as a straightforward application of game-based design elements and evolve into a strategic management tool. It will involve reshaping the cognitive processes, fostering cooperation, and promoting collaboration while innovating novel approaches to learning and community advancement [52].

Artificial Intelligence Component of the GAIN MHI App

The majority of the HCPs enrolled in this study showed a high desire to recommend the GAIN MHI app to their colleagues and suggested its implication in other medical domains, which highlights the fact that continuing education is essential and applies to different health care practitioners [53]. The potential of AI to customize and personalize the learning experience stands out as one of the most significant potential developments in medical education [53]. A study by Mir et al [53] showed that the implementation of AI-based intervention among medical students empowered them to identify their areas of knowledge gaps and act accordingly. Comparable findings were reported in our study, with 75% of the GAIN MHI app users agreeing on tailoring their learning experience. AI has proved to enhance the HCPs' learning process and promote continuous advancement within the field [54,55], given that medical education is perceived as a lifetime learning journey [53]. These preliminary evaluations emphasize the importance of integrating AI into interventions targeting the HCPs. Such integration may serve as a valuable tool for their ongoing professional development through continuing education. The results of these early assessments highlight the potential benefits and effectiveness of using AI-driven approaches to enhance the learning experiences and knowledge acquisition of health care professionals by identifying their knowledge gaps. As a result, AI could play a crucial role in shaping how HCPs engage with educational content and adapt their learning strategies based on personalized insights provided by AI technologies. This not only supports their professional growth but also ensures that they stay up-to-date with the latest advancements in their field.

Lessons Learned and Potential for Scaling Up

The pilot test of the GAIN MHI app yielded insightful lessons with respect to user satisfaction and engagement. A pivotal observation pertained to the app's innate ability, driven by the synergy of AI and gamification components, to enhance user engagement. The amalgamation of these elements not only captured users' interest but also fostered a sustained and enthusiastic interaction with the app, primarily attributable to the intrinsic rewards offered through gamified experiences. This underscores the importance of cultivating a personalized and engaging environment to ensure continued user participation and commitment, in order to contribute to the professional development of the HCPs. Furthermore, the pilot test revealed that users were not only inclined to use the GAIN MHI app for their own professional development but also recommended it to their colleagues within the health care community, which underscores its potential as a valuable tool for broader adoption. Users appreciated the convenience of the mobile platform, the data-driven insights offered by the AI component, and the motivation instilled by the gamified structure. Moreover, their expressed interest in exploring additional topics beyond the current scope of the app indicates a desire for diverse and comprehensive educational content. This highlights the app's potential to serve as a versatile and adaptable platform catering to a wide spectrum of health care professionals' learning needs. This preliminary assessment also garnered important feedback regarding the improvement of the content of the app, both in terms of the asked MCQs and the associated Knowledge Resource Center. This feedback is guiding improvements in the design and content of the app, which would be vital for future upscaling. Future work will focus on refining the content and design of the GAIN MHI app based on user feedback, particularly in enhancing the quality of MCQs. This process is crucial for ensuring the continued engagement and satisfaction of users, ultimately contributing to the professional development of HCPs and facilitating broader adoption of the app within the health care community in Lebanon.

Limitations

While this pilot study offers valuable insights into the satisfaction of HCPs using the GAIN MHI app in improving maternal health outcomes among pregnant women residing in a fragile setting, it is important to acknowledge several limitations that were encountered during the course of this investigation. One primary constraint revolved around the limited sample size, consisting of merely 12 [12] HCPs located within the selected centers. This constraint may shed light on the possibility of a shortage of HCPs in these centers, potentially influencing the representation of provider perspectives and experiences in the study. Another limitation faced during the study was the consecutive malfunction and slowness of the app due to network issues. To address this concern, the implementation of an offline version of the app could potentially offer a viable solution for improving display speed and ensuring seamless accessibility to the app, especially in remote areas.

XSL•FO
RenderX

Last, this GAIN MHI app version was developed in 2 languages: Arabic and English. However, it did not take into consideration French-educated HCPs. This concern could be rectified by upgrading the GAIN MHI app to include other languages in the future.

Conclusions

HCPs, especially those working in remote and underserved areas, find it increasingly difficult to engage in professional development activities. Embarking on modern techniques and pedagogies is essential to empower them to learn at a time and place of their convenience. Personalizing the learning experience and enhancing participation through gamification are the future of education.

The consensus among HCPs emphasizes a high level of satisfaction with the GAIN MHI app and its integrated gamification and AI components. The positive reception reaffirms the potential of incorporating such innovative approaches to facilitate the advancement of their professional skills. The findings of this study build a strong case for upscaling similar apps into various contexts and specialties. In doing that, it is recommended that future studies delve deeper into assessing the app's feasibility and reception across a broader spectrum of HCPs and contexts, alongside its effect on maternal health outcomes.

Acknowledgments

The authors would like to thank the health care providers of the primary health care centers of the Lebanese Ministry of Public Health and the United Nations Relief and Works Agency for Palestine refugees in the Near East (UNRWA) for their participation in the GAIN MHI intervention, and the GAIN MHI app developer at the Global Health Institute at the American University of Beirut for designing and developing the mobile app. This study was funded by the International Development Research Centre (IDRC).

Conflicts of Interest

None declared.

References

1. Maternal mortality. WHO. 2023. URL: https://www.who.int/news-room/fact-sheets/detail/maternal-mortality [accessed 2023-08-01]
2. El‐Kak F, Kabakian‐Khasholian T, Ammar W, Nassar A. A review of maternal mortality trends in Lebanon, 2010–2018. Int J Gynaecol Obstet 2020 Jan;148(1):14-20. [doi: 10.1002/ijgo.12994]
3. Geller SE, Koch AR, Garland CE, MacDonald EJ, Storey F, Lawton B. A global view of severe maternal morbidity: moving beyond maternal mortality. Reprod Health 2018 Jun 22;15(Suppl 1):98. [doi: 10.1186/s12978-018-0527-2] [Medline: 29945657]
4. Reese Masterson A, Usta J, Gupta J, Ettinger AS. Assessment of reproductive health and violence against women among displaced Syrians in Lebanon. BMC Womens Health 2014 Feb 20;14(1):25. [doi: 10.1186/1472-6874-14-25] [Medline: 24552142]
5. Talhouk R, Mesmar S, Thieme A, et al. Syrian refugees and digital health in Lebanon: opportunities for improving antenatal health. 2016 Presented at: CHI '16: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems; May 7-12, 2016; San Jose, CA.
6. Carmichael SL, Mehta K, Srikantiah S, et al. Use of mobile technology by frontline health workers to promote reproductive, maternal, newborn and child health and nutrition: a cluster randomized controlled trial in Bihar, India. J Glob Health 2019 Dec;9(2):0204249. [doi: 10.7189/jogh.09.020424] [Medline: 31788233]
7. Jaeger FN, Bechir M, Harouna M, Moto DD, Utzinger J. Challenges and opportunities for healthcare workers in a rural district of Chad. BMC Health Serv Res 2018 Jan 8;18(1):7. [doi: 10.1186/s12913-017-2799-6] [Medline: 29310644]
8. Mormina M, Pinder S. A conceptual framework for training of trainers (ToT) interventions in global health. Global Health 2018 Oct 22;14(1):100. [doi: 10.1186/s12992-018-0420-3] [Medline: 30348183]
9. El-Jardali F, Alameddine M, Jamal D, et al. A national study on nurses' retention in healthcare facilities in underserved areas in Lebanon. Hum Resour Health 2013 Sep 30;11:1-13. [doi: 10.1186/1478-4491-11-49] [Medline: 24079458]
10. Jaana M, Majdalani M, Tamim H, Rahbany R. Perceived healthcare workforce needs in Lebanon: a step towards informed human resources planning and professional development. East Mediterr Health J 2018 Dec 9;24(9):855-865. [doi: 10.26719/2018.24.9.855] [Medline: 30570118]
11. Bou Sanayeh E, El Chamieh C. The fragile healthcare system in Lebanon: sounding the alarm about its possible collapse. Health Econ Rev 2023 Apr 4;13(1):21. [doi: 10.1186/s13561-023-00435-w] [Medline: 37014485]
12. Hachoumi N, Eddabbah M, El Adib AR. Health sciences lifelong learning and professional development in the era of artificial intelligence. Int J Med Inform 2023 Oct;178:105171. [doi: 10.1016/j.ijmedinf.2023.105171] [Medline: 37573636]

XSL•FO
RenderX

13. Maheu-Cadotte MA, Cossette S, Dubé V, et al. Effectiveness of serious games and impact of design elements on engagement and educational outcomes in healthcare professionals and students: a systematic review and meta-analysis protocol. BMJ Open 2018 Mar 16;8(3):e019871. [doi: 10.1136/bmjopen-2017-019871] [Medline: 29549206]
14. Gentry SV, Gauthier A, L'Estrade Ehrstrom B, et al. Serious gaming and gamification education in health professions: systematic review. J Med Internet Res 2019 Mar 28;21(3):e12994. [doi: 10.2196/12994] [Medline: 30920375]
15. Feng JY, Chang YT, Chang HY, Erdley WS, Lin CH, Chang YJ. Systematic review of effectiveness of situated e‐learning on medical and nursing education. Worldviews Evid Based Nurs 2013 Aug;10(3):174-183. [doi: 10.1111/wvn.12005]
16. Panda M, Desbiens NA. An "education for life" requirement to promote lifelong learning in an internal medicine residency program. J Grad Med Educ 2010 Dec;2(4):562-565. [doi: 10.4300/JGME-D-09-00068.1] [Medline: 22132278]
17. Rugh JD, Hendricson WD, Hatch JP, Glass BJ. The San Antonio CATs initiative. J Am Coll Dent 2010;77(2):16-21. [Medline: 20836411]
18. Weisman A, Lin E, Yona T, Gottlieb U, Impellizzeri FM, Masharawi Y. Healthcare providers have insufficient up-to-date knowledge of lower limb sports injuries, and their knowledge is similar to that of athletes. Musculoskelet Sci Pract 2023 Jun;65:102750. [doi: 10.1016/j.msksp.2023.102750] [Medline: 37003161]
19. Nilsson C, Sørensen BL, Sørensen JL. Comparing hands-on and video training for postpartum hemorrhage management. Acta Obstet Gynecol Scand 2014 May;93(5):517-520. [doi: 10.1111/aogs.12372] [Medline: 24754607]
20. Sondaal SFV, Browne JL, Amoakoh-Coleman M, et al. Assessing the effect of mHealth interventions in improving maternal and neonatal care in low- and middle-income countries: a systematic review. PLoS One 2016;11(5):e0154664. [doi: 10.1371/journal.pone.0154664] [Medline: 27144393]
21. Bolan NE, Sthreshley L, Ngoy B, et al. mLearning in the Democratic Republic of the Congo: a mixed-methods feasibility and pilot cluster randomized trial using the Safe Delivery App. Glob Health Sci Pract 2018 Dec 27;6(4):693-710. [doi: 10.9745/GHSP-D-18-00275] [Medline: 30591577]
22. Lund S, Boas IM, Bedesa T, Fekede W, Nielsen HS, Sørensen BL. Association between the Safe Delivery App and quality of care and perinatal survival in Ethiopia: a randomized clinical trial. JAMA Pediatr 2016 Aug 1;170(8):765-771. [doi: 10.1001/jamapediatrics.2016.0687] [Medline: 27322089]
23. Thomsen CF, Barrie AMF, Boas IM, et al. Health workers' experiences with the Safe Delivery App in West Wollega Zone, Ethiopia: a qualitative study. Reprod Health 2019 May 9;16(1):50. [doi: 10.1186/s12978-019-0725-6] [Medline: 31072399]
24. Maternity foundation and UNFPA launch new Arabic language version of a popular digital health tool to improve quality training of midwives in the Arab region. UNFPA. 2022. URL: https://arabstates.unfpa.org/en/news/maternity-foundation-and-unfpa-launch-new-arabic-language-version-popular-digital-health-tool [accessed 2023-08-01]
25. Singh Sodha T, Grønbæk A, Bhandari A, Mary B, Sudke A, Smith LT. mHealth learning tool for skilled birth attendants: scaling the Safe Delivery App in India. BMJ Open Qual 2022 Aug;11(Suppl 1):e001928. [doi: 10.1136/bmjoq-2022-001928] [Medline: 35977730]
26. Al-Rayes S, Al Yaqoub FA, Alfayez A, et al. Gaming elements, applications, and challenges of gamification in healthcare. Inform Med Unlock 2022;31:100974. [doi: 10.1016/j.imu.2022.100974]
27. Akl EA, Sackett KM, Erdley WS, et al. Educational games for health professionals. Cochrane Database Syst Rev 2013 Jan 31(1):CD006411. [doi: 10.1002/14651858.CD006411.pub3] [Medline: 23440807]
28. Damaševičius R, Maskeliūnas R, Blažauskas T. Serious games and gamification in healthcare: a meta-review. Information 2023;14(2):105.
29. Mesko B. The role of artificial intelligence in precision medicine. Expert Rev Precis Med Drug Dev 2017 Sep 3;2(5):239-241. [doi: 10.1080/23808993.2017.1380516]
30. Johnson KB, Wei WQ, Weeraratne D, et al. Precision medicine, AI, and the future of personalized health care. Clin Transl Sci 2021 Jan;14(1):86-93. [doi: 10.1111/cts.12884] [Medline: 32961010]
31. Brault N, Saxena M. For a critical appraisal of artificial intelligence in healthcare: the problem of bias in mHealth. J Eval Clin Pract 2021 Jun;27(3):513-519. [doi: 10.1111/jep.13528] [Medline: 33369050]
32. Zhang S, Bamakan SMH, Qu Q, Li S. Learning for personalized medicine: a comprehensive review from a deep learning perspective. IEEE Rev Biomed Eng 2019;12:194-208. [doi: 10.1109/RBME.2018.2864254] [Medline: 30106692]
33. Haenlein M, Kaplan A. A brief history of artificial intelligence: on the past, present, and future of artificial intelligence. Calif Manage Rev 2019 Aug;61(4):5-14. [doi: 10.1177/0008125619864925]
34. Mishra S, Tripathi AR. AI business model: an integrative business approach. J Innov Entrep 2021 Dec;10(1):18. [doi: 10.1186/s13731-021-00157-5]
35. Mustak M, Salminen J, Plé L, Wirtz J. Artificial intelligence in marketing: topic modeling, scientometric analysis, and research agenda. J Bus Res 2021 Jan;124:389-404. [doi: 10.1016/j.jbusres.2020.10.044]
36. Shaheen MY. Applications of artificial intelligence (AI) in healthcare: a review. ScienceOpen Preprints. Preprint posted online on Sep 25, 2021. [doi: 10.31219/osf.io/mjthd]
37. Amann J, Blasimme A, Vayena E, Frey D, Madai VI, Precise4Q consortium. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. BMC Med Inform Decis Mak 2020 Nov 30;20(1):310. [doi: 10.1186/s12911-020-01332-6] [Medline: 33256715]

XSL•FO
RenderX

38. Hosny A, Aerts HJWL. Artificial intelligence for global health. Science 2019 Nov 22;366(6468):955-956. [doi: 10.1126/science.aay5189] [Medline: 31753987]
39. Rittenhouse KJ, Vwalika B, Keil A, et al. Improving preterm newborn identification in low-resource settings with machine learning. PLoS One 2019;14(2):e0198919. [doi: 10.1371/journal.pone.0198919] [Medline: 30811399]
40. Araya J, Rodriguez A, Lagos-SanMartin K, et al. Maternal thyroid profile in first and second trimester of pregnancy is correlated with gestational diabetes mellitus through machine learning. Placenta 2021 Jan 1;103:82-85. [doi: 10.1016/j.placenta.2020.10.015] [Medline: 33099203]
41. Debata PP, Mohapatra P. Diagnosis of diabetes in pregnant woman using a Chaotic-Jaya hybridized extreme learning machine model. J Integr Bioinform 2020 Aug 13;18(1):81-99. [doi: 10.1515/jib-2019-0097] [Medline: 32790643]
42. Jiménez-Serrano S, Tortajada S, García-Gómez JM. A mobile health application to predict postpartum depression based on machine learning. Telemed J E Health 2015 Jul;21(7):567-574. [doi: 10.1089/tmj.2014.0113] [Medline: 25734829]
43. Natarajan S, Prabhakar A, Ramanan N, Bagilone A, Siek K, Connelly K. Boosting for postpartum depression prediction. Presented at: 2017 IEEE/ACM International Conference on Connected Health; Jul 17-19, 2017; Philadelphia, PA. [doi: 10.1109/CHASE.2017.82]
44. Chen M, Decary M. Artificial intelligence in healthcare: an essential guide for health leaders. Healthc Manage Forum 2020 Jan;33(1):10-18. [doi: 10.1177/0840470419873123] [Medline: 31550922]
45. Braun V, Clarke V. Toward good practice in thematic analysis: avoiding common problems and be(com)ing a knowing researcher. Int J Transgend Health 2023;24(1):1-6. [doi: 10.1080/26895269.2022.2129597] [Medline: 36713144]
46. Hennink M, Kaiser BN. Sample sizes for saturation in qualitative research: a systematic review of empirical tests. Soc Sci Med 2022 Jan;292:114523. [doi: 10.1016/j.socscimed.2021.114523]
47. Dunleavy G, Nikolaou CK, Nifakos S, Atun R, Law GCY, Tudor Car L. Mobile digital education for health professions: systematic review and meta-analysis by the digital health education collaboration. J Med Internet Res 2019 Feb 12;21(2):e12937. [doi: 10.2196/12937] [Medline: 30747711]
48. Walker DM, Holme F, Zelek ST, et al. A process evaluation of PRONTO simulation training for obstetric and neonatal emergency response teams in Guatemala. BMC Med Educ 2015 Jul 24;15(1):117. [doi: 10.1186/s12909-015-0401-7] [Medline: 26206373]
49. King D, Greaves F, Exeter C, Darzi A. "Gamification": influencing health behaviours with games. J R Soc Med 2013 Mar;106(3):76-78. [doi: 10.1177/0141076813480996] [Medline: 23481424]
50. Wiley K, Vedress S, Mandryk RL. How points and theme affect performance and experience in a gamified cognitive task. 2020 Apr 21 Presented at: CHI '20: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems; Apr 25-30, 2020; Honolulu, HI p. 1-15. [doi: 10.1145/3313831.3376697]
51. Fiş Erümit S, Karakuş Yılmaz T. Gamification design in education: what might give a sense of play and learning? Technol Know Learn 2022 Dec;27(4):1039-1061. [doi: 10.1007/s10758-022-09604-y]
52. Ansari S, Khot R, Shaikh Z, Yadav P, Pirani Z. A comprehensive review of gamification in healthcare: incentives in mobile healthcare app. Int Res J Eng Technol 2022.
53. Mir MM, Mir GM, Raina NT, et al. Application of artificial intelligence in medical education: current scenario and future perspectives. J Adv Med Educ Prof 2023 Jul;11(3):133-140. [doi: 10.30476/JAMP.2023.98655.1803] [Medline: 37469385]
54. Dave M, Patel N. Artificial intelligence in healthcare and education. Br Dent J 2023 May;234(10):761-764. [doi: 10.1038/s41415-023-5845-2] [Medline: 37237212]
55. Jacobs SM, Lundy NN, Issenberg SB, Chandran L. Reimagining core entrustable professional activities for undergraduate medical education in the era of artificial intelligence. JMIR Med Educ 2023 Dec 19;9:e50903. [doi: 10.2196/50903] [Medline: 38052721]

Abbreviations:

AI: artificial intelligence
GAIN MHI: gamification, artificial intelligence, and mHealth network for maternal health improvement
HCP: health care provider
MCQ: multiple-choice question
mHealth: mobile health
MIP: most improved player
OBGYN: obstetrics and gynecology
PHC: primary health care center
SDA: Safe Delivery App

XSL•FO
RenderX

Edited by T Leung; submitted 19.10.23; peer-reviewed by AZ Yang, C Ochoa-Zezzatti; revised version received 11.12.24; accepted 11.02.25; published 10.09.25.

Please cite as:
Alameddine M, Sabra N, El Arnaout N, El Dakdouki A, El Jaouni M, Hamadeh R, Shanaa A, Saleh S
Utilizing Gamification, Artificial Intelligence, and mHealth for the Professional Development of Maternal Care Providers: Exploratory Pilot Cross-Sectional Study Assessing Providers' Satisfaction in Primary Health Care Centers in Lebanon
JMIR Serious Games 2025;13:e53735
URL: https://games.jmir.org/2025/1/e53735
doi:10.2196/53735

© Mohamad Alameddine, Nadine Sabra, Nour El Arnaout, Asmaa El Dakdouki, Mahmoud El Jaouni, Randa Hamadeh, Abed Shanaa, Shadi Saleh. Originally published in JMIR Serious Games (https://games.jmir.org), 10.9.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Gamified Adaptive Approach Bias Modification in Individuals With Methamphetamine Use History From Communities in Sichuan: Pilot Randomized Controlled Trial

Danlin Shen[1], MD; Jianping Jiao[2], BSc; Liqun Zhang[3], BS; Yanru Liu[3], MS; Xiang Liu[3], PhD; Yuanhui Li[3], MS; Tianjiao Zhang[3], MS; Dai Li[3], BS; Wei Hao[4], MD

[1]Mental Health Center, West China Hospital, Sichuan University, Chengdu, China

[2]Kunshan Mental Health Center, Kunshan, China

[3]Adai Technology (Beijing) Co Ltd, 2808 Guojichuangzhanzhongxin, Beijing, China

[4]National Clinical Research Center on Mental Disorders and Mental Health Institute of the Second Xiangya Hospital, Central South University, 139 Renmin Middle Rd, Changsha, China

Corresponding Author:
Wei Hao, MD
National Clinical Research Center on Mental Disorders and Mental Health Institute of the Second Xiangya Hospital, Central South University, 139 Renmin Middle Rd, Changsha, China

Abstract

Background: Cognitive bias modification (CBM) programs have shown promise in treating psychiatric conditions, but they can be perceived as boring and repetitive. Incorporating gamified designs and adaptive algorithms in CBM training may address this issue and enhance engagement and effectiveness.

Objectives: This study aims to gather preliminary data and assess the preliminary efficacy of an adaptive approach bias modification (A-ApBM) paradigm in reducing cue-induced craving in individuals with methamphetamine use history.

Methods: A randomized controlled trial with 3 arms was conducted. Individuals aged 18 - 60 years with methamphetamine dependence and at least 1 year of methamphetamine use were recruited from 12 community-based rehabilitation centers in Sichuan, China. Individuals with the inability to fluently operate a smartphone and the presence of mental health conditions other than methamphetamine use disorder were excluded. The A-ApBM group engaged in ApBM training using a smartphone app for 4 weeks. The A-ApBM used an adaptive algorithm to dynamically adjust the difficulty level based on individual performance. Cue-induced craving scores and relapses were assessed using a visual analogue scale at baseline, postintervention, and at week-16 follow-up.

Results: A total of 136 participants were recruited and randomized: 48 were randomized to the A-ApBM group, 48 were randomized to the static approach bias modification (S-ApBM) group, and 40 were randomized to the no-intervention control group. The A-ApBM group showed a significant reduction in cue-induced craving scores at postintervention compared with baseline (Cohen d=0.34; P<.01; 95% CI 0.03-0.54). The reduction remained significant at the week-16 follow-up (Cohen d=0.40; P=.01; 95% CI 0.18-0.57). No significant changes were observed in the S-ApBM and control groups.

Conclusions: The A-ApBM paradigm with gamified designs and dynamic difficulty adjustments may be an effective intervention for reducing cue-induced craving in individuals with methamphetamine use history. This approach improves engagement and personalization, potentially enhancing the effectiveness of CBM programs. Further research is needed to validate these findings and explore the application of A-ApBM in other psychiatric conditions.

Trial Registration: ClinicalTrials.gov NCT05794438; https://clinicaltrials.gov/study/NCT05794438

(JMIR Serious Games 2025;13:e56978) doi:10.2196/56978

KEYWORDS

digital therapeutics; substance use disorder; game; pilot RCT; randomized controlled trial; methamphetamine; psychiatric; gamified design; engagement; effectiveness; smartphone app; cognitive bias modification

Introduction

Cognitive biases, such as attention bias, approach bias, and interpretation bias, are intricately linked to the psychopathologies of conditions, including anxiety, depression, posttraumatic disorders, and addiction. Typically, individuals with psychiatric disorders exhibit a distinctive cognitive bias toward certain types of stimuli. Researchers have developed

XSL•FO
RenderX

cognitive bias modification (CBM) training schemes aimed at redirecting the bias from negative to positive stimuli. The mechanisms underlying CBM's efficacy are grounded in theories of automaticity and habit formation, where repeated exposure to corrective training diminishes the influence of negative stimuli over time. For example, pertaining to addiction, approach bias modification (ApBM) trainings aim to correct the cognitive bias of approaching addictive substances, and for treating anxiety and depression, attention bias modification and interpretation bias modification aim to correct the cognitive bias of paying excessive attention to negative emotions and interpretations. CBM interventions have shown potential in treating addiction, anxiety, and depression [1,2].

CBM training, although effective, has been criticized for being monotonous and unengaging [3]. Several studies have reported the issue of boredom with CBM training [4-6]. Qualitative interviews with patients [5] have revealed that participants often found CBM to be "boring," "repetitive," and "tedious," and that they "just tried to get through it as quickly as possible." This may be attributed to the nature of CBM's design, where each session typically consists of repetitive trials, and the rules and instructions remain the same throughout the intervention program. Consequently, the tedium and boredom associated with CBM training may result in a loss of attention and interest in the program, thereby hindering its efficacy. However, given the evidence of the effectiveness of CBM in treating various psychopathologies, it is crucial to address the issue of boredom and make the training more engaging and fun for participants to further enhance efficacy.

In treating substance user disorders, ApBM is a type of CBM training method aimed at counteracting approach biases and ultimately reducing the craving and usage of addictive substances. Through repeated exercises, participants are exposed to substance-related images and trained to make avoidance movements (eg, pushing an image away) while associating non–substance-related images with approach movements (eg, pulling the image closer). This retraining helps shift automatic tendencies, encouraging avoidance of substance-related cues. Research, including several randomized controlled trials [7-11], has shown that incorporating ApBM into treatment programs for substance use disorders can reduce relapse rates, even with brief training sessions of 10‐15 minutes conducted over multiple sessions [11].

To make CBM training more engaging and effective, efforts have been made to incorporate gamification into CBM, with some studies demonstrating improved engagement and outcomes through the use of animations, sounds, feedback, and point-scoring systems [12]. The integration of gamification in digital therapeutics, particularly in CBM, is becoming more prevalent, suggesting a promising avenue for enhancing the efficacy of these interventions [12-15]. This study introduces an innovative approach: an adaptive approach bias modification (A-ApBM) scheme designed to counter the monotony of traditional CBM. This scheme not only includes gamified elements but also features an adaptive algorithm that adjusts the training difficulty based on individual performance to enhance engagement [16,17]. The key idea of A-ApBM is to use an algorithm to dynamically estimate individual performance and adjust the training difficulty accordingly. In contrast to traditional static approach bias modification (S-ApBM), where the difficulty level is fixed throughout the intervention, A-ApBM continuously adapts to maintain an optimal level of challenge. This ensures personalized engagement, keeping participants motivated, and maximizing the intervention's effectiveness. Difficulty adjustment is a crucial concept in game design because it provides participants with a sense of challenge that matches their skill level. By striking a balance between ease and difficulty, the intervention avoids boredom from being too predictable and frustration from being too demanding. This sense of challenge promotes a state of "flow," where participants are fully immersed and engaged in the task. By making the training feel more like a dynamic and rewarding game, A-ApBM not only enhances the overall experience but also increases adherence and improves therapeutic outcomes. This ensures personalized engagement, keeping participants motivated and maximizing the intervention's effectiveness. This is expected to maintain engagement and challenge throughout the program. While previous studies have attempted to modify CBM difficulty for better outcome, our approach is distinguished by its analytical and quantitative emphasis on algorithm-based adjustment [18,19].

The study is designed to be a randomized controlled trial with 3 arms: the A-ApBM as the intervention arm, an S-ApBM as the active control arm, and a no-intervention control arm. The inclusion of the S-ApBM group was crucial for evaluating the effectiveness of the adaptive intervention. The S-ApBM group served as an active control to isolate the impact of the adaptivity—the dynamic difficulty adjustments in the A-ApBM group. Both groups participated in bias modification training, but the S-ApBM group's static nature allowed us to assess whether adaptivity specifically contributed to greater reductions in cue-induced cravings. This control also accounted for general engagement effects, ensuring that any observed differences were due to the adaptive algorithm rather than task participation alone. By including both versions, we tested whether personalized, adaptive training improved outcomes compared with static training, thus helping to clarify the role of adaptivity in enhancing cognitive interventions. Given the significant public health burden posed by methamphetamine addiction [20-23], this study's findings may inform the development of more effective therapeutic tools that engage users and sustain their motivation throughout treatment.

Methods

Study Design

This study was a 3-armed randomized controlled trial, including 1 intervention group (A-ApBM) and 2 control groups (an S-ApBM group as active control and a no-intervention control group). Participants were randomly assigned (using computer-generated random numbers) to the 3 groups using a computer random number generator. This method ensured that each participant had an equal chance of being placed in any of the groups. While the trial incorporated blinding, due to the presence of a no-intervention control group, it is more accurately described as partially double-blinded. Specifically, the

XSL•FO
RenderX

participants in the intervention groups were unaware of whether they were receiving adaptive or static training, which helped reduce bias from the participant side. However, those in the no-intervention control group likely knew that they were not receiving an active treatment, which limits full blinding. The participants in the A-ApBM and S-ApBM groups were asked to engage in daily trainings for 4 weeks.

Ethical Considerations

This study was conducted in accordance with the Declaration of Helsinki and approved by the ethics committee of the Second Xiangya Hospital of Central South University (2021 - 076). Informed consent was obtained from all participants prior to their inclusion in the study. Participants were provided with detailed information about the study's purpose, procedures, potential risks, and benefits, and they were informed of their right to withdraw at any time without consequence. All data collected during the study were anonymized before analysis. Participants' identifiable information was securely stored and accessible only to authorized personnel. Deidentified datasets were used for all analyses to ensure the protection of participants' privacy.

Sample Size Calculation

The sample size for this pilot trial was determined based on practical considerations rather than formal statistical power calculations, as there were insufficient data to estimate the effect size and variance of the proposed A-ApBM intervention. We aimed to recruit 150 participants (50 per group), based on preliminary communications in December 2022, with 12 community rehabilitation centers. These centers' rosters indicated that approximately 150 individuals met the inclusion criteria for the study. This sample size allowed for preliminary analyses and to assess the feasibility of the study procedures and intervention, while acknowledging the logistical constraints associated with recruitment, time, and resources.

Participants

Participants were recruited from 12 community-based rehabilitation centers in Sichuan, China, all having a history of methamphetamine use and undergoing community-based rehabilitation. Participants were individuals aged between 18 and 55 years with a history of methamphetamine use for at least 1 year. The diagnosis of methamphetamine use disorder was confirmed based on community rehabilitation center records and self-reports. Furthermore, we excluded individuals with any other mental health disorders, as well as those who could not fluently operate a smartphone, ensuring that all participants could engage with the smartphone-based intervention. These revisions provide a clearer understanding of the participant demographics and eligibility criteria. Participants' characteristics were collected using survey, which is shown in Multimedia Appendix 1.

Intervention: A-ApBM

The A-ApBM group underwent a specialized intervention using the WonderLab Harbour smartphone app (Adai Technology [Beijing] Co Ltd; Multimedia Appendix 2) for 4 weeks. This intervention is rooted in the principles of ApBM training, as delineated in previous studies on substance use disorders [10,11]. In each session, participants were instructed to swipe upward (or downward) when shown images in portrait (or landscape) orientation. After swiping upward (downward), an animation shrank (grew) to create the effect of distancing (approaching) the object. The images depicted either methamphetamine-related items (such as crystals, powders, or paraphernalia) or representations of healthy living (eg, wealth, sports, gourmet food, family activities, etc).

The A-ApBM group's training focused on congruency, as per the model by Kruijt and Carlbring [24], where congruent trials aligned with the training's intention, such as avoiding drug cues. Our study adopts a modified approach for calculating the intended training ratio (ITR), using the ratio of congruent trials to the total number of trials. This adjustment was made to enhance mathematical robustness and to directly correlate the ITR with the training's intensity and specificity. In A-ApBM, the ITR dynamically varied based on an algorithm (described in more detail in subsection "Design of the A-ApBM Algorithm") that monitored the user's performance, adjusting the difficulty index and ITR accordingly. This adaptive method is intended to individualize the training's intensity to the user's performance level, offering a personalized and dynamic approach to CBM.

Controls: S-ApBM Active Control and No-Intervention Control

The study included 2 control groups: the S-ApBM group and a no-intervention control group. Participants in the S-ApBM group also used the same app for a similar duration but experienced a static form of ApBM training. Each session in the S-ApBM group consisted of a constant number of drug and nondrug trials, with a high proportion of congruent trials, resulting in a stable ITR of 92.3% (144/156 trials) throughout the program. This static training model provided a control baseline to compare against the dynamic, A-ApBM intervention. The no-intervention control group did not receive any interventions, serving as an additional baseline to measure the effects of engaging in any structured intervention versus no intervention.

Design of the A-ApBM Algorithm

Algorithmic Components and Metrics

Intended Training Ratio

The ITR is crucial in the A-ApBM, representing the proportion of trials within a session that align with core training objectives, specifically avoiding drug cues and approaching non–drug-related cues. Calculated as the ratio of congruent trials to total trials, this measure directly correlates with training intensity and contingency.

Performance Index

This index reflects the participant's ability to meet the training's objectives. Defined as pt=Correct response rateMedianlog(Response times), it combines the correct response rate with the median of the logarithmically transformed response times, offering a comprehensive measure of a user's performance in an ApBM training session t (the index *t* indicates the sequential session number within the intervention. The first

XSL•FO
RenderX

training session would have t=1, and the second training session would have t=2, and so on).

Difficulty Index

We hypothesized that the perceived difficulty of an ApBM session by users is influenced mainly by 2 factors: the ITR and the variation in the ITR across sessions. The relationship between the ITR and the difficulty is proposed to be an inverted U-shape. This shape is chosen because it effectively captures the nuanced way in which the ITR influences the cognitive processing required during the task. At an ITR of 100% or 0%, the task becomes predictable, with all trials being congruent, allowing users to form automatic associations between the stimulus type and the format. However, at an ITR of 50%, where half of the trials are congruent and the other half are not, the task reaches peak difficulty. The 50% ITR creates maximum randomness in stimulus-type and format associations, making it challenging for users to discern a clear pattern, thus requiring more cognitive effort to perform correctly. The second crucial factor is the variability of the ITR in consecutive sessions. This aspect of difficulty is grounded in the concept of cognitive adaptability. If the ITR remains constant (or shows minimal variation) across sessions, participants may quickly adapt to the pattern, reducing cognitive load and perceived difficulty. In contrast, a high variation in the ITR across sessions introduces unpredictability, challenging the participant to continuously adapt to new patterns. This unpredictability requires greater cognitive flexibility and is hypothesized to increase perceived difficulty. To incorporate this factor, our model includes the SD of the ITRs from recent sessions: SD(ITRt−k+1,…,ITRt). A higher SD indicates greater variability in ITR from one session to the next, thus contributing to an increased difficulty index. This dynamic adjustment ensures that the training remains challenging yet achievable, adapting to the individual's learning curve and maintaining engagement throughout the intervention.

To model the difficulty level, we use a quadratic term (ITR^2): $dt=\alpha0+\alpha1ITRt2+\alpha2SD(ITRt-k+1,\ldots,ITRt)$, where α values are to be estimated dynamically.

Relationship Between Performance and Difficulty

The adaptive algorithm within A-ApBM is designed to learn a dynamic relationship between the participant's performance index and the difficulty index. We use a linear model to model the relationship between performance pt and the difficulty index dt in session t. Specifically, for each user i, we model the relationship using the equation $pt=\beta i0+\beta1dt+\beta2t+\beta3t2+$ t. The terms $\beta2t$ and $\beta3t2$ account for the nonlinear (quadratic) learning curve as the user furthers into the intervention program and becomes more familiar with the ApBM training. The term t represents the idiosyncratic error or residual error, which captures the random shocks that can affect performance.

Sketch of the Algorithm

The A-ApBM algorithm initiates with 3 warm-up sessions, during which no ITR adjustments are made. Performance and difficulty indices for each participant are dynamically calculated at the end of each session, starting from the fourth session. From this point, a linear regression model is fitted to each user's cumulative data, estimating the relevant parameters (αs and βs). Based on these estimations, the algorithm then adjusts the difficulty level of subsequent sessions to align with a predetermined difficulty curve. Building upon the analysis presented by Li et al [17], 4 difficulty curve shapes are considered: U-shape, inverted U-shape, N-shape, and inverted N-shape. For our study, we have identified an inverted U-shaped difficulty curve as the most optimal for long-duration games with a general reward structure. This curve provides a balanced challenge, gradually increasing in difficulty to a peak before tapering off, effectively maintaining user engagement and motivation. A proprietary algorithm is used to ensure that the user's difficulty level closely follows this inverted U-shaped trajectory, adapting the training intensity in real time to the user's performance and learning curve. An implementation of the A-ApBM algorithm is available upon request.

Outcome Measures

Primary Outcome: Self-Reported Cue-Induced Craving Score

Self-reported cue-induced craving scores were used as the primary outcome measure. Participants were asked to review images that depict methamphetamine crystals, powders, and paraphernalia, and rate their cravings using a 0 - 100 visual analogue scale. To improve the accuracy and reliability of participants' self-reported cravings, they were instructed to rate the images twice upon enrollment: first on a smartphone and then again, 1 day later, on a printed paper. The average score of these 2 assessments was used as the baseline measure. The follow-up assessment took place in week 16 (12 weeks after intervention completion).

Secondary Outcome: Relapse

Relapse was defined as the resumption of methamphetamine use following a period of abstinence. This was monitored through self-report and urine drug-screening tests at week 4 and week 16 of the study. Participants were asked to provide urine samples at regular intervals during the follow-up period, which were tested for methamphetamine metabolites. Furthermore, they were required to report any drug use incidents, including the date, amount, and context of use.

Statistical Analysis

Data were analyzed using a mixed-design ANOVA, with time as the within-subjects factor and group as the between-subjects factor. The dependent variable was the mean craving score. The ANOVA adhered to an intention-to-treat approach, with missing outcomes imputed using the Last Observation Carried Forward method. To address the skewed distribution of craving scores and ensure robustness, a nonparametric bootstrap with 10,000 resamples was applied to calculate F statistics and effect sizes for pairwise comparisons. CIs were derived using the bias-corrected and accelerated method.

For relapse, we also used an intention-to-treat approach by assuming that the missing observations meant relapse. A 3-sample test for equality of proportions (using the prop.test function in R) was conducted by assuming the missings to have relapsed. The P value was .23, indicating no significant differences.

XSL•FO
RenderX

To assess the proposed relationship between performance and difficulty indices, their association was retrospectively analyzed using data from the A-ApBM group. The performance index for the ApBM sessions was calculated. For the computation of the SD of the past 3 ITRs, a rolling window of 3 was used, excluding the initial 3 sessions due to their constant ITRs. A linear mixed model was fitted to predict the performance index, incorporating ITR, ITR^2, t, and t^2, and Past3ITRSD as predictors. The model accounted for subject variability through a random-effects parameter. All statistical analyses were performed using R 4.3.2. Significance levels for all tests were set at an α value of .05.

Results

Participants

In March 2023, 136 participants were screened and assessed for eligibility, and all participants were eligible and randomized into 3 groups using a computer random number generator: the A-ApBM group (n=48), the S-ApBM group (n=48), and the no-intervention control group (n=40). The trial aimed to recruit 150 participants; however, after successfully recruiting 136 participants, the pool of eligible individuals from the centers' rosters was exhausted. No new participants were available within the short-term time frame, making further recruitment infeasible. Given that we had nearly reached our target and that 50 participants per group were considered sufficient for feasibility and preliminary analysis, the recruitment phase was concluded at 136 participants. Throughout the course of the study, dropout rates were monitored. All 136 participants completed the intervention (or control) at week 4. At week 16, 3 A-ApBM participants dropped out and were not evaluable due to missing follow-up outcomes. The no-intervention control group had 2 dropouts whose outcome were not evaluable at the 16-week follow-up. The S-ApBM group maintained full participation throughout the intervention phase. Overall, the dropout rate was 4% (5/136). Table 1 shows the baseline characteristics of the participants by group. We conducted statistical tests among the groups. Specifically, Pearson χ^2 tests were applied for categorical variables (eg, sex, dominant hand, marital status, education, smoker, and drinker), and the Kruskal-Wallis rank sum test was chosen for continuous variables (eg, age, meth use history, and craving at baseline) due to the nonnormality of the data. As indicated in the table, all P values are above .05, demonstrating that there are no statistically significant differences in participant characteristics across the 3 groups.

Figure 1 displays the CONSORT (Consolidated Standards of Reporting Trails) flow diagram illustrating the participants' progression throughout the study.

XSL•FO
RenderX

Table . Participant baseline characteristics.

Characteristics	No-intervention control (n=40)	S-ApBM[a] active control (n=48)	A-ApBM[b] intervention (n=48)	*P* value
Sex, n (%)				.50[c]
Female	10 (25)	12 (25)	8 (17)	
Male	30 (75)	36 (75)	40 (83)	
Age (years), mean (SD)	31.71 (8.34)	33.78 (5.32)	34.51 (6.20)	.14[d]
Dominant hand, n (%)				.70[c]
Left	4 (10)	7 (15)	8 (17)	
Right	36 (90)	41 (85)	40 (83)	
Marital status, n (%)				.70[c]
Divorced	10 (25)	14 (29)	14 (29)	
Married	15 (38)	19 (40)	13 (27)	
Unmarried	15 (38)	15 (31)	21 (44)	
Education, n (%)				.20[c]
College or above	3 (7.5)	0 (0)	4 (8.3)	
Elementary school	3 (7.5)	3 (6.3)	3 (6.3)	
Junior high school	22 (55)	26 (54)	32 (67)	
Senior high school	12 (30)	19 (40)	9 (19)	
Smoker, n (%)	32 (80)	43 (90)	43 (90)	.30[c]
Drinker, n (%)	24 (60)	27 (56)	37 (77)	.08[c]
Methamphetamine use history (years), mean (SD)	2.80 (2.40)	2.55 (2.25)	3.17 (2.87)	.60[d]
Heroin history, n (%)	1 (2.5)	1 (2.1)	2 (4.2)	.80[c]
Ketamine history, n (%)	0 (0)	1 (2.1)	0 (0)	.40[c]
Craving (baseline), mean (SD)	7.66 (11.79)	6.33 (9.28)	8.65 (13.44)	.80[d]

[a]S-ApBM: static approach bias modification.

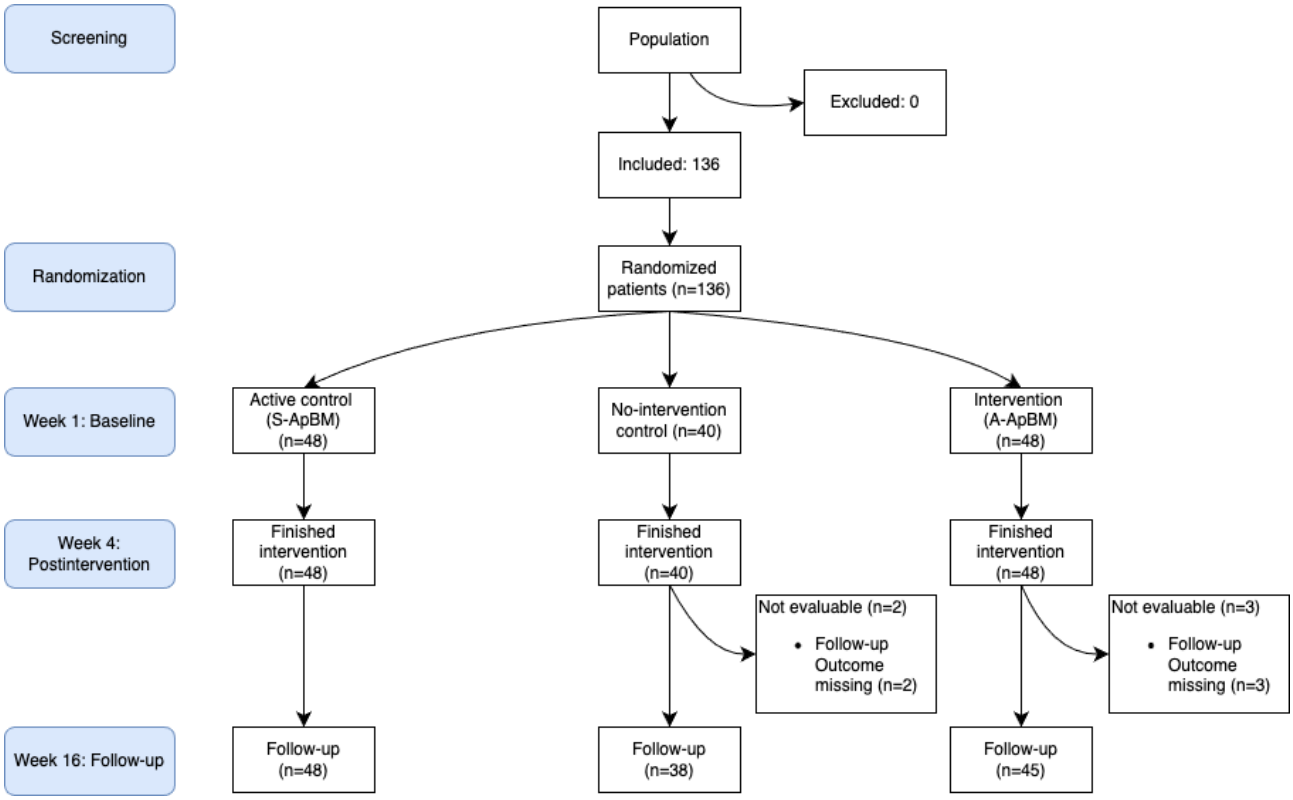
[b]A-ApBM: adaptive approach bias modification.

[c]Pearson χ^2 test.

[d]Kruskal-Wallis rank sum test.

XSL•FO
RenderX

Figure 1. CONSORT (Consolidated Standards of Reporting Trails) diagram. A-ApBM: adaptive approach bias modification; S-ApBM: static approach bias modification.



Training Time

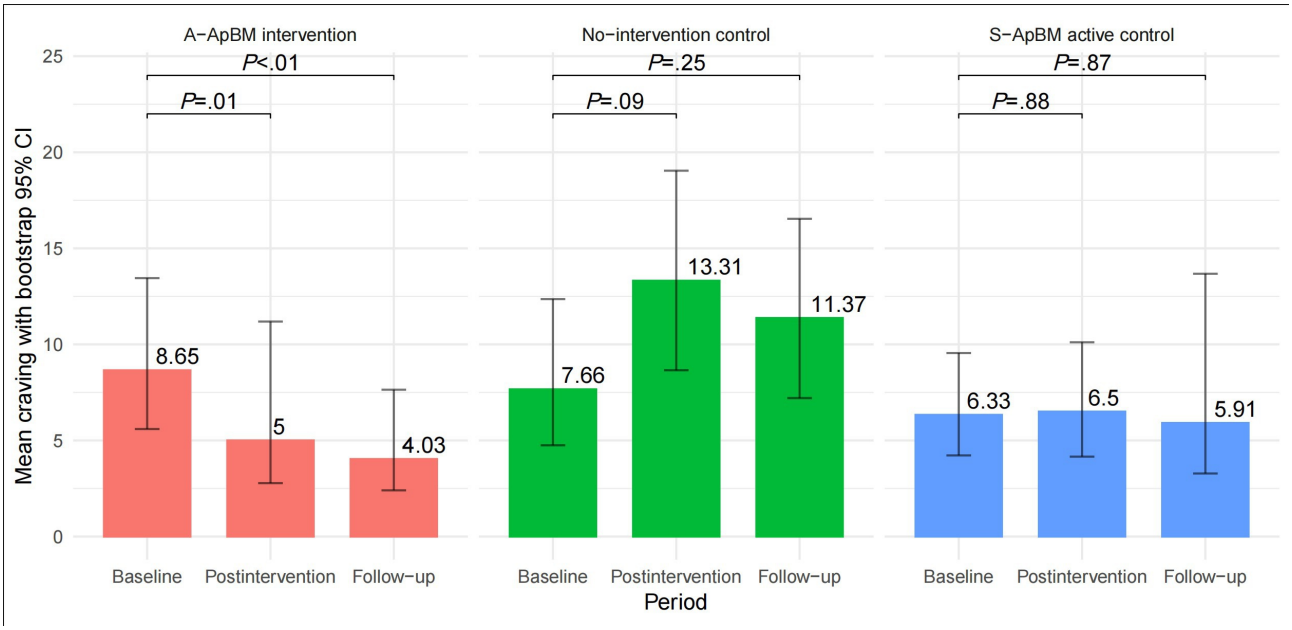
We calculated the daily training time for each user, which is calculated by dividing the total duration of time each user spent in the training by the number of days that each user was engaged in the training. The A-ApBM group had an average daily training time of 5.32 (SD 1.43) minutes, and the S-ApBM group had an average daily training time of 5.09 (SD 1.53) minutes.

Change in Cue-Induced Craving

There was a statistically significant interaction between group and time in explaining the craving score ($F_{4,266}$=3.02, 95% CI 1.18-11.20; P=.02; ηð2=0.043, 95% CI 0.016-0.141). Bonferroni-corrected pairwise comparison showed that the mean craving score was significantly different at postintervention between no-intervention control and A-ApBM (P=.01) and between no-intervention control and S-ApBM (P=.05). At follow-up the mean craving score was significantly different between no-intervention control and A-ApBM (P=.03). Bootstrapped pairwise 2-tailed t tests revealed that only the A-ApBM group showed differences between baseline and postintervention (Cohen d=0.34; P<.01; 95% CI 0.03-0.54) and between baseline and follow-up (Cohen d=0.40; P=.01; 95% CI 0.18-0.57) (Figure 2).

XSL•FO
RenderX

Figure 2. Change in cue-induced craving. A-ApBM: adaptive approach bias modification; S-ApBM: static approach bias modification.



Relapse

Throughout the study duration, including at the 4-week and 16-week follow-up points, there were no instances of self-reported relapse among participants across all study groups. This observation was further corroborated by objective measures: urine drug tests conducted at weeks 4 and 16 yielded uniformly negative results for all participants, irrespective of their group assignment.

Relationship Between Performance and Difficulty Indices

Table 2 tabulates the estimates of the fixed effects of the linear mixed-effects model. The explanatory power is substantial (conditional R^2=0.82). The 95% CIs and *P* values were computed using a Wald *t*-distribution approximation. The effect of ITR was significant and positive (β=2.18; *P*=.03), indicating that as the ITR increased, participants' performance improved. The quadratic term for ITR was negative and highly significant (β=–2.47; *P*<.001), suggesting an inverted U-shaped relationship between ITR and performance. This means that performance improved with increasing difficulty up to a certain point, after which further increases in difficulty resulted in diminished performance, supporting the theory that moderate difficulty is most effective for learning. Time (*t*) also had a significant and positive effect (β=2.95; *P*<.001), indicating that performance improved over time as participants became more familiar with the task. The nonsignificant quadratic term for time (*P*=.70) suggests that while participants' performance improved over time, the learning curve did not exhibit a significant acceleration or deceleration during the intervention. Finally, the SD of the ITR from recent sessions (Past3ITRSD) was negatively associated with performance, although this effect was not statistically significant (*P*=.10). This indicates that greater variability in difficulty across sessions tended to reduce performance, although this finding warrants further investigation.

Table . Mixed-effect linear regression results.

Characteristic	β	95% CI	*P* value
ITR[a]			
ITR	2.18	0.25 to 4.1	.03
ITR²	−2.47	−3.6 to −1.3	<.001
Time (*t*)			
t	2.95	2.2 to 3.8	<.001
*t*²	0.13	−0.58 to 0.84	.70
Past3ITRSD[b]	−6.30	−14 to 1.3	.10

[a]ITR: intended training ratio.

[b]Past3ITRSD: SD of the intended training ratio from recent sessions.

XSL•FO
RenderX

Discussion

Principal Findings

The findings of this pilot randomized controlled trial are 2-fold. First, there was a significant reduction in craving scores in the A-ApBM group and a lack of significant change in the control groups. Second, a significant relationship between performance and difficulty indices within the A-ApBM intervention indicates the merit of the proposed adaptive algorithm. These findings underscore the potential effectiveness of an A-ApBM intervention in reducing cue-induced cravings among individuals with a history of methamphetamine use.

One of the primary outcomes observed was a significant reduction in craving scores in the A-ApBM group, both immediately after the intervention and at the 16-week follow-up. This suggests that the A-ApBM intervention may have a lasting impact, which could be particularly beneficial for interventions aimed at behavior change or addiction treatment. The observed reduction in cravings is likely attributable to the adaptive and personalized nature of the A-ApBM intervention, which dynamically adjusts the difficulty level based on the user's performance. This approach appears to maintain engagement and challenge, potentially enhancing the overall effectiveness of CBM programs.

In contrast, the control groups, including the S-ApBM group and the no-intervention control group, did not exhibit significant changes in craving scores over time. The use of an active control such as S-ApBM is crucial in isolating the effect of specific intervention components. The lack of significant findings in the S-ApBM group suggests that the specific elements of the A-ApBM intervention, rather than mere participant engagement, may be driving the observed effects. This lack of significant change in the control groups highlights the importance of the specific elements incorporated into the A-ApBM intervention. The static nature of the S-ApBM, which did not adapt to individual performance, was insufficient to produce a meaningful reduction in cravings, emphasizing that engagement alone, without the dynamic adjustment of training difficulty, may not be enough to achieve the desired therapeutic outcomes.

Moreover, the study revealed a significant relationship between performance and difficulty indices within the A-ApBM intervention. The adaptive algorithm used in this study was designed to maintain an optimal level of difficulty by continuously adjusting based on real-time performance data. This relationship between performance and difficulty underscores the importance of personalized interventions in maintaining user engagement and enhancing training efficacy. By leveraging this relationship, the adaptive algorithm helps ensure that the training remains challenging yet achievable, thereby maximizing its therapeutic impact.

These findings have broader implications for the development of future interventions for substance use disorders and other psychiatric conditions. The success of the A-ApBM intervention in this study suggests that incorporating gamification and adaptive elements into CBM programs may be a promising strategy for improving outcomes in various therapeutic contexts. The effect size, measured by Cohen *d*, suggests a small to moderate effect, which could imply that patients might continue to benefit even after the intervention has ended. This could reduce the need for continuous treatment and potentially lower the costs and resource use in clinical settings. Future research could explore the application of similar adaptive algorithms in other types of addiction treatments or in interventions targeting different psychiatric disorders.

A study by Smith [25] has found that monotony, repetition, lack of novelty, and lack of complexity cause boredom. Boredom can negatively impact engagement, attention, and interest, which can ultimately reduce the effectiveness of CBM programs. The mixed-effect linear regression indicates that the posited relationship between difficulty and performance was evident. The inverted U-shape effect of the individualized task difficulty on performance and the negative correlation with the SD of past difficulty indices suggest that the algorithm makes the ApBM training more engaging, effective, and personalized for individuals with substance use disorder. This adaptive approach has the potential to improve treatment outcomes and contribute to the field of clinical psychology and addiction treatment.

Limitations

We acknowledge the following limitations of the study. First, the sample size was relatively small and homogenous in race and ethnicity, which may limit the generalizability. The originally planned sample size of 300 participants across 6 groups was revised to 150 participants across 3 groups due to logistical constraints and feasibility considerations. Specifically, the limited number of eligible individuals in the rosters of the 12 participating community rehabilitation centers necessitated this adjustment. While we successfully enrolled 136 participants, nearly reaching our revised target, the recruitment phase was concluded when the available pool of eligible participants was exhausted. This reduction in sample size may limit the generalizability of our findings. Furthermore, the decision to exclude additional algorithmic variants (U-shape, N-shape, and inverted N-shape difficulty curves) reduced the scope of the study. Future research should aim to replicate these findings with larger sample sizes and include evaluations of the excluded algorithmic variants to validate and expand upon our results. Second, the duration of the intervention program was relatively short at 4-week intervention and 16-week follow-up. Third, the relapse outcome might have been influenced by participant behavior due to their awareness of the study. Furthermore, the loss of 5 participants from the control and A-ApBM groups at week 16, who did not complete the follow-up assessments, raises unconfirmed concerns about possible relapse. This attrition and the potential for behavior modification among participants highlight the need for longer-term studies with strategies to ensure sustained engagement for a more accurate assessment of relapse prevention in methamphetamine use. In considering the relapse rates observed in our study, it is important to contextualize these findings within the framework of participant behavior and study design. The participants were aware that the study duration was 4 weeks, which might have influenced their commitment to abstain from methamphetamine use during this period. This awareness could have motivated participants to consciously refrain from relapsing, in an effort

XSL•FO
RenderX

to demonstrate their engagement and adherence to the study protocols. Furthermore, it is noteworthy that in both the no-intervention control group and the A-ApBM group, there were instances of participant attrition at the 16-week mark—2 and 3 participants, respectively, were lost to follow-up and did not complete the craving assessment or the urine tests. While the reasons for this attrition are unknown, one cannot rule out the possibility of relapse among these individuals. However, this remains speculative in the absence of concrete evidence. The loss of these participants from the follow-up assessment phase underlines the complexities involved in conducting long-term studies in populations with substance use disorders and highlights the potential challenges in maintaining consistent participant engagement over extended periods.

Conclusions

This study highlights the promise of an adaptive and gamified approach to CBM in mitigating cue-induced cravings among individuals with a history of methamphetamine use. The dynamic adjustment of training difficulty based on individual performance is crucial in maximizing intervention effectiveness. These findings emphasize the need for engagement and personalization in CBM programs, suggesting that adaptive strategies could significantly enhance treatment outcomes for substance use disorders. Future research should continue to explore these adaptive elements to further validate their benefits in diverse therapeutic contexts.

Acknowledgments

The study was funded in part by National Key Research and Development Program of China (grant 2023YFC3304200). The authors would like to thank the participating social workers and administrative staffs at the participating communities.

Data Availability

Example data and the code for replicating the algorithm are available at GitHub [26].

Authors' Contributions

LZ contributed to conceptualization, writing—original draft preparation, investigation, data curation, and software; YL contributed to conceptualization, validation, software, writing—original draft preparation, writing—review, and editing; XL contributed to conceptualization, formal analysis, validation, and supervision; TZ contributed to data curation and software; DL contributed to conceptualization and project administration; and DS, JJ, and WH contributed to supervision, writing—review, and editing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
User survey questionnaire.
[PDF File, 36 KB - games_v13i1e56978_app1.pdf]

Multimedia Appendix 2
Description of WonderLab Harbour.
[DOCX File, 637 KB - games_v13i1e56978_app2.docx]

Checklist 1
CONSORT (Consolidated Standards of Reporting Trails) checklist.
[DOC File, 224 KB - games_v13i1e56978_app3.doc]

References

1. Cristea IA, Kok RN, Cuijpers P. The effectiveness of cognitive bias modification interventions for substance addictions: a meta-analysis. PLoS One 2016;11(9):e0162226. [doi: 10.1371/journal.pone.0162226] [Medline: 27611692]
2. Jones EB, Sharpe L. Cognitive bias modification: a review of meta-analyses. J Affect Disord 2017 Dec 1;223:175-183. [doi: 10.1016/j.jad.2017.07.034] [Medline: 28759865]
3. Boendermaker WJ, Sanchez Maceiras S, Boffo M, Wiers RW. Attentional bias modification with serious game elements: evaluating the shots game. JMIR Serious Games 2016 Dec 6;4(2):e20. [doi: 10.2196/games.6464] [Medline: 27923780]
4. Lisk SC, Pile V, Haller SPW, Kumari V, Lau JYF. Multisession cognitive bias modification targeting multiple biases in adolescents with elevated social anxiety. Cognit Ther Res 2018;42(5):581-597. [doi: 10.1007/s10608-018-9912-y] [Medline: 30237649]
5. Blackwell SE, Holmes EA. Modifying interpretation and imagination in clinical depression: a single case series using cognitive bias modification. Appl Cogn Psychol 2010 Apr;24(3):338-350. [doi: 10.1002/acp.1680]

XSL•FO
RenderX

6. Beard C, Weisberg RB, Primack J. Socially anxious primary care patients' attitudes toward cognitive bias modification (CBM): a qualitative study. Behav Cogn Psychother 2012 Oct;40(5):618-633. [doi: 10.1017/S1352465811000671] [Medline: 22127022]
7. Kakoschke N, Kemps E, Tiggemann M. Approach bias modification training and consumption: a review of the literature. Addict Behav 2017 Jan;64:21-28. [doi: 10.1016/j.addbeh.2016.08.007] [Medline: 27538198]
8. Schumacher SE, Kemps E, Tiggemann M. Bias modification training can alter approach bias and chocolate consumption. Appetite 2016 Jan 1;96:219-224. [doi: 10.1016/j.appet.2015.09.014] [Medline: 26375357]
9. Eberl C, Wiers RW, Pawelczack S, Rinck M, Becker ES, Lindenmeyer J. Approach bias modification in alcohol dependence: do clinical effects replicate and for whom does it work best? Dev Cogn Neurosci 2013 Apr;4:38-51. [doi: 10.1016/j.dcn.2012.11.002] [Medline: 23218805]
10. Zhang L, Li N, Li Y, et al. Preliminary efficacy of a digital therapeutics smartphone application for methamphetamine use disorder: an experimental study. Front Psychiatry 2022;13:1027695. [doi: 10.3389/fpsyt.2022.1027695]
11. Manning V, Piercy H, Garfield JBB, Clark SG, Andrabi MN, Lubman DI. A personalized approach bias modification smartphone app ("SWiPE") to reduce alcohol use: open-label feasibility, acceptability, and preliminary effectiveness study. JMIR Mhealth Uhealth 2021 Dec 10;9(12):e31353. [doi: 10.2196/31353] [Medline: 34890355]
12. Zhang M, Ying J, Wing T, Song G, Fung DSS, Smith H. A systematic review of attention biases in opioid, cannabis, stimulant use disorders. Int J Environ Res Public Health 2018 Jun 1;15(6):1138. [doi: 10.3390/ijerph15061138] [Medline: 29857586]
13. Zhang M, Heng S, Song G, Fung DS, Smith HE. Co-designing a mobile gamified attention bias modification intervention for substance use disorders: participatory research study. JMIR Mhealth Uhealth 2019;7(10):e15871. [doi: 10.2196/15871]
14. Vermeir JF, White MJ, Johnson D, Crombez G, Van Ryckeghem DML. Gamified web-delivered attentional bias modification training for adults with chronic pain: protocol for a randomized, double-blind, placebo-controlled trial. JMIR Res Protoc 2022 Jan 27;11(1):e32359. [doi: 10.2196/32359] [Medline: 35084344]
15. Schmidt-Kraepelin M, Thiebes S, Stepanovic S, Mettler T, Sunyaev A. Gamification in health behavior change support systems-a synthesis of unintended side effects. Presented at: Proceedings of the 14th International Conference on Wirtschaftsinformatik; Feb 24-27, 2019; Siegen, Germany.
16. Cowley B, Charles D, Black M, Hickey R. Toward an understanding of flow in video games. Comput Entertain 2008 Jul;6(2):1-27. [doi: 10.1145/1371216.1371223]
17. Li Y, Ryan CT, Sheng L. Optimal sequencing in single-player games. Manage Sci 2023 Oct;69(10):6057-6075. [doi: 10.1287/mnsc.2022.4665]
18. Amir N, Kuckertz JM, Strege MV. A pilot study of an adaptive, idiographic, and multi-component attention bias modification program for social anxiety disorder. Cognit Ther Res 2016 Oct;40(5):661-671. [doi: 10.1007/s10608-016-9781-1] [Medline: 27795598]
19. Amir N, Holbrook A, Kallen A, et al. Multiple adaptive attention-bias-modification programs to alter normative increase in the error-related negativity in adolescents. Clin Psychol Sci 2024 May;12(3):447-467. [doi: 10.1177/21677026231170563] [Medline: 39040548]
20. Piper BJ, Ogden CL, Simoyan OM, et al. Trends in use of prescription stimulants in the United States and Territories, 2006 to 2016. PLoS One 2018;13(11):e0206100. [doi: 10.1371/journal.pone.0206100] [Medline: 30485268]
21. Hedegaard H, Bastian BA, Trinidad JP, Spencer M, Warner M. Drugs most frequently involved in drug overdose deaths: United States, 2011-2016. Natl Vital Stat Rep 2018 Dec;67(9):1-14. [Medline: 30707673]
22. Office of China National Narcotics Control Commission. Drug situation in China (2019). China Anti-drug Network. 2020. URL: http://www.nncc626.com/2020-06/25/c_1210675877.htm [accessed 2025-02-11]
23. Wang X, Li Y, Li J, Hao W. Emerging patterns of substance abuse and related treatment in China. Curr Opin Psychiatry 2023 Jul 1;36(4):277-282. [doi: 10.1097/YCO.0000000000000878] [Medline: 37191649]
24. Kruijt AW, Carlbring P. Mechanics of contingency-based cognitive bias modification: pre-existing bias affects potency of active training but not placebo conditions. PsyArXiv. Preprint posted online on 2018. [doi: 10.31234/osf.io/scqf3]
25. Smith RP. Boredom: a review. Hum Factors 1981 Jun;23(3):329-340. [doi: 10.1177/001872088102300308] [Medline: 7016721]
26. liuxiang91/A-apbm. GitHub. URL: https://github.com/liuxiang91/A-ApBM/ [accessed 2025-02-11]

Abbreviations

A-ApBM: adaptive approach bias modification
ApBM: approach bias modification
CBM: cognitive bias modification
ITR: intended training ratio
Past3ITRSD: SD of the intended training ratio from recent sessions.
S-ApBM: static approach bias modification

XSL•FO
RenderX

Edited by A Coristine, T Leung; submitted 31.01.24; peer-reviewed by K Lee, M Jia, R Wiers; revised version received 12.12.24; accepted 12.12.24; published 10.03.25.

Please cite as:
Shen D, Jiao J, Zhang L, Liu Y, Liu X, Li Y, Zhang T, Li D, Hao W
Gamified Adaptive Approach Bias Modification in Individuals With Methamphetamine Use History From Communities in Sichuan: Pilot Randomized Controlled Trial
JMIR Serious Games 2025;13:e56978
URL: https://games.jmir.org/2025/1/e56978
doi:10.2196/56978

© Danlin Shen, Jianping Jiao, Liqun Zhang, Yanru Liu, Xiang Liu, Yuanhui Li, Tianjiao Zhang, Dai Li, Wei Hao. Originally published in JMIR Serious Games (https://games.jmir.org), 10.3.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Considering Theory-Based Gamification in the Co-Design and Development of a Virtual Reality Cognitive Remediation Intervention for Depression (bWell-D): Mixed Methods Study

Mark Hewko[1], MEng; Vincent Gagnon Shaigetz[2], MSc; Michael S Smith[1], MSc; Elicia Kohlenberg[1], DEC; Pooria Ahmadi[3], MD; Maria Elena Hernandez Hernandez[3], PhD; Catherine Proulx[2], MSc; Anne Cabral[2], MEng; Melanie Segado[2], PhD; Trisha Chakrabarty[3*], MD; Nusrat Choudhury[2*], MEng

[1]National Research Council, Winnipeg, MB, Canada

[2]National Research Council, Boucherville, QC, Canada

[3]Department of Psychiatry, Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada

[*]these authors contributed equally

Corresponding Author:
Mark Hewko, MEng
National Research Council
435 Ellice Avenue
Winnipeg, MB, R3B 1Y6
Canada
Phone: 1 204 984 4625
Email: mark.hewko@nrc-cnrc.gc.ca

Abstract

Background: In collaboration with clinical domain experts, we developed a prototype of immersive virtual reality (VR) cognitive remediation for major depressive disorder (bWell-D). In the development of a new digital intervention, there is a need to determine the effective components and clinical relevance using systematic methodologies. From an implementation perspective, the effectiveness of digital intervention delivery is challenged by low uptake and high noncompliance rates. Gamification may play a role in addressing this as it can boost adherence. However, careful consideration is required in its application to promote user motivation intrinsically.

Objective: We aimed to address these challenges through an iterative process for development that involves co-design for developing content as well as in the application of gamification while also taking into consideration behavior change theories. This effort followed the methodological framework guidelines outlined by an international working group for development of VR therapies.

Methods: In previously reported work, we collected qualitative data from patients and care providers to understand end-user perceptions on the use of VR technologies for cognitive remediation, reveal insights on the drivers for behavior change, and obtain suggestions for changes specific to the VR program. In this study, we translated these findings into concrete representative software functionalities or features and evaluated them against behavioral theories to characterize gamification elements in terms of factors that drive behavior change and intrinsic engagement, which is of particular importance in the context of cognitive remediation. The implemented changes were formally evaluated through user trials.

Results: The results indicated that feedback from end users centered on using gamification to add *artificial challenges*, *personalization* and *customization* options, and *artificial assistance* while focusing on *capability* as the behavior change driver. It was also found that, in terms of promoting intrinsic engagement, the need to meet *competence* was most frequently raised. In user trials, bWell-D was well tolerated, and preliminary results suggested an increase in user experience ratings with high engagement reported throughout a 4-week training program.

Conclusions: In this paper, we present a process for the application of gamification that includes characterizing what was applied in a standardized way and identifying the underlying mechanisms that are targeted. Typical gamification elements such as *points* and *scoring* and *rewards* and *prizes* target *motivation* in an extrinsic fashion. In this work, it was found that modifications suggested by end users resulted in the inclusion of gamification elements less commonly observed and that tend to focus more on individual

XSL•FO
RenderX

ability. It was found that the incorporation of end-user feedback can lead to the application of gamification in broader ways, with the identification of elements that are potentially better suited for mental health domains.

(JMIR Serious Games 2025;13:e59514) doi:10.2196/59514

KEYWORDS

virtual reality; clinical psychology; cognitive assessment; neuropsychology; mental health; cognitive training; cognitive remediation; cognitive rehabilitation; digital therapeutics

Introduction

Background

Major depressive disorder (MDD) is the leading cause of disability worldwide [1]. Current treatment regimens, while effective in improving core symptoms of MDD, fall short in addressing MDD-related disability [2-4]. Cognitive dysfunction—encompassing deficits in memory, processing speed, attention, and executive functioning—has emerged as an important determinant of functional outcomes in MDD [5-7]. Advancing cognitive-targeting therapies such as cognitive remediation has accordingly become a critical treatment priority. Cognitive remediation is a behavioral-based therapy aimed at enhancing neurocognitive skills, typically through a combination of "drill-and-practice" cognitive games and implementation of compensatory strategies [8]. Traditional cognitive remediation programs delivered in paper-and-pencil or computerized formats have demonstrated robust improvements in both cognition and real-world functioning in schizophrenia [9], attention-deficit/hyperactivity disorder, and autism spectrum disorder [10]. In MDD, results to date have been equivocal. Although cognitive remediation in patients with MDD appears to improve performance in select cognitive domains, this has not translated to appreciable improvements in real-world functioning [11,12].

Immersive virtual reality (VR) has emerged as a promising modality to enhance the real-world effectiveness of cognitive remediation. Such technologies allow for almost complete sensory immersion with extensive design possibilities. By creating simulations that engage users via multiple senses (sight, hearing, and touch) while permitting natural movement, immersive VR creates a sense of presence that can elicit a genuine response in the individual. The use of VR environments also permits tight experimental control, which makes it feasible to measure and study everyday functioning that can be otherwise prohibitive in real-world settings. The ability of VR to deliver and control stimuli while capturing responses with high fidelity during an exercise provides a controlled and repeatable tool that is unavailable in traditional testing or remediation methods and makes it ideal for assessing and training cognitive skills in the performance of simulated real-life tasks. There has been growing interest in the use of VR for both assessment and rehabilitation from the research and clinical neuropsychology communities in recent years [13,14]. The feasibility of using VR for cognitive assessment and care has been demonstrated across various cognitive domains, such as attention [15-21], executive functions [22-25], memory [26-31], and spatial abilities [32-34].

The National Research Council Canada (NRC) has developed bWell, a VR-based immersive platform for cognitive assessment and remediation [35]. bWell is a foundational platform featuring configurable exercises and designed as a broadly applicable toolkit targeting general aspects of cognition that are commonly impacted across many disorders. A video sample of bWell's interface and laboratory exercise are presented in Multimedia Appendix 1. Preliminary acceptability studies in healthy individuals have indicated that the platform is enjoyable, engaging, and well tolerated [36,37]. These studies have also shown that using immersive VR for clinical applications not only is technically feasible but also has advantages over traditional 2D equivalents. The first study showed that subjective reports of engagement when performing a task in bWell (an immersive environment) were greater than those when performing the same task on a tablet (a nonimmersive environment). The second study showed that the use of more complex types of immersion (eg, the user was not static) did not increase subjective reports of cybersickness during a visual attention task and provided the required support for testing in clinical populations.

bWell is currently being incorporated into clinical settings using an interdisciplinary approach successfully developed by the NRC [38-41]. This method, which has proven effective in creating surgical simulation platforms, most notably for neurosurgery (eg, NeuroTouch, distributed worldwide as NeuroVR by CAE Healthcare), ensures a comprehensive and collaborative framework for implementation. Specifically, the team established long-term collaborative research agreements with 5 world-renowned clinical sites across Canada and 1 in Japan—the NRC's cognitive care network (CCN). The CCN consists of an early adopter group that provided domain-specific perspectives as well as feedback on iterative deliveries of the platform as bWell development progressed. The CCN is currently launching several studies investigating the content of bWell exercises, as well as its use, specific to target populations.

In one of these collaborations, with the University of British Columbia, we customized bWell to deficits observed in depression as bWell-D, a prototype of immersive VR cognitive remediation for MDD. bWell-D targets cognitive and real-world challenges germane to MDD depression [6]. It uses previously identified common components of successful cognitive remediation in psychiatric populations [42]. These include (1) *errorless learning*, wherein participants are provided support in learning all task components, with supports gradually removed as skills and confidence increase; and (2) *adaptivity*, where tasks become increasingly complex to match the participants' growing competence.

In the development of a new digital technology, there is a need to determine the effective components and clinical relevance using systematic methodologies. A working group of

XSL•FO
RenderX

international experts has published the first consensus guidelines for best practices in the design and testing of VR clinical applications [43]. In these guidelines, incorporation of qualitative feedback from multiple end users is identified as a critical component of VR program content development. In addition, the Medical Research Council (United Kingdom), in a guidance document on the development of complex interventions to improve health and health care [44], recommends including prevailing theories and frameworks in a given domain to anchor the feedback obtained.

From an implementation perspective, the effectiveness of digital intervention delivery is challenged by low uptake and high noncompliance rates, which gamification may play a role in addressing. A systematic study of technology use in the treatment of mental health disorders found that, specific to depression, gamification was being used as an approach aiming to improve adherence and engagement [45]. The application of gamification to mental health domains is still a new research field with limited studies [46] that tend to be heterogeneous [47]. The results are inconsistent; for example, it was observed that its application to web-based interventions did not lead to a significant increase in adherence [48]. There is also a lack of comparison between gamified and nongamified interventions, which makes it difficult to evaluate the impact of such features on adherence and engagement. Finally, it has been suggested that gamification might be used without fully understanding the behavioral theories or mechanisms behind it [46,49]. For instance, adding too many gamification elements that rely on external rewards can overwhelm the user with unnecessary cognitive demands, potentially invalidating the task [47].

Objectives

Using behavioral theory as a framework for understanding likely processes for change induced by an intervention is hypothesized to enhance its effectiveness [40,50]. The capability, opportunity, and motivation–behavior (COM-B) model [51] is one such theory that provides valuable insights. It is widely used and identifies *capability*, *motivation*, and *opportunity* as the 3 essential elements for all behavior. Effecting behavior change, such as improving compliance with a digital intervention, requires addressing at least one of these drivers of behavior. Another relevant behavioral theory is self-determination theory, which delves deeper into the aspects of motivation. This theory posits that people are intrinsically motivated to grow and change when their needs for *autonomy*, *competence*, and *connection* are met [52].

Methods

Overview

In this work, we aimed to address these challenges through an iterative process for the development of bWell-D that involves co-design for the development of content and in the application of gamification using behavior change theories as a framework.

Following best practice guidelines, qualitative data were collected from patients and care providers to understand end-user perceptions on the use of VR technologies for cognitive remediation and obtain suggestions for changes specific to the bWell-D program. This qualitative study phase was reported in a previous paper from the authors [53]. Briefly, participants comprising individuals with lived experience of depression (n=15) and clinicians (n=12) were interviewed, and the results were transcribed and coded followed by thematic analysis. The demographic and clinical characteristics of these end users are described in our previous paper. Interviewees were asked questions to probe drivers of behavior change related to the use of VR (eg, perceptions and experiences with VR, desired outcomes from a workplace cognitive and functional remediation program, and perceived barriers). Participants were also asked for feedback specific to the bWell-D program itself, for which they were shown previously recorded video samples of the tasks and probed for potential relevance to daily life and how to improve the tasks to enhance engagement and real-world applicability. Emerging themes from patients and clinicians were identified and ranked based on prevalence.

The focus of this paper was to structure these previously collected qualitative data into a theoretically grounded model to guide software and functionality development for the next iterations of bWell-D. Specifically, we aimed to use these findings to implement features for the next iterations of bWell-D specific to its use in cognitive remediation while promoting the required motivation, engagement, and ease of use or access to obtain favorable outcomes. Following best practice guidelines for human-centered VR content development [43], the themes suggested by end users to improve the bWell-D program were discussed among the research team and translated into representative software functionalities during team brainstorming sessions. It was found that the functionalities that could be implemented could also be described in many ways. In an effort to standardize the categorization, all functionalities were encoded against gamification elements based on the taxonomy by Cheng et al [49], who describe 18 elements used in digital technologies for 17 mental health and well-being domains, including depressive disorders.

The gamification elements were then evaluated against behavioral theories to characterize each element in terms of (1) factors that drive behavior change and (2) potential to intrinsically promote user engagement. The COM-B factors and psychological needs of self-determination are presented in Textbox 1 [51,52] to provide a clearer definition.

Taking into consideration the factors identified from behavioral theories (eg, opportunity), the ideation included both updating the concepts for each bWell-D exercise as well as the overall delivery and use of the bWell-D program itself. The modifications were assigned priority for implementation based on (1) how frequently these changes were suggested by end users and (2) whether these changes were feasible for development in a VR environment.

The steps of the methodology used are summarized in Figure 1. End-user feedback was thematically grouped and developed into software functionalities. These functionalities were categorized using a taxonomy for digital technologies used in mental health and well-being applications. Furthermore, the functionalities were assessed using selected behavior change theories to view the distribution of needs identified by end users

XSL•FO
RenderX

for platform improvement. The last step involves validating the implemented software functionalities through an iterative development process that gathers new feedback from end users for further refinement. Multiple rounds of iteration will help create a comprehensive overview of the functionalities that end users want and refine these implementations to ensure that they effectively meet users' needs. This overview will highlight the range of engagement methods preferred by users organized according to a taxonomy of gamification elements for digital health applications. In addition, selected behavior change theories will provide insights into what motivates users to engage with and adhere to treatments.

Finally, in this study, new participants were recruited for a pilot feasibility trial with a healthy population. With the aim of evaluating the impact of the gamification elements, participants trialed the original version for 40 minutes, after which they were given the option to either try out the expanded features of the bWell-D platform for up to 15 minutes or embark on a training program for 8 sessions (occurring 2 times a week for 4 weeks). Feedback regarding tolerability and cybersickness was obtained using the Simulator Sickness Questionnaire (SSQ) [54] before and after each session. User experience feedback was obtained using the User Engagement Scale (UES) [55] and Game User Experience Satisfaction Scale (GUESS) [56] after the initial assessment session and after the first, fourth, and last training sessions. Ad hoc comments made by the participants during the trials were also noted. For participants who chose to try out the expanded feature without attending the training sessions, feedback was obtained via a semistructured interview while performing the exercises.

Textbox 1. Categorization in terms of factors that drive behavior change and potential to intrinsically promote engagement.

Categorization for effective behavior change

- The capability, opportunity, and motivation–behavior model posits that there are 3 drivers for behavior change. (1) *Capability* is the ability to engage in an activity; (2) *opportunity* is factors that lie outside the individual that can make behavior possible, prompt it, or present a barrier to it; and (3) *motivation* is the brain processes that energize and direct behavior.

Categorization for intrinsic engagement

- According to self-determination theory, 3 needs must be met to achieve psychological (inner) growth. (1) *Autonomy* is when a person feels in control of their own behaviors and that their action will result in real change, it plays a major part in helping them feel self-determined; (2) *competence* is when a person feels equipped with the skills needed, they are more likely to take action; (3) *connection or relatedness* is a sense of belonging and attachment to others promotes the actualization of inherent potential. Social environments supporting these needs enhance one's internal motivation.

Figure 1. Iterative software development process using end-user feedback, thematic segmentation of feedback, software functionality development, taxonomy assignment, assessment using selected behavior change theories, distribution review, and functionality validation for improvement.



Ethical Considerations

The research and studies were reviewed and approved by the institutional research ethics board at the University of British Columbia (H21-00028 and H20-00746) and the NRC (2020-122). Informed consent was obtained before the start of the study. Data were deidentified before any analysis. Participants were compensated for taking part with CAD $100 (US $70).

Results

Overview

Results from the previously published qualitative study [53] showed that, overall, patients and clinicians considered bWell-D to be interesting, acceptable, and potentially feasible and provided suggestions to enhance its applicability. Five main themes were identified from the patient interviews: (1) patients would prefer to undergo VR treatment at home, (2) the VR treatment should emulate real-life settings, (3) the VR intervention needs to be multisensory, (4) the VR treatment should be challenging, and (5) the VR treatment should be engaging. Three main themes were identified from the clinician interviews: (1) the VR treatment should be multi-domain, (2) the VR tasks should become more challenging over time, and (3) remote treatment would be more convenient for patients. In addition, end users suggested having a user-friendly interface and tutorials to have an opportunity to practice the tasks, which they considered especially important for at-home treatment. Participants also recommended including an additional VR activity that specifically targets their mood.

In this study, these themes were evaluated against the identified behavioral theories, and it was found that the themes encompassed all 3 COM-B drivers for behavior change. Regarding *motivation*, VR is seen as a potential, interesting, but underexplored treatment option for remediating cognitive

XSL•FO
RenderX

deficits in depression. Many suggestions for bWell were related to better harnessing the affordances of VR to deliver multisensorial and multimodal exercises. Regarding *capability*, VR was seen as "the second-best thing to real life" as it allows for immersion in highly realistic settings. Some skepticism about effectiveness in the short and long term was also expressed, especially when the purpose and real-world applications of practiced skills were not made explicit. Regarding *motivation*, the gamified experience of VR adds fun, engaging, and immersive elements to the treatment. There is a need for a good balance in challenge (motivating but not too hard) for continued engagement. Regarding *opportunity*, there is a need for inclusivity and customizability in treatment delivery. Participants believed that the option for home-based treatment provided a sense of comfort and control. However, it was not the preference for everyone. Specifically, patients with limited space or suboptimal conditions for treatment at home would prefer having the option to receive treatment at a clinic if necessary. Concerns about equipment accessibility were also highlighted.

In terms of feedback specific to bWell-D, the main areas of improvement suggested by both clinicians and patients centered on (1) *motivation*—taking greater advantage of VR multisensorial affordances, (2) *capability*—the inclusion of more realistic VR environments, and (3) *motivation*—providing sufficient challenge for fun and engaging tasks.

It was observed that patient and clinician themes overlapped on several common topics, and thus, they were combined. Similarly, the feedback specific to bWell was related to that for the general use of VR. However, while the suggestions specific to bWell addressed 2 of the drivers—*motivation* and *capability*—*opportunity* was not captured. In an aim to capture all drivers for behavior change as well as all the major suggestions from end users, the following overall themes emerged as priorities for the development of bWell-D: (1) provide sufficient *challenge* over time for fun and engaging tasks (patient themes 4 and 5 and clinician theme 2), (2) take greater advantage of the potential of VR to create tasks that are more *multi-domain* and *multisensorial* (visual, auditory, and tactile stimuli; patient theme 3 and clinician theme 1), (3) include more *representative* VR elements (to reflect more realistic VR environments; patient theme 2), (4) increase *autonomy*, opportunity, or enablement with remote or *in-home* use (patient theme 1 and clinician theme 3), and (5) provide *mood-focused* features and tasks (participant recommendation).

The gamification elements for each exercise, including relevant original elements and the newly implemented modifications based on qualitative feedback, are summarized in the upcoming sections. Each element is classified according to gamification taxonomy and characterized according to the drivers of behavior change from the COM-B framework it relates to (*motivation*, *capability*, or *opportunity*), as well as the need from self-determination theory that it most addresses (*autonomy*, *competence*, or *connection*). Each element is also associated with overall themes summarized from the qualitative analysis.

General Interface, Tutorial, and bWell-D Administration

The general modifications to bWell-D are described in this section, and Figure 2 provides a summary of the categorization.

bWell-D has 3 distinct modes to optimize user engagement and outcomes. (1) Tutorial mode is for initiation with the virtual environment, providing a comprehensive walkthrough of the tasks to be undertaken. It ensures that users are comfortable and well prepared to navigate the tasks effectively. (2) Assessment mode serves as a diagnostic tool to evaluate the users' abilities by monitoring their performance at a consistent level of difficulty. The primary objective is to pinpoint the users' unique requirements and areas where they may need additional support or intervention. (3) Training mode is tailored to foster user growth. The training mode dynamically adjusts the difficulty level in response to individual performance. This personalized approach ensures that the tasks remain challenging yet achievable, promoting continuous skill development. Users receive instantaneous feedback on progress via level or score change notifications.

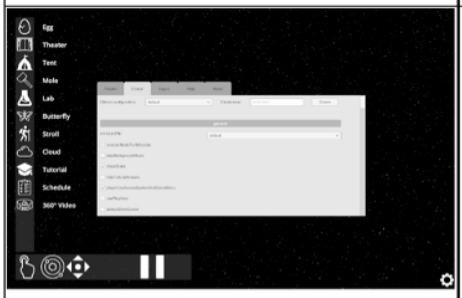
The modes can be configured and customized via a clinical interface as bWell-D is intended for administration by a facilitator.

On the basis of the emerging theme of increased opportunity with at-home use, a new user interface was implemented to support autonomous use. As a result, bWell-D's hardware-agnostic core has now been leveraged to integrate stand-alone headsets such as the PICO Neo3 Eye and the Oculus Quest 2. Such headsets have come a long way, from clunky, tethered devices to performant and comfortable stand-alone units aimed at the mass market that enable seamless onboarding of nontechnical users new to VR.

Existing features of bWell did not include the emerging theme of representative real-world elements (Figure 2). With the aim to make the purpose and real-world applications of practiced skills more explicit, bWell-D administration now includes "bridge exercises" as a supplement to the VR program. These exercises consist of asking participants to reflect on how they think the VR tasks will apply to their personal contexts to help them understand how practiced skills could be related to their real life.

XSL•FO
RenderX

Figure 2. General interface, tutorial, and program administration, including modifications made after end-user feedback. Checkmarks (✓) denote existing elements and plus signs (+) denote new elements that were added in the updated version of bWell-D. Gray shading indicates that the theme does not apply. VR: virtual reality.

General interface, tutorial, and program administration								
Functionality or feature	Categorization			Theme addressed				
	Gamification element	Behavior change factor	Intrinsic need met	Challenging	Multisensorial or multidomain	Ecological	Use at home or autonomy	Target mood
Inclusion of tutorials to provide opportunities to practice	Artificial assistance	Capability	Competence				✓	
Patient preferences menu allows user to choose among tutorial, assessment, or training	Customization	Opportunity	Autonomy				✓	
User interface supporting autonomous use	Customization	Capability	Autonomy				+	
Change to administration protocol to include bridging exercises after the VR session	None	Capability	Competence			+		
VR program can be delivered on tethered or stand-alone headsets, which more suited for use at home	None	Opportunity	Autonomy				+	

Egg Scene

The egg scene takes place in an office work environment. The user is tasked with scanning the environment for eggs, drawing on visual processing and sustained attention. They are then required to direct and hold their gaze on a located egg long enough to make it hatch. Modifications to the scene and a summary of their categorization are presented in Figure 3, where it can be observed that only the theme of representative real-world elements has been addressed in the existing features (ie, creation of a simulated real-world environment).

In the feedback obtained, end users recommended including more real-world VR elements. To address this, the egg target was changed to new, more context-appropriate ones for the office-themed virtual environment. These included staplers, legal pads, and coffee mugs.

An in-game user interface was added to the player's wrist during the egg scene to display important information, including the current score, target object, and target location. This type of interface takes advantage of the user's natural arm movements, allowing them to easily access information or controls by looking at or interacting with their wrist, much like checking a smartwatch. It is designed to be intuitive and accessible without interrupting the immersive experience. This *artificial assistance* element addressed the need for *autonomy.*

Ambient, multisensorial distractors were added as options to the scene. Visual distractors include videos playing on computer screens, coworkers walking around the office, and fake eggs that resemble the real target eggs. Audio distractors include phone rings, coworkers talking, and a buzzing sound that emanates from randomized locations within the office. These distractors add *artificialchallenge* as well as addressing the mood theme by introducing an element of hot cognition.

A *customization* option was added in which users follow auditory rather than the usual visual instructions. This addresses the user's *motivation*, providing an alternative sensorial interaction pathway to engage with the user.

An adaptive difficulty progression including multisensorial stimuli was developed for the scene. To address the user's *capability*, difficulty adjusts based on user performance, making this a *personalization* element. Difficulty adjustment includes adjusting the number of targets, the reaction delay allowed to obtain bonus points, the time it takes to hatch an egg, and the number and frequency of audio distractors. The user's performance is assessed based on how they respond to multiple stimuli: a visual (egg turns green), tactile (controller vibrates), or audio (clock ticking sound clip) stimulus.

The last modification to the scene was the addition of a preset difficulty progression. The user is given tasks, starting off simple at lower levels and becoming increasingly complex at higher levels. These tasks include finding a new target, going to a target location, and pressing a button sequence to obtain bonus points. Because these levels are chosen by the user or clinician rather than being adaptive, they are considered a *customization* element.

XSL•FO
RenderX

Figure 3. Egg exercise (targeting states of attention), including modifications made after end-user feedback. Checkmarks (✓) denote existing elements and plus signs (+) denote new elements that were added in the updated version of bWell-D.

Egg exercise (states of attention)									
Functionality or feature	Categorization			Theme addressed					
	Gamification element	Behavior change factor	Intrinsic need met	Challenging	Multisensorial or multidomain	Ecological	Use at home or autonomy	Target mood
Choice of environment: office versus park	Customization	Capability	Competence			✓		
Addition of realistic (office-related) targets	Narrative/theme	Capability	Competence			+		
In-game information displayed to user as a menu on their wrist (score, game mode, and current target)	Artificial assistance	Capability	Autonomy				+	
Ambient distractors (visual, audio, and emotional charge)	Artificial challenge	Capability	Competence	+	+			+
Preset variant of task: version where user follows auditory, rather than visually presented instructions	Customization	Motivation	Competence		+			
Adaptive difficulty progression modified to include visual, audio, and tactile elements	Personalization	Capability	Competence	+	+			
Preset difficulty progression: simple to more complex real-world mini-tasks (eg, find new target, go to target location, and press button sequence to obtain bonus points)	Customization	Capability	Competence	+		+		

Laboratory Scene

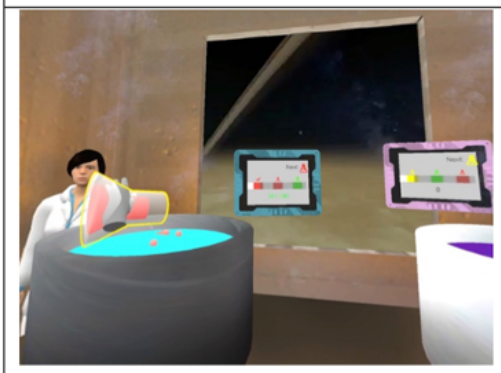
In the laboratory scene, the user must complete 2 recipes simultaneously by pouring the correct ingredients at the correct times into 2 different tubs. The scene challenges the user's executive function and attention across several dimensions, including divided, selective, and sustained attention. A supervisor avatar is present in the scene and adds *artificialassistance* by giving auditory feedback on the user's performance and advice on how to improve.

Modifications to the scene and a summary of their categorization are presented in Figure 4. On the basis of the identified theme of added challenge and the inclusion of more representative elements, a *customization* option was added in which the supervisor avatar gives the user auditory instructions rather than having the recipes visually presented on tablets. An aspect of hot cognition was also added by modifying the avatar behavior so that it can make distracting or irrelevant comments to the user.

In total, 2 modifications were made to the scene explicitly to add *artificialchallenge*. The first one simply added 3 new nonrelevant ingredient bottles as distractions while the user attempts to complete the recipes. The second modification was the introduction of new secondary ingredients to recipes. The user must first create the secondary ingredient by combining 2 primary ingredients in a beaker before pouring the beaker in the tub, adding complexity to the task. A chart showing possible ingredient combinations was added to the scene.

XSL•FO
RenderX

Figure 4. Laboratory exercise (targeting multitasking), including modifications made after end-user feedback. Checkmarks (✓) denote existing elements and plus signs (+) denote new elements that were added in the updated version of bWell-D.

Laboratory exercise (multitasking)									
Functionality or feature	Categorization			Theme addressed					
	Gamification element	Behavior change factor	Intrinsic need met	Challenging	Multisensorial or multidomain	Ecological	Use at home or autonomy	Target mood
Supervisor present, observing and providing audio feedback on performance	Artificial assistance	Capability	Competence			✓	✓	✓
Supervisor can make distracting or nonrelevant comments or give negative feedback	Artificial challenge	Capability	Competence					+
Addition of distractor beakers and target colors	Artificial challenge	Capability	Competence	+		+		
Preset variant of task: version where user follows auditory rather than visually presented instructions	Customization	Motivation	Competence		+			
New component in preset difficulty progression: mix ingredients to obtain a secondary ingredient according to chart	Artificial challenge	Capability	Competence	+		+		

Mole Scene

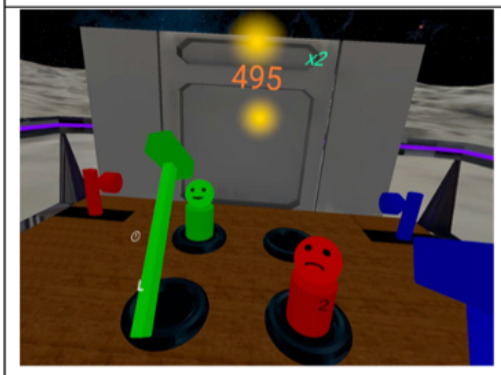
The mole scene is a *Whac-A-Mole* variation in which the user has a colored hammer in each hand with which they are to hit matching colored cylinders (moles) that pop up in front of them. Hammer colors change unexpectedly from time to time. Hitting valid moles exercises the user's impulse control with selective attention, visual processing, and executive functioning.

A score streak animation (visual reward) is used to provide *level* and *progressfeedback*, serving to *motivate* the user to maintain a fast pace as the difficulty increases. This occurs when multiple moles are correctly hit in sequence. In addition, to enable investigations of potential bias in MDD, the moles were modified to have a simple facial expression, either a smile or a frown. The program checks whether the particular facial expression influences the accuracy of the hits. In one mode, valid moles (ones with a color matching a hammer) are assigned a smiling expression, and invalid moles are frowning. In another mode, the expressions are randomized.

Modifications to the scene and a summary of their categorization are presented in Figure 5. From Figure 5, it can be observed that the existing features did not include the emerging theme of more multi-domain and multisensorial elements. The changes focused on exercise variants to address other domains. The multiple-hit mole *customization* variant was added to target mood. The multiple-hit mole is a variant taken from the bWell core battery of exercises designed to frustrate the user by requiring repeated hits for the mole to go down. In total, 2 other bWell core *customization* variants were added to the mole scene. The side moles and directional hit variants require additional physical effort or dexterity. Side moles appear to the side of the main table, requiring the user to turn their head and extend their arms further to hit them. In the directional hit variant, moles must be hit on the top of their head to score points. This is more challenging compared to the usual case in which moles can be hit from any angle and still score points. It was hoped that the additional physical engagement required by these variations would increase *motivation* for users to continue using the program.

XSL•FO
RenderX

Figure 5. Mole exercise (targeting response inhibition and cognitive control, including modifications made after end-user feedback). Checkmarks (✓) denote existing elements and plus signs (+) denote new elements that were added in the updated version of bWell-D.

Mole exercise									
Functionality or feature	Categorization			Theme addressed					
	Gamification element	Behavior change factor	Intrinsic need met	Challenging	Multisensorial or multidomain	Ecological	Use at home or autonomy	Target mood
Visual graphics to signal a score streak to user	Level/progress feedback	Motivation	None				✓	
Smiling/frowning moles to test for bias for negative stimuli	Artificial challenge	Capability	Competence					✓
Preset variants of task: versions designed to require additional physical effort (side moles and directional hits)	Customization	Motivation	Competence	+	+			
Preset variant of task: version that includes frustrating elements (multiple-hit moles)	Customization	Motivation	Competence		+			+

Theater Scene

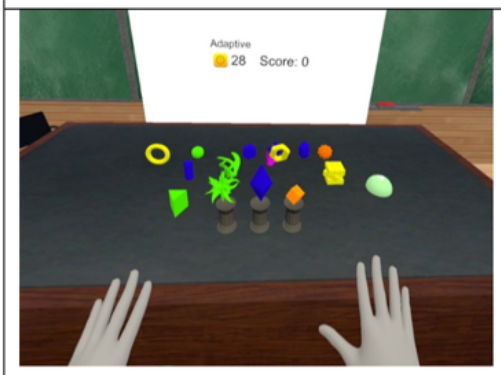
The theater scene is a matching task for visual and short-term memory. The user is presented with a set of target shapes for a short period and asked to remember it. The target shapes are then hidden from view. After a set time, random shapes start falling into view, and the user must reproduce the initial series of shapes. The scene was modeled after a university classroom, with the teacher's desk in front and the students' desks laid out in a semicircle in front of it. This setting provides an ecological environment for the sample population, which were recruited on a university campus. The user is placed in front of the teacher, who is seated at their desk. A few student avatars who may be enabled or disabled are seated in the classroom behind the user. This arrangement was intended to be somewhat daunting. For additional distraction, the teacher in front can create noise by taking down notes, and the student avatars can make loud whispering noises. The arrangement and the distractions add a level of *artificialchallenge* for the user as well as hot cognition.

Further *artificialchallenge* was added by including distracting shapes that are not part of the correct answer. The pool size of distracting shapes was increased to further enhance the challenge.

Modifications to the scene and a summary of their categorization are presented in Figure 6, where it can be observed that the existing features did not include the emerging theme of multisensorial and multi-domain elements. A *customization* option was added to the scene in which a virtual voice gives instructions about which shape to place in which order instead of presenting the target shapes visually. Anticipating that this audio mode would be too difficult when the presented shapes were complex or unfamiliar, a new collection of number-shaped objects was created for use in the auditory mode.

XSL•FO
RenderX

Figure 6. Theater exercise (targeting visual memory), including modifications made after end-user feedback. Checkmarks (✓) denote existing elements and plus signs (+) denote new elements that were added in the updated version of bWell-D.

Theater exercise (visual memory)									
Functionality or feature	Categorization			Theme addressed					
	Gamification element	Behavior change factor	Intrinsic need met	Challenging	Multisensorial or multidomain	Ecological	Use at home or autonomy	Target mood
Classroom scene environment with distractor avatars	Customization	Capability	Competence			✓		
Teacher present in front and students seated behind the user	Artificial challenge	Capability	Competence			✓		✓
Teacher and students can be distracting	Artificial challenge	Capability	Competence					✓
Inclusion of distractor objects (nontarget shapes)	Artificial challenge	Capability	Competence	+				
Preset variant of task: version where user follows auditory rather than visual presentation	Customization	Motivation	Competence		+			

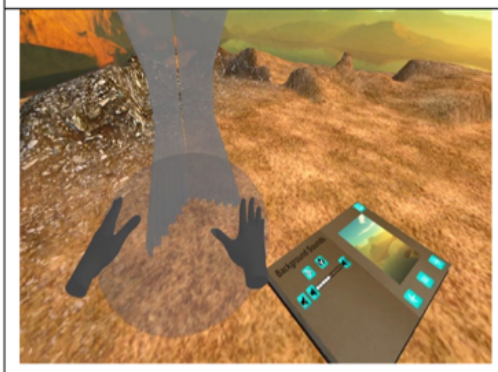
Tent Scene

The tent scene is intended as a space for relaxation and sensory immersion. The user is placed in a relaxing nature environment and allowed to freely explore. A ball that grows and shrinks to guide the user's breathing, as well as additional user *customization* promoting relaxation, is included. Specifically, the user was given control of the selection of the environment and background music that plays. A virtual book placed in front of the user includes controls that allow them to navigate through a variety of virtual environments. Giving users control over their environment is expected to add motivation for continued use and address their need for autonomy.

While this scene was part of the bWell platform, it was not initially included in bWell-D. Feedback from end users about the need for activities that target mood prompted its inclusion. The scene was designed to follow an *exploratoryopen-world approach*, allowing users to freely look around the environment for as much or as little time as they wished. A summary of the features and their categorization is presented in Figure 7.

XSL•FO
RenderX

Figure 7. Tent exercise (targeting relaxation). A task to target mood was added to bWell-D based on end-user recommendations. This task was taken as is from the bWell platform core battery of exercises. Checkmarks (✓) denote existing elements.

Tent exercise (relaxation)									
Functionality or feature	Categorization			Theme addressed					
	Gamification element	Behavior change factor	Intrinsic need met	Challenging	Multisensorial or multidomain	Ecological	Use at home or autonomy	Target mood	
Immersion in nature with self-guided exploration and duration	Exploratory/ open-world approach	Capability	Competence					✓	
Choice of environment: the nature scene and accompanying music can be changed according to user preference by flipping through a catalog	Customization	Motivation	Autonomy				✓		

Iterative Feedback

In the previously published qualitative study [53], a second round of interviews was conducted with 4 patients, during which they were shown the original bWell-D videos and updated videos that highlighted the added functionalities and features. All the participants reacted positively to these changes, felt that their input was considered, and thought that the initial version of bWell-D was improved.

In this study, the updated version was formally evaluated in participant trials with the original version in a single standardized session for initial assessment and as multi-session cognitive training for the improvements and gamification elements. In both the assessment and training sessions, the participants carried out the exercises for a total of 40 minutes. Of the 20 participants, 15 (75%) completed the assessment session, 4 (20%) opted to do the training sessions, and 1 (5%) tried the expanded features. The demographics and clinical characteristics of the participants of this study are summarized in Table 1 for reference. It was planned for each participant to interact with bWell for 1 assessment session, and those who opted in for training would have 8 more sessions. However, 25% (1/4) of the participants who elected to do the training sessions withdrew from the study after the second training session. User experience scales (GUESS and UES) were administered on training sessions 1, 4, and 8; therefore, 75% (3/4) of the participants rated the bWell-D expanded features 3 times, and 25% (1/4) of the participants rated bWell-D only once, for a total of 10 measures taken.

The results indicated that bWell-D was well tolerated by the tested participants in the initial and training sessions, with scores within the acceptable range in the literature (total SSQ score of <5 after the session) [57]. The SSQ scores, as summarized in Table 2, showed no significant increases (after vs before the session) in cybersickness symptoms in the initial assessment session or in the training sessions for the 20% (4/20) of the participants who opted in (the scores for the training sessions were measured as average SSQ score changes across all sessions). The GUESS, a 7-point Likert scale and whose results are summarized in Table 3, demonstrated high scores for user experience. In addition, for the improved version of bWell-D, an increase in score in 4 subscales (usability, enjoyment, audio aesthetics, and visual aesthetics) was observed. The effect size was large for usability and audio aesthetics, medium for visual aesthetics, and small for enjoyment [58]. While all subscales were higher for the improved bWell-D version, the small sample size and limited effect size do warrant further studies with larger populations. Although not statistically significant, engagement scores measured using the UES, a 5-point Likert scale with 4 subscales, remained high, with an average score of 15.07 (SD 1.53) for the initial session and 15.48 (SD 0.42) for the training sessions out of a possible maximum score of 20, as shown in Table 4.

Regarding the unstructured feedback during the training sessions, one participant mentioned that they did not feel the passage of time while playing the mole game and was surprised when their 10-minute game session was over. Another participant compared the mole game to the original arcade games they used to play years before, and their intense engagement with the game led to a noticeable perspiration and higher heartbeat levels. The theater game was mentioned as the most liked game by a participant, whereas a different person stated that they did not like the game in the initial session as it was hard to obtain a sense of achievement. One person specifically

XSL•FO
RenderX

enjoyed the relaxation scene (tent), in particular the ability to freely explore the environment.

The participant that tried 1 session of bWell-D with expanded features stated that moving within the office space was interesting and led to a better experience compared to the original version of bWell-D. Regarding the laboratory scene, they reacted positively to the addition of ingredient mixing, believed that it would also make it more difficult, and made additional suggestions to further increase the complexity. In the mole game, they recognized the addition of faces to the moles and the multiple-hit moles as variants. Furthermore, they approved of the option to have moles coming from the sides more than an expanding table as it involves peripheral vision. However, they also mentioned that these elements should be introduced after practicing the initial task to avoid possible frustration. They also noted that the tent scene was indeed a relaxing environment.

Table 1. Characteristics of the nonclinical participants who took part in this study (N=20).

Participant characteristics	Values
Age (y), mean (SD)	41.8 (12.58)
Age (y), median (IQR)	43.5 (34.25-53.25)
Sex (female), n (%)	12 (60)
BDI[a] (score range 0-63), mean (SD)	3.7 (5.55)
BDI (score range 0-63), median (IQR; SEM)	2 (0-4.25; 1.24)
STAI[b], mean (SD)	59.85 (20.3)
STAI, median (IQR; SEM)	50.5 (46.50-68.25; 4.54)

[a]BDI: Beck Depression Inventory.

[b]STAI: State-Trait Anxiety Inventory.

XSL•FO
RenderX

Table 2. Scores for before and after the virtual reality session for the initial assessment and training sessions (Simulator Sickness Questionnaire [SSQ] scores averaged over all sessions).[a]

	Assessment session—original version		2-sided paired *t* test *P* value	Training sessions—improved version		2-sided paired *t* test *P* value
	Postsession SSQ score, mean (SD)	Presession SSQ score, mean (SD)		Postsession SSQ score, mean (SD)	Presession SSQ score, mean (SD)	
Total score	4.533 (9.672)	5.733 (13.014)	.14	1.550 (1.909)	2.125 (2.604)	.27
General discomfort	0.200 (0.400)	0.200 (0.561)	>.99	0.050 (0.112)	0.100 (0.163)	.18
Fatigue	0.267 (0.573)	0.533 (0.915)	.16	0.100 (0.105)	0.375 (0.468)	.18
Boredom	0.133 (0.340)	0.400 (0.828)	.30	0.050 (0.112)	0.025 (0.056)	.37
Drowsiness	0.067 (0.249)	0.333 (0.900)	.16	0.100 (0.163)	0.175 (0.274)	.21
Headache	0.267 (0.680)	0.067 (0.258)	.19	0.000 (0.000)	0.025 (0.056)	.37
Eyestrain	0.333 (0.596)	0.333 (0.617)	>.99	0.200 (0.381)	0.225 (0.437)	.37
Difficulty focusing	0.200 (0.542)	0.333 (0.816)	.16	0.050 (0.112)	0.050 (0.068)	>.99
Salivation increase	0.067 (0.249)	0.133 (0.516)	.33	0.000 (0.000)	0.025 (0.056)	.37
Salivation decrease	0.067 (0.249)	0.067 (0.258)	>.99	0.000 (0.000)	0.025 (0.056)	.37
Sweating	0.333 (0.471)	0.200 (0.561)	.43	0.225 (0.335)	0.125 (0.280)	.59
Nausea	0.133 (0.340)	0.067 (0.258)	.33	0.000 (0.000)	0.000 (0.000)	>.99
Difficulty concentrating	0.267 (0.573)	0.333 (0.816)	.33	0.025 (0.056)	0.050 (0.112)	.70
Mental depression	0.067 (0.249)	0.200 (0.561)	.16	0.000 (0.000)	0.000 (0.000)	>.99
“Fullness of the head”	0.267 (0.680)	0.267 (0.594)	>.99	0.000 (0.000)	0.025 (0.056)	.37
Blurred vision	0.200 (0.542)	0.333 (0.724)	.16	0.225 (0.503)	0.250 (0.492)	.37
Dizziness—eyes open	0.200 (0.542)	0.133 (0.516)	.33	0.075 (0.168)	0.050 (0.112)	.37
Dizziness—eyes closed	0.067 (0.249)	0.200 (0.561)	.16	0.200 (0.447)	0.225 (0.503)	.37
Vertigo	0.133 (0.499)	0.133 (0.516)	>.99	0.000 (0.000)	0.000 (0.000)	>.99
Visual flashbacks	0.133 (0.499)	0.267 (0.799)	.16	0.000 (0.000)	0.000 (0.000)	>.99
Faintness	0.067 (0.249)	0.133 (0.516)	.33	0.000 (0.000)	0.000 (0.000)	>.99
Awareness of breathing	0.067 (0.249)	0.333 (0.816)	.10	0.000 (0.000)	0.000 (0.000)	>.99
Stomach awareness	0.133 (0.499)	0.067 (0.258)	.33	0.100 (0.224)	0.200 (0.447)	.37
Loss of appetite	0.133 (0.340)	0.200 (0.561)	.33	0.000 (0.000)	0.000 (0.000)	>.99
Increased appetite	0.467 (0.718)	0.067 (0.258)	.05	0.000 (0.000)	0.050 (0.112)	.37
Desire to move bowels	0.067 (0.249)	0.200 (0.561)	.33	0.100 (0.224)	0.100 (0.224)	>.99
Confusion	0.133 (0.499)	0.200 (0.775)	.33	0.000 (0.000)	0.000 (0.000)	>.99
Burping	0.067 (0.249)	0.000 (0.000)	.33	0.000 (0.000)	0.000 (0.000)	>.99
Vomiting	0.000 (0.000)	0.000 (0.000)	>.99	0.000 (0.000)	0.000 (0.000)	>.99
Other	0.000 (0.000)	0.000 (0.000)	>.99	0.050 (0.112)	0.025 (0.056)	.37

[a]Total SSQ scores were computed based on the study by Bouchard et al [59]. Comparing post- versus presession scores, it was observed that there were no significant increases in SSQ scores in the initial session (15 participants) and the training sessions (4 participants).

XSL•FO
RenderX

Table 3. Comparison of Game User Experience Satisfaction Scale results across 15 participants for the assessment session (original version) and across 4 participants who opted for the training sessions (improved version), averaging 3 recordings per participant.[a]

	Usability/playability	Play engrossment	Enjoyment	Audio aesthetics	Personal gratification	Visual aesthetics
Assessment session (original version), mean (SD)	5.399 (0.573)	4.925 (0.887)	5.363 (1.050)	4.783 (1.372)	5.544 (0.717)	5.422 (1.080)
Multiple training sessions (improved version), mean (SD)	6.053 (0.500)	4.965 (0.647)	5.583 (0.514)	5.708 (0.241)	5.528 (0.718)	6.028 (0.048)
Effect size (Cohen *d*)	1.216	0.052	0.266	0.939	−0.022	0.793

[a]All scores were, on average, higher in the training sessions than in the assessment session. This trend of high training means relative to assessment means is captured in the effect size, with the *usability/playability* and *audio aesthetics* subscales having large effect sizes (<0.80) [58].

Table 4. Comparison of User Engagement Scale results across 15 participants for the assessment session (original version) and across 4 participants who opted for the training sessions (improved version), averaging 3 recordings per participant.[a]

	Focused attention subscale	Perceived usability subscale	Esthetic appeal subscale	Reward factor subscale	Overall engagement score (out of 20)
Assessment session (original version), mean (SD)	3.486 (0.438)	3.667 (0.452)	3.907 (0.696)	4.007 (0.465)	15.07 (1.526)
Training sessions (improved version), mean (SD)	3.345 (0.238)	3.938 (0.172)	4.200 (0.452)	4.000 (0.141)	15.483 (0.415)

[a]High engagement was observed for all sessions (overall engagement score of >15 out of 20).

Discussion

Principal Findings

Overview

In this paper, we propose a 2-pronged approach using end-user feedback for the co-design of content as well as in the application of gamification. This process considers behavior change theories to structure feedback obtained and provide insights in terms of drivers that potentially promote engagement and enhance intervention efficacy. Ultimately, the aim was to improve the compliance rate of a VR cognitive remediation intervention for depression.

Several interesting findings emerged with implications for technology designers involved in developing interventions for people with depression, as well as for the implementation of such interventions. Optimizing engagement is critical for the success of digital interventions for any patient cohort. However, it is an especially vital consideration for interventions targeting depression, which is defined by decreased ability to experience enjoyment or interest, decreased motivation, and a lack of self-efficacy [6]. Grounded in feedback obtained from persons with lived experience, this analysis provides useful guidance for developing intrinsically engaging, motivating digital interventions that can overcome these hallmark features of depression.

Identification of Gamification Elements

The classification of the gamification elements added to bWell-D indicated that feedback from end users centered on using gamification to add *personalization* and *customization* options, *artificialchallenges*, and *artificialassistance*. In defining *customization* versus *personalization* elements, the former consist of settings controlled by the clinician or user, whereas the latter are settings automatically imposed by bWell-D. In defining *personalization* versus *artificialchallenge* elements, the former are adaptations applied by the software in response to user characteristics and performance, whereas the latter are applied regardless of the user. Inclusion of these elements indicates that users expressed a need for challenges, tasks, and settings that are generic, as well as challenges and options that either are automatically tailored to the individual or the individual themselves chooses.

In an example of *customization*, the tent scene provides a variety of nature scenes that the user can choose from. It is worthwhile to note that, although it may seem interesting to develop additional virtual environments to reach more user-specific contexts, this requires careful consideration. For instance, one specific piece of feedback received from participants was that, although the office environment was created to represent a real-world setting, it still did not look like their workplace. Rather than creating a vast catalog of representative workplace settings, it was instead decided to include *bridge exercises* to provide the opportunity for the user to reflect on their specific context. Hence, it may be good practice to identify and consider the processes for which the intervention is intended to work and identify different potential options that can be implemented to support these processes. This would be of particular benefit in instances requiring substantial developmental effort.

Looking at the types of gamification elements included overall, while *personalization* and *customization* are among the most commonly observed gamification elements [48,49], the feedback obtained from end users also resulted in the inclusion of gamification elements in bWell-D that are less commonly observed, such as *artificialassistance*, *exploratory open-world approaches*, and *artificialchallenge*.

XSL•FO
RenderX

COM-B Model Insights

On the basis of the results of the classification in terms of the driver of behavior according to the COM-B model, it is clear that the feedback received from our clinical collaborators focused on implementing features related to *capability* (19/28, 68% of the features). Interestingly, most of the gamification elements that were implemented centered on adding variety to the VR experiences (choice of environment, difficulty level adjustment, and types of stimuli and distractors). These features were implemented so that they can be turned on or off to promote engagement over time to enable new and more challenging features and game modes as the participant progresses throughout the intervention. This keeps the experience fresh, with the goal to help users stay the course of the remediation program. Evidence in support of this goal was provided by the UES results, where it was observed that participants maintained a high user engagement throughout the multiple sessions of the 8-week training program. Another advantage is that bWell-D offers a great deal of flexibility, permitting clinicians to adjust the configuration to the specific needs of different and potentially vulnerable populations. On that matter, Koivisto and Malik [60] present the importance of developing features that target a specific population and note that very few gamification studies focus on needs specific to older adults.

The elements that focused on *motivation* generally related to providing exercise variants based on the identified theme of leveraging the multisensorial affordances of VR for more fun and engaging experiences. For example, audio versions of the tasks were included as an alternative to visual presentation of instructions. It is likely that such elements contributed to the increase in GUESS scores for *usability/playability* and *audio and visual aesthetics* in the improved version of bWell-D.

Of the modifications to bWell-D, only two related to the *opportunity* aspect of the COM-B model: (1) providing the option for use at home and (2) inclusion of tutorials. From the feedback received, it was made clear that compliance and adherence depend not only on user engagement with the exercises but also on other factors such as ease of use and accessibility. Making bWell-D available for at-home use and accompanying it with tutorials in support has the potential to increase compliance by reducing the barriers to access it. Users can perform the intervention from the comfort of their own home and on their own schedule rather than having to travel to a clinic at a specific appointment time. The improved access is especially important for people with depression, who may find it *difficult to get out of bed*, and also extends the reach to diverse patient populations (eg, older adults with decreased mobility and those in remote or rural settings). While the inclusion of tutorials is a form of *artificialassistance*, providing the option for use at home was not classified as an element of gamification. It was observed that the typical gamification elements that target *opportunity* (cues and reminders or social mechanics to prompt users) were not included in the modifications. While the application of such gamification elements can be readily integrated into mobile interventions, in the case of a VR intervention, this is less straightforward. However, future iterations of bWell-D development will nonetheless aim to incorporate these. A potential and interesting avenue could be the development of a companion mobile app to prompt users at opportune moments by sending notifications.

Promoting Intrinsic Engagement

Efforts were intentionally made to choose modifications to bWell-D that enhance intrinsic motivation. Of the changes implemented, only 1 did not address the 3 fundamental needs outlined in self-determination theory, which are essential for psychological growth. This exception was the score streak indicator in the mole scene, which offers extrinsic motivation. Of the modifications made, most addressed the need for *competence*, which is not surprising given the context of cognitive remediation. Many of these modifications were related to increasing user engagement through the introduction of new features, adding multisensorial elements, or varying the difficulty of the exercise. The need for *autonomy* was addressed by 4 of the modifications. These were primarily related to providing user-selectable options and menus for customization and enabling at-home use. In total, bWell-D offers 3 types of gamified menus that users can access in different ways: by clicking on option tabs, by viewing their wrist in the game, and by flipping through the pages of a book provided in the scene. Each menu provides different types of information to the user.

Analysis of the results indicated that none of the modifications addressed the third need according to self-determination theory—*socialconnection*. Currently, social elements are only implemented in the exercises via nonplayable characters. From the taxonomy, it was observed that there are quite a few gamification elements related to social connection, specifically *socialcomparison*, *socialcompetition*, *socialcooperation*, and *social networking*. Given that the integration of social mechanics favors the *opportunity* domain of the COM-B model and that self-determination theory states that a sense of connection with others promotes self-actualization, this aspect will be prioritized in the next iterations of bWell-D. However, feedback from end users and domain experts is required to identify which of these elements are most appropriate for mental health domains.

Limitations and Next Steps

A limitation in our previously published qualitative study is that, due to COVID-19–related restrictions, users who provided feedback did not experience the VR implementation firsthand as originally intended but instead watched videos of the implementation. This limited the type of feedback that the end users could provide on aspects such as game mechanics and usability. This study addressed these limitations by recruiting healthy participants to trial both the original and improved versions of bWell-D. Both structured and unstructured feedback was collected to evaluate the impact of the gamification elements. Another limitation is that the implemented features have not yet been tested on a clinical population. Therefore, it is not currently possible to evaluate whether the changes to our software have the expected impact on user engagement or intervention efficacy in this clinical population. We plan to obtain this feedback and measure adherence in upcoming 8-week bWell-D remediation trials with people with MDD. It is hoped that the modifications based on end-user feedback will lead to good compliance with the program. An additional limitation is

XSL•FO
RenderX

the relatively small sample size of participants who completed the initial user trials and provided feedback on the optimized features of bWell-D.

Current literature exposes possibilities for gamification that are yet to be explored and evaluated. For instance, *randomness*, *artificial challenge*, *artificial assistance*, *exploratory or open-world approaches*, and *socialcooperation* are reported to be underused for applications in the mental health domain, and further research is required to determine how these elements can best be used, if at all [49]. The mechanisms through which gamification impacts mental health in the real world and which gamification elements are associated with more benefits should also be further explored. bWell-D has integrated a number of these underused elements based on the iterative feedback obtained, and with the iterative development process in place, it will be feasible to take on such investigations.

Conclusions

The bWell-D prototype was developed following best practices in VR clinical application design using iterative feedback from both patients and health care providers, resulting in the development of a VR cognitive remediation tool that was demonstrated to be acceptable and provided a favorable user experience with sustained engagement. We presented a process for the application of gamification that includes characterizing modifications in a standardized way and identifying the underlying mechanisms that are targeted. This classification revealed that the gamification elements implemented in the prototype included some of those less commonly observed in other applications and tended to address intrinsic motivations. In addition, the modifications made were found to address all 3 drivers of behavior identified in the COM-B model. As such, the latest iteration of bWell-D appears to be well poised to promote user uptake and compliance. In future developments, we plan to incorporate elements that are currently missing to ensure a balanced representation of the factors identified from behavioral theories.

Acknowledgments

This project was supported in part by funding from the National Research Council Canada's Canadian Digital Technology Supercluster support program (CDTS-103). TC has received grant funding from the Michael Smith Foundation for Health Research and the National Research Council Canada.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Video of the bWell interface and laboratory scene of the original prototype version.
[MP4 File (MP4 Video), 100868 KB - games_v13i1e59514_app1.mp4]

References

1. Depressive disorder (depression). World Health Organization. 2023 Mar 31. URL: https://www.who.int/news-room/fact-sheets/detail/depression [accessed 2025-01-22]
2. Woo YS, Rosenblat JD, Kakar R, Bahk WM, McIntyre RS. Cognitive deficits as a mediator of poor occupational function in remitted major depressive disorder patients. Clin Psychopharmacol Neurosci 2016 Feb 29;14(1):1-16 [FREE Full text] [doi: 10.9758/cpn.2016.14.1.1] [Medline: 26792035]
3. Sheehan DV, Harnett-Sheehan K, Spann ME, Thompson HF, Prakash A. Assessing remission in major depressive disorder and generalized anxiety disorder clinical trials with the discan metric of the Sheehan disability scale. Int Clin Psychopharmacol 2011 Mar;26(2):75-83. [doi: 10.1097/YIC.0b013e328341bb5f] [Medline: 21102344]
4. Bortolato B, Miskowiak KW, Köhler CA, Maes M, Fernandes BS, Berk M, et al. Cognitive remission: a novel objective for the treatment of major depression? BMC Med 2016 Jan 22;14:9 [FREE Full text] [doi: 10.1186/s12916-016-0560-3] [Medline: 26801406]
5. Evans VC, Iverson GL, Yatham LN, Lam RW. The relationship between neurocognitive and psychosocial functioning in major depressive disorder: a systematic review. J Clin Psychiatry 2014 Dec 24;75(12):1359-1370. [doi: 10.4088/jcp.13r08939]
6. Chakrabarty T, Hadjipavlou G, Lam RW. Cognitive dysfunction in major depressive disorder: assessment, impact, and management. Focus (Am Psychiatr Publ) 2016 Apr;14(2):194-206 [FREE Full text] [doi: 10.1176/appi.focus.20150043] [Medline: 31975803]
7. Matcham F, Simblett SK, Leightley D, Dalby M, Siddi S, Haro JM, et al. The association between persistent cognitive difficulties and depression and functional outcomes in people with major depressive disorder. Psychol Med 2022 Dec 13;53(13):6334-6344. [doi: 10.1017/s0033291722003671]
8. Medalia A, Choi J. Cognitive remediation in schizophrenia. Neuropsychol Rev 2009 Sep;19(3):353-364. [doi: 10.1007/s11065-009-9097-y] [Medline: 19444614]

XSL•FO
RenderX

9. Vita A, Barlati S, Ceraso A, Nibbio G, Ariu C, Deste G, et al. Effectiveness, core elements, and moderators of response of cognitive remediation for schizophrenia: a systematic review and meta-analysis of randomized clinical trials. JAMA Psychiatry 2021 Aug 01;78(8):848-858 [FREE Full text] [doi: 10.1001/jamapsychiatry.2021.0620] [Medline: 33877289]
10. Keshavan MS, Vinogradov S, Rumsey J, Sherrill J, Wagner A. Cognitive training in mental disorders: update and future directions. Am J Psychiatry 2014 May;171(5):510-522 [FREE Full text] [doi: 10.1176/appi.ajp.2013.13081075] [Medline: 24700194]
11. Legemaat AM, Semkovska M, Brouwer M, Geurtsen GJ, Burger H, Denys D, et al. Effectiveness of cognitive remediation in depression: a meta-analysis. Psychol Med 2021 Apr 14;52(16):4146-4161. [doi: 10.1017/s0033291721001100]
12. Thérond A, Pezzoli P, Abbas M, Howard A, Bowie CR, Guimond S. The efficacy of cognitive remediation in depression: a systematic literature review and meta-analysis. J Affect Disord 2021 Apr 01;284:238-246. [doi: 10.1016/j.jad.2021.02.009] [Medline: 33631438]
13. Slater M. Immersion and the illusion of presence in virtual reality. Br J Psychol 2018 Aug 21;109(3):431-433. [doi: 10.1111/bjop.12305] [Medline: 29781508]
14. Jahn FS, Skovbye M, Obenhausen K, Jespersen AE, Miskowiak KW. Cognitive training with fully immersive virtual reality in patients with neurological and psychiatric disorders: a systematic review of randomized controlled trials. Psychiatry Res 2021 Jun;300:113928 [FREE Full text] [doi: 10.1016/j.psychres.2021.113928] [Medline: 33857847]
15. Wiebe A, Kannen K, Selaskowski B, Mehren A, Thöne AK, Pramme L, et al. Virtual reality in the diagnostic and therapy for mental disorders: a systematic review. Clin Psychol Rev 2022 Dec;98:102213 [FREE Full text] [doi: 10.1016/j.cpr.2022.102213] [Medline: 36356351]
16. Voinescu A, Petrini K, Stanton Fraser D, Lazarovicz RA, Papavă I, Fodor LA, et al. The effectiveness of a virtual reality attention task to predict depression and anxiety in comparison with current clinical measures. Virtual Real 2021 Apr 17;27(1):119-140. [doi: 10.1007/s10055-021-00520-7]
17. Schroeder AH, Bogie BJ, Rahman TT, Thérond A, Matheson H, Guimond S. Feasibility and efficacy of virtual reality interventions to improve psychosocial functioning in psychosis: systematic review. JMIR Ment Health 2022 Feb 18;9(2):e28502 [FREE Full text] [doi: 10.2196/28502] [Medline: 35179501]
18. Rizzo AA, Buckwalter JG, Bowerly T, Van Der Zaag C, Humphrey L, Neumann U, et al. The virtual classroom: a virtual reality environment for the assessment and rehabilitation of attention deficits. CyberPsychol Behav 2000 Jun;3(3):483-499. [doi: 10.1089/10949310050078940]
19. Parsons TD, Bowerly T, Buckwalter JG, Rizzo AA. A controlled clinical comparison of attention performance in children with ADHD in a virtual reality classroom compared to standard neuropsychological methods. Child Neuropsychol 2007 Jul 11;13(4):363-381. [doi: 10.1080/13825580600943473] [Medline: 17564852]
20. Law AS, Logie RH, Pearson DG. The impact of secondary tasks on multi-tasking in a virtual environment. Acta Psychol (Amst) 2006 May;122(1):27-44. [doi: 10.1016/j.actpsy.2005.09.002] [Medline: 16307711]
21. Kourtesis P, MacPherson SE. An ecologically valid examination of event-based and time-based prospective memory using immersive virtual reality: the influence of attention, memory, and executive function processes on real-world prospective memory. Neuropsychol Rehabil 2023 Mar;33(2):255-280. [doi: 10.1080/09602011.2021.2008983] [Medline: 34856886]
22. Zhu K, Zhang Q, He B, Huang M, Lin R, Li H. Immersive virtual reality-based cognitive intervention for the improvement of cognitive function, depression, and perceived stress in older adults with mild cognitive impairment and mild dementia: pilot pre-post study. JMIR Serious Games 2022 Feb 21;10(1):e32117 [FREE Full text] [doi: 10.2196/32117] [Medline: 35188466]
23. Pedroli E, La Paglia F, Cipresso P, La Cascia C, Riva G, La Barbera D. A computational approach for the assessment of executive functions in patients with obsessive-compulsive disorder. J Clin Med 2019 Nov 14;8(11):1975 [FREE Full text] [doi: 10.3390/jcm8111975] [Medline: 31739514]
24. McLean E, Cornwell MA, Bender HA, Sacks-Zimmerman A, Mandelbaum S, Koay JM, et al. Innovations in neuropsychology: future applications in neurosurgical patient care. World Neurosurg 2023 Feb;170:286-295. [doi: 10.1016/j.wneu.2022.09.103] [Medline: 36782427]
25. McGeorge P, Phillips LH, Crawford JR, Garden SE, Sala SD, Milne AB, et al. Using virtual environments in the assessment of executive dysfunction. Presence (Camb) 2001 Aug;10(4):375-383. [doi: 10.1162/1054746011470235]
26. Thapa N, Park HJ, Yang JG, Son H, Jang M, Lee J, et al. The effect of a virtual reality-based intervention program on cognition in older adults with mild cognitive impairment: a randomized control trial. J Clin Med 2020 Apr 29;9(5):1283 [FREE Full text] [doi: 10.3390/jcm9051283] [Medline: 32365533]
27. Parsons TD, Rizzo AA. Initial validation of a virtual environment for assessment of memory functioning: virtual reality cognitive performance assessment test. Cyberpsychol Behav 2008 Feb;11(1):17-25. [doi: 10.1089/cpb.2007.9934] [Medline: 18275308]
28. Knight RG, Titov N. Use of virtual reality tasks to assess prospective memory: applicability and evidence. Brain Impairment 2012 Feb 21;10(1):3-13. [doi: 10.1375/brim.10.1.3]
29. Faria AL, Andrade A, Soares L, I Badia SB. Benefits of virtual reality based cognitive rehabilitation through simulated activities of daily living: a randomized controlled trial with stroke patients. J Neuroeng Rehabil 2016 Nov 02;13(1):96 [FREE Full text] [doi: 10.1186/s12984-016-0204-z] [Medline: 27806718]

XSL•FO
RenderX

30. Chan KC, Hui CL, Suen YN, Lee EH, Chang WC, Chan SK, et al. Application of immersive virtual reality for assessment and intervention in psychosis: a systematic review. Brain Sci 2023 Mar 10;13(3):471 [FREE Full text] [doi: 10.3390/brainsci13030471] [Medline: 36979281]
31. Astur RS, Tropp J, Sava S, Constable RT, Markus EJ. Sex differences and correlations in a virtual Morris water task, a virtual radial arm maze, and mental rotation. Behav Brain Res 2004 May 05;151(1-2):103-115. [doi: 10.1016/j.bbr.2003.08.024] [Medline: 15084426]
32. Maeng S, Hong JP, Kim WH, Kim H, Cho SE, Kang JM, et al. Effects of virtual reality-based cognitive training in the elderly with and without mild cognitive impairment. Psychiatry Investig 2021 Jul;18(7):619-627 [FREE Full text] [doi: 10.30773/pi.2020.0446] [Medline: 34265200]
33. Faria AL, Pinho MS, Bermúdez I Badia S. A comparison of two personalization and adaptive cognitive rehabilitation approaches: a randomized controlled trial with chronic stroke patients. J Neuroeng Rehabil 2020 Jun 16;17(1):78 [FREE Full text] [doi: 10.1186/s12984-020-00691-5] [Medline: 32546251]
34. Beck L, Wolter M, Mungard NF, Vohn R, Staedtgen M, Kuhlen T, et al. Evaluation of spatial processing in virtual reality using functional magnetic resonance imaging (FMRI). Cyberpsychol Behav Soc Netw 2010 Apr;13(2):211-215. [doi: 10.1089/cyber.2008.0343] [Medline: 20528281]
35. Gagnon Shaigetz V, Proulx C, Cabral A, Choudhury N, Hewko M, Kohlenberg E, et al. An immersive and interactive platform for cognitive assessment and rehabilitation (bWell): design and iterative development process. JMIR Rehabil Assist Technol 2021 Nov 03;8(4):e26629 [FREE Full text] [doi: 10.2196/26629] [Medline: 34730536]
36. Proulx C, Cabral A, Choudhury N, Debergue P. Acceptability study of a novel immersive cognitive care platform for remediation of cognitive deficits. Annu Rev CyberTherapy Telemed 2018;16:43-49.
37. Cabral A, Choudhury N, Proulx C, Harmouche R, Kohlenberg E, Debergue P. Evaluation of cybersickness in a passive walking virtual reality cognitive exercise. Annu Rev CyberTherapy Telemed 2019;17:59-64.
38. De Luca G, Choudhury N, Pagiatakis C, Laroche D. A multi-procedural virtual reality simulator for orthopaedic training. In: Proceedings of the 11th International Conference on Virtual, Augmented and Mixed Reality. Applications and Case Studies. 2019 Presented at: VAMR 2019; July 26-31, 2019; Orlando, FL. [doi: 10.1007/978-3-030-21565-1_17]
39. De Croon R, Geuens J, Verbert K, Abeele VV. A systematic review of the effect of gamification on adherence across disciplines. In: Proceedings of the Third International Conference on HCI in Games: Experience Design and Game Mechanics. 2021 Presented at: HCII 2021; July 24-29, 2021; Virtual Event. [doi: 10.1007/978-3-030-77277-2_14]
40. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ 2008 Sep 29;337:a1655 [FREE Full text] [doi: 10.1136/bmj.a1655] [Medline: 18824488]
41. Choudhury N, Gélinas-Phaneuf N, Delorme S, Del Maestro R. Fundamentals of neurosurgery: virtual reality tasks for training and evaluation of technical skills. World Neurosurg 2013 Nov;80(5):e9-19. [doi: 10.1016/j.wneu.2012.08.022] [Medline: 23178917]
42. Wykes T, Spaulding WD. Thinking about the future cognitive remediation therapy--what works and could we do better? Schizophr Bull 2011 Sep 22;37 Suppl 2(Suppl 2):S80-S90 [FREE Full text] [doi: 10.1093/schbul/sbr064] [Medline: 21860051]
43. Birckhead B, Khalil C, Liu X, Conovitz S, Rizzo A, Danovitch I, et al. Recommendations for methodology of virtual reality clinical trials in health care by an international working group: iterative study. JMIR Ment Health 2019 Jan 31;6(1):e11973 [FREE Full text] [doi: 10.2196/11973] [Medline: 30702436]
44. O'Cathain A, Croot L, Duncan E, Rousseau N, Sworn K, Turner KM, et al. Guidance on how to develop complex interventions to improve health and healthcare. BMJ Open 2019 Aug 15;9(8):e029954 [FREE Full text] [doi: 10.1136/bmjopen-2019-029954] [Medline: 31420394]
45. Dias LP, Barbosa JL, Vianna HD. Gamification and serious games in depression care: a systematic mapping study. Telemat Inform 2018 Apr;35(1):213-224. [doi: 10.1016/j.tele.2017.11.002]
46. Johnson D, Deterding S, Kuhn KA, Staneva A, Stoyanov S, Hides L. Gamification for health and wellbeing: a systematic review of the literature. Internet Interv 2016 Nov 02;6:89-106 [FREE Full text] [doi: 10.1016/j.invent.2016.10.002] [Medline: 30135818]
47. Lumsden J, Edwards EA, Lawrence NS, Coyle D, Munafò MR. Gamification of cognitive assessment and cognitive training: a systematic review of applications and efficacy. JMIR Serious Games 2016 Jul 15;4(2):e11 [FREE Full text] [doi: 10.2196/games.5888] [Medline: 27421244]
48. Brown M, O'Neill N, van Woerden H, Eslambolchilar P, Jones M, John A. Gamification and adherence to web-based mental health interventions: a systematic review. JMIR Ment Health 2016 Aug 24;3(3):e39 [FREE Full text] [doi: 10.2196/mental.5710] [Medline: 27558893]
49. Cheng VW, Davenport T, Johnson D, Vella K, Hickie IB. Gamification in apps and technologies for improving mental health and well-being: systematic review. JMIR Ment Health 2019 Jun 26;6(6):e13717 [FREE Full text] [doi: 10.2196/13717] [Medline: 31244479]

XSL•FO
RenderX

50. Baranowski T, Anderson C, Carmack C. Mediating variable framework in physical activity interventions. How are we doing? How might we do better? Am J Prev Med 1998 Nov;15(4):266-297. [doi: 10.1016/s0749-3797(98)00080-4] [Medline: 9838973]
51. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 Apr 23;6(1):42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
52. Manninen M, Dishman R, Hwang Y, Magrum E, Deng Y, Yli-Piipari S. Self-determination theory based instructional interventions and motivational regulations in organized physical activity: a systematic review and multivariate meta-analysis. Psychol Sport Exerc 2022 Sep;62:102248. [doi: 10.1016/j.psychsport.2022.102248]
53. Hernandez Hernandez ME, Michalak E, Choudhury N, Hewko M, Torres I, Menon M, et al. Co-design of a virtual reality cognitive remediation program for depression (bWell-D) with patient end users and clinicians: qualitative interview study among patients and clinicians. JMIR Serious Games 2023 Apr 07;11:e43904 [FREE Full text] [doi: 10.2196/43904] [Medline: 37027183]
54. Kennedy RS, Lane NE, Berbaum KS, Lilienthal MG. Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. Int J Aviat Psychol 1993;3(3):203-220. [doi: 10.1207/s15327108ijap0303_3]
55. O'Brien HL, Cairns P, Hall M. A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form. Int J Hum Comput Stud 2018 Apr;112:28-39. [doi: 10.1016/j.ijhcs.2018.01.004]
56. Keebler JR, Shelstad WJ, Smith DC, Chaparro BS, Phan MH. Validation of the GUESS-18: a short version of the game user experience satisfaction scale (GUESS). J Usability Stud 2020 Nov 1;16(1):49-62 [FREE Full text]
57. Bouchard S, St-Jacques J, Renaud P, Wiederhold BK. Side effects of immersions in virtual reality for people suffering from anxiety disorders. J Cyber Ther Rehabil 2009 Jan;2(2):127-137.
58. Calin-Jageman RJ. The new statistics for neuroscience majors: thinking in effect sizes. J Undergrad Neurosci Educ 2018 Jun 15;16(2):E21-E25 [FREE Full text] [Medline: 30057503]
59. Bouchard S, Robillard G, Renaud P. Revising the factor structure of the simulator sickness questionnaire. Annu Rev CyberTherapy Telemed 2007 Jan;5:128-137.
60. Koivisto J, Malik A. Gamification for older adults: a systematic literature review. Gerontologist 2021 Sep 13;61(7):e360-e372 [FREE Full text] [doi: 10.1093/geront/gnaa047] [Medline: 32530026]

Abbreviations

CCN: cognitive care network
COM-B: capability, opportunity, and motivation–behavior
GUESS: Game User Experience Satisfaction Scale
MDD: major depressive disorder
NRC: National Research Council Canada
SSQ: Simulator Sickness Questionnaire
UES: User Engagement Scale
VR: virtual reality

Edited by A Bustillo; submitted 15.04.24; peer-reviewed by T Davidson, SA Triantafyllou, C Abbatantuono; comments to author 11.09.24; revised version received 20.12.24; accepted 08.01.25; published 31.03.25.

Please cite as:
Hewko M, Gagnon Shaigetz V, Smith MS, Kohlenberg E, Ahmadi P, Hernandez Hernandez ME, Proulx C, Cabral A, Segado M, Chakrabarty T, Choudhury N
Considering Theory-Based Gamification in the Co-Design and Development of a Virtual Reality Cognitive Remediation Intervention for Depression (bWell-D): Mixed Methods Study
JMIR Serious Games 2025;13:e59514
URL: https://games.jmir.org/2025/1/e59514
doi:10.2196/59514
PMID:

©Mark Hewko, Vincent Gagnon Shaigetz, Michael S Smith, Elicia Kohlenberg, Pooria Ahmadi, Maria Elena Hernandez Hernandez, Catherine Proulx, Anne Cabral, Melanie Segado, Trisha Chakrabarty, Nusrat Choudhury. Originally published in JMIR Serious Games (https://games.jmir.org), 31.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic

XSL•FO
RenderX

information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Research Hotspots and Trends of Virtual Reality Intervention for Stroke: Bibliometric Analysis

Yixin Wei[1,2*], MPT; Yuan Chen[3*], PhD; Runting Ma[1,2], MPT; Yitong Qiu[1,2], MPT; Wei Su[1,2], MPT; Li Zhang[1,2], MPT, DPT; Qiang Gao[1,2], Prof Dr, PhD

[1]Rehabilitation Medicine Center and Institute of Rehabilitation Medicine, West China Hospital, Sichuan University, 37 Guoxue Lane, Chengdu, China

[2]Key Laboratory of Rehabilitation Medicine in Sichuan Province, West China Hospital, Sichuan University, Chengdu, China

[3]Department of Pharmacy, Evidence-Based Pharmacy Center, West China Second Hospital, Sichuan University, Chengdu, China

[*]these authors contributed equally

Corresponding Author:
Qiang Gao, Prof Dr, PhD
Rehabilitation Medicine Center and Institute of Rehabilitation Medicine, West China Hospital, Sichuan University, 37 Guoxue Lane, Chengdu, China

Abstract

Background: Virtual reality (VR) is a rapidly developing technology that has gained significant traction in the treatment and rehabilitation of individuals with stroke. Research on VR-based stroke treatment has garnered increasing attention.

Objective: The aim of this study is to present a bibliometric analysis of VR for stroke studies to identify the application status, research hotspots, and emerging trends and guide future scientific research.

Methods: We included studies and reviews on the topic of VR-based stroke treatment and rehabilitation from 1999 to 2023 were retrieved from Web of Science Core Collection database. Citespace 6.3.1 and VOSviewer 1.6.20 software was used for the visual analysis of publications, institutions, authors, journals, citations, and Scimago Graphica software was used for the geographic visualization of published countries or regions.

Results: Our study analyzed 1171 papers on VR-based stroke rehabilitation published between 1999 and 2023, revealing a gradual increase in annual publications over the past 2 decades, peaking at 154 in 2022. North America and Western Europe were identified as major contributors, with significant input from their institutions, researchers, and publications. The Journal of NeuroEngineering and Rehabilitation emerged as the leading journal in this field, while Calabrò Rocco Salvatore was recognized as the most prolific author, focusing on the neurophysiological impacts of VR on patients with stroke. Keywords with notable citation bursts, such as “environment,” “trial,” “arm,” and “motor learning,” highlighted the core research themes in this domain.

Conclusions: Our study provides valuable insights into the current research hotspots and emerging trends in VR-based stroke treatment and rehabilitation. Current research primarily focuses on evaluating the effectiveness of VR in improving upper limb function and balance in patients with stroke. Future directions are shifting towards integrating VR with rehabilitation techniques, such as physiotherapy and occupational therapy, while advancements in VR technology continue to garner increasing attention.

(JMIR Serious Games 2025;13:e65993) doi:10.2196/65993

KEYWORDS

virtual reality; stroke; bibliometric analysis; hotspots; research trends

Introduction

Stroke is a significant contributor to mortality and morbidity globally, ranking as the leading cause of permanent acquired disability [1]. In 2016, there were 80.1 million cases of stroke worldwide, and this figure is likely to increase further in the future due to the growth of the aging population [2]. A significant proportion of patients who have suffered from stroke report a range of impairments to motor, sensory, cognitive, and visual function [3]. This results in a considerable number of stroke survivors being unable to perform activities of daily living, participate fully in society, or interact socially in an effective manner [3]. The accumulating evidence from research indicates that rehabilitation is the most effective means of reducing disability and a vital component of stroke management [4,5]. The conventional neurorehabilitation approach, which involves high-intensity, repetitive, and task-specific practice, has demonstrated a curative effect on enhancing the functional abilities of individuals with stroke [6,7]. However, these rehabilitation techniques are costly and inconvenient due to their reliance on professional expertise and substantial resource requirements, particularly in the context of the global pandemic of COVID-19 [8,9]. It follows that there is a pressing

XSL•FO
RenderX

requirement for a novel and efficacious approach to stroke management.

Virtual reality (VR) can be defined as an interactive, computer-generated environment that simulates the real world [10]. This enables interaction with VR images and sounds, which prompt user responses and deliver immediate performance feedback. This interaction is facilitated through a variety of devices, including computers, mobile screens, and head-mounted displays, which collectively enhance user engagement and experience. With the continuous advancement of health technologies, VR is anticipated to play a more prominent role in clinical rehabilitation [11,12]. The application of VR technology in the rehabilitation of a diverse range of diseases that emerged in the 1990s has become increasingly prevalent [10,13-15]. Over the past 60 years, the technology has undergone significant transformations. In recent years, VR has emerged as a vital tool for both assessment and intervention in clinical rehabilitation, driven by extensive research and a reduction in the costs of VR technologies [10,16,17].

VR represents a novel technology with the potential to be used with greater frequency in clinical stroke rehabilitation [6,16]. Common examples of VR on stroke neurorehabilitation are as follows: (1) manipulating VR objects exercises for upper limb motor function [18,19], (2) VR-based training system for regaining gait, balance and daily function [20], and (3) cognitive-motor exergame training based on VR program for cognitive rehabilitation [17,20]. The therapeutic benefits of VR in stroke rehabilitation are attributed to mechanisms such as task-specific repetition, strategic feedback, and embodied simulation. In addition, the enriching environmental context provided by VR therapy has been substantiated in numerous studies, highlighting its efficacy in promoting neuroplasticity and functional recovery.

In light of the growing demand for stroke rehabilitation, the increasing application of VR therapy for stroke has positioned it as a significant area of research [16,18,20,21]. Guo et al conducted a 25-year visualization analysis of VR for neurorehabilitation, which indicated that VR for stroke is a promising avenue for future research [22]. However, existing studies often suffer from complexity, lack of structure, and an incomplete understanding of the underlying therapeutic mechanisms. These limitations, coupled with the absence of an up-to-date and comprehensive bibliometric analysis, underscore the need for a systematic investigation into the scientific landscape of VR for stroke rehabilitation. Bibliometric analysis, as a cross-disciplinary field of quantitative analysis, applies mathematical and statistical methods to examine knowledge carriers. It is a widely used methodology for identifying the developmental trajectory of specific research areas [23]. By analyzing citation networks, co-citation patterns, and the distribution of academic output across regions and institutions, bibliometric analysis allows researchers to quickly grasp current research hotspots, identify influential studies, and discern emerging trends [24]. In this context, bibliometric analysis provides a valuable tool for synthesizing large volumes of literature and offering a structured, data-driven perspective on the research landscape. This bibliometric study aims to provide a comprehensive overview of the global scientific output on the use of VR for stroke rehabilitation, spanning from its inception to 2023. We used Citespace and VOSviewer software to map the intellectual structure of the field and visualize citation patterns, co-citations, and the distribution of research across countries, authors, and journals.

Methods

Data Collection

Web of Science (WOS) database, a renowned scientific data services platform developed by Clarivate was used to retrieve the Clarivate Journal Impact Factor (IF) for the last 5 years [25]. Meanwhile, publications with related themes from the inception (1900) to 2023 were searched from the Science Citation Index Expanded (SCIE) of the Web of Science Core Collection (WOSCC) database. SCIE is a subdatabase of WOSCC, which consists of global journals of basic science research, covering neuroscience and medical research related to the theme of this study, “VR for stroke.”

The data were obtained on March 30, 2024, from SCIE. To obtain documents explicitly using the concerning terms we performed a topical search with the query TS=(((“Virtual reality”) OR (VR)))) AND TS=(((stroke* OR brain vascular accident* OR cerebrovascular accident*) OR ((ischemi* OR infarct* OR thrombo* OR haematoma* OR Fhematoma* OR haemorrhag* OR hemorrhag*) AND (brain* OR cerebr* OR cerebell* OR intracran*))). We only selected articles and reviews in English, other document types, such as letters, commentaries, and meeting abstracts, were excluded. Finally, a total of literature records was included.

Data Import and Deduplication

All included documents underwent peer review. We imported all bibliometric data into Endnote 20 (Clarivate) to remove duplicates. Then, we independently screened the titles, abstracts, and full texts of the included papers based on the exclusion criteria to identify eligible studies. The exclusion criteria were (1) VR intervention not present, (2) targeted conditions not related to stroke, and (3) paper theme not correlated with VR implementation in stroke neuromodulation. Finally, 1171 articles (924 articles and 247 reviews) meeting the criteria were included.

Data Merging

After screening and verification in Endnote, relevant literature was manually selected from the WOS database. The plain text containing information about these documents was downloaded from the WOS database. After downloading the data, field items need to be manually merged. We identify 3 common scenarios requiring data merging and propose solutions for each. These scenarios include (1) different spellings or formats of the same country name, for example, “USA” and “United States of America” will be merged into “USA”; and (2) different abbreviations or variations of the same author’s name, resolved using ORCID information and author affiliation.

Data Analysis and Visualization

After data deduplication and merging, the plain text will import into VOSviewer 1.6.18, Citespace 6.2.4, Scimago Graphica,

XSL•FO
RenderX

and GraphPad Prism 10.4.0 for further analysis and visualization. Quantitative and qualitative analyses were performed and presented in this study, including the following:

First, the time trends of publication were analyzed based on the annual number of publications after applying the screening criteria. The number of publications per year and the cumulative number of publications were calculated and visualized using GraphPad Prism software. A second-order polynomial (quadratic) fit was applied to the cumulative publication data to model its trajectory and predict future trends in research publication output [26]. This approach enables the identification of growth patterns and provides insights into the potential future development of the VR for stroke [27].

Second, the country distribution of publications was analyzed using Scimago Graphica software. After data extraction and screening, the countries contributing to the research publications were identified and their publication counts were aggregated. The results were visualized using Scimago Graphica to generate intuitive and visually appealing geographic maps [28]. These maps illustrated the spatial distribution and relative contribution of each country to the field, enabling a clear understanding of the global research landscape.

Third, the distribution of journals, institutions, and authors contributing to the research was analyzed using VOSviewer. The software was used to process co-authorship, co-citation, and bibliographic coupling relationships among journals, institutions, and authors. VOSviewer extracted bibliometric metadata, including titles, authors, affiliations, and journal information. Applied clustering algorithms in VOSviewer to group related entities, identifying patterns and prominent contributors. Generated network visualizations representing nodes (journals, institutions, and authors) and links (relationships such as citations or collaborations). These visualizations provided insights into the academic landscape, highlighting significant journals, leading institutions, and key authors in the research field [29].

Fourth, to analyze keyword clustering and detect emerging trends, CiteSpace was used to process and visualize bibliometric data. A visual network of keywords was generated, with nodes representing keywords and edges representing co-occurrence relationships [30]. The combined analysis of keyword clustering and citation bursts provided a comprehensive understanding of the intellectual structure, research hotspots, and evolving trends in the field [30,31].

Ethical Considerations

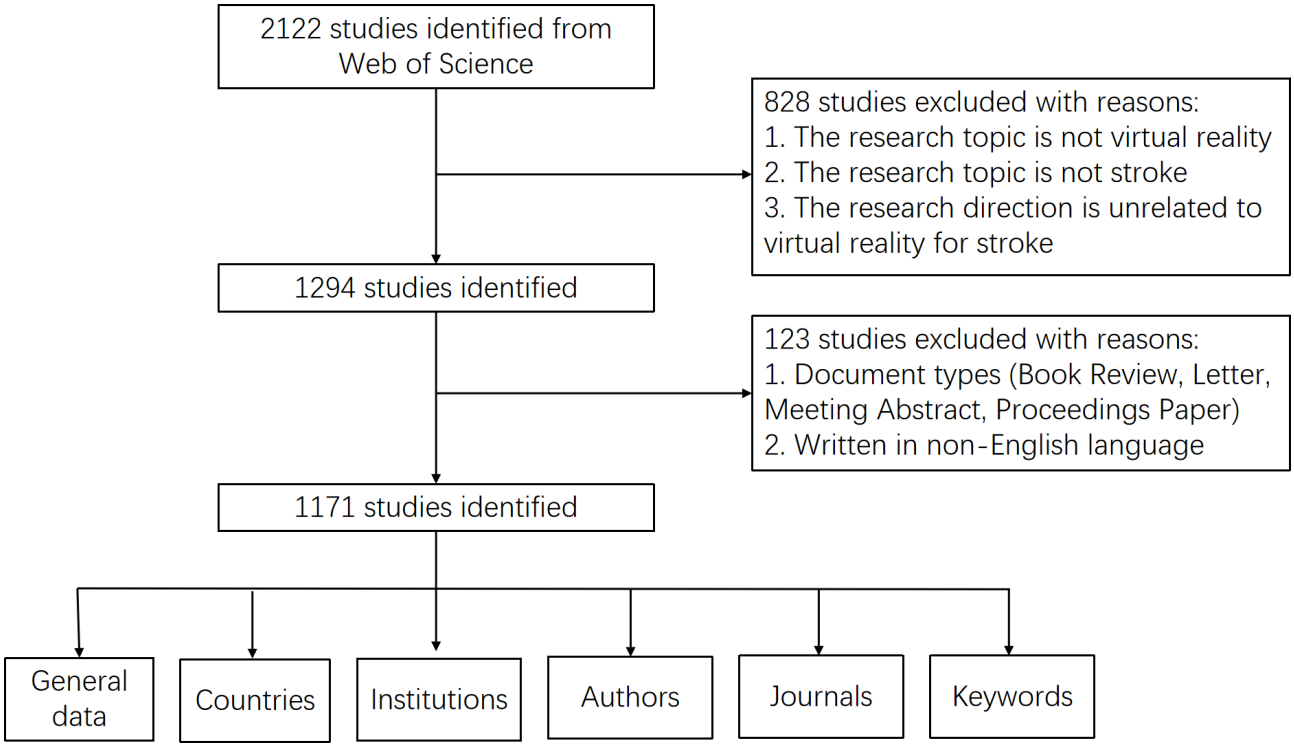
Our study falls under the category of bibliometric research. As this study is classified as bibliometric analysis, it did not require formal ethics approval. Bibliometric research involves the quantitative examination of publicly accessible scholarly publications (eg, journal articles, patents, and conference proceedings) and does not entail (1) direct interaction with human participants, (2) collection of personal, sensitive, or identifiable data, and (3) interventions in human behavior or health.

Under international guidelines (eg, the Helsinki Declaration) and institutional policies, studies of this nature are typically exempt from ethics review, as they pose no risks to privacy, autonomy, or welfare.

Results

Figure 1 shows the bibliometric search and analysis flowchart.

Figure 1. Flow chart of the bibliometric search and analysis process.

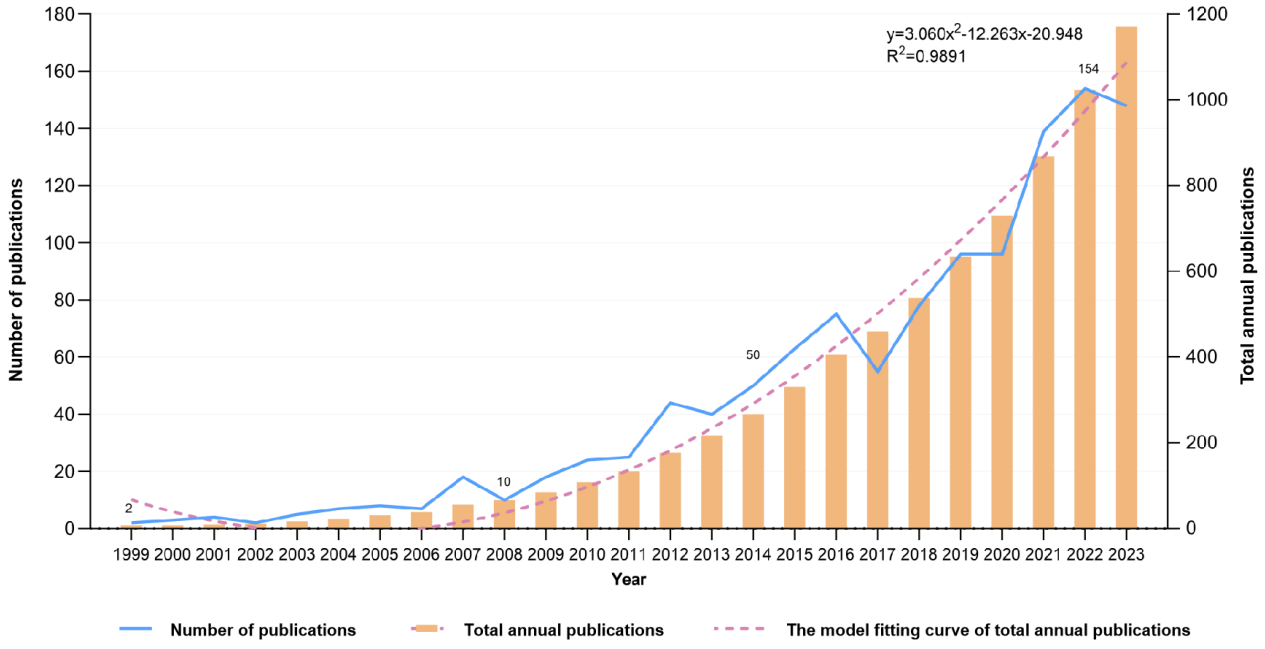


XSL•FO
RenderX

Time Trends of Publication

The initial search within the WOSCC identified a total of 2122 publications. Subsequent filtration to exclude nonacademic literature types, publications not written in English, and studies outside the scope of the targeted research theme narrowed the field to 1171 eligible papers. This corpus includes 844 research articles and 327 review articles, spanning a publication period from 1999 to 2023. The temporal distribution and evolution of the annual publication volume from the inception of the database until 2022 are illustrated in Figure 2. The publication timeline has been segmented into 3 distinct phases for analysis: the infancy phase (1999 - 2006), during which annual publications did not exceed 10; the slow-growth phase (2007 - 2014), characterized by a steady increment in publications, rising from 10 papers in 2008 to 50 papers by 2014; and the high-growth phase (2015 - 2023), where the publication rate consistently surpassed 50 papers annually, demonstrating a robust upward trajectory in research output. Furthermore, a linear regression analysis was performed to assess the growth trends in publication output, revealing a significant positive correlation with the model $y=3.060x^2-12.263x-20.948$ ($R^2=0.9891$; the closer R^2 is to 1, the better the fit of the trendline), underscoring a robust expansion in scholarly activities. The analysis forecasts a sustained increase in research productivity in the area, suggesting a thriving academic interest and a promising future for further investigations.

Figure 2. Trend of publication outputs from 1999 to 2023 on virtual reality (VR) for stroke topic.

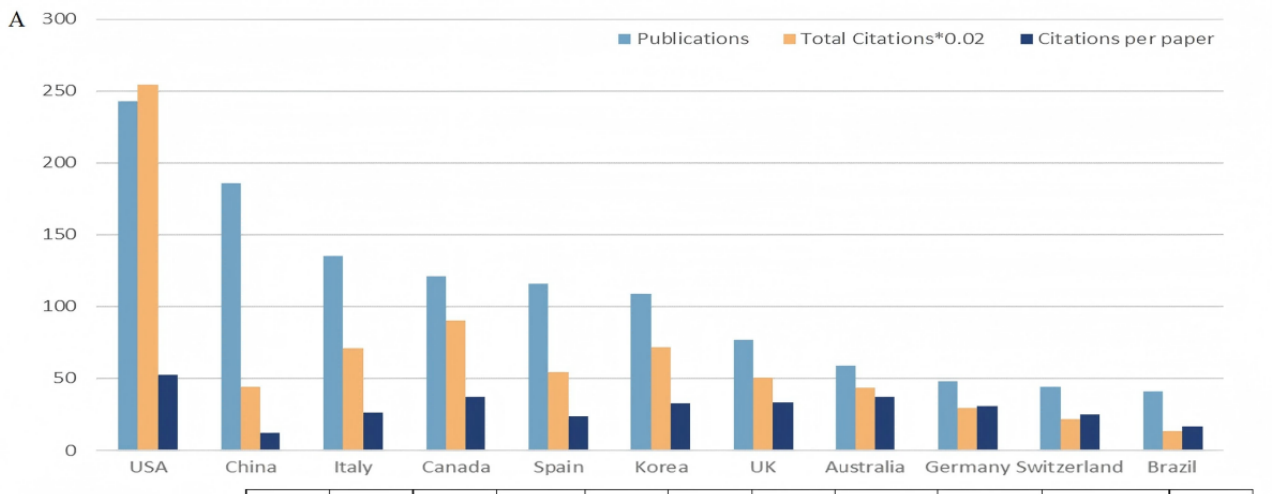


Distribution of Countries

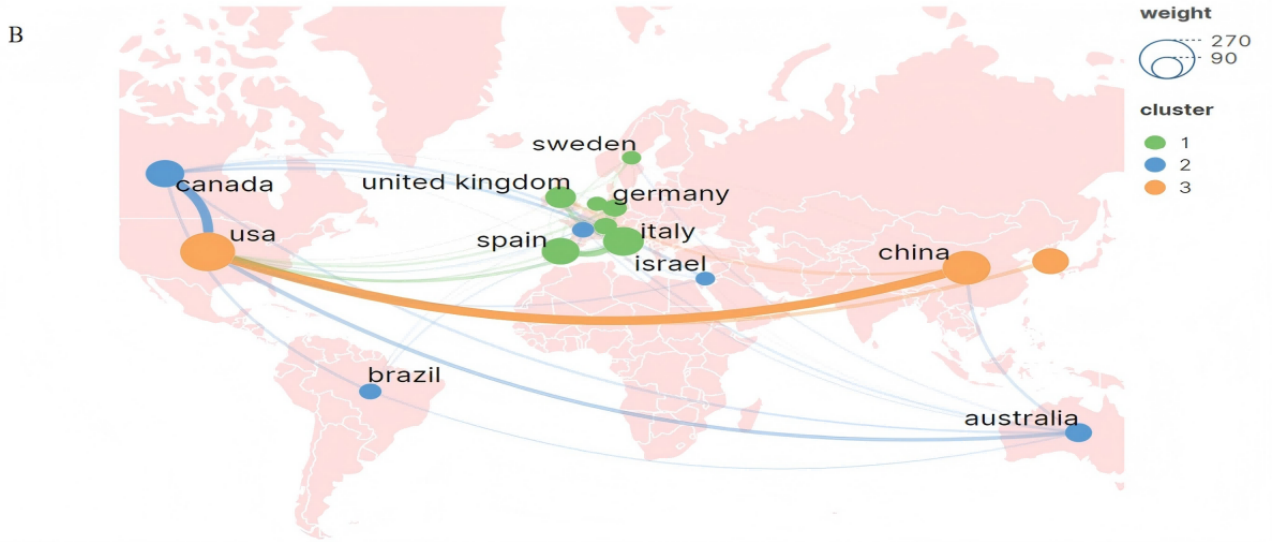
Figure 3A delineates the leading 11 nations or regions based on their scholarly output in the field. The United States spearheads this list with 243 publications, closely followed by China with 186, and Italy with 135. Canada, Spain, and Korea each have contributed over 100 publications to this domain. Notably, within this group of prolific publishers, only China and Brazil are classified as developing nations, with the remainder being developed nations. Research originating from the United States not only leads in quantity but also in scholarly impact, amassing 12,713 citations, with Canada and Italy following with 4502 and 3555 citations, respectively. When examining citation efficiency, the United States remains at the forefront with an average of 52.32 citations per publication, with Canada at 37.21 and Australia close behind at 37.08. The geographical distribution of the included studies was analyzed using VOSviewer and further visualized through Scimago Graphica, facilitating the creation of a cooperative network depicted in Figure 3B. This network prominently features 3 principal clusters located in North America, Western Europe, and Eastern Asia. Among these, the strongest collaborative ties are observed between the United States and China, highlighting a significant axis of academic partnership within the field.

XSL•FO
RenderX

Figure 3. The top 11 prolific countries and international collaboration network on virtual reality (VR) for stroke research. (**A**) The number of publications, total citations, and citations per paper in the top 11 countries. (**B**) The cooperative network visualization map of countries.



	USA	China	Italy	Canada	Spain	Korea	UK	Australia	Germany	Switzerland	Brazil
Publications	243	186	135	121	116	109	77	59	48	44	41
Total citations*0.02	254.26	44.38	71.10	90.04	54.44	71.84	50.78	43.76	29.70	22.02	13.68
Citations per paper	52.32	11.93	26.33	37.21	23.47	32.95	32.97	37.08	30.94	25.02	16.68



Analysis of Institutions

The dataset reveals participation from 1916 distinct institutions in the authorship of 1171 scholarly articles. Table 1 enumerates the top 10 contributing institutions, with McGill University in Canada leading the cohort with a total of 49 publications. This is closely followed by the University of Ottawa, also in Canada, with 29 papers, and Rutgers University in the United States, contributing 22 papers. Notably, Rutgers University also boasts the highest citation count, amassing a total of 2210 citations, and an impressive average of 100.45 citations per paper. Figure 4 elucidates the interinstitutional collaborative network among the leading entities applying virtual reality to stroke research. these collaborative ties, 4 institutions represented by McGill University, University of Ottawa, Hong Kong Polytechnic University, and Istituto di Ricovero e Cura a Carattere Scientifico (IRCCS) Centro Neurolesi Bonino Pulejo have formed 4 representative networks. The connections between these networks are robust, underscoring a highly cohesive and dynamic collaborative framework. These institutions collectively exemplify a strong and well-integrated network of cooperation in advancing this field of research.

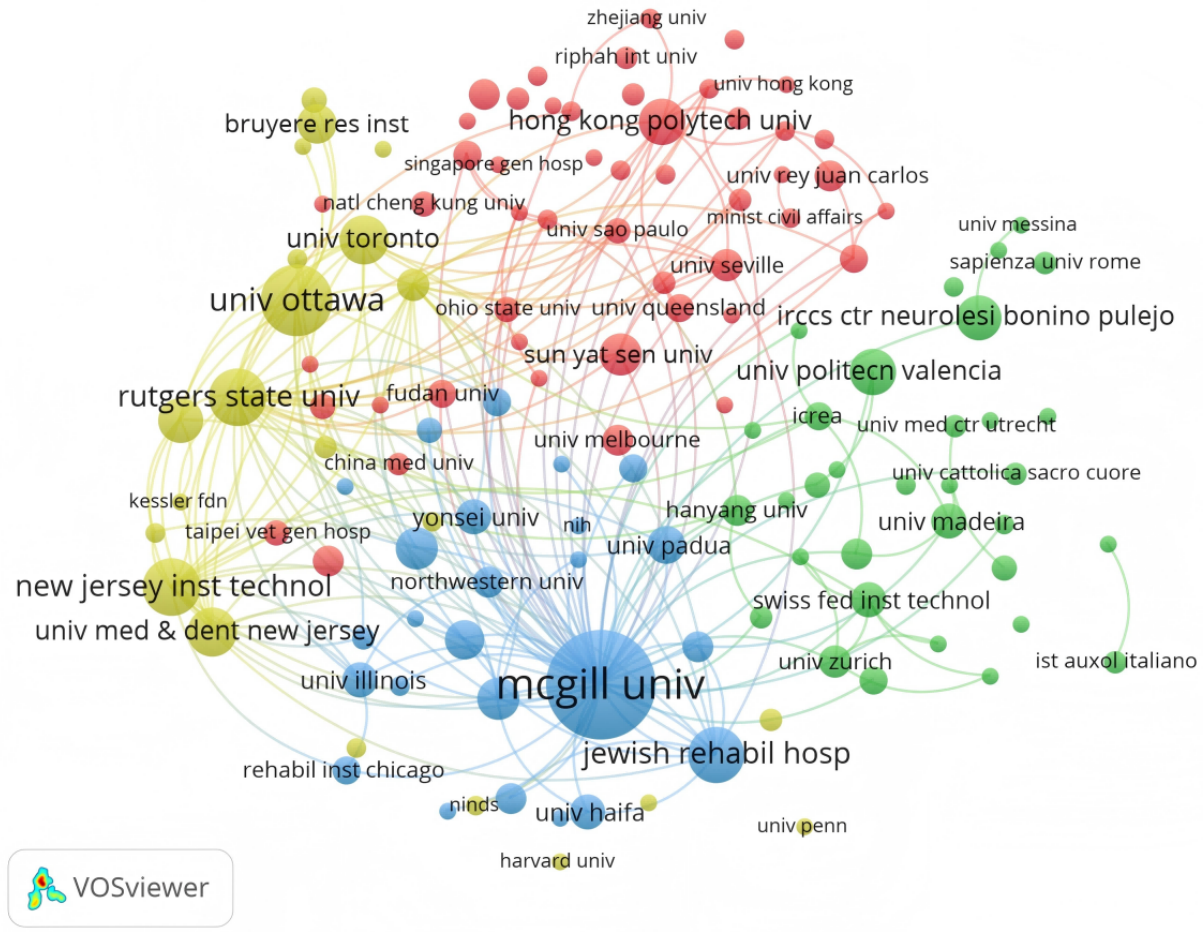
XSL•FO
RenderX

Table . The top 10 productive institutions regarding the research on virtual reality (VR) for stroke.

Rank	Institution	Country	Count	Citations
1	McGill University	Canada	49	2166
2	University of Ottawa	Canada	29	602
3	Rutgers University	United States	22	2210
4	New Jersey Institute of Technology	United States	21	1529
5	Jewish Rehabilitation Hospital	Canada	21	708
6	University of Toronto	Canada	18	1526
7	Polytechnic University of Valencia	Spain	17	436
8	Hong Kong Polytechnic University	China	17	144
9	IRCCS[a] Centro Neurolesi Bonino Pulejo	Italy	16	453
10	Northeastern University	United States	16	389

[a]IRCCS: Istituto di Ricovero e Cura a Carattere Scientifico.

Figure 4. The VOSviewer density visualization map of institutions in the field of virtual reality (VR) for stroke.



Analysis of Journals

This study analyzed 1171 scholarly articles published across 344 academic journals. According to Bradford's Law, core journals are identified as those publishing more than one-third of all journal articles within a specific field. Within this area of research, 12 journals were classified as core, while 332 were deemed noncore. Table 2 indicates that the top 10 journals collectively accounted for 29.97% of the publications, contributing 351 articles. The *Journal of Neuroengineering and*

XSL•FO
RenderX

Rehabilitation led with 79 publications, followed by the *International Journal of Stroke* with 42 papers, and *Institute of Electrical and Electronics Engineers (IEEE) Transactions on Neural Systems and Rehabilitation Engineering* with 37 papers. The highest IF observed over the past 5 years was 8.8, associated with the journal *Stroke*. Among the journals, 5 boasted an IF greater than 5.000, 3 had IF ranging from 3.000 to 5.000, and 2 journals recorded an IF below 3.000. The impact of scholarly publications in a research area is quantified by the number of co-citations they receive. The co-citation analysis was conducted using the VOSviewer and is presented in Figure 5. The size of the nodes represents the number of co-citations, and the lines connecting the nodes indicate co-citation relationships. With regard to the color coding used in the cluster analysis, the red cluster represents journals specializing in rehabilitation engineering, such as the *Journal of Neuroengineering and Rehabilitation* and the *IEEE Transactions on Neural Systems and Rehabilitation Engineering*. The green cluster represents academic journals in the field of neurology research, with *Stroke* and *Frontiers in Neurology* being representative examples. General medical journals, represented by *Medicine*, are classified as the blue cluster.

Table . The top 10 journals that published articles regarding the research on virtual reality (VR) for stroke.

Rank	Journal title	Country	Count	Citations	JCR[a]	IF[b] 5 year
1	*Journal of Neuroengineering and Rehabilitation*	England	79	3145	Q2[c]	5.8
2	*International Journal of Stroke*	England	42	170	Q1	6.5
3	*IEEE* [d] *Transactions on Neural Systems and Rehabilitation Engineering*	United States	37	1603	Q1	5.5
4	*Frontiers in Neurology*	Switzerland	35	418	Q2	3.9
5	*Topics in Stroke Rehabilitation*	England	29	1491	Q2	2.6
6	*Archives of Physical Medicine and Rehabilitation*	United States	29	1215	Q1	4.4
7	*Disability and Rehabilitation*	England	28	836	Q2	2.5
8	*Sensors*	Switzerland	25	209	Q2	4.1
9	*Stroke*	United States	24	1617	Q1	8.8
10	*Virtual Reality*	England	23	203	Q2	5.5

[a]JCR: Journal Citation Reports.

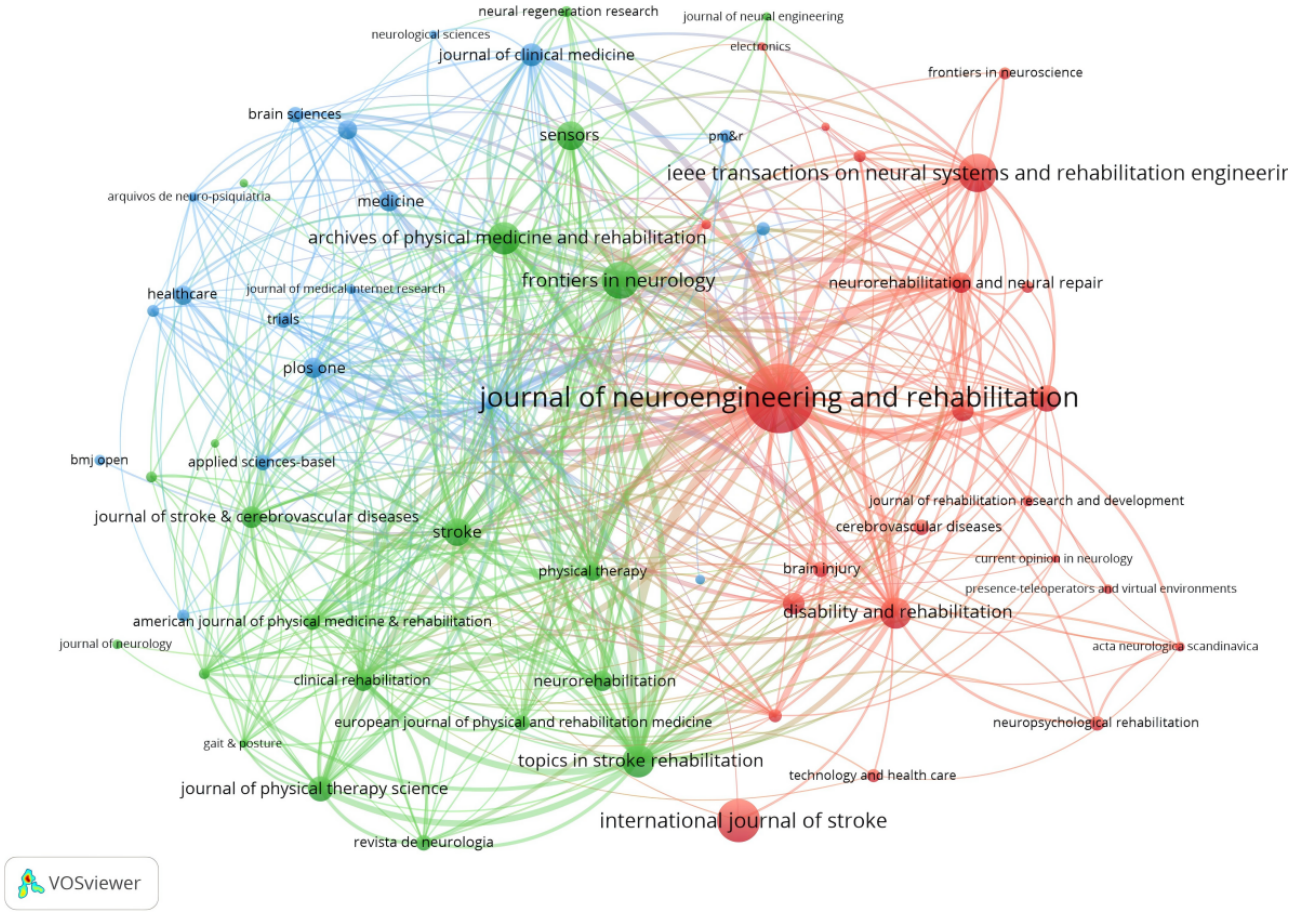
[b]IF: impact factor.

[c]Q: Quartile.

[d]IEEE: Institute of Electrical and Electronics Engineers.

XSL•FO
RenderX

Figure 5. VOSviewer co-citation analysis clustering map of journal in the field of virtual reality (VR) for stroke.



Analysis of Authors

A total of 5319 authors were engaged in the documents on VR in the field of stroke. Table 3 shows the top 10 most active authors and their related information. Calabrò Rocco Salvatore published 22 publications, ranking first among all authors. The majority of the top 10 authors are affiliated with Italian and Canadian research institutions. As illustrated in Table 3, Andrea Turolla from the University of Bologna was the most cited author, with 896 citations. The average number of citations per paper, Tonin Paolo from the Sant'Anna Institute was the most cited author. Furthermore, the H-index can also accurately reflect the academic achievements of authors. Levin Mindy was ranked first in H-index (50) and has the largest impact in the field. Figure 6 shows an overlay visualization of the author co-occurrence analysis generated by VOSviewer. The graph forms major clusters centered on the top 10 authors. The collaboration between them is particularly strong and occurs predominantly within the same cluster.

XSL•FO
RenderX

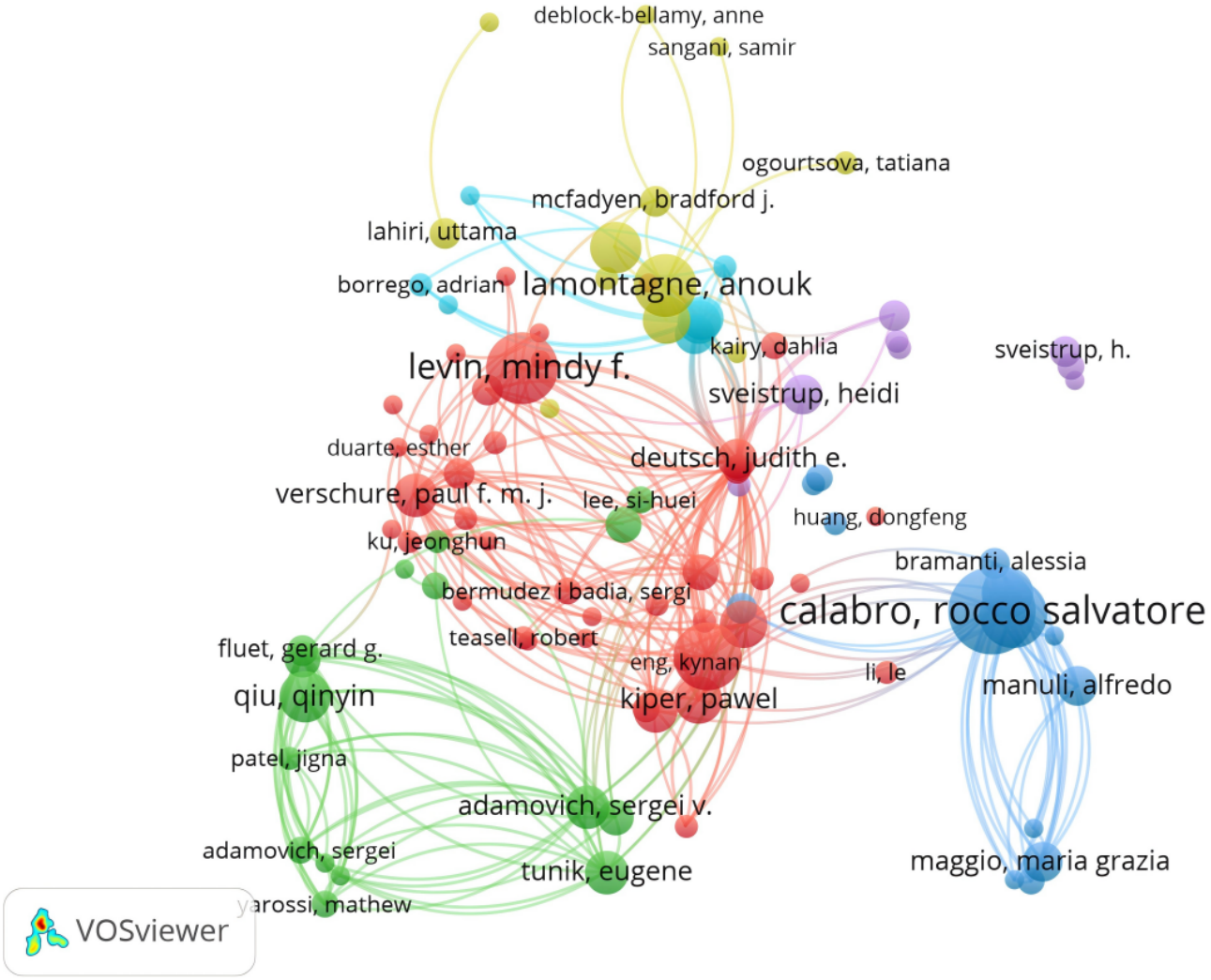
Table . The top 10 active authors in virtual reality (VR) for stroke research.

Rank	Author	Institution	Country	Publications	Citations	H-index
1	Calabrò, Rocco Salvatore	Scientific Institute for Research, Hospitalization and Healthcare	Italy	22	565	33
2	Levin, Mindy	McGill University	Canada	18	816	50
3	Andrea, Turolla	University of Bologna	Italy	17	896	26
4	De Luca, Rosaria	IRCCS[a] Centro Neurolesi Bonino Pulejo	Italy	17	524	28
5	Lamontagne, Anouk	McGill University	Canada	16	304	26
6	Qiu, Qinyin	Rutgers University	United States	13	392	14
7	Naro, Antonino	IRCCS Centro Neurolesi Bonino Pulejo	Italy	13	328	29
8	Fung, Joyce	McGill University	Canada	13	246	33
9	Tonin, Paolo	Sant'Anna Institute	Italy	12	828	33
10	Pawel Kiper	IRCCS San Camillo Hospital	Italy	12	472	16

[a]IRCCS: Istituto di Ricovero e Cura a Carattere Scientifico.

XSL•FO
RenderX

Figure 6. Network map of active authors contributed to virtual reality (VR) for stroke research.



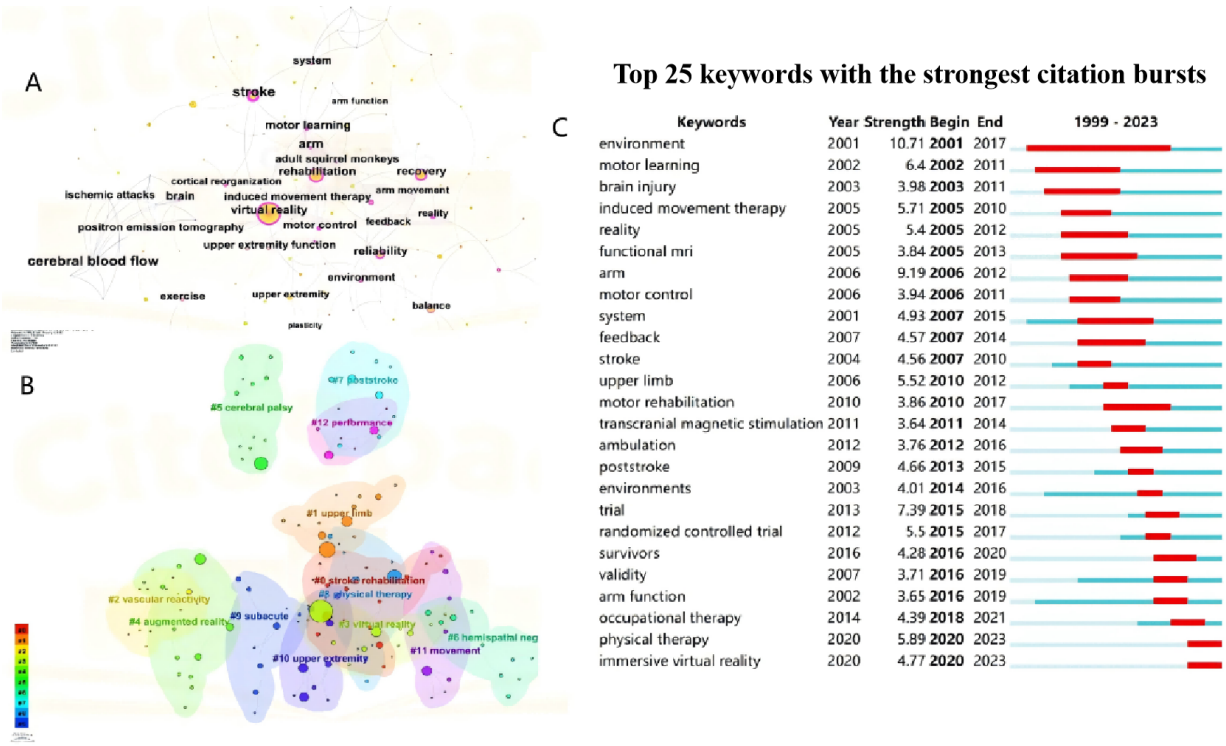
Analysis of Keywords

Keywords are words that have a significant meaning within the context of a paper. The identification of high-frequency or emergent keywords can provide insight into the current topics of interest and potential future research frontiers. As shown in Figure 7A, the top 5 most frequent keywords are virtual reality, rehabilitation, recovery, stroke, and upper limb. According to the different types of keywords, the keywords can be categorized into 13 clusters, as shown in Figure 7B. The #0, #3, #4, and #8 clusters mainly describe the main forms and means of using VR in stroke rehabilitation. The #1, #10, #11, and #12 clusters primarily delineate the principal objective of VR in stroke rehabilitation is the enhancement of upper limb function and performance. The other clusters represent the application of VR to different periods of stroke, as well as other dysfunctions of stroke.

Citation burst analysis is a valuable tool that can provide insightful analysis of specific areas of research for specific year hotspots and trends. Citespace is used to generate the top 25 keywords with the strongest bursts from 1999 to 2023, with the results presented in Figure 7C. Among these keywords, environment exhibits the highest burst intensity. Arm function, physical therapy, occupational therapy, and immersive virtual reality are the most recent emergent keywords, which indicates the recent research direction.

XSL•FO
RenderX

Figure 7. Analysis of keywords related to publications on virtual reality (VR) for stroke field from 1999 to 2023. (**A**) The keyword co-occurrence network map. (**B**) The keyword cluster map. (**C**) The top 25 keywords with the strongest citation bursts.



Discussion

Overview of the Results

In this bibliometric study, 1171 papers focusing on VR for stroke were included and visualized by VOSviewer and Citespace to display the research hotspots and trends of this field. The publications on the related topic might reveal variations in research activity and productivity, which could be classified into 3 phases. In the early stages, the integration of VR technology into stroke rehabilitation was minimal. The primary reasons for this included the high costs associated with VR systems and the nascent state of the technology [10,32]. VR hardware and software were not yet sophisticated or affordable enough to be widely adopted in medical settings [33]. During this period, research output remained relatively static, with few papers being published on the subject. With the reduced costs and the availability of high-quality technologies in the 2010s, there was a steady increase in the number of publications, reflecting the growing interest and investment in this area. The COVID-19 pandemic acted as a catalyst for the adoption of VR in stroke rehabilitation. With the need for remote health care solutions, telerehabilitation gained prominence [34,35]. VR, being a versatile and effective tool for remote rehabilitation, saw a dramatic increase in use and research interest. By 2021, the number of annual publications exceeded 100, indicating a significant surge in research activity. It is foreseeable that this research area will maintain its popularity in the near future.

In terms of the national distribution of researchers, more than one-third of the world's countries have published papers on the application of VR technology for stroke rehabilitation. Notably, the United States, Italy, Canada, the United Kingdom, and Spain have emerged as predominant leaders in this field, significantly contributing to telemedicine research facilitated by VR. This dominance can be attributed to the substantial gross national product of these Western European and North American countries, which provides robust support for clinical research, in addition to the higher incidence and prevalence of stroke within these regions [36]. Furthermore, in the last five years, only two developing nations, China and Brazil, have made notable contributions in this domain. Particularly, Chinese ongoing investment in the application of novel smart technologies to health care management, along with the increasing availability of more portable, personalized, and affordable VR devices, is becoming increasingly viable with continuous innovations in smart technologies such as AI and the advent of the 5G era [37]. Nevertheless, the relatively low average citations per article for Chinese research indicates a need for improvement in the quality of their publications. The Chinese government is also actively seeking to enhance the quality of its research output in the future.

The analysis of researchers and institutions, as presented in Table 1 and Figure 4, reveals that the most active institutions are primarily from prominent universities in developed countries, particularly Canada and the United States. These institutions have substantial academic resources, evident in the authors' co-occurrence network. As shown in Table 3, half of the top 10 researchers is from Italian institutions, yet their citation counts are lower than those of North American researchers due to the recency of their publications. The distribution of core journals publishing virtual reality therapy research for stroke highlights that the top 10 journals publish less than a third of the total papers, indicating a broad dissemination across many journals. The IF of these journals is generally modest, with only three having an IF of 5 or higher.

XSL•FO
RenderX

This suggests that the research quality in this field needs improvement, necessitating increased international collaboration for higher-quality clinical research outcomes.

Hotspots Analysis

Author clusters are instrumental in identifying significant research trends, offering insight into the past, present, and future hotspots within a specific scientific domain [38]. As depicted in Figure 5 , the authors responsible for the publication of 1171 papers can be categorized into 3 distinct groups. The leading contributors currently driving the field forward are Levin Mindy, Calabrò Rocco Salvatore, and Qiu Qinyin, respectively. This categorization not only highlights the dominant figures but also provides insight into the evolving dynamics and key influencers in the research area.

The team of Levin Mindy has been at the forefront of investigating the efficacy of VR in improving upper limb function in stroke survivors [39,40]. Their research found that both healthy volunteers and stroke survivors used comparable movement strategies when grasping and placing a ball within VR environments [41]. These findings suggest that VR can simulate realistic tasks and provide an immersive rehabilitation experience, making it a promising and effective tool for stroke recovery [41]. By replicating the mechanics of daily life tasks, VR environments offer a unique opportunity to bridge the gap between clinical therapy and real-world applications, potentially accelerating motor recovery and improving functional performance in stroke survivors [42]. The immersive nature of VR not only engages patients in a meaningful way but also helps overcome some of the limitations of traditional rehabilitation methods, particularly by offering repetitive and task-specific exercises that are critical for motor learning and neuroplasticity [43,44].

The research group led by Calabrò Rocco Salvatore investigates the neurophysiology effects of VR on individuals with stroke. Their studies revealed significant increases in alpha-band power in the occipital regions and beta-band power in the frontal regions of cognitively impaired individuals with stroke following VR-based cognitive rehabilitation [45]. These results highlight the potential of VR not only for enhancing motor function but also for influencing brain activity patterns associated with cognition and sensory-motor integration. Further expanding on this, their research demonstrated that robotic-assisted gait training in combination with VR can enhance lower limb function in stroke survivors with lower limb paresis. Notably, this approach activates key brain regions that are essential for gait control and rehabilitation, suggesting that VR may facilitate neuroplastic changes by stimulating the brain's motor networks [46]. This dual benefit—improving both motor and cognitive recovery—positions VR as a powerful tool in stroke rehabilitation, capable of promoting neuroplasticity and accelerating recovery across multiple domains of function [45-47].

Qiu Qinyin's research team has made significant strides in developing home-based VR rehabilitation systems designed specifically for patients with chronic stroke. Their empirical studies have demonstrated that patients can use this system with minimal supervision, achieving significant enhancements in upper limb functionality. By empowering patients to perform exercises at home, these home-based VR systems not only reduce the strain on health care systems but also provide patients with more flexibility, potentially leading to better adherence and outcomes in the long term [48]. Furthermore, the team has integrated haptically rendered VR with robotic assistance to facilitate upper limb functional training in stroke survivors [49,50]. Comparative analyses indicate that this combined modality yields superior clinical outcomes relative to conventional physiotherapy methods. The findings suggest that haptically rendered VR and robotics not only augment patient engagement but also enable more precise and effective rehabilitation exercises, thereby advancing the standard of care in stroke rehabilitation [49-51].

Keywords and Trend Analysis

The cluster analysis of co-occurrence and burst keywords can identify frontier topics and emerging trends in a field. In the case of VR stroke treatment, the analysis highlighted that upper limb function is a primary research area and direction. According to Figure 7C, the evolution of topics over time can be divided into three phases:

1. Phase I (2001 - 2006): Early literature focused on the therapeutic modalities and rationale for VR therapy in stroke treatment. Key topics included the use of VR environments, motor learning, and induced movement therapy [52,53]. Researchers also used functional magnetic resonance imaging to explore the association between VR therapy and brain injury [54].
2. Phase II (2007 - 2014): This period saw a significant increase in the global use of VR for stroke rehabilitation, particularly for upper limb motor function and balance transfer [55]. The research expanded beyond initial therapeutic modalities to explore broader applications of VR in stroke recovery.
3. Phase III (2015 - 2020): With advancements in VR technology, especially immersive VR, research designs became more standardized, incorporating randomized controlled trials. This phase marked an increased focus on the efficacy of VR in combination with traditional physiotherapy and occupational therapy [56,57]. Studies validating the effectiveness of VR therapies for neurological disorders, primarily through RCTs, emerged as a major trend.

Overall, the transition from exploratory studies to more rigorous and standardized research reflects the growing maturity and credibility of VR in neurorehabilitation. This progression underscores the importance of ongoing innovation in VR technology and the need for high-quality research to validate its clinical benefits.

Strength and Limitations

This review is the first to comprehensively summarize current publications and development trends in VR for stroke from a bibliometric perspective. To ensure a systematic and thorough evaluation, the study collected 1171 related papers published over the last two decades from the WOSCC database. Using popular bibliometric tools such as Citespace and VOSviewer,

XSL•FO
RenderX

the study conducted a quantitative analysis of VR therapy in the stroke field, focusing on data related to countries, institutions, authors, journals, citations, and keywords.

Despite the comprehensive approach, certain limitations of this work need to be acknowledged: First, due to limitations of the analysis tools, we only analyzed publications in the WOSCC. Although WOSCC offers robust coverage of high-impact journals, our exclusive reliance on this platform excluded potentially relevant studies from PubMed, Scopus, and Google Scholar, which may skew the geographical and methodological trends. Second, the language restriction to English-language publications in our inclusion criteria risks introducing publication bias, as significant research outputs in other languages were systematically excluded. Third, inherent limitations exist in our search strategy: only peer-reviewed articles and review papers were considered, thereby excluding potentially valuable academic books, conference proceedings, and gray literature. These methodological constraints underscore the necessity for future bibliometric investigations in VR for stroke rehabilitation to implement multi-database search architectures, incorporate multilingual literature, and develop more inclusive document type classification frameworks to provide a more inclusive and accurate representation of the research landscape in VR for stroke rehabilitation to provide a more inclusive and accurate representation of the research landscape in VR for stroke rehabilitation.

Conclusion

This bibliometric study offers a comprehensive analysis of research trends in VR therapy for stroke rehabilitation over the past three decades. The data reveal a consistent upward trajectory in the number of studies, with North America and Western Europe leading in both publication volume and citation impact. The majority of these studies have been published in journals with relatively low IF, indicating a need for heightened focus on this area in future research. The author co-occurrence analysis identified 3 predominant clusters centering on neurophysiological effects of VR for stroke by Calabro RS, upper limb motor rehabilitation by Levin Mindy, and telerehabilitation of VR for stroke by Qiu Qinyin, respectively. The analysis also highlights the most recent research trends, which include the optimization of VR devices and the integration of VR with traditional physiotherapy and occupational therapy. These insights can guide future researchers in understanding the current focal points and potential directions for further development in VR therapy for stroke rehabilitation.

Acknowledgments

We would like to thank the researchers involved who work at the Department of Rehabilitation Medicine of West China Hospital at Sichuan University. This study is funded by the 2023-YF09-00044-SN from Chengdu Science and Technology Bureau.

Data Availability

The datasets generated during this study are available from the corresponding author on reasonable request.

Authors' Contributions

YW and YC contributed to the concept, idea, research design, data collection, data analysis, and writing the original draft. RM and YQ contributed to research design, data collection, and data analysis. YQ, WS, LZ, and RM contributed to revising and approving the manuscript. All authors contributed to the paper and research.

Conflicts of Interest

None declared.

References

1. Kuriakose D, Xiao Z. Pathophysiology and treatment of stroke: present status and future perspectives. Int J Mol Sci 2020 Oct 15;21(20):7609. [doi: 10.3390/ijms21207609] [Medline: 33076218]
2. Saini V, Guada L, Yavagal DR. Global epidemiology of stroke and access to acute ischemic stroke interventions. Neurology (ECronicon) 2021 Nov 16;97(20 Suppl 2):S6-S16. [doi: 10.1212/WNL.0000000000012781] [Medline: 34785599]
3. Campbell BCV, De Silva DA, Macleod MR, et al. Ischaemic stroke. Nat Rev Dis Primers 2019 Oct 10;5(1):70. [doi: 10.1038/s41572-019-0118-8] [Medline: 31601801]
4. Everard G, Declerck L, Detrembleur C, et al. New technologies promoting active upper limb rehabilitation after stroke: an overview and network meta-analysis. Eur J Phys Rehabil Med 2022 Aug;58(4):530-548. [doi: 10.23736/S1973-9087.22.07404-4] [Medline: 35666491]
5. Meng L, Liang Q, Yuan J, et al. Vestibular rehabilitation therapy on balance and gait in patients after stroke: a systematic review and meta-analysis. BMC Med 2023 Aug 25;21(1):322. [doi: 10.1186/s12916-023-03029-9] [Medline: 37626339]
6. Gittler M, Davis AM. Guidelines for adult stroke rehabilitation and recovery. JAMA 2018 Feb 27;319(8):820-821. [doi: 10.1001/jama.2017.22036] [Medline: 29486016]
7. Kundi MK, Spence NJ. Efficacy of mirror therapy on lower limb motor recovery, balance and gait in subacute and chronic stroke: a systematic review. Physiother Res Int 2023 Apr;28(2):e1997. [doi: 10.1002/pri.1997] [Medline: 36880119]

XSL•FO
RenderX

8. Bernhardt J, Urimubenshi G, Gandhi DBC, Eng JJ. Stroke rehabilitation in low-income and middle-income countries: a call to action. Lancet 2020 Oct 31;396(10260):1452-1462. [doi: 10.1016/S0140-6736(20)31313-1] [Medline: 33129396]
9. Moncion K, Rodrigues L, MacKay-Lyons M, et al. Exercise-based stroke rehabilitation: clinical considerations following the COVID-19 pandemic. Neurorehabil Neural Repair 2022 Jan;36(1):3-16. [doi: 10.1177/15459683211054175] [Medline: 34711094]
10. Andreae MH. Virtual reality in rehabilitation. BMJ 1996 Jan 6;312(7022):4-5. [doi: 10.1136/bmj.312.7022.4] [Medline: 8555863]
11. Bruno RR, Wolff G, Wernly B, et al. Virtual and augmented reality in critical care medicine: the patient's, clinician's, and researcher's perspective. Crit Care 2022 Oct 25;26(1):326. [doi: 10.1186/s13054-022-04202-x] [Medline: 36284350]
12. Percy D, Phillips T, Torres F, Chaleunphonh M, Sung P. Effectiveness of virtual reality-based balance and gait in older adults with fear of movement: a systematic review and meta-analysis. Physiother Res Int 2023 Oct;28(4):e2037. [doi: 10.1002/pri.2037] [Medline: 37427539]
13. Steffin M. Computer assisted therapy for multiple sclerosis and spinal cord injury patients application of virtual reality. Stud Health Technol Inform 1997;39:64-72. [Medline: 10168956]
14. Riva G, Bacchetta M, Baruffi M, et al. The use of PC based VR in clinical medicine: the VREPAR projects. Technol Health Care 1999;7(4):261-269. [Medline: 10461791]
15. Johnson DA, Rose FD, Rushton S, Pentland B, Attree EA. Virtual reality: a new prosthesis for brain injury rehabilitation. Scott Med J 1998 Jun;43(3):81-83. [doi: 10.1177/003693309804300307] [Medline: 9682294]
16. Laver KE, Lange B, George S, Deutsch JE, Saposnik G, Crotty M. Virtual reality for stroke rehabilitation. Cochrane Database Syst Rev 2017 Nov 20;11(11):CD008349. [doi: 10.1002/14651858.CD008349.pub4] [Medline: 29156493]
17. Maggio MG, Latella D, Maresca G, et al. Virtual reality and cognitive rehabilitation in people with stroke: an overview. J Neurosci Nurs 2019 Apr;51(2):101-105. [doi: 10.1097/JNN.0000000000000423] [Medline: 30649091]
18. Wu J, Zeng A, Chen Z, et al. Effects of virtual reality training on upper limb function and balance in stroke patients: systematic review and meta-meta-analysis. J Med Internet Res 2021 Oct 12;23(10):e31051. [doi: 10.2196/31051] [Medline: 34636735]
19. Chen J, Or CK, Chen T. Effectiveness of using virtual reality-supported exercise therapy for upper extremity motor rehabilitation in patients with stroke: systematic review and meta-analysis of randomized controlled trials. J Med Internet Res 2022 Jun 20;24(6):e24111. [doi: 10.2196/24111] [Medline: 35723907]
20. Zhang B, Li D, Liu Y, Wang J, Xiao Q. Virtual reality for limb motor function, balance, gait, cognition and daily function of stroke patients: a systematic review and meta-analysis. J Adv Nurs 2021 Aug;77(8):3255-3273. [doi: 10.1111/jan.14800] [Medline: 33675076]
21. Demeco A, Zola L, Frizziero A, et al. Immersive virtual reality in post-stroke rehabilitation: a systematic review. Sensors (Basel) 2023 Feb 3;23(3):1712. [doi: 10.3390/s23031712] [Medline: 36772757]
22. Guo QF, He L, Su W, et al. Virtual reality for neurorehabilitation: a bibliometric analysis of knowledge structure and theme trends. Front Public Health 2022;10:1042618. [doi: 10.3389/fpubh.2022.1042618] [Medline: 36438265]
23. Chen C, Song M. Visualizing a field of research: a methodology of systematic scientometric reviews. PLoS ONE 2019;14(10):e0223994. [doi: 10.1371/journal.pone.0223994] [Medline: 31671124]
24. Guler AT, Waaijer CJF, Palmblad M. Scientific workflows for bibliometrics. Scientometrics 2016;107:385-398. [doi: 10.1007/s11192-016-1885-6] [Medline: 27122644]
25. Fu BSC, Yu Z, Tiwari K. New release of 2020 Impact Factor (Web of Science) and Cite Score (Scopus). J Orthop Translat 2021 Jul;29:A3. [doi: 10.1016/j.jot.2021.07.003] [Medline: 34466383]
26. Nelis JLD, Rosas da Silva G, Ortuño J, et al. The general growth tendency: a tool to improve publication trend reporting by removing record inflation bias and enabling quantitative trend analysis. PLoS ONE 2022;17(5):e0268433. [doi: 10.1371/journal.pone.0268433] [Medline: 35594252]
27. He D, Cao S, Le Y, Wang M, Chen Y, Qian B. Virtual reality technology in cognitive rehabilitation application: bibliometric analysis. JMIR Serious Games 2022 Oct 19;10(4):e38315. [doi: 10.2196/38315] [Medline: 36260388]
28. Peng Y, Liu Y, Lai S, et al. Global trends and prospects in health economics of robotic surgery: a bibliometric analysis. Int J Surg 2023 Dec 1;109(12):3896-3904. [doi: 10.1097/JS9.0000000000000720] [Medline: 37720937]
29. van Eck NJ, Waltman L. Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics 2010 Aug;84(2):523-538. [doi: 10.1007/s11192-009-0146-3]
30. Chen C, Chen Y. Searching for clinical evidence in CiteSpace. AMIA Annu Symp Proc 2005;2005:121-125. [Medline: 16779014]
31. Chen C. Searching for intellectual turning points: progressive knowledge domain visualization. Proc Natl Acad Sci U S A 2004 Apr 6;101 Suppl 1(Suppl 1):5303-5310. [doi: 10.1073/pnas.0307513100] [Medline: 14724295]
32. Kaltenborn KF, Rienhoff O. Virtual reality in medicine. Methods Inf Med 1993 Nov;32(5):407-417. [Medline: 8295549]
33. Berra K. Computers in medicine. Virtual rehabilitation: dream or reality? Clin Invest Med 2006 Aug;29(4):187-192. [Medline: 16986481]

XSL•FO
RenderX

34. Lloréns R, Noé E, Colomer C, Alcañiz M. Effectiveness, usability, and cost-benefit of A virtual reality-based telerehabilitation program for balance recovery after stroke: a randomized controlled trial. Arch Phys Med Rehabil 2015 Mar;96(3):418-425. [doi: 10.1016/j.apmr.2014.10.019] [Medline: 25448245]
35. Putrino D. Telerehabilitation and emerging virtual reality approaches to stroke rehabilitation. Curr Opin Neurol 2014 Dec;27(6):631-636. [doi: 10.1097/WCO.0000000000000152] [Medline: 25333603]
36. GBD 2019 Stroke Collaborators. Global, regional, and national burden of stroke and its risk factors, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet Neurol 2021 Oct;20(10):795-820. [doi: 10.1016/S1474-4422(21)00252-0] [Medline: 34487721]
37. Bai Y, Liu F, Zhang H. Artificial intelligence limb rehabilitation system on account of virtual reality technology on long-term health management of stroke patients in the context of the internet. Comput Math Methods Med 2022;2022:2688003. [doi: 10.1155/2022/2688003] [Medline: 35651925]
38. Synnestvedt MB, Chen C, Holmes JH. CiteSpace II: visualization and knowledge discovery in bibliographic databases. AMIA Annu Symp Proc 2005;2005:724-728. [Medline: 16779135]
39. Henderson A, Korner-Bitensky N, Levin M. Virtual reality in stroke rehabilitation: a systematic review of its effectiveness for upper limb motor recovery. Top Stroke Rehabil 2007;14(2):52-61. [doi: 10.1310/tsr1402-52] [Medline: 17517575]
40. Levin MF, Baniña MC, Frenkel-Toledo S, et al. Personalized upper limb training combined with anodal-tDCS for sensorimotor recovery in spastic hemiparesis: study protocol for a randomized controlled trial. Trials 2018 Jan 4;19(1):7. [doi: 10.1186/s13063-017-2377-6] [Medline: 29301545]
41. Viau A, Feldman AG, McFadyen BJ, Levin MF. Reaching in reality and virtual reality: a comparison of movement kinematics in healthy subjects and in adults with hemiparesis. J Neuroeng Rehabil 2004 Dec 14;1(1):11. [doi: 10.1186/1743-0003-1-11] [Medline: 15679937]
42. Saposnik G, Levin M, Outcome Research Canada (SORCan) Working Group. Virtual reality in stroke rehabilitation: a meta-analysis and implications for clinicians. Stroke 2011 May;42(5):1380-1386. [doi: 10.1161/STROKEAHA.110.605451] [Medline: 21474804]
43. Levin MF, Snir O, Liebermann DG, Weingarden H, Weiss PL. Virtual reality versus conventional treatment of reaching ability in chronic stroke: clinical feasibility study. Neurol Ther 2012 Dec;1(1):3. [doi: 10.1007/s40120-012-0003-9] [Medline: 26000209]
44. Subramanian SK, Levin MF. Viewing medium affects arm motor performance in 3D virtual environments. J Neuroeng Rehabil 2011 Jun 30;8:36. [doi: 10.1186/1743-0003-8-36] [Medline: 21718542]
45. Gangemi A, De Luca R, Fabio RA, et al. Effects of virtual reality cognitive training on neuroplasticity: a quasi-randomized clinical trial in patients with stroke. Biomedicines 2023 Dec 6;11(12):3225. [doi: 10.3390/biomedicines11123225] [Medline: 38137446]
46. Calabrò RS, Naro A, Russo M, et al. The role of virtual reality in improving motor performance as revealed by EEG: a randomized clinical trial. J Neuroeng Rehabil 2017 Jun 7;14(1):53. [doi: 10.1186/s12984-017-0268-4] [Medline: 28592282]
47. Torrisi M, Maggio MG, De Cola MC, et al. Beyond motor recovery after stroke: The role of hand robotic rehabilitation plus virtual reality in improving cognitive function. J Clin Neurosci 2021 Oct;92:11-16. [doi: 10.1016/j.jocn.2021.07.053] [Medline: 34509235]
48. Qiu Q, Cronce A, Patel J, et al. Development of the Home based Virtual Rehabilitation System (HoVRS) to remotely deliver an intense and customized upper extremity training. J Neuroeng Rehabil 2020 Nov 23;17(1):155. [doi: 10.1186/s12984-020-00789-w] [Medline: 33228709]
49. Patel J, Fluet G, Qiu Q, et al. Intensive virtual reality and robotic based upper limb training compared to usual care, and associated cortical reorganization, in the acute and early sub-acute periods post-stroke: A feasibility study. J Neuroeng Rehabil 2019 Jul 17;16(1):92. [doi: 10.1186/s12984-019-0563-3] [Medline: 31315612]
50. Merians AS, Fluet GG, Qiu Q, et al. Robotically facilitated virtual rehabilitation of arm transport integrated with finger movement in persons with hemiparesis. J Neuroeng Rehabil 2011 May 16;8:27. [doi: 10.1186/1743-0003-8-27] [Medline: 21575185]
51. Merians AS, Fluet GG, Qiu Q, Lafond I, Adamovich SV. Learning in a virtual environment using haptic systems for movement re-education: can this medium be used for remodeling other behaviors and actions? J Diabetes Sci Technol 2011 Mar 1;5(2):301-308. [doi: 10.1177/193229681100500215] [Medline: 21527097]
52. Deutsch JE, Merians AS, Adamovich S, Poizner H, Burdea GC. Development and application of virtual reality technology to improve hand use and gait of individuals post-stroke. Restor Neurol Neurosci 2004;22(3-5):371-386. [Medline: 15502277]
53. Sisto SA, Forrest GF, Glendinning D. Virtual reality applications for motor rehabilitation after stroke. Top Stroke Rehabil 2002;8(4):11-23. [doi: 10.1310/YABD-14KA-159P-MN6F] [Medline: 14523727]
54. You SH, Jang SH, Kim YH, et al. Virtual reality-induced cortical reorganization and associated locomotor recovery in chronic stroke: an experimenter-blind randomized study. Stroke 2005 Jun;36(6):1166-1171. [doi: 10.1161/01.STR.0000162715.43417.91] [Medline: 15890990]
55. Turolla A, Dam M, Ventura L, et al. Virtual reality for the rehabilitation of the upper limb motor function after stroke: a prospective controlled trial. J Neuroeng Rehabil 2013 Aug 1;10:85. [doi: 10.1186/1743-0003-10-85] [Medline: 23914733]

XSL•FO
RenderX

56. Marques-Sule E, Arnal-Gómez A, Buitrago-Jiménez G, Suso-Martí L, Cuenca-Martínez F, Espí-López GV. Effectiveness of Nintendo Wii and physical therapy in functionality, balance, and daily activities in chronic stroke patients. J Am Med Dir Assoc 2021 May;22(5):1073-1080. [doi: 10.1016/j.jamda.2021.01.076] [Medline: 33639116]
57. Shin JH, Bog Park S, Ho Jang S. Effects of game-based virtual reality on health-related quality of life in chronic stroke patients: a randomized, controlled study. Comput Biol Med 2015 Aug;63:92-98. [doi: 10.1016/j.compbiomed.2015.03.011] [Medline: 26046499]

Abbreviations

IEEE: Institute of Electrical and Electronics Engineers
IF: Impact Factor
IRCCS: Istituto di Ricovero e Cura a Carattere Scientifico
SCIE: Science Citation Index Expanded
VR: virtual reality
WOS: Web of Science
WOSCC: Web of Science Core Collection

Edited by A Coristine; submitted 31.08.24; peer-reviewed by B Liu, PC Wang; revised version received 12.03.25; accepted 13.03.25; published 16.04.25.

Please cite as:
Wei Y, Chen Y, Ma R, Qiu Y, Su W, Zhang L, Gao Q
Research Hotspots and Trends of Virtual Reality Intervention for Stroke: Bibliometric Analysis
JMIR Serious Games 2025;13:e65993
URL: https://games.jmir.org/2025/1/e65993
doi:10.2196/65993

©Yixin Wei, Yuan Chen, Runting Ma, Yitong Qiu, Wei Su, Li Zhang, Qiang Gao. Originally published in JMIR Serious Games (https://games.jmir.org), 16.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Feasibility, Usability, and Effects of Leisure-Based Cognitive Training Using a Fully Immersive Virtual Reality System in Older Adults: Single-Arm Pretest-Posttest Pilot Study

I-Ching Chuang[1,2], PhD; Xiao-Ting Huang[1], MD; I-Chen Chen[3], PhD; Yih-Ru Wu[2,4], MD; Ching-Yi Wu[1,5,6], ScD

[1]Department of Occupational Therapy and Graduate Institute of Behavioral Sciences, College of Medicine, Chang Gung University, No. 259 Wen-hwa First Road, Taoyuan, Taiwan

[2]Department of Neurology, Linkou Chang Gung Memorial Hospital, Taoyuan, Taiwan

[3]Department of Occupational Therapy, College of Nursing and Health Sciences, Da-Yeh University, Changhua, Taiwan

[4]School of Medicine, Chang Gung University, Taoyuan, Taiwan

[5]Department of Physical Medicine and Rehabilitation, Linkou Chang Gung Memorial Hospital, Taoyuan, Taiwan

[6]Healthy Aging Research Center, Chang Gung University, Taoyuan, Taiwan

Corresponding Author:
Ching-Yi Wu, ScD
Department of Occupational Therapy and Graduate Institute of Behavioral Sciences, College of Medicine, Chang Gung University, No. 259 Wen-hwa First Road, Taoyuan, Taiwan

Abstract

Background: Cognitive training is an effective approach to support cognitive function in older adults. Incorporating meaningful leisure activities, such as gardening, may enhance both engagement and training outcomes. While fully immersive virtual reality (VR) offers ecologically valid and engaging environments that can further boost motivation, limited research has explored the combination of VR-based cognitive training and leisure activities for older adults.

Objective: This study aims to assess the feasibility, usability, and preliminary effectiveness of leisure-based VR cognitive training for community-dwelling older adults.

Methods: A fully immersive VR cognitive training system, controlled via a head-mounted display, was developed, incorporating gardening-themed activities such as planting, fertilizing, watering, and harvesting. These tasks were designed to engage multiple cognitive domains, including memory, attention, executive function, processing speed, and visuospatial abilities. The program consisted of 16 sessions delivered over 8 weeks (twice weekly, 1 hour per session). Cognitive outcomes were assessed before and after training using the Montreal Cognitive Assessment, the digit symbol substitution test, word list immediate and delayed recall, spatial span, and the Stroop Color and Word Test. Feasibility, acceptance, and usability were evaluated using the System Usability Scale and a posttraining questionnaire. Licensed occupational therapists from both community and institutional settings assessed the training system's usability.

Results: All 41 participants (mean age 69.79, SD 5.05 y) completed the training with 100% adherence and no serious adverse events. Feasibility ratings—particularly for perceived usefulness, intention to use, and subjective norms—reflected strong acceptance. Usability ratings from older adults indicated high ease of use, enjoyment, and positive experience, while professionals rated the system as moderately usable (mean System Usability Scale score 68.01, SD 8.38). Statistically significant improvements were observed in general cognition (P=.004), processing speed (P=.049), immediate and delayed memory (P<.001), and executive function (P=.002). No significant changes were found in visuospatial memory (P=.29).

Conclusions: This study provides preliminary evidence supporting the feasibility and usability of a gardening-based VR cognitive training program for older adults. Feasibility was demonstrated through full adherence, absence of major adverse events, and high participant acceptance. Usability feedback was favorable from both older adults and professionals across community and long term care settings. Additionally, improvements in multiple cognitive domains, including general cognition, processing speed, memory, and executive function, suggest potential cognitive benefits. Future randomized controlled trials with more diverse samples and extended follow-up are warranted to confirm and expand upon these findings.

Trial Registration: ClinicalTrials.gov NCT05227495; https://clinicaltrials.gov/study/NCT05227495

(JMIR Serious Games 2025;13:e66673) doi:10.2196/66673

KEYWORDS

cognitive training; leisure activities; feasibility; usability; immersive virtual reality; older adults

XSL•FO
RenderX

Introduction

Background

The rapid growth of the global aging population is expected to lead to an escalation of age-related cognitive decline, making it a noteworthy health concern. Cognitive decline may lead to dependence on others in engaging in daily activities [1], contributing to a growing need for social care [2]. Providing early preventive training to preserve and maintain cognitive functions is crucial for older adults in the community.

Cognitive training, involving the repetitive practice of standardized tasks [3], is grounded in the principles of neuroplasticity. Cognitive training often targets various cognitive domains, including attention, memory, visuospatial function, processing speed, and executive function [4,5]. The results of previous studies have shown that cognitive training has a small to moderate effect on attention, memory, visuospatial abilities, and executive function in older adults [5-8]. Despite the demonstrated effectiveness of cognitive training, its impact remains constrained [9].

Engaging in leisure activities, commonly perceived as enjoyable and easily accessible, may provide protective effects against age-related cognitive decline and dementia [10]. Integrating leisure activities into cognitive training has the potential for enhancing the effectiveness of the enjoyment and meaningfulness of interventions for older adults, with relatively straightforward implementation in community settings [11]. Horticultural therapy is one of the favorite leisure activities for older adults because gardening-related activities resonate with their past experiences [12]. Growing evidence supports the cognitive benefits of gardening in older adults [13-17], as demonstrated in both systematic reviews and empirical studies showing improvements in memory, attention, and overall cognitive function, alongside enhanced physical health and well-being [16,17]. Physiologically, gardening has been associated with increased levels of brain-derived neurotrophic factor and platelet-derived growth factor, which are linked to neurogenesis and memory enhancement [13,14]. Furthermore, large-scale population-based and intervention studies suggest that gardening may help reduce the risk of cognitive decline and improve cognitive performance, including in individuals with mild to moderate dementia [15,16]. The integration of horticultural activities into cognitive training may be a strategy for enhancing cognitive function in older adults.

Fully immersive virtual reality (VR) augments ecological validity, provides immediate performance feedback, and facilitates customization of environments and activities, thereby conferring advantages over conventional cognitive training [18]. Feasibility studies on cognitive training using VR technology have demonstrated that older adults accepted and were motivated by VR-based tasks [19-21]. VR cognitive training has been effective in enhancing cognitive function in older adults, including global cognition, visuospatial processing, visual memory, and executive function, both in individuals with and without cognitive decline [22-27]. Recent studies have shown that immersive VR significantly enhances cognitive performance in individuals with mild cognitive impairment or dementia, particularly in executive function and global cognition [28]. Randomized trials also indicate that VR-based cognitive-motor training yields cognitive improvements comparable to traditional approaches, with the added benefits of greater task engagement and user interaction [29]. Furthermore, immersive simulations of daily tasks can stimulate multiple cognitive domains such as attention, memory, and planning [30]. These findings underscore the potential of immersive VR as an effective and practical cognitive training tool for older adults. A recent review also highlights that relevance of content, ease of use, and user support are key to successful implementation in aged care settings [31].

Most previous studies of immersive VR cognitive training for older adults have focused on daily activities [21,24,25,32] or, for convenience, have used commercially available gaming scenarios, such as Jet Run [33], rather than designing serious gaming activities [22]. This might easily result in older adults perceiving the training process as less enjoyable and meaningful, thereby limiting the effectiveness of the cognitive training. Incorporating leisure activities into immersive VR cognitive training could enhance the enjoyment and meaningfulness of the training for older adults, leading to an increase in its effectiveness [9]. However, studies on incorporating leisure activities into immersive VR cognitive training for older adults remain limited.

Conducting feasibility and usability testing for a VR training program is considered a prerequisite for its successful implementation [34]. For the implementation of VR training in older adults, the significance of usability is underscored by age-related factors that contribute to barriers in use [34]. Consequently, considering the perspective of professionals may improve the applicability of VR training for older adults [35]. However, few studies have comprehensively investigated the usability of VR cognitive training for professionals.

Objectives

This study aims to investigate the feasibility and usability of leisure-based cognitive training using a VR system for community-dwelling older adults and evaluate its preliminary effectiveness. We hypothesize that leisure-based cognitive training with VR would be feasible, acceptable, and enjoyable for older adults. We also expect a good acceptance of the VR system. On the basis of the reviewed literature, we further hypothesize that leisure-based VR cognitive training may enhance cognitive domains such as global cognition, memory, and executive function in older adults [22,23,25,27].

Methods

Participants

The inclusion criteria were (1) age ≥60 years, (2) a Montreal Cognitive Assessment (MoCA) score of ≥21, and (3) the ability to follow instructions. A MoCA threshold of 21 was selected to include older adults with mild cognitive decline who retained sufficient cognitive capacity for interacting with VR systems, while avoiding the exclusion of those with subtle impairments who may still benefit from feasibility testing. Exclusion criteria included a diagnosis of dementia or other major psychiatric

XSL•FO
RenderX

disorders, a history of motion sickness, or psychosis-related symptoms that could interfere with VR training.

Demographic information, including education, age, and sex, was collected using a structured questionnaire at the time of participant enrollment. Data were gathered through face-to-face interviews after obtaining participant consent.

In addition to the older adult participants, 40 licensed occupational therapists (n=20, 50% from day care centers or long-term care homes and n=20, 50% from community centers) were recruited to evaluate the usability of the VR cognitive training system. These professionals had at least one year of relevant clinical experience. Owing to their backgrounds, they were familiar with individuals both with and without mild or severe cognitive impairments and were therefore well positioned to assess usability across older adults with varying cognitive profiles.

Sample Size Estimation

This pilot study included 41 participants, a sample size determined based on recruitment feasibility and literature guidelines. Prior studies suggest that 24 to 55 participants are adequate for exploratory trials [36,37], and a review reported a median of 30 participants in feasibility studies conducted in the United Kingdom [38].

Materials

The VR content was custom designed using Unity 3D software (Unity Technologies) and integrated with the HTC Vive Pro head-mounted display. The system supported full-body interaction using handheld controllers, allowing participants to perform tasks in either a seated or standing position depending on their physical condition (Figure 1). The training content featured realistic horticultural tasks with embedded cognitive demands and provided immediate visual and auditory feedback. The VR-based cognitive training program comprised more than 10 gardening-themed activities designed to engage multiple cognitive domains (Figure 2). Each gardening activity was designed to target specific cognitive domains based on its task demands; for example, the sowing task required participants to recall the color and position of illuminated flowerpots and match them with corresponding seed packets, thereby engaging working memory. The harvesting task involved monitoring multiple flowerpots for signs of blooming or wilting, selecting appropriate tools, and executing timely actions to either harvest blooming flowers or remove wilted plants, thereby training executive function. The watering task, in which participants monitored plant hydration levels and adjusted watering accordingly, targeted attention and processing speed. The fertilizing task required recalling and applying a sequence of fertilizer colors to the correct plants, enhancing working memory. These activities collectively supported the training of attention, memory, executive function, processing speed, and visuospatial abilities in an ecologically meaningful and engaging context. In each training session, at least 2 different cognitive domains were addressed to ensure comprehensive cognitive training throughout the program.

Figure 1. Community-dwelling older adults interacting with the immersive virtual reality cognitive training system during the intervention period. (A) Participant using the system in a seated position. (B) Participant using the system in a standing position.



XSL•FO
RenderX

Figure 2. Screenshots from the immersive virtual reality cognitive training system showing examples of 4 gardening-related tasks used in the intervention: (A) planting, (B) harvesting, (C) watering, and (D) fertilizing.



Procedure

The VR cognitive training program consisted of 16 sessions delivered over 8 weeks (twice weekly, 1 hour per session). In previous studies on the application of VR cognitive training for older adults, the average number of training sessions was approximately 17, with each session lasting between 20 and 60 minutes [30]. The total training duration per participant was approximately 13 hours. Moreover, some studies adopted a training frequency of 2 to 3 sessions per week [25,39,40]. Therefore, we adopted a training schedule of 2 sessions per week, each lasting 1 hour, over an 8-week period. Each session consisted of 4 to 5 training segments, interspersed with breaks to prevent motion sickness. Training was guided by occupational therapists who received standardized instruction to ensure fidelity and consistency. Each therapist had more than two years of experience providing cognitive interventions for older adults in community or day care centers, enabling them to adjust the training based on participants' responses as needed. When older adults had difficulty understanding the operations or tasks in the early stages of training, the therapists provided verbal and physical assistance. The data collection process was conducted by trained research assistants under the supervision of the principal investigator. Prior to data collection, all research assistants received standardized training to ensure consistency in administering both questionnaires and assessments. Data were collected between 2021 and 2022. To maintain procedural fidelity across sites, uniform protocols were implemented throughout the study.

Outcome Measures

Feasibility Measures

Feasibility was evaluated through adherence (ie, attendance and completion rates), adverse events (eg, dizziness or discomfort), and participant acceptance of the VR training. Adherence was evaluated based on the attendance rate of older adults, defined as the number of participants who completed the programs divided by the number of participants at baseline. Therapists were required to report any adverse events related to the use of the VR system, such as headaches, head fullness, blurred vision, dizziness, and vertigo, that occurred during the study. To assess acceptance, participants were asked to complete selected items from an acceptance questionnaire (Multimedia Appendix 1) based on constructs from the unified theory of acceptance and use of technology [41] and the technology acceptance model [42]. The following subscales were used: perceived usefulness, subjective norms, and intention to use. These reflect the feasibility dimensions of acceptability and adoption potential [41,43]. Participants rated all items on a 5-point Likert scale ranging from 1=strongly disagree to 5=strongly agree [44].

Usability Measures

Usability was assessed from the perspectives of both participants and professionals. Participants completed the following additional subscales from the same acceptance questionnaire, focusing on user-system interaction: perceived ease of use, perceived enjoyment, and user experience. These dimensions align with the ISO (International Organization for Standardization) 9241‐11 criteria (effectiveness, efficiency, and satisfaction) and the usability heuristics formulated by

XSL•FO
RenderX

Nielsen [45]. Participants rated these items on a 5-point Likert scale ranging from 1=strongly disagree to 5=strongly agree.

Professionals evaluated usability using the System Usability Scale (SUS; Multimedia Appendix 2) [46] after a single VR session. The SUS consists of 10 questions. Professionals rated each question on a scale ranging from 1=strong disagreement to 5=strong agreement. These ratings were then processed through an algorithm to produce a score out of 100 points.

Cognitive Assessments

For cognitive function, assessors conducted all outcome assessments before and immediately after the training. The reliable and valid assessments used included the MoCA [47,48], the digit symbol substitution test (DSST) from the Wechsler Adult Intelligence Scale [49-51], word list immediate and delayed recall from the Wechsler Memory Scale–Third Edition (WMS-III) [52,53], and the Stroop Color Word Test (SCWT) [54,55].

The MoCA is a widely used tool for assessing global cognitive function and is known for its high sensitivity in detecting cognitive decline [47]. The MoCA had high internal consistency (Cronbach α=0.86) [48]. Global cognitive functioning encompasses a range of cognitive abilities, such as orientation, attention, memory, naming, visuospatial skills, executive function, language, and abstraction. The assessment yields a total score of 30 points, with higher scores indicating superior global cognitive function.

The DSST, a subtest from the Wechsler Adult Intelligence Scale, measures information processing speed [49]. During the test, participants are required to match symbols to corresponding numbers and replicate these symbols in designated spaces within 120 seconds. The number of correctly completed symbols determines the DSST score. This test demonstrates high test-retest reliability [50] and strong sensitivity to changes in cognitive function [51]. Higher DSST scores reflect superior information processing speed.

The word list subtest from the WMS-III is a widely used tool for assessing working memory. The word list test comprises 2 parts: word list immediate recall and word list delayed recall. In word list immediate recall, participants are presented with 12 unrelated words over 4 trials and asked to recall them immediately after each presentation. Word list delayed recall asks participants to recall the same 12 words again 25 to 35 minutes after completing word list immediate recall. Higher word list immediate recall or word list delayed recall scores signify better immediate or delayed recall, respectively. The spatial span subtest from the WMS-III is used to assess visuospatial memory. During the spatial span subtest, participants are instructed to touch sequentially positioned blocks in either the same order as demonstrated by the instructor or in reverse order. The word list and spatial span subtests from the WMS-III demonstrate good internal consistency, with reported Cronbach α coefficients ranging from 0.78 to 0.94 [52]. The reliability and validity of the WMS-III have been established in community-dwelling adults, making it a valid tool for assessing cognitive improvements after training [53].

The SCWT is a widely used assessment of inhibitory executive function. Its internal consistency has been demonstrated to be reliable, with Cronbach α coefficients averaging 0.80 [54]. The SCWT is sensitive in distinguishing individuals with cognitive decline from those without [55]. Performance differences between congruent and incongruent tasks indicate response interference. We measured inhibitory executive function by calculating the time differences between the congruent and incongruent tasks, with higher SCWT scores indicating poorer inhibitory executive function.

Ethical Considerations

This study was approved by the institutional review board of Chang Gung Memorial Hospital (202100553B0). All participants provided written informed consent prior to enrollment. To protect participants' privacy and confidentiality, all data were deidentified prior to analysis. Participants received a gift valued at approximately NT $50 (US $1.65) as a token of appreciation. No identifiable images of participants are included in the manuscript.

Statistical Analysis

The demographic characteristics of the participants are reported using descriptive statistics, including means and SDs. A paired 2-tailed *t* test was conducted to evaluate the effect of the leisure-based VR cognitive training program. The significance level was set at .05 for all statistical tests. SPSS software (version 27.0; IBM Corp) was used for data analysis.

Results

Participant Characteristics

The study recruited 41 older adults from local community facilities. Participants ranged in age from 62 to 82 (mean 69.79, SD 5.05) years. Of the 41 participants, 6 (15%) were male, and 35 (85%) were female. They had an average of 12.11 (SD 3.80) years of education. The participants' demographic data and baseline characteristics are presented in Table 1.

Most of the participants experienced no adverse events during the VR training, except for 1 older adult who reported experiencing dizziness once during the middle phase of the training, which resolved after a brief period of rest. All 41 older adults completed the 16-session program, yielding a 100% adherence rate.

XSL•FO
RenderX

Table . Demographic data and baseline characteristics of community-dwelling older adults participating in the virtual reality cognitive training program. Attendance was 100%, with all older adults completing the 16 sessions of the program (N=41).

Characteristics	Participants
Age (years), mean (SD)	69.79 (5.05)
Sex, n (%)	
Male	6 (15)
Female	35 (85)
Education (years), mean (SD)	12.11 (3.80)
Montreal Cognitive Assessment score, mean (SD)	27.82 (2.44)

Intervention Acceptability

Regarding intervention acceptability, older adult participants reported high scores on items reflecting perceived usefulness, subjective norms, and intention to use, with a mean feasibility score of 4.11 (SD 0.61) on a 5-point Likert scale. Notably, perceived usefulness and intention to use both exceeded a mean of 4, suggesting strong acceptance and adoption potential (Table 2).

Older adult participants reported favorable experiences regarding system usability. Subscale scores for perceived ease of use, enjoyment, and user experience yielded a mean usability score of 4.23 (SD 0.58) on a 5-point Likert scale, indicating that the system was generally perceived as intuitive, enjoyable, and user-friendly (Table 2).

Table . Results of the posttraining acceptance questionnaire assessing the feasibility and usability of the virtual reality cognitive training experience from the perspective of older adults (N=41).

Variables	Scores, mean (SD)
Feasibility subscale	
Perceived usefulness	4.42 (0.49)
Intention to use	4.34 (0.59)
Subjective norms	3.57 (0.73)
Usability subscale	
Perceived ease of use	3.93 (0.65)
Perceived enjoyment	4.34 (0.58)
User experience	4.42 (0.49)

From the professionals' perspective, usability was evaluated using the SUS. A total of 40 occupational therapists participated in the evaluation (n=20, 50% from day care centers or long-term care homes and n=20, 50% from community centers). The overall SUS mean score was 68.01 (SD 8.38), indicating an acceptable level of usability [56]. An independent samples *t* test was conducted to compare SUS scores between professionals from day care centers or long-term care homes and those from community centers. No significant difference was found between the 2 groups (t_{38}=0.78; *P*=.45), with institution-based professionals reporting a mean score of 69.63 (SD 1.83) and community-based professionals a mean score of 66.40 (SD 8.44).

After 16 training sessions, older adult participants showed significant improvements in the MoCA scores (Cohen *d*=0.47 [medium effect]; *P*=.004), DSST (Cohen *d*=0.16 [small effect]; *P*=.049), word list immediate recall (Cohen *d*=0.58 [medium effect]; *P*<.001), word list delayed recall (Cohen *d*=0.39 [small to medium effect]; *P*<.001), and the SCWT (Cohen *d*=0.41 [medium effect]; *P*=.002). No significant pretest-posttest differences were found for the spatial span (Cohen *d*=0.15 [small effect]; *P*=.29; Table 3).

XSL•FO
RenderX

Table . Results of paired sample *t* tests, with mean (SD) values, for pretest-posttest outcome measures in older adults (N=41).

Outcome measures	Baseline scores, mean (SD)	Posttraining scores, mean (SD)	*t* test (*df*)	*P* value	Effect size (Cohen *d*)
MoCA[a]	27.76 (2.55)	28.76 (1.55)	3.03 (40)	.004	0.47
DSST[b]	62.02 (15.48)	64.54 (15.27)	1.05 (40)	.049	0.16
Wechsler Memory Scale–Third Edition					
Word list immediate recall	29.76 (7.33)	34.10 (7.64)	3.71 (40)	<.001	0.58
Word list delayed recall	12.73 (3.0)	13.93 (3.14)	2.5 (40)	<.001	0.39
Spatial span	14.24 (3.02)	14.71 (3.32)	−2.65 (40)	.29	0.15
SCWT[c]	16.55 (8.62)	13.39 (6.52)	0.95 (40)	.002	0.41

[a]MoCA: Montreal Cognitive Assessment.

[b]DSST: digit symbol substitution test.

[c]SCWT: Stroop Color Word Test.

Discussion

Principal Findings

To the best of our knowledge, this is the first study to examine the feasibility, usability, and preliminary cognitive outcomes of a fully immersive VR cognitive training system incorporating gardening-themed leisure activities for community-dwelling older adults. The findings support the feasibility of the intervention, as evidenced by full attendance, minimal adverse events, and high levels of acceptance reflected in participants' responses to perceived usefulness, subjective norms, and intention to use. Both older adults and health professionals provided favorable impressions of the system's usability. Among older adults, high ratings on perceived ease of use, perceived enjoyment, and user experience indicated that the system was intuitive and engaging. Likewise, professionals from both community and institutional settings rated the system as acceptably usable based on the SUS. Moreover, the 8-week program led to significant improvements in general cognition, processing speed, immediate recall, delayed recall, and executive function. These findings support the notion that incorporating meaningful leisure activities, such as gardening, into VR-based training offers an engaging and ecologically valid strategy to enhance cognitive function among community-dwelling older adults.

Feasibility Perspectives Among Older Adults

Feasibility was supported by excellent adherence (100% program completion) and minimal adverse events, consistent with prior VR studies involving older adults [24,57]. Participant acceptance of the VR system was high, particularly in terms of perceived usefulness and intention to use, with both averaging above 4 on a 5-point Likert scale. These results suggest that the leisure-based and culturally familiar content helped foster engagement and adoption potential. The relatively lower score in subjective norms may reflect neutral or uncertain perceptions among family members or caregivers regarding the necessity of VR training, which aligns with previous research noting that social influence is often less pronounced in early-stage adoption [41]. Nevertheless, the overall acceptance profile indicates strong feasibility for broader implementation in community settings.

Usability Perspectives From Older Adults and Professionals

The usability of the leisure-based VR cognitive training system was supported by both older adult participants and professionals from both community and institutional settings. Older adults reported high scores across usability-related domains, including perceived ease of use, perceived enjoyment, and user experience, indicating that the system was generally intuitive, engaging, and user-friendly. These findings suggest that the training interface and task flow were well suited for older users, even among those with limited prior exposure to immersive technologies.

From the professionals' perspective, system usability was evaluated using the SUS, which yielded an average score indicative of acceptable usability [58]. Ratings were generally positive, with no significant differences observed between professionals working in day care centers or long-term care homes and those working in community centers. Despite minor variations in mean scores, the results suggest that the system was perceived as similarly usable across settings. This consistency underscores the flexibility and applicability of the VR training system in both long-term care homes and community-based environments.

Effects on Cognitive Function

Our study demonstrated that the 16-session leisure-based VR cognitive training program led to improvements in general cognition, processing speed, immediate and delayed recall, and executive function in older adults. These findings extend previous research on VR showing significant enhancements in general cognition, executive function, and memory [59,60], suggesting that leisure-based activities, especially gardening, might improve various aspects of cognitive ability. Gardening procedures such as sowing seeds, deciding the amount of watering, and making embossed flowers require different

XSL•FO
RenderX

cognitive abilities to complete the task successfully. Practicing such tasks might directly benefit cognitive performance. Effect sizes ranged from small to moderate, reflecting varying impacts of the training on cognitive outcomes.

Our results are consistent with earlier studies demonstrating the cognitive benefits of VR interventions for older adults [59,60]. The incorporation of gardening elements, such as sequencing planting tasks, adjusting watering levels, and managing virtual tools, appears to offer additional ecological validity and engagement. However, our findings diverged from previous work in visuospatial memory [60], which did not improve significantly. One possible explanation is that our VR training primarily emphasized object-to-object spatial relations. The assessment and management of self-to-object spatial relations in spatial relationships typically require enhanced perceptual and adjustment abilities from individuals and may therefore be more challenging than assessing and managing object-to-object spatial relations [61]. As a result, no significant enhancement was observed in visuospatial memory.

Strengths and Limitations

This study has several strengths, including the development of a fully interactive and task-oriented VR system, excellent adherence (100% program completion), and the integration of feedback from professionals across diverse care settings.

However, several limitations should be noted. First, the participants were community-dwelling older adults who regularly engaged in community activities. These individuals were likely more motivated and cognitively active, which may have contributed to more favorable training outcomes. This relatively homogeneous and proactive sample may limit the generalizability of the findings to the broader population of older adults, particularly those with lower levels of engagement or residing in institutional settings. Future studies may aim to recruit a more diverse population to enhance the external validity and applicability of the results across various subgroups of older adults.

Second, the history of motion sickness in older adults was assessed via self-report rather than using a standardized questionnaire. Although therapists continuously monitored and documented adverse events related to the use of the VR system during training, the study did not include a formal assessment of the incidence of cybersickness resulting from VR exposure. Future research could address these limitations by incorporating standardized tools to assess both the history of motion sickness and the occurrence of cybersickness, thereby enhancing the accuracy and reliability of the findings.

Third, participants were informally asked about their gardening experience—approximately one-third reported prior gardening activity—but no formal data were collected regarding their familiarity with gardening tasks or prior use of VR technologies. The lack of detailed assessment of prior experience with gardening or technology may have influenced participants' engagement and performance during the training. Future studies might consider systematically capturing such background characteristics to better interpret individual differences in outcomes.

Fourth, the absence of a control group limits the ability to attribute the observed cognitive and functional improvements solely to the VR-based intervention. Without a comparison group, it is challenging to rule out the influence of confounding variables such as spontaneous recovery, learning effects, or engagement in external training activities. Furthermore, this study could not determine whether VR training outperforms traditional cognitive training. To validate the findings of this study and demonstrate the possible superiority of VR training, future studies may use a randomized controlled trial design with different control groups, including an active control group (eg, traditional cognitive training), a passive control group (eg, watching leisure videos), and a no-intervention group.

Finally, this study only examined immediate pretest-posttest intervention effects. Without a follow-up assessment, it is not possible to determine the long-term sustainability of the observed cognitive improvements. Future studies may incorporate longitudinal follow-up assessments to determine whether the benefits of VR-based cognitive training are sustained over time and to further explore its potential for promoting long-term cognitive improvement.

In summary, addressing these limitations through rigorous study designs, including randomized controlled trials, standardized tools to assess cybersickness, detailed participant profiling, and longitudinal tracking, will strengthen the evidence base and clarify the potential of VR-based leisure training for cognitive function among community-dwelling older adults.

Conclusions

We developed a gardening-based, leisure-oriented cognitive training program using a VR system tailored for older adults in community settings. Our findings support the feasibility of this program, as demonstrated by full participant adherence, the absence of major adverse events, and high levels of acceptance. Older adults perceived the program as useful, expressed willingness to use it in the future, and felt supported by others. Usability ratings were also favorable, with both older adults and professionals in the community and in long-term care homes reporting positive user experiences. Preliminary cognitive outcomes showed statistically significant improvements in general cognition, processing speed, immediate and delayed recall, and executive function. However, due to the absence of a control group and the limited sample size, causal inferences cannot be drawn. These findings should be interpreted as initial evidence of potential benefit. Future research using randomized controlled trial designs, long-term follow-up assessments, and more diverse participant groups is needed to validate and extend these results.

XSL•FO
RenderX

Acknowledgments

The authors are grateful to all study participants. This study was supported by the Ministry of Science and Technology, Taiwan (MOST 110-2314-B-182-014 and MOST 111-2314-B-182-039-MY3); Chang Gung Memorial Hospital (CMRPD1L0192, CMRPD1L0193, BMRP553, and BMRPJ29); Healthy Aging Research Center, Chang Gung University—Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education in Taiwan (URRPD1Q0181); and Chang Gung University—University Social Responsibility by the Ministry of Education in Taiwan (EORPD3Q0011). This study was also supported by grants from Chang Gung Memorial Hospital (XMRPG3L1021 and XMRPG3L1022), sponsored by the Wang Chang Gung Charity Trust.

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

I-Ching Chuang and YW contributed to the study design, coordination, and project management. XH and YW interpreted the data and drafted and revised the manuscript. I Chen performed the data analyses and helped draft the manuscript. CW participated in subject recruitment and provided consultation. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questionnaire: acceptance of the virtual reality experience among older adults.
[DOCX File, 14 KB - games_v13i1e66673_app1.docx]

Multimedia Appendix 2
The System Usability Scale [46].
[DOCX File, 20 KB - games_v13i1e66673_app2.docx]

References

1. Bamidis PD, Vivas AB, Styliadis C, et al. A review of physical and cognitive interventions in aging. Neurosci Biobehav Rev 2014 Jul;44:206-220. [doi: 10.1016/j.neubiorev.2014.03.019] [Medline: 24705268]
2. Wimo A, Jönsson L, Bond J, Prince M, Winblad B, International AD. The worldwide economic impact of dementia 2010. Alzheimers Dement 2013 Jan;9(1):1-11. [doi: 10.1016/j.jalz.2012.11.006] [Medline: 23305821]
3. Gates NJ, Valenzuela MJ. Cognitive exercise and its role in cognitive function in older adults. Curr Psychiatry Rep 2010 Feb;12(1):20-27. [doi: 10.1007/s11920-009-0085-y] [Medline: 20425306]
4. Chiu HL, Chu H, Tsai JC, et al. The effect of cognitive-based training for the healthy older people: a meta-analysis of randomized controlled trials. PLoS ONE 2017;12(5):e0176742. [doi: 10.1371/journal.pone.0176742] [Medline: 28459873]
5. Martin M, Clare L, Altgassen AM, Cameron MH, Zehnder F. Cognition-based interventions for healthy older people and people with mild cognitive impairment. Cochrane Database Syst Rev 2011 Jan 19(1):CD006220. [doi: 10.1002/14651858.CD006220.pub2] [Medline: 21249675]
6. Gates NJ, Rutjes AW, Di Nisio M, et al. Computerised cognitive training for maintaining cognitive function in cognitively healthy people in late life. Cochrane Database Syst Rev 2019 Mar 13;3(3):CD012277. [doi: 10.1002/14651858.CD012277.pub2] [Medline: 30864187]
7. Reijnders J, van Heugten C, van Boxtel M. Cognitive interventions in healthy older adults and people with mild cognitive impairment: a systematic review. Ageing Res Rev 2013 Jan;12(1):263-275. [doi: 10.1016/j.arr.2012.07.003] [Medline: 22841936]
8. Reynolds GO, Willment K, Gale SA. Mindfulness and cognitive training interventions in mild cognitive impairment: impact on cognition and mood. Am J Med 2021 Apr;134(4):444-455. [doi: 10.1016/j.amjmed.2020.10.041] [Medline: 33385339]
9. Belleville S, Moussard A, Ansaldo AI, et al. Rationale and protocol of the ENGAGE study: a double-blind randomized controlled preference trial using a comprehensive cohort design to measure the effect of a cognitive and leisure-based intervention in older adults with a memory complaint. Trials 2019 May 22;20(1):282. [doi: 10.1186/s13063-019-3250-6] [Medline: 31118095]
10. Verghese J, Lipton RB, Katz MJ, et al. Leisure activities and the risk of dementia in the elderly. N Engl J Med 2003 Jun 19;348(25):2508-2516. [doi: 10.1056/NEJMoa022252] [Medline: 12815136]
11. Park DC, Lodi-Smith J, Drew L, et al. The impact of sustained engagement on cognitive function in older adults: the Synapse Project. Psychol Sci 2014 Jan;25(1):103-112. [doi: 10.1177/0956797613499592] [Medline: 24214244]

XSL•FO
RenderX

12. Lu LC, Lan SH, Hsieh YP, Yen YY, Chen JC, Lan SJ. Horticultural therapy in patients with dementia: a systematic review and meta-analysis. Am J Alzheimers Dis Other Demen 2020;35:1533317519883498. [doi: 10.1177/1533317519883498] [Medline: 31690084]
13. Wang D, MacMillan T. The benefits of gardening for older adults: a systematic review of the literature. Act Adapt Aging 2013 Apr;37(2):153-181. [doi: 10.1080/01924788.2013.784942]
14. Park SA, Lee AY, Park HG, Lee WL. Benefits of gardening activities for cognitive function according to measurement of brain nerve growth factor levels. Int J Environ Res Public Health 2019 Mar 2;16(5):760. [doi: 10.3390/ijerph16050760] [Medline: 30832372]
15. Park SA, Son SY, Lee AY, Park HG, Lee WL, Lee CH. Metabolite profiling revealed that a gardening activity program improves cognitive ability correlated with BDNF levels and serotonin metabolism in the elderly. Int J Environ Res Public Health 2020 Jan 15;17(2):541. [doi: 10.3390/ijerph17020541] [Medline: 31952145]
16. Tseng WSW, Ma YC, Wong WK, Yeh YT, Wang WI, Cheng SH. An indoor gardening planting table game design to improve the cognitive performance of the elderly with mild and moderate dementia. Int J Environ Res Public Health 2020 Feb 25;17(5):1483. [doi: 10.3390/ijerph17051483] [Medline: 32106583]
17. Wang K, Li Y, Chen X, et al. Gardening and subjective cognitive decline: a cross-sectional study and mediation analyses of 136,748 adults aged 45+ years. Nutr J 2024 Jun 4;23(1):59. [doi: 10.1186/s12937-024-00959-9] [Medline: 38834985]
18. Manera V, Chapoulie E, Bourgeois J, et al. A feasibility study with image-based rendered virtual reality in patients with mild cognitive impairment and dementia. PLoS ONE 2016;11(3):e0151487. [doi: 10.1371/journal.pone.0151487] [Medline: 26990298]
19. Arlati S, Di Santo SG, Franchini F, et al. Acceptance and usability of immersive virtual reality in older adults with objective and subjective cognitive decline. J Alzheimers Dis 2021;80(3):1025-1038. [doi: 10.3233/JAD-201431] [Medline: 33646164]
20. Bauer ACM, Andringa G. The potential of immersive virtual reality for cognitive training in elderly. Gerontology 2020;66(6):614-623. [doi: 10.1159/000509830] [Medline: 32906122]
21. Zhu K, Zhang Q, He B, Huang M, Lin R, Li H. Immersive virtual reality-based cognitive intervention for the improvement of cognitive function, depression, and perceived stress in older adults with mild cognitive impairment and mild dementia: pilot pre-post study. JMIR Serious Games 2022 Feb 21;10(1):e32117. [doi: 10.2196/32117] [Medline: 35188466]
22. Huang KT. Exergaming executive functions: an immersive virtual reality-based cognitive training for adults aged 50 and older. Cyberpsychol Behav Soc Netw 2020 Mar;23(3):143-149. [doi: 10.1089/cyber.2019.0269] [Medline: 31794673]
23. Kim H, Hong JP, Kang JM, et al. Cognitive reserve and the effects of virtual reality-based cognitive training on elderly individuals with mild cognitive impairment and normal cognition. Psychogeriatrics 2021 Jul;21(4):552-559. [doi: 10.1111/psyg.12705] [Medline: 33934441]
24. Kwan RYC, Liu JYW, Fong KNK, et al. Feasibility and effects of virtual reality motor-cognitive training in community-dwelling older people with cognitive frailty: pilot randomized controlled trial. JMIR Serious Games 2021 Aug 6;9(3):e28400. [doi: 10.2196/28400] [Medline: 34383662]
25. Makmee P, Wongupparaj P. Virtual reality-based cognitive intervention for enhancing executive functions in community-dwelling older adults. Psychosoc Interv 2022 Jul;31(3):133-144. [doi: 10.5093/pi2022a10] [Medline: 37361011]
26. Park JS, Jung YJ, Lee G. Virtual reality-based cognitive-motor rehabilitation in older adults with mild cognitive impairment: a randomized controlled study on motivation and cognitive function. Healthcare (Basel) 2020 Sep 11;8(3):335. [doi: 10.3390/healthcare8030335] [Medline: 32932997]
27. Zając-Lamparska L, Wiłkość-Dębczyńska M, Wojciechowski A, et al. Effects of virtual reality-based cognitive training in older adults living without and with mild dementia: a pretest-posttest design pilot study. BMC Res Notes 2019 Nov 27;12(1):776. [doi: 10.1186/s13104-019-4810-2] [Medline: 31775849]
28. Ren Y, Wang Q, Liu H, Wang G, Lu A. Effects of immersive and non-immersive virtual reality-based rehabilitation training on cognition, motor function, and daily functioning in patients with mild cognitive impairment or dementia: a systematic review and meta-analysis. Clin Rehabil 2024 Mar;38(3):305-321. [doi: 10.1177/02692155231213476] [Medline: 38289618]
29. Buele J, Avilés-Castillo F, Del-Valle-Soto C, Varela-Aldás J, Palacios-Navarro G. Effects of a dual intervention (motor and virtual reality-based cognitive) on cognition in patients with mild cognitive impairment: a single-blind, randomized controlled trial. J Neuroeng Rehabil 2024 Aug 1;21(1):130. [doi: 10.1186/s12984-024-01422-w] [Medline: 39090664]
30. Buele J, Varela-Aldás JL, Palacios-Navarro G. Virtual reality applications based on instrumental activities of daily living (iADLs) for cognitive intervention in older adults: a systematic review. J Neuroeng Rehabil 2023 Dec 19;20(1):168. [doi: 10.1186/s12984-023-01292-8] [Medline: 38110970]
31. Holloway H, Conroy B, Isbel S, D'Cunha NM. Immersive virtual reality in the promotion of health and well-being for people in residential aged care without cognitive impairment: a scoping review. Digit Health 2024;10:20552076241249568. [doi: 10.1177/20552076241249568] [Medline: 38680734]
32. Arlati S, Greci L, Mondellini M, et al. A virtual reality-based physical and cognitive training system aimed at preventing symptoms of dementia. In: Wireless Mobile Communication and Healthcare. MobiHealth 2017. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol 247: Springer; 2018. [doi: 10.1007/978-3-319-98551-0_14]

XSL•FO
RenderX

33. Park JH, Liao Y, Kim DR, et al. Feasibility and tolerability of a culture-based virtual reality (VR) training program in patients with mild cognitive impairment: a randomized controlled pilot study. Int J Environ Res Public Health 2020 Apr 27;17(9):3030. [doi: 10.3390/ijerph17093030] [Medline: 32349413]
34. Kraaijkamp JJM, van Dam van Isselt EF, Persoon A, Versluis A, Chavannes NH, Achterberg WP. eHealth in geriatric rehabilitation: systematic review of effectiveness, feasibility, and usability. J Med Internet Res 2021 Aug 19;23(8):e24015. [doi: 10.2196/24015] [Medline: 34420918]
35. Yun SJ, Kang MG, Yang D, et al. Cognitive training using fully immersive, enriched environment virtual reality for patients with mild cognitive impairment and mild dementia: feasibility and usability study. JMIR Serious Games 2020 Oct 14;8(4):e18127. [doi: 10.2196/18127] [Medline: 33052115]
36. Julious SA. Sample size of 12 per group rule of thumb for a pilot study. Pharm Stat 2005 Oct;4(4):287-291 [FREE Full text] [doi: 10.1002/pst.185]
37. Sim J, Lewis M. The size of a pilot study for a clinical trial should be calculated in relation to considerations of precision and efficiency. J Clin Epidemiol 2012 Mar;65(3):301-308. [doi: 10.1016/j.jclinepi.2011.07.011] [Medline: 22169081]
38. Totton N, Lin J, Julious S, Chowdhury M, Brand A. A review of sample sizes for UK pilot and feasibility studies on the ISRCTN registry from 2013 to 2020. Pilot Feasibility Stud 2023 Nov 21;9(1):188. [doi: 10.1186/s40814-023-01416-w] [Medline: 37990337]
39. Liao YY, Tseng HY, Lin YJ, Wang CJ, Hsu WC. Using virtual reality-based training to improve cognitive function, instrumental activities of daily living and neural efficiency in older adults with mild cognitive impairment. Eur J Phys Rehabil Med 2020 Feb;56(1):47-57. [doi: 10.23736/S1973-9087.19.05899-4] [Medline: 31615196]
40. Liao YY, Chen IH, Lin YJ, Chen Y, Hsu WC. Effects of virtual reality-based physical and cognitive training on executive function and dual-task gait performance in older adults with mild cognitive impairment: a randomized control trial. Front Aging Neurosci 2019;11:162. [doi: 10.3389/fnagi.2019.00162] [Medline: 31379553]
41. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003;27(3):425. [doi: 10.2307/30036540]
42. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319. [doi: 10.2307/249008]
43. Proctor E, Silmere H, Raghavan R, et al. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. Adm Policy Ment Health 2011 Mar;38(2):65-76. [doi: 10.1007/s10488-010-0319-7] [Medline: 20957426]
44. Syed-Abdul S, Malwade S, Nursetyo AA, et al. Virtual reality among the elderly: a usefulness and acceptance study from Taiwan. BMC Geriatr 2019 Aug 19;19(1):223. [doi: 10.1186/s12877-019-1218-8] [Medline: 31426766]
45. Nielsen J. Usability Engineering: Academic Press; 1994. [doi: 10.1016/B978-0-08-052029-2.50007-3]
46. Brooke J. SUS: a quick and dirty usability scale. In: Jordan PW, Thomas PW, Weerdmeester BA, McClelland IL, editors. Usability Evaluation in Industry: Taylor & Francis; 1996:189-194.
47. Nasreddine ZS, Phillips NA, Bédirian V, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. J Am Geriatr Soc 2005 Apr;53(4):695-699. [doi: 10.1111/j.1532-5415.2005.53221.x] [Medline: 15817019]
48. Tsai CF, Lee WJ, Wang SJ, Shia BC, Nasreddine Z, Fuh JL. Psychometrics of the Montreal Cognitive Assessment (MoCA) and its subscales: validation of the Taiwanese version of the MoCA and an item response theory analysis. Int Psychogeriatr 2012 Apr;24(4):651-658. [doi: 10.1017/S1041610211002298] [Medline: 22152127]
49. Wechsler D. Adult Intelligence Scale-Revised: Psychological Corporation; 1981.
50. Matarazzo JD, Herman DO. Base rate data for the WAIS-R: test-retest stability and VIQ-PIQ differences. J Clin Neuropsychol 1984 Nov;6(4):351-366. [doi: 10.1080/01688638408401227] [Medline: 6501578]
51. Jaeger J. Digit symbol substitution test: the case for sensitivity over specificity in neuropsychological testing. J Clin Psychopharmacol 2018 Oct;38(5):513-519. [doi: 10.1097/JCP.0000000000000941] [Medline: 30124583]
52. Wechsler D. Wechsler Memory Scale–Third Edition (WMS–III) Technical Manual: The Psychological Corporation; 1997.
53. Iverson GL. Interpreting change on the WAIS-III/WMS-III in clinical samples. Arch Clin Neuropsychol 2001 Feb;16(2):183-191. [doi: 10.1016/S0887-6177(00)00060-3] [Medline: 14590186]
54. Wang RY, Zhou JH, Huang YC, Yang YR. Reliability of the Chinese version of the trail making test and stroop color and word test among older adults. Int J Gerontol 2018 Dec;12(4):336-339. [doi: 10.1016/j.ijge.2018.06.003]
55. Seo EH, Lee DY, Choo IH, et al. Normative study of the Stroop Color and Word Test in an educationally diverse elderly population. Int J Geriatr Psychiatry 2008 Oct;23(10):1020-1027. [doi: 10.1002/gps.2027] [Medline: 18425990]
56. Sauro J. A Practical Guide to the System Usability Scale: Background, Benchmarks, & Best Practices: Measuring Usability LLC; 2011.
57. Niki K, Yahara M, Inagaki M, et al. Immersive virtual reality reminiscence reduces anxiety in the oldest-old without causing serious side effects: a single-center, pilot, and randomized crossover study. Front Hum Neurosci 2020;14:598161. [doi: 10.3389/fnhum.2020.598161] [Medline: 33536887]
58. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the System Usability Scale. Int J Hum Comput Interact 2008 Jul 29;24(6):574-594. [doi: 10.1080/10447310802205776]

XSL•FO
RenderX

59. Gamito P, Oliveira J, Alves C, Santos N, Coelho C, Brito R. Virtual reality-based cognitive stimulation to improve cognitive functioning in community elderly: a controlled study. Cyberpsychol Behav Soc Netw 2020 Mar;23(3):150-156. [doi: 10.1089/cyber.2019.0271] [Medline: 32031888]
60. Maeng S, Hong JP, Kim WH, et al. Effects of virtual reality-based cognitive training in the elderly with and without mild cognitive impairment. Psychiatry Investig 2021 Jul;18(7):619-627. [doi: 10.30773/pi.2020.0446] [Medline: 34265200]
61. Xiao C, Mou W, McNamara TP. Use of self-to-object and object-to-object spatial relations in locomotion. J Exp Psychol Learn Mem Cogn 2009 Sep;35(5):1137-1147. [doi: 10.1037/a0016273] [Medline: 19686010]

Abbreviations

DSST: digit symbol substitution test
ISO: International Organization for Standardization
MoCA: Montreal Cognitive Assessment
SCWT: Stroop Color Word Test
SUS: System Usability Scale
VR: virtual reality
WMS-III: Wechsler Memory Scale–Third Edition

Edited by A Coristine; submitted 19.09.24; peer-reviewed by B Najand, L Mendes, R Saryazdi; revised version received 12.08.25; accepted 12.08.25; published 09.10.25.

Please cite as:
Chuang IC, Huang XT, Chen IC, Wu YR, Wu CY
Feasibility, Usability, and Effects of Leisure-Based Cognitive Training Using a Fully Immersive Virtual Reality System in Older Adults: Single-Arm Pretest-Posttest Pilot Study
JMIR Serious Games 2025;13:e66673
URL: https://games.jmir.org/2025/1/e66673
doi:10.2196/66673

© I-Ching Chuang, Xiao-Ting Huang, I-Chen Chen, Yih-Ru Wu, Ching-Yi Wu. Originally published in JMIR Serious Games (https://games.jmir.org), 9.10.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Evaluating the Effectiveness of Immersive Virtual Reality Rehabilitation Games With Enhanced Visual Training for Hand Motor Function Improvement Using Electromyography: Randomized Controlled Trial

Faisal Amin[1], PhD; Asim Waris[1], PhD; Muhammad Jawad Khan[1], PhD; Muhammad Adeel Ijaz[1], PhD; Hammad Nazeer[2], PhD; Syed Omer Gilani[3], PhD; Fawwaz Hazzazi[4], PhD; Umer Hameed Shah[5], PhD

[1]Department of Biomedical Engineering and Sciences, School of Mechanical and Manufacturing Engineering, National University of Sciences and Technology, NUST, H-12, Islamabad, Pakistan

[2]Department of Mechatronics Engineering, Air University, Islamabad, Pakistan

[3]Department of Electrical, Computer, and Biomedical Engineering, Abu Dhabi University, Abu Dhabi, United Arab Emirates

[4]Department of Electrical Engineering, College of Engineering, Prince Sattam bin Abdulaziz University, Al-Kharj, Saudi Arabia

[5]Department of Mechanical Engineering and Artificial Intelligence Research Center, College of Engineering and Information Technology, Ajman University, Ajman, United Arab Emirates

Corresponding Author:
Asim Waris, PhD
Department of Biomedical Engineering and Sciences, School of Mechanical and Manufacturing Engineering, National University of Sciences and Technology, NUST, H-12, Islamabad, Pakistan

Abstract

Background: Hand motor dysfunction greatly reduces the performance of stroke survivors. This affects their ability to perform hand motor tasks effectively. Patients receive slow interventions due to interventional limitations in stroke rehabilitation, which can pose challenges for sustaining enduring improvements. We developed immersive virtual reality (VR) games that used an innovative approach to cognitive engagement within visual training feedback for achieving long-lasting improvements.

Objective: This study aimed to evaluate the effectiveness of fully immersive VR-based hand games compared with conventional physical therapy and to assess the correlations between electromyographic data and clinical outcome measures for improving hand motor function in patients with subacute stroke.

Methods: A randomized controlled study was conducted among 52 patients with subacute stroke who met the inclusion criteria. These patients were equally allocated to an experimental group (n=26) and a control group (n=26). The experimental group received both fully immersive VR-based hand game therapy and conventional physical therapy, whereas the control group received only conventional physical therapy. Owing to the nature of the intervention, the study was unblinded, and both therapists and patients were aware of the intervention. Both groups participated in intervention sessions 4 days a week for 6 weeks (24 sessions in total). Moreover, both groups underwent 2 weeks of follow-up. Clinical outcome measures, including the Fugl-Meyer Assessment-upper extremity (FMA-UE), Action Research Arm Test (ARAT), and Box and Block Test (BBT), were used to assess motor recovery and functional performance. The minimal clinically meaningful difference (MCID) was used for comparing clinical outcome measures to examine clinically meaningful improvements. Furthermore, the correlation between electromyography data and clinical outcome measures, and the weekly progression in movement performance were evaluated to identify improvements in hand motor function.

Results: After the intervention, there were significant differences in FMA-UE, ARAT, and BBT scores (all $P<.001$) between the experimental and control groups. The MCID findings illustrated that the experimental group had clinically meaningful improvements compared to the control group. There were significant correlations between electromyography signal features and clinical outcome measures (all $P<.05$) in both groups after rehabilitation. However, the experimental group exhibited strong positive correlations, while the control group exhibited moderate positive correlations. At follow-up, the mean movement accuracy was notably higher in the experimental group than in the control group (mean 83.59%, SD 1.1% vs mean 79.20%, SD 0.8%), indicating that hand motor function was effectively sustained through the use of the VR-based intervention in the experimental group.

Conclusions: The findings of this study revealed that VR-based hand games with enhanced visual training feedback substantially improved hand motor function in patients with subacute stroke.

Trial Registration: ClinicalTrials.gov NCT06582303; https://clinicaltrials.gov/study/NCT06582303

XSL•FO
RenderX

(JMIR Serious Games 2025;13:e74314) doi:10.2196/74314

KEYWORDS

electromyography; virtual reality; games; visual feedback; hand motor function

Introduction

Background

A stroke results in physical impairments that restrict a person's ability to engage in everyday activities. Stroke survivors need their upper extremities to be more controllable and adaptive for improving their motor recovery [1,2]. Conventional physical therapy for patients with stroke has evolved into advanced rehabilitation because of technological advancements [3]. The brain, through motor learning, adjusts to new situations, which results in neuronal plasticity and the creation of new neural networks [4]. To improve plasticity, consideration is given to movements that are task-oriented and recurring [5,6].

Virtual reality (VR) studies have implied that VR-based therapy improves neuroplasticity through motor movements in a VR environment [7]. The primary improvement in motor performance has been attributed to the strengthening of neural pathways, which can be achieved through visual feedback training that includes task-specific and repetitive workouts based on VR games [8]. Because of the interactive environment and the visual feedback, patients with stroke continue performing task-oriented VR game–based exercises that allow them to track their progress, making it feasible to achieve functional goals. Recent scientific developments have led to the creation of fully immersive VR devices, which can overcome the limitations of semi-immersive VR technologies [9].

Although traditional physical therapy and robotics are applied in stroke rehabilitation, they often cause limitations in achieving long-term improvements. These methods emphasize task-specific and physical exercises, which become tedious, reduce therapy motivation, and reduce patient involvement. Another main obstacle to effective stroke recovery is the repetitiveness of traditional exercises without any interactive engagement and motivation. This leads to decreasing repetitions of exercise that can ultimately limit the effectiveness of rehabilitation. Consequently, this results in slow progress in improving hand motor function, leading to inadequate compliance with therapy.

Furthermore, it has been revealed that VR-based rehabilitation increases the motivation and enjoyment of patients with stroke during therapy [9]. However, these studies concentrated mainly on motivation for improvement and neglected to sustain long-lasting improvements in hand motor function. Therefore, sustaining enduring improvements in motor function necessitates the potential integration of cognitive engagement within visual feedback. Implementing such approaches can expedite motor learning, substantially enhancing hand motor function.

Literature Review

Electromyography (EMG) has been extensively used in clinical research involving neurorehabilitation as an objective measure for functional status assessment and muscle strength [10]. Ying et al [5] created an exoskeleton arm for stroke survivors, and surface EMG signals were used for repeated training to examine distinctions in muscle amplitude. Classifier performance has been improved and developed for different features of EMG. Feature extraction techniques allow EMG signals to characterize and interpret the information for intentions and motor tasks associated with stroke rehabilitation [11]. Zhang and Zhou [12] suggested an investigative approach for intention potential in pattern recognition–based rehabilitation, using the performance of classification and high-density surface EMG.

When comparing fully immersive VR systems to semi-immersive VR systems, it can be observed that the latter systems are much more advanced and provide 3D graphics [3,13,14]. Rodríguez-Hernández et al [15] used Rheometric software to assess the combined effects of traditional therapy and VR on the motor functions of the upper extremities in patients with stroke. Moreover, Shahmoradi et al [16] used VR games in chronic stroke survivors.

According to Brannstrom recovery phases, findings indicate that playing games can improve participants' range of motion (ROM) [17]. A significant role in the process of adaptation to movements is played by the enhancement of neuroplasticity that occurs because of the successful use of visual feedback [17].

Subjective assessments in clinical settings and scenarios mostly fall into several scales based on clinical experience measurement techniques [3]. These include the Barthel Index [18], Action Research Arm Test (ARAT) [19], Box and Block Test (BBT) [20], and Fugl-Meyer Assessment-upper extremity (FMA-UE) [1]. The minimal clinically meaningful difference (MCID) signifies the minimal change in an outcome measure considered meaningful in clinical and research practice [21]. The MCID ascertains whether intervention score changes reflect a clinically meaningful patient improvement [22]. In subjective assessment, a doctor's training and expertise determine the reliability of the results of the clinical outcome measure examination [23]. Stroke survivors' physiological data [24], such as EMG signals, are used to conduct a functional assessment of the movements of patients with stroke [25].

Study Aim

This research aimed to evaluate the effectiveness of VR-based games for achieving enduring improvements in hand motor function, using EMG signals. This longitudinal assessment provides insights into the correlation among EMG signal features, clinical outcome measures (FMA-UE, ARAT, and BBT), and weekly progression in movement performance. The following VR-based hand games were developed: hit a rolling ball, grasp a balloon, swap hands, and grip a pencil. The movements in the VR-based hand games included flexion and extension, opening and closing, supination and pronation, and pinching. Correlations were evaluated before and after the intervention and at follow-up. Movement classification was

XSL•FO
RenderX

investigated for movement performance progression at baseline, week 4, week 6, and week 9 (follow-up) using k-nearest neighbor (KNN), random forest (RF), and support vector machine (SVM) classifiers.

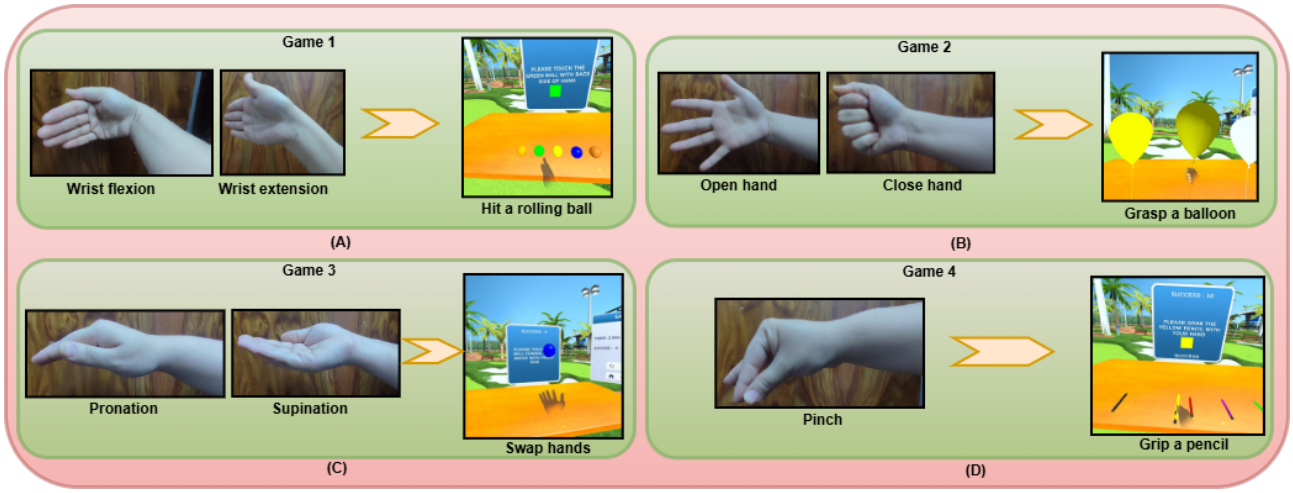
This research marks a valuable contribution to the EMG-based assessment of the effectiveness of customized, fully immersive VR-based hand games. These games used an innovative approach of integrating cognitive engagement within visual training feedback to attain long-lasting improvements.

Methods

Implementation of Rehabilitation Games Using Unity3D and a VR Headset

VR games were created for individuals with subacute stroke to help with hand rehabilitation. Four VR games were designed around hand movements, specifically flexion and extension (game 1: hit a rolling ball), opening and closing (game 2: grasp a balloon), supination and pronation (game 3: swap hands), and pinching (game 4: grip a pencil) (Figure 1). Visuals of the VR games are provided in Multimedia Appendix 1. The Unity3D game engine (version 2021.2.12f1; long-term support release) was used to develop the rehabilitation games. The Oculus Quest 2 VR device (processor: Qualcomm Snapdragon XR2; RAM: 6 GB; display: 1832×1920 pixels per eye; refresh rate: supports up to 120 Hz; field of view: approximately 100°; tracking: inside-out, 6 degrees of freedom, with hand-tracking capability via integrated cameras and controller input; operating system: Android-based Oculus OS; wireless connectivity: supports Wi-Fi 6 [802.11ax]) was used to run the games. Regarding system resource usage during gameplay, GPU utilization averaged 72% - 80% and CPU utilization averaged 60% - 70%. Moreover, the headset maintained a stable 72 - 90 frames per second without any visual discomfort or motion sickness, and memory utilization stayed within safe levels (less than 75% of the available RAM).

Figure 1. Virtual reality–based rehabilitation games based on hand movements: (A) wrist flexion and extension, (B) hand opening and closing, (C) pronation and supination, and (D) pinch.



Unity3D was selected for its reliability, enduring upgrades, high-quality physics and rendering, and robust compatibility with the Oculus Quest 2 VR device. Unity3D features native VR integration with the Oculus software development kit (SDK), ensuring dependable hand tracking and minimal latency, which are crucial for real-time rehabilitation activities. Moreover, Unity's C# environment made it possible to efficiently implement custom interaction logic and adaptive game mechanics. In addition, Unity offers a free license for academic and research purposes. Regarding runtime performance, no runtime crashes, overheating, or motion-latency issues were observed, which could adversely impact patient experience. Minor frame drops occurred when rendering multiple complex objects simultaneously; however, these were infrequent and did not interfere with the operations of the rehabilitation tasks.

Patients with stroke experienced cognitive engagement paradigms within visual feedback that involved attention, with patients selecting an object's color while focusing on visual cues, and decision-making engagement involved correctly choosing the color of an object following the appearance of objects randomly with varying colors on a table. In the games, patients used hand-tracking technology within the VR headset and observed their affected hand moving and interacting with virtual objects.

Game Development Process

Game development was carried out using Unity3D (2021.2.12f1) integrated with the Oculus SDK, which targeted the Oculus Quest 2 platform. Unity's XR Hand Tracking API was used to provide naturalistic hand interactions, which were enabled by implementing an interaction model based on inside-out tracking with hand tracking. Custom C# control scripts were used to map kinematic inputs, such as hand flexion, pinch detection, and rotation angles, to interaction states, such as grasp, release, and hit. The built-in PhysX engine was used to simulate object dynamics, such as gravity, collision, and resistance. To minimize clutter and maintain immersion, the user interface was created with only a few world-space overlays, such as progress bars and score counters. To enhance user involvement, instant feedback mechanisms were incorporated, which combined audio signals, such as success tones, to reward proper motions with visual cues, including object color changes and animations. To minimize GPU load and preserve shape recognition, optimization was given priority while rendering objects with a

XSL•FO
RenderX

low polygon count. This ensured fluid performance at a consistent frame rate of 72 frames per second in all games. Ultimately, the project was built and deployed as an Android package kit (APK) to run on the stand-alone VR headset Oculus Quest 2.

Immersive VR Game Design Mechanics

The developed VR-based hand games for hand rehabilitation are presented in Figure 1. Descriptions of the VR games are provided below.

Game 1: Hit a Rolling Ball

The patient approaches randomly generated rolling balls of various colors on a virtual table and flexes the affected wrist to strike a specific ball with the palm of their hand. Subsequently, the patient extends the same wrist to strike a specific ball with the back of their hand. This game challenges the patient to have exact control over wrist flexion and extension to time the strike properly. Faster speeds at more challenging levels require enhanced motor coordination, reaction time, and sustained attention.

Game 2: Grasp a Balloon

The patient approaches randomly generated balloons of various colors on a virtual table. The patient closes the affected hand to pick up a specific balloon (grasping movement). Once the hand is closed, the patient holds the balloon for 5 seconds. This holding of the balloon helps to strengthen the hand muscles. After the grasping time, the balloon is released, and the hand is considered to be open and ready for the subsequent interaction. This game challenges the patient to sustain grasping without exerting excessive effort and to pay selective attention to the target. It emphasizes grasping precision and decision-making.

Game 3: Swap Hands

The patient extends the affected hand in a supinated position, and a randomly generated virtual ball descends from the top with appropriate speed and touches the palm of the patient's hand. The patient then instantly swaps their hand from the supinated position to a pronated position, and another virtual ball descends and touches the back of the patient's hand. This game challenges the patient to rotate their forearm quickly, precisely, and with coordination. At higher levels, the quick transition demand makes it more challenging to sustain controlled and smooth movements. The game emphasizes movement adaptability and responsiveness.

Game 4: Grip a Pencil

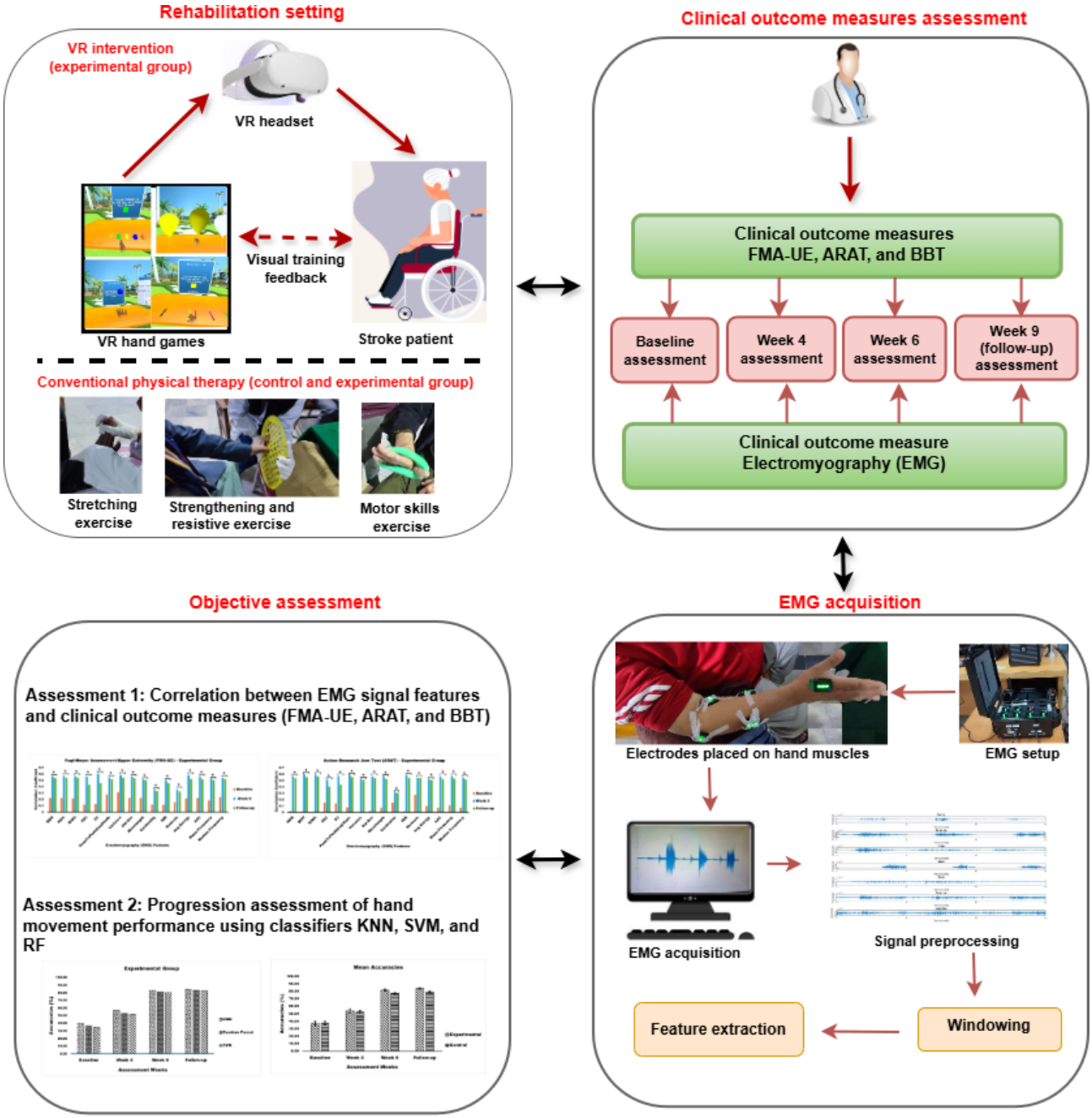
The patient approaches randomly generated pencils of various colors with wide gaps on a virtual table. The patient pinches a specific pencil with the affected hand to pick it up. The pencil is held for 5 seconds before placing it back on the virtual table. This game challenges the patient to develop stability when gripping a small object, fine motor skills, and selective color recognition. Narrow spacing between pencils requires more dexterity and increases the risk of errors. The game emphasizes pinch grip strength and accuracy.

Ethical Considerations

This study was carried out in the physiotherapy department of a hospital. The study was approved by the research ethics committee (reference number: BMES/REC/22/027), and it followed the institutional guidelines and the principles of the Declaration of Helsinki. The study was registered on ClinicalTrials.gov (NCT06582303). Registration was completed retrospectively with respect to participant enrollment but prospectively with respect to data analysis and outcome assessment. Written informed consent was obtained from each patient for participation in the study. No compensation was provided to the participants. All collected data have been treated anonymously, and identifiable information has not been linked to research data. Figure 2 illustrates the rehabilitation setting and assessment framework used in this research.

XSL•FO
RenderX

Figure 2. The rehabilitation settings and assessment framework used in this research. ARAT: Action Research Arm Test; BBT: Box and Block Test; FMA-UE: Fugl-Meyer Assessment-upper extremity; KNN: k-nearest neighbor; RF: random forest; SVM: support vector machine; VR: virtual reality.



Inclusion and Exclusion Criteria

Based on the inclusion and exclusion criteria, patients were registered in this study. The inclusion criteria were as follows: (1) Modified Ashworth Scale score <4, (2) Fugl-Meyer Assessment-upper limb score between 25 and 55, (3) Montreal Cognitive Assessment score ≥21, (4) subacute stroke type according to the American Stroke Association (ASA) and existing literature (the subacute phase of stroke is generally described as the period ranging from 2 weeks to 6 months after stroke onset), and (5) age ≥18 years.

The exclusion criteria were as follows: (1) contractures resulting from burns, (2) wrist impairments caused by muscular disorders, (3) presence of external fixation devices, and (4) presence of vestibular problems.

Sample Size and Recruitment

G*Power software 3.9.1.4 (Heinrich-Heine-Universität Düsseldorf) was used to evaluate the sample size for this study. The parameters included an effect size (d) of 0.80, α error of .05, and power of 0.80. Considering a dropout rate of 10%, 56 individuals with subacute stroke were included. Four patients who did not meet the eligibility criteria were excluded from the trial.

Patients were recruited from the physiotherapy department of the hospital. During the initial screening, patients with a diagnosis of subacute stroke were first identified by a physiotherapist. Clinical coordinators who were trained physiotherapists on our team then directly approached these patients in the outpatient waiting areas. The coordinators explained the study's purpose, procedures, potential risks, and

XSL•FO
RenderX

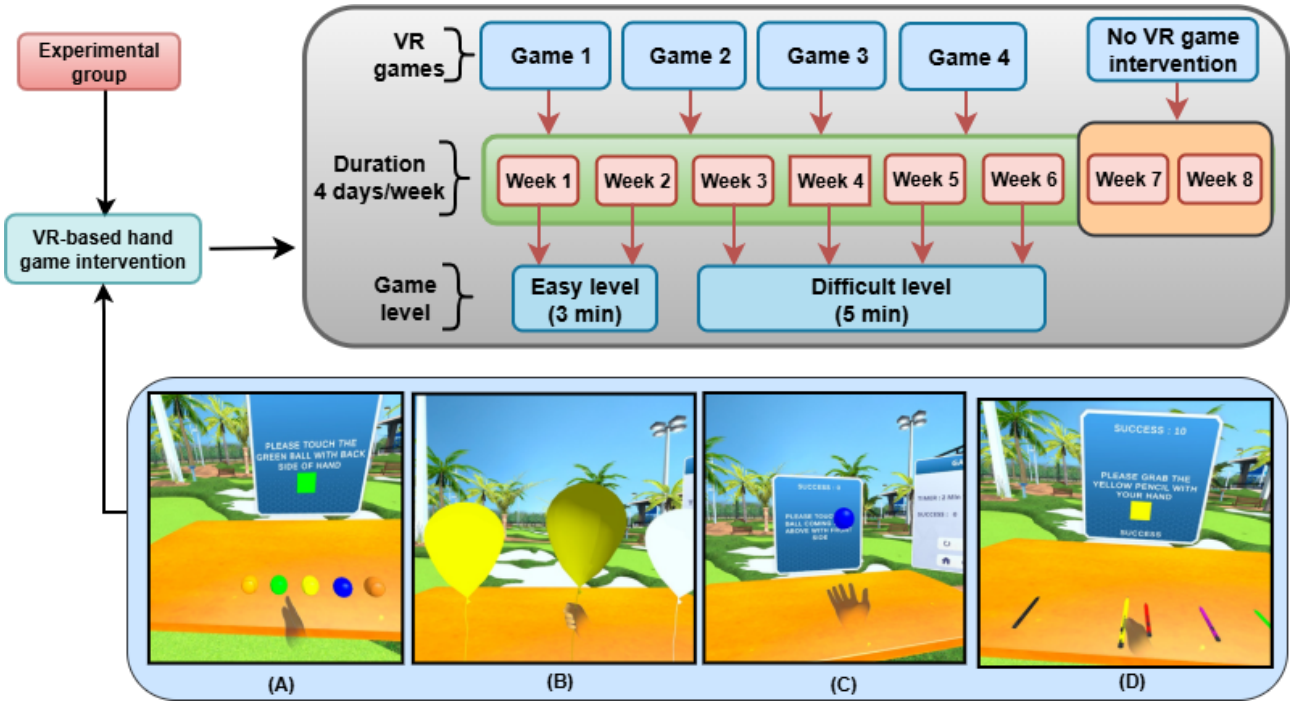
benefits. Patients who expressed interest in participating underwent screening against the predefined inclusion and exclusion criteria. Those who met all the eligibility criteria were invited to participate. Written informed consent was obtained before enrollment, and participants underwent randomization.

This study used the sealed envelope method to ensure proper allocation concealment. The randomization procedure was conducted by an independent individual who was not involved in any aspect of the study. This person used the sealed envelope method to allocate participants equally into the experimental and control groups. Before the trial began, a randomization sequence was created to determine the assignment of participants to either the experimental or control group. Each assignment was placed inside an identical, sequentially numbered, opaque, and sealed envelope to prevent prior knowledge of the allocation. Patients were assigned the next available sequentially numbered envelope when they were confirmed eligible and provided informed consent. The envelope was only opened after recording participants' details, ensuring the allocation process remained concealed.

Interventions for the Experimental and Control Groups

The experimental group received both VR game therapy and conventional physical therapy, while the control group only received conventional physical therapy. The experimental group used the VR game intervention protocol paradigm for hand rehabilitation games, as shown in Figure 3. Conventional physical therapy included ROM exercises, stretching, resistance exercises, and strengthening exercises. Both groups participated in intervention sessions 4 days a week for 6 weeks (24 sessions in total). Moreover, both groups completed 2 weeks of follow-up. The experimental group received 24 minutes of VR game therapy and 24 minutes of conventional physical therapy per session in weeks 1 and 2, and 40 minutes of VR game therapy and 40 minutes of conventional physical therapy per session in weeks 3-6. The control group received 48 minutes of conventional physical therapy per session in weeks 1 and 2, and 80 minutes of conventional physical therapy per session in weeks 3-6. The conventional physical therapy protocol included ROM exercises for the joints (shoulder, elbow, and wrist), strengthening and resistance exercises for weak muscles (using a power web and gym equipment), muscle stretching (shoulder flexors, abductors, external rotators, elbow and wrist extensors, and hand musculature), and motor skills training for the upper limb (Thera putty and occupational therapy tools). The amount of time spent on VR game therapy and conventional therapy in the experimental group was equivalent to the amount of time spent on conventional therapy in the control group.

Figure 3. Virtual reality (VR) game intervention protocol in the experimental group. The games are as follows: (A) hit a rolling ball, (B) grasp a balloon, (C) swap hands, and (D) grip a pencil [26].



After week 2, the extended intervention duration in the experimental group allowed patients to increase their repetition volume in more challenging conditions, which in turn improved their intensity and functional engagement. Similarly, the longer duration in the control group allowed for increased repetition of conventional exercises at higher volume. Consequently, to ensure that any disparities in outcomes were not the result of unequal therapy exposure but rather the result of the qualitative content of the intervention, all groups were subjected to the same escalation in therapy time.

Due to the nature of the intervention, which necessitated active participation and interaction with the VR system, blinding participants and therapists was not feasible. To minimize potential bias, we standardized training techniques and provided comprehensive guidelines to physiotherapists to ensure uniformity in implementation across groups. Patients were also

XSL•FO
RenderX

provided with uniform instructions to ensure consistency in engagement across participants.

Outcome Measures

Overview

The primary outcome measures included the correlation between EMG signal features and clinical outcome measures (FMA-UE, ARAT, and BBT), and the weekly progression of movement performance. FMA-UE, ARAT, and BBT outcome measures have been evaluated and discussed in [26] for the experimental group and the control group. These were used to evaluate the efficacy of the hand's motor function. The MCID values of the FMA-UE, ARAT, and BBT reported in previous studies [21,27,28] serve as benchmarks for interpreting the progression of the clinical outcome measures of subacute stroke. The clinical outcome measures (FMA-UE, ARAT, and BBT) are explained below.

FMA-UE Outcome Measure

The FMA-UE was used to assess stroke survivors' upper extremities for the recovery of motor function, which involved performance-based measurement. The FMA-UE had high inter- and intrarater reliability (r=0.98 - 0.99), responsiveness, and construct validity in previous studies [21,22]. FMA-UE construct validity has shown significant correlations with the Functional Independence Measure and the ARAT, with reported correlation values of 0.61 - 0.94 [22]. This scale includes 33 items used to evaluate motor function. The grading criteria range from 0 to 66, with a score of 0 denoting severe impairment and 66 denoting normal motor function. A higher score is indicative of an improvement in motor function for the individual. The FMA-UE was administered by a physical therapist in a clinical setting. The reported MCID of the FMA-UE for the subacute stroke stage is 9 points [21,29].

ARAT Outcome Measure

The ARAT evaluates the functional ability performance of the upper limbs of stroke survivors. This test emphasizes reaching, grasping, pinching, and general movement activities. ARAT performance was found to be significantly correlated with FMA-UE and BBT performance. Prior studies revealed high intra- and interrater reliability (intercorrelation coefficient >0.98) and validity (r=0.81 - 0.97) of the ARAT to evaluate the functional ability performance of stroke survivors [30,31]. The overall score ranges from 0 to 57, with each item rated on a 4-point scale. Severe disability is indicated by the lowest score of 0, while normal upper extremity function is determined by the highest score of 57 [30]. A higher score on the ARAT is indicative of enhanced functional capabilities of the upper extremities. A physical therapist administered the ARAT in a clinical setting. The reported MCID of the ARAT for stroke is 6 points [27,32].

BBT Outcome Measure

The BBT is used to evaluate the hand dexterity of patients who have experienced a stroke. The BBT has been identified as a reliable and valid assessment method for hand dexterity, including fine motor skills, in prior investigations [33]. Patients are instructed to transfer wooden blocks from one compartment to another within a time limit of 1 minute. The final score is determined by the number of wooden blocks transferred during the 1-minute trial [27]. The BBT was administered by a physical therapist in a clinical setting. The reported MCID of the BBT for stroke is 6 points [27,28,32].

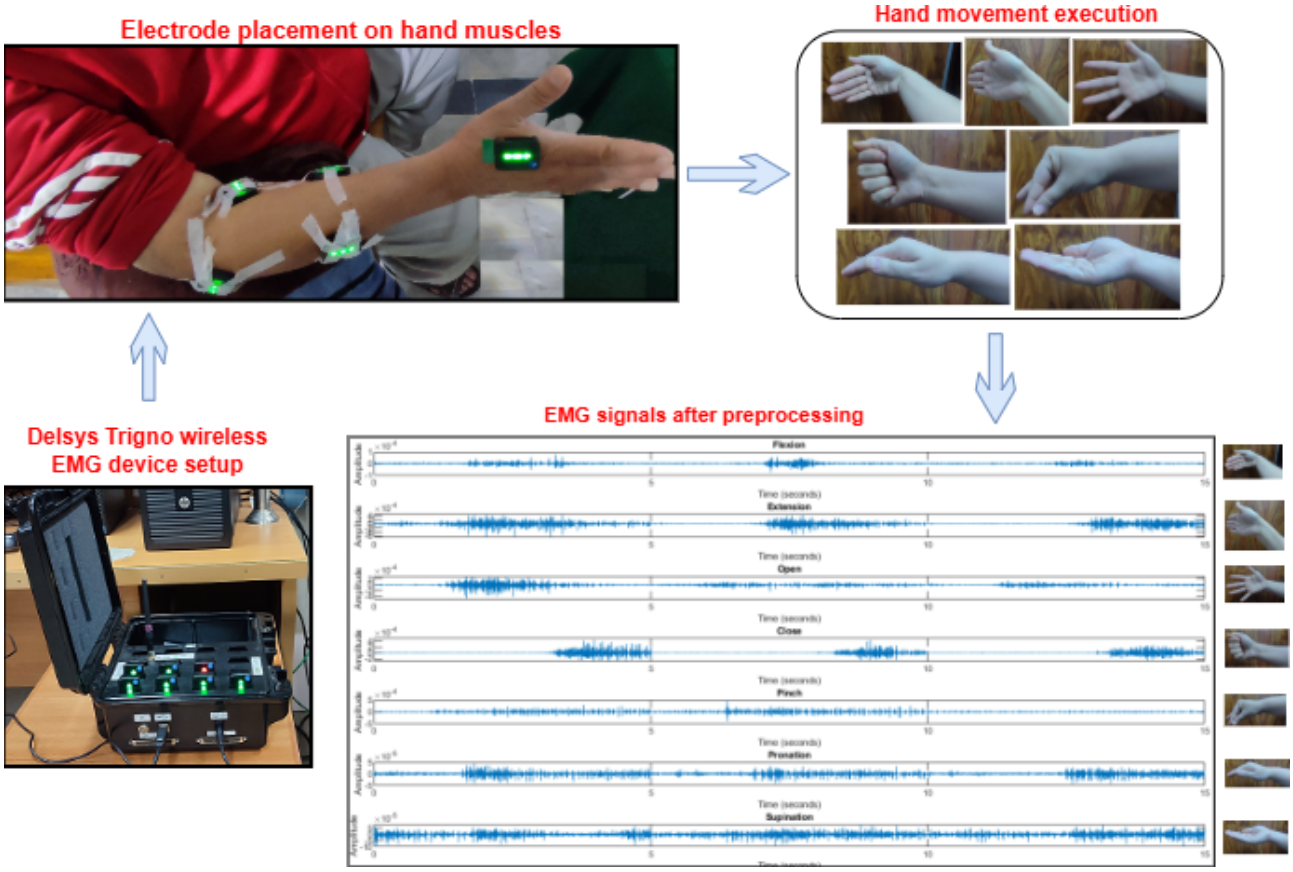
EMG Assessment

EMG Acquisition

EMG signals were recorded for 7 movements by using the Delsys Trigno Wireless EMG system, which has a sampling frequency of 1926 Hz. EMG electrodes were placed on 5 forearm muscles, including the extensor digitorum communis, flexor digitorum superficialis, supinator, pronator teres, and dorsal interossei. The hand movements included flexion, extension, opening, closing, pinch, pronation, and supination. EMG signals were recorded at baseline, week 4, week 6, and week 9 (follow-up). Each patient was positioned upright on a chair in front of an LCD screen. A MATLAB graphical user interface was developed to assist patients in performing the required movements with the help of visual cues. Every patient in the EMG recording session executed the protocol by initially keeping their affected hand at rest for 5 seconds and then using their muscles to contract for a further 5 seconds to complete the movement. Each movement was performed 3 times with relaxation intervals between the repetitions. Each movement and relaxation repetition lasted for 5 seconds. EMG acquisition is presented in Figure 4.

XSL•FO
RenderX

Figure 4. Electromyography (EMG) acquisition. EMG signals are recorded using the Delsys Trigno wireless EMG system. EMG electrodes are placed on 5 forearm muscles, including the extensor digitorum communis, flexor digitorum superficialis, supinator, pronator teres, and dorsal interossei. The EMG signals show hand movements that include flexion, extension, opening, closing, pinch, pronation, and supination.



EMG Preprocessing

EMG signals are intrinsically contaminated with noise due to power line interference and motion artifacts. Therefore, it is of utmost importance to remove the noise. Power line noise was removed by using a 50-Hz notch filter. A fourth-order Butterworth bandpass filter of 20 - 450 Hz was applied to remove motion artifacts and noise from the acquired EMG signal data. In the next step, EMG signals corresponding to movements were separated from the rest of the EMG signals to ensure increased prediction accuracy and boost training. The windowing technique ensures a continuous distribution of pertinent data [34,35]. A window in the overlap windowing approach is described by both the window size and the step size, as opposed to the disjoint windowing technique, where a window is characterized by its size. Following preprocessing, this assessment used overlapping windowing segmentation with a window size of 250 ms and an overlapping size of 50% (125 ms). Since it contains the most significant information in the signal, most researchers have included it in their studies [36].

EMG Feature Extraction

The EMG signal features used in this study are based on both time-domain features and frequency-domain features. The time-domain features include root mean square, mean absolute value, integrated EMG, slope sign change, zero crossing, peak-to-peak amplitude, variance, SD, waveform length, cardinality, IQR, kurtosis, average energy, and average amplitude change. The frequency-domain features include mean frequency and median frequency. Both feature groups are being widely used in EMG research and practice [35,37-41]. These features were extracted from each window segment obtained in the previous step.

Statistical Analysis of Correlations and Movement Performance

Data from the experimental and control groups were analyzed using SPSS 21 statistical software (IBM Corp). Variables were tested for normality using the Shapiro-Wilk test. The correlation between EMG signal features and clinical outcome measures was examined using the Spearman rank correlation coefficient. The statistical significance was set at $P<.05$. Machine learning techniques were applied to the EMG data of stroke patients to evaluate the movement performance of the experimental and control groups at assessment points, including baseline, week 4, week 6, and week 9 (follow-up). Three classifiers (KNN, RF, and SVM) were used for movement performance. These classifiers have been extensively used on EMG data for different movement classifications and muscle activities [39,42-44] and have therefore been adopted in this study. KNN assigns class labels according to the Euclidean distance to the closest neighbors in the feature space. RF constructs ensembles of decision trees by partitioning the feature space into several segments. SVM identifies an ideal hyperplane, whether linear or kernel-based, to distinguish between classes. The classifiers were tested for normality using the Shapiro-Wilk test. Movement performance was tested using ANOVA and the paired 2-tailed *t* test.

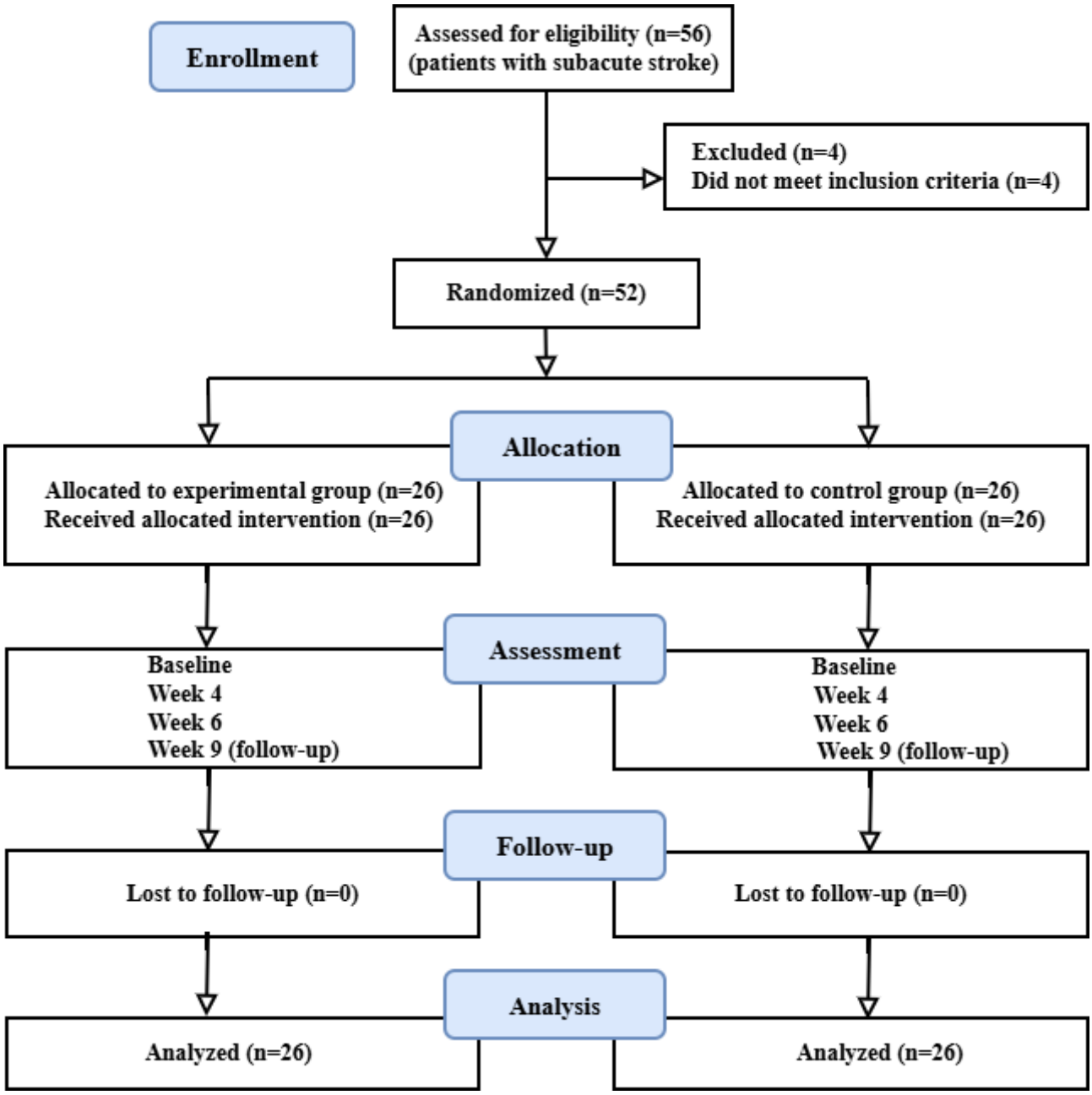
XSL•FO
RenderX

Results

Analysis Structure

The participants in this research were 52 patients with subacute stroke who met the eligibility criteria. These patients were randomly allocated to the experimental and control groups, with 26 patients in each group. The results have been divided into three parts: (1) comparison of clinical outcome measure scores with the MCID, (2) examination of the correlation between muscle activity (EMG) and hand function clinical outcome measures, and (3) assessment of the weekly progress of hand movement performance using EMG-based classification accuracies. Figure 5 presents the CONSORT (Consolidated Standards of Reporting Trials) [45] flow diagram. The CONSORT-EHEALTH checklist is provided in Checklist 1.

Figure 5. CONSORT (Consolidated Standards of Reporting Trials) flowchart.



Demographic and Clinical Data

Demographic and clinical data are presented in Table 1. The Mann-Whitney *U* test revealed no significant differences in characteristics between the experimental and control groups (all $P>.05$). Furthermore, the *P* values for demographic and clinical data were high, indicating that the groups exhibited comparable variations.

XSL•FO
RenderX

Table . Demographic and clinical data.

Characteristic	Control group (n=26)	Experimental group (n=26)	*P* value[a]
Gender, n (%)			.56
Male	18 (69)	16 (62)	
Female	8 (31)	10 (38)	
Age (years), mean (SD)	49.8 (9.9)	51.8 (12.9)	.78
Affected hand, n (%)			.27
Right	15 (58)	11 (42)	
Left	11 (42)	15 (58)	
MoCA[b], mean (SD)	22.5 (1.9)	22.6 (1.9)	.58
FMA-UE[c], mean (SD)	36.6 (7.4)	40.84 (8.6)	.19
MAS[d], n (%)			.32
Grade 0	7 (27)	10 (39)	
Grade 1	7 (27)	7 (27)	
Grade 1+	6 (23)	5 (19)	
Grade 2	6 (23)	4 (15)	

[a]Mann-Whitney *U* test.

[b]MoCA: Montreal Cognitive Assessment.

[c]FMA-UE: Fugl-Meyer Assessment-upper extremity.

[d]MAS: Modified Ashworth Scale.

Comparison of Clinical Outcome Measure Scores With the MCID

The mean differences in the clinical outcome measure instruments (FMA-EU, ARAT, and BBT) between groups and their comparisons with the MCID are presented in Table 2 [26]. At baseline (preintervention), there were no significant differences in the scores of FMA-UE (*P*=.19), ARAT (*P*=.13), or BBT (*P*=.09) between the experimental and control groups. However, at follow-up (postintervention), there were significant differences in the scores of all outcome measures (all *P*<.001) between the experimental and control groups. All mean differences in FMA-UE, ARAT, and BBT scores were greater than the corresponding MCIDs (9, 6, and 6 points, respectively).

XSL•FO
RenderX

Table . Clinical outcome measure scores and comparison with the MCID[a].

Clinical outcome measure and assessment time point	Experimental group, mean (SD)	Control group, mean (SD)	*P* value[b]	Mean difference between groups	Comparison with the MCID
FMA-UE[c] score				9.46	Greater than the MCID of 9 points
Preintervention (baseline)	40.84 (8.6)	36.61 (7.4)	.19		
Postintervention (follow-up)	55.92 (7.3)	42.23 (7.9)	<.001		
ARAT[d] score				8.96	Greater than the MCID of 6 points
Preintervention (baseline)	24.23 (4.9)	21.23 (8.2)	.13		
Postintervention (follow-up)	37.46 (9.1)	25.50 (6.7)	<.001		
BBT[e] score				20.46	Greater than the MCID of 6 points
Preintervention (baseline)	15.11 (6.7)	11.76 (8.6)	.09		
Postintervention (follow-up)	39.50 (12.8)	15.69 (8.4)	<.001		

[a]MCID: minimal clinically meaningful difference.

[b]Mann-Whitney *U* test for comparison between the experimental and control groups.

[c]FMA-UE: Fugl-Meyer Assessment-upper extremity.

[d]ARAT: Action Research Arm Test.

[e]BBT: Box and Block Test.

Correlations Between EMG Signal Features and Clinical Outcome Measures

The Spearman rank correlation coefficient was used to determine the correlation between EMG signal features and clinical outcome measures. The correlation between EMG signal features and clinical outcome measures provides insights into the association of muscle activity with motor function recovery, functional abilities, and hand dexterity variability among stroke survivors. The correlation was analyzed at baseline, week 6, and week 9 (follow-up). Figures 6-11 illustrate the results of the correlations between the experimental and control groups. Both groups exhibited a weak and nonsignificant (all $P>.05$) relationship between EMG features and clinical outcome measures at baseline. Results at follow-up revealed significant (all $P<.05$) strong positive correlations in the experimental group and significant (all $P<.05$) moderate positive correlations in the control group.

XSL•FO
RenderX

Figure 6. Correlation between electromyography (EMG) signal features and Fugl-Meyer Assessment-upper extremity scores at each time measurement in the experimental group. AAC: average amplitude change; IEMG: integrated EMG; MAV: mean absolute value; RMS: root mean square; SSC: slope sign change; ZC: zero crossing. *P<.05.

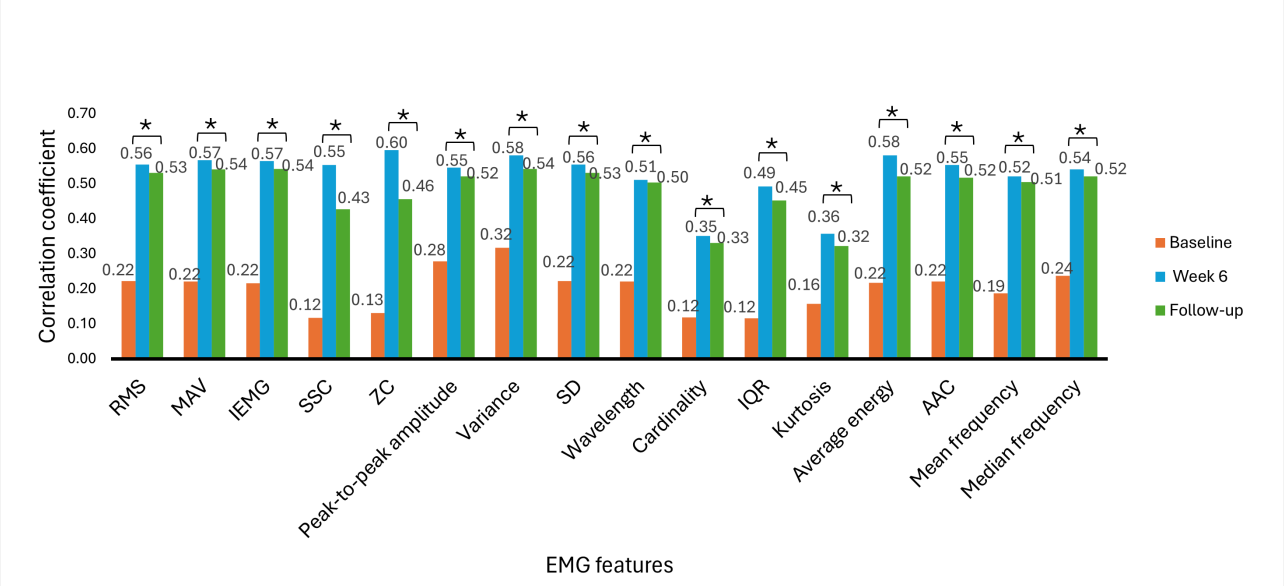
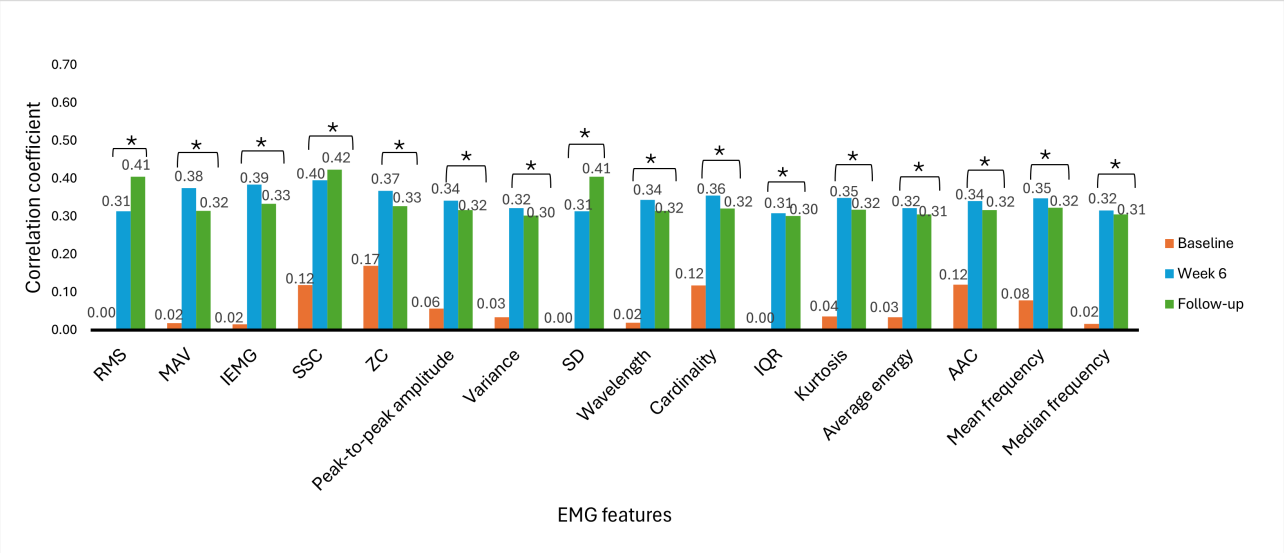


Figure 7. Correlation between electromyography (EMG) signal features and Fugl-Meyer Assessment-upper extremity scores at each time measurement in the control group. AAC: average amplitude change; IEMG: integrated EMG; MAV: mean absolute value; RMS: root mean square; SSC: slope sign change; ZC: zero crossing. *P<.05.



XSL•FO
RenderX

Figure 8. Correlation between electromyography (EMG) signal features and Action Research Arm Test scores at each time measurement in the experimental group. AAC: average amplitude change; IEMG: integrated EMG; MAV: mean absolute value; RMS: root mean square; SSC: slope sign change; ZC: zero crossing. *P<.05.

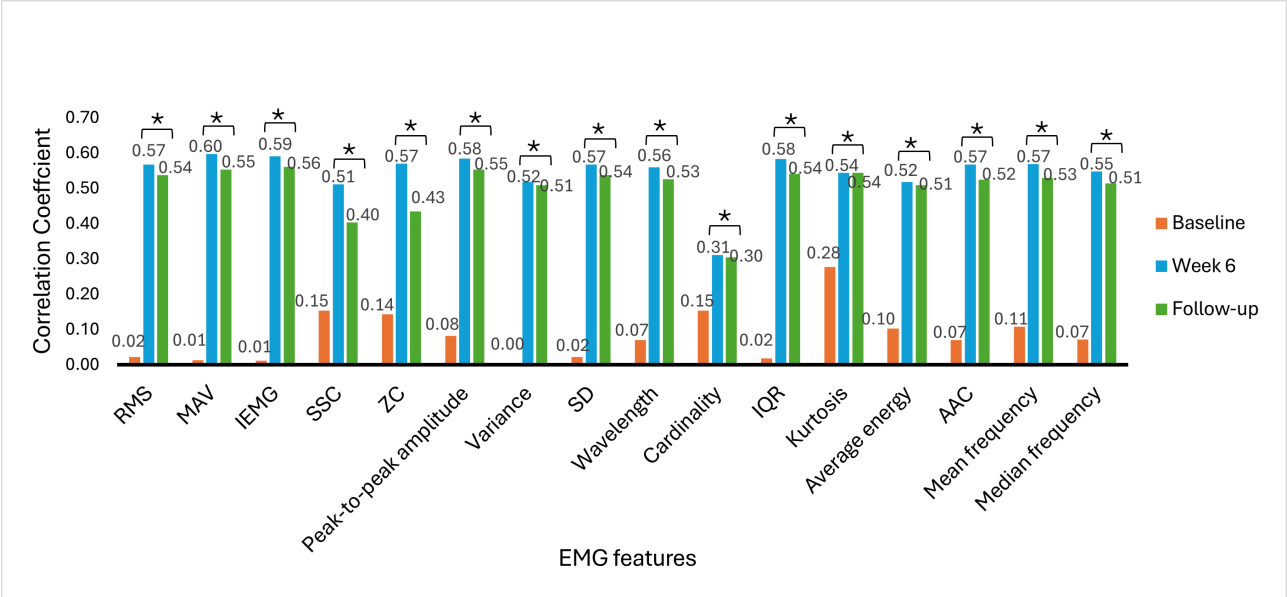
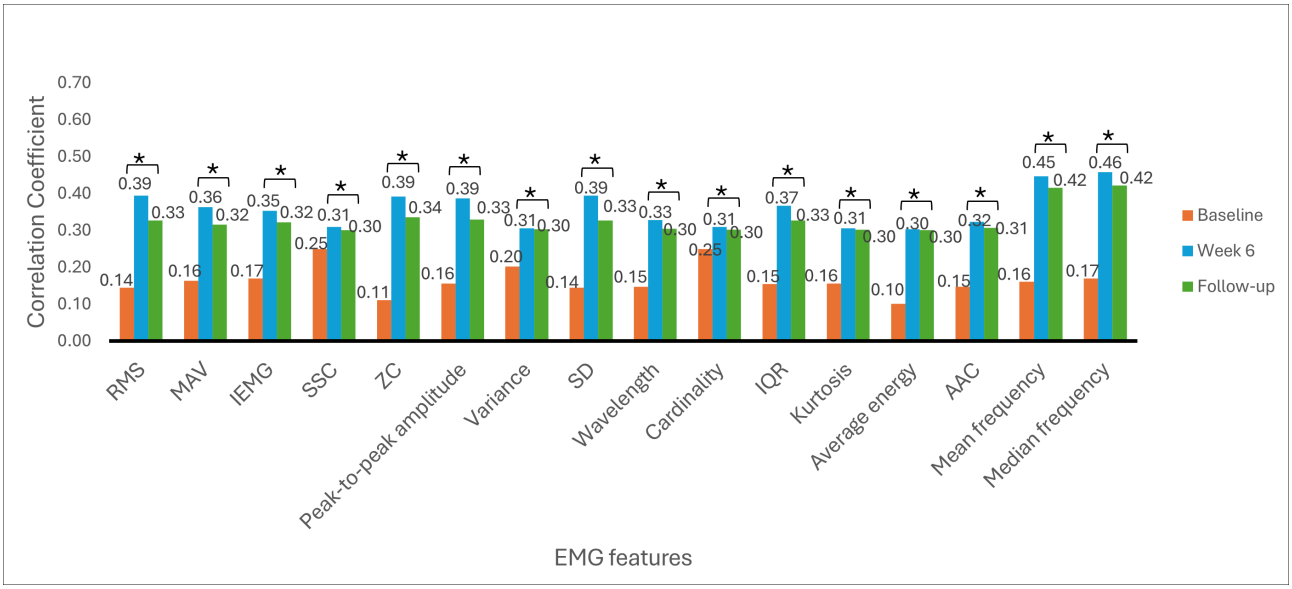


Figure 9. Correlation between electromyography (EMG) signal features and Action Research Arm Test scores at each time measurement in the control group. AAC: average amplitude change; IEMG: integrated EMG; MAV: mean absolute value; RMS: root mean square; SSC: slope sign change; ZC: zero crossing. *P<.05.



XSL•FO
RenderX

Figure 10. Correlation between electromyography (EMG) signal features and Box and Block Test scores at each time measurement in the experimental group. AAC: average amplitude change; IEMG: integrated EMG; MAV: mean absolute value; RMS: root mean square; SSC: slope sign change; ZC: zero crossing. *P<.05.

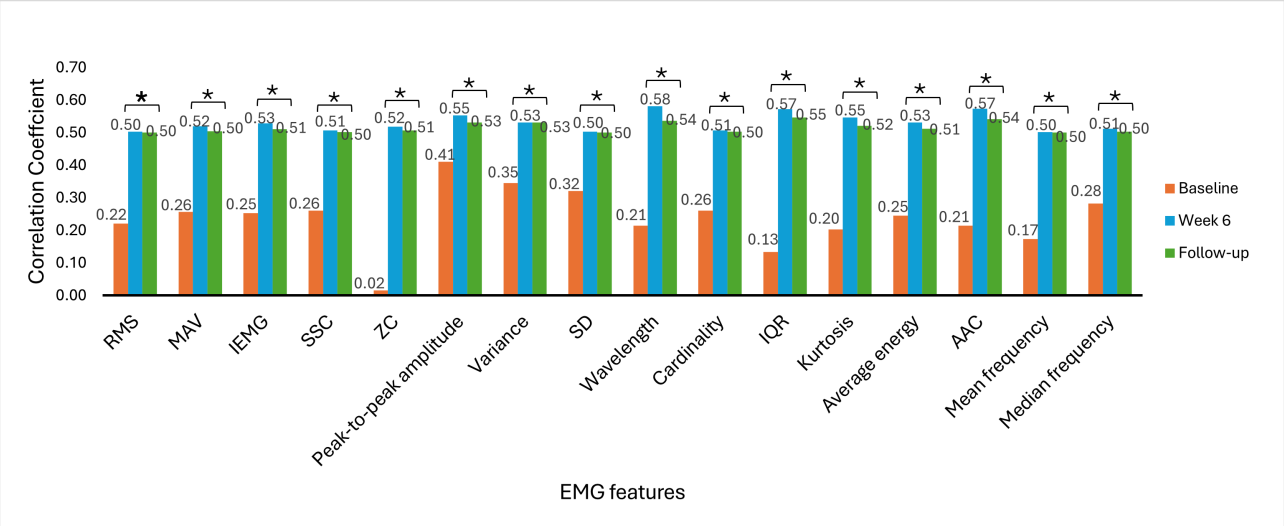
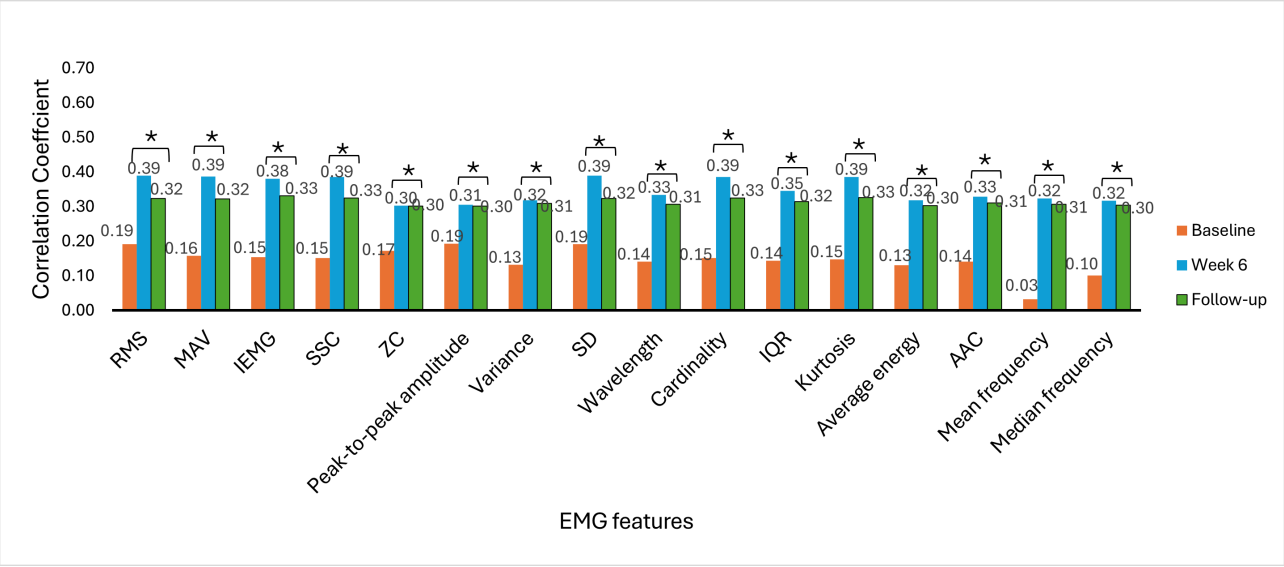


Figure 11. Correlation between electromyography (EMG) signal features and Box and Block Test scores at each time measurement in the control group. AAC: average amplitude change; IEMG: integrated EMG; MAV: mean absolute value; RMS: root mean square; SSC: slope sign change; ZC: zero crossing. *P<.05.



Progression Analysis of Movement Performance

Machine learning algorithms were used on the EMG data of patients with stroke to assess the progression of movement performance at different time points in the experimental and control groups. The analysis focused on the weekly movement performance measured by the classification accuracy of EMG signals. There were significant differences between the paired weeks in KNN, RF, and SVM values (all P<.05) (Table 3). Furthermore, there were significant differences between the 2 groups in KNN, RF, and SVM values (all P<.001) (Table 3).

XSL•FO
RenderX

Table . Effect of virtual reality games on the progression of weekly movement performance.

Classifier	Assessment time points				*P* value[a] for difference between groups
	Baseline and week 4	Week 4 and week 6	Week 6 and week 9	Week 9 (follow-up) and baseline	
KNN[b]					<.001
Value, mean (SD)	15.93 (1.6)	24.61 (2.1)	1.59 (0.1)	42.14 (3.6)	
P value[c]	.045	.04	.04	.04	
RF[d]					<.001
Value, mean (SD)	15.26 (1.2)	27.34 (1.6)	1.94 (0.1)	44.50 (2.9)	
P value[c]	.04	.03	.002	.03	
SVM[e]					<.001
Value, mean (SD)	16.16 (1.3)	27.20 (1.7)	1.99 (0.1)	45.37 (3.2)	
P value[c]	.04	.03	.02	.03	

[a]ANOVA for mean difference comparison between the experimental and control groups.

[b]KNN: k-nearest neighbor.

[c]Paired *t* test for mean difference comparison between time points within the experimental and control groups.

[d]RF: random forest.

[e]SVM: support vector machine.

The progressions of classification accuracies (KNN, RF, and SVM) obtained from EMG signal features are depicted in Figures 12 and 13. These accuracies were evaluated at baseline, week 4, week 6, and week 9 (follow-up) assessments in both the experimental and control groups. For KNN, RF, and SVM, the experimental group had baseline accuracies of 39.98%, 36.85%, and 34.91%, respectively, while the control group had baseline accuracies of 40.05%, 37.28%, and 35.15%, respectively. Due to antagonist co-contraction, reduced selective muscle activation, and excessive variability, these low values revealed the complications in recognizing movement-specific EMG signals among patients who have experienced a subacute stroke. At week 4, both groups showed some signs of improvement. However, the experimental group showed a marginal increase in movement performance (KNN: 57.05%; RF: 53.03%; SVM: 52.05%) compared to the control group (KNN: 54.85%; RF: 51.63%; SVM: 50.33%). In comparison to traditional training, this implied that VR games produced earlier improvements in neuromuscular consistency and movement performance. At week 6, a notable improvement in movement performance was observed. The experimental group exhibited higher classification accuracies for movement performance (KNN: 83.18%; RF: 81.57%; SVM: 80.50%), while the control group plateaued (KNN: 77.95%; RF: 77.78%; SVM: 76.32%). This notable improvement in movement performance in the experimental group indicated that patients receiving VR games with enhanced visual feedback showed improved dexterity and voluntary motor control. During the follow-up, the experimental group exhibited sustained high accuracies (KNN: 84.68%; RF: 83.52%; SVM: 82.56%), indicating the strength of their recovery, while the control group plateaued at lower levels (KNN: 79.64%; RF: 79.72%; SVM: 78.25%). This indicates that VR games with enhanced visual feedback effectively sustain hand motor function.

XSL•FO
RenderX

Figure 12. Classification accuracies in the experimental group. KNN: k-nearest neighbor; SVM: support vector machine.

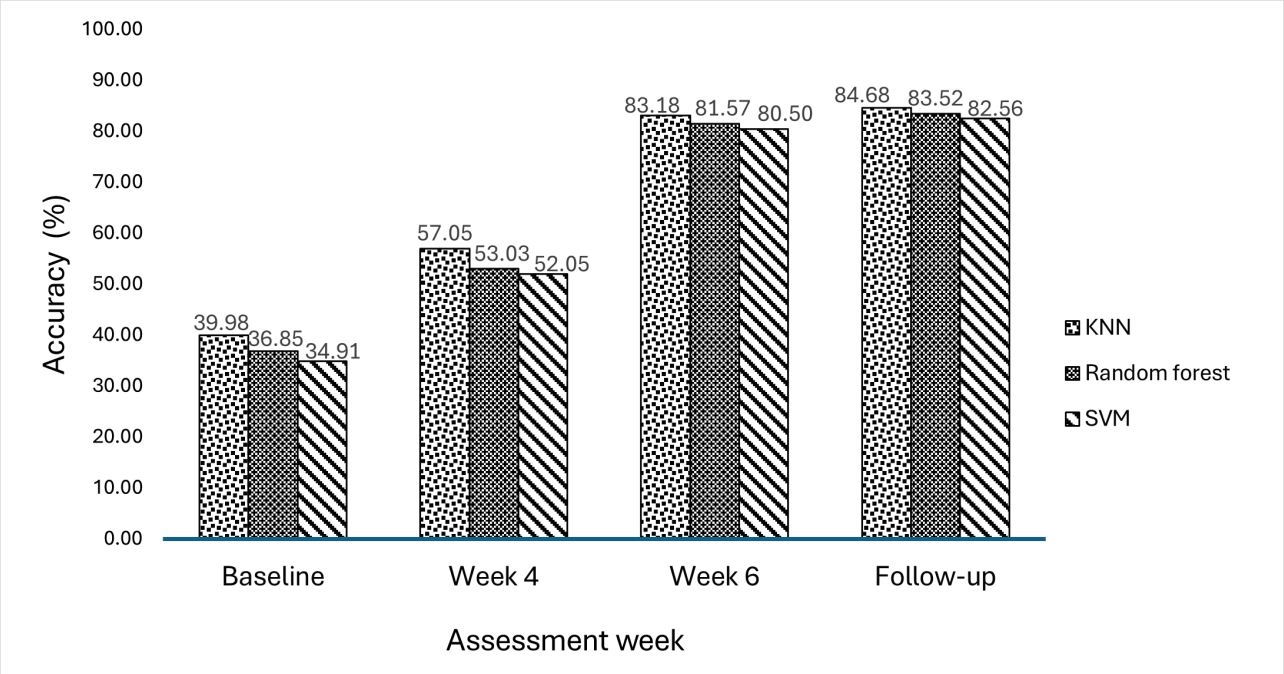
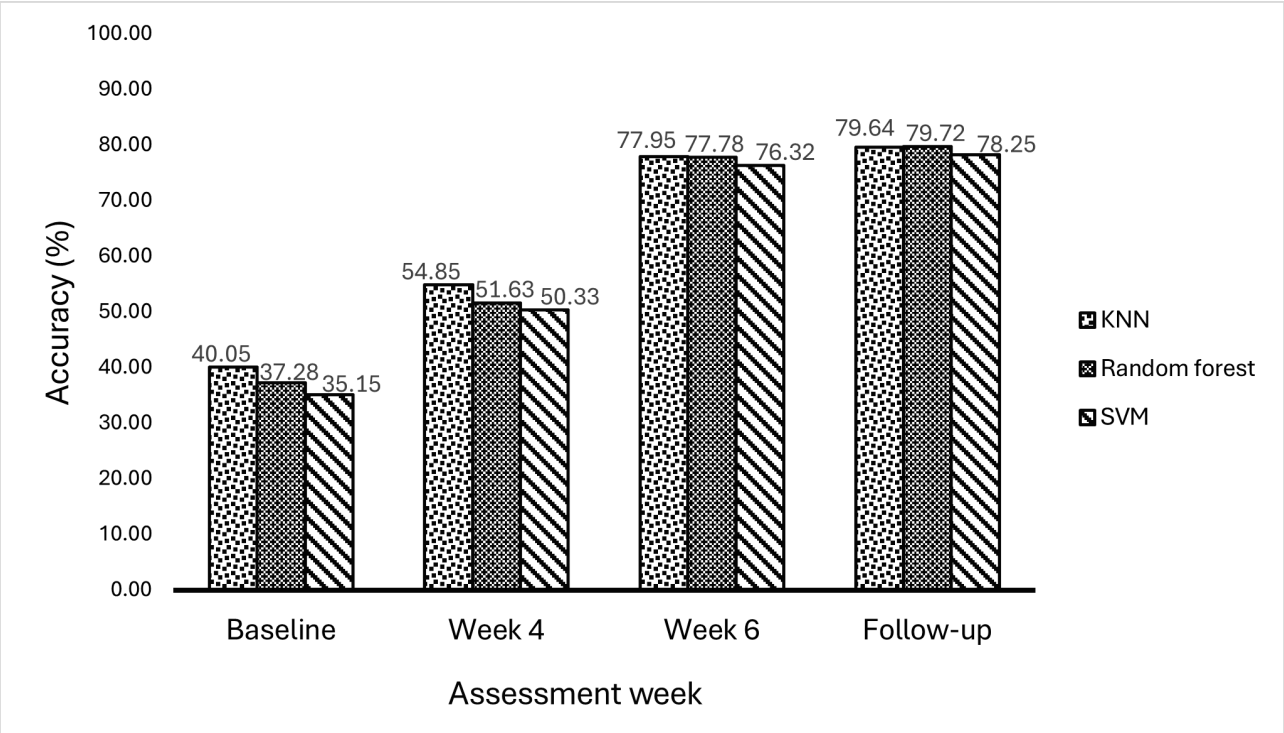


Figure 13. Classification accuracies in the control group. KNN: k-nearest neighbor; SVM: support vector machine.



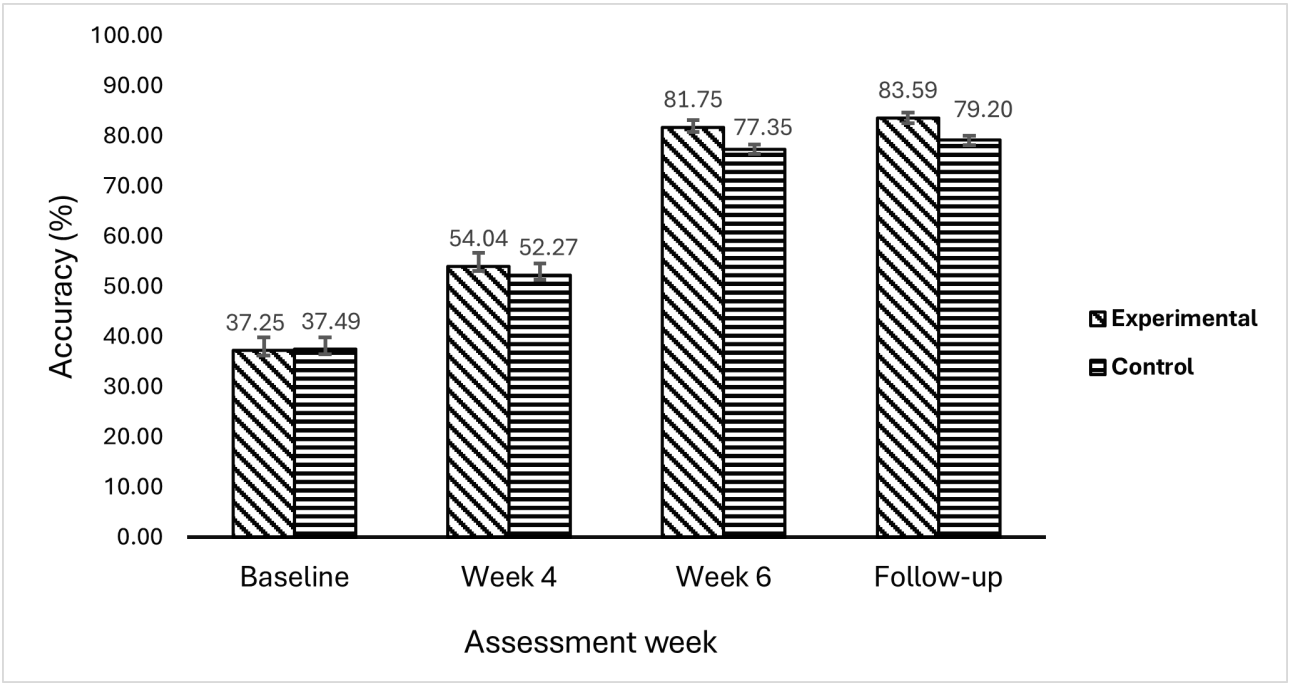
The results in Figure 12 indicate that the experimental group had better improvements in motor functional abilities due to better progression in weekly movement performance. The weekly accuracy increased in all classifiers; however, the highest movement classification accuracy of 84.68% was obtained in the KNN classifier. As shown in Figure 13, the control group also progressed in weekly movement performance, with the highest movement classification accuracy of 79.72% in the RF classifier. Comparing the movement classification accuracies in both groups, it was found that the experimental group had higher weekly movement classification accuracies than the control group. This implied that the functional abilities of the hand had been effectively promoted through the VR-based game intervention in the experimental group.

The mean classification accuracies of the 3 classifiers (KNN, RF, and SVM) are shown in Figure 14. These accuracies are presented for both the experimental group and the control group across all evaluation weeks (baseline, week 4, week 6, and follow-up). The mean accuracies of the 2 groups were nearly identical at baseline (experimental group: 37.24%, SD 2.6%; control group: 37.49%, SD 2.4%), indicating that patients commenced the study at a comparable degree of impairment

XSL•FO
RenderX

with weak EMG signal separability. At week 4, both groups exhibited modest improvements in movement performance, with the experimental group experiencing slightly greater improvements in terms of accuracy (mean 54.04%, SD 2.6%) compared to the control group (mean 52.27%, SD 2.3%). This revealed that EMG signals showed slight improvements in movement performance due to the use of VR games in the experimental group. At week 6, the experimental group achieved a higher accuracy (mean 81.75%, SD 1.3%) compared to the control group (mean 77.35%, SD 0.9%). This indicated that EMG signals showed increased improvements in movement performance because of the effective use of VR-based hand games for enhancing hand motor function in the experimental group. At follow-up, the mean movement accuracy was notably higher in the experimental group than in the control group (mean 83.59%, SD 1.1% vs mean 79.20%, SD 0.8%) (Figure 14). This revealed that the hand motor function of patients with subacute stroke was effectively sustained through the use of the VR-based intervention in the experimental group.

Figure 14. Mean classification accuracies of classifiers in the experimental and control groups. The error bars show the SD.



Discussion

Impact of Enhanced Visual Feedback on Hand Motor Recovery

This research assessed the effectiveness of customized VR-based hand rehabilitation games with an innovative approach of integrating cognitive engagement within visual training feedback. Patients with subacute stroke performed activities in the VR-based hand games that corresponded to their functional tasks. For instance, the game “hit a rolling ball” involved extending and grabbing, the game “grasp a balloon” involved holding and releasing, the game “swap hands” imitated forearm rotation, and the game “grip a pencil” simulated the activities that necessitate fine motor control. These tasks were selected to perceive the achievability of game design contents in real-life scenarios. They were adopted owing to their applicability to movements that occur in real-world environments. The patients with subacute stroke who performed the activities exhibited promising outcomes and improved hand motor function, including increased dexterity, improved ROM, increased hand strength and grip, and enhanced quality of life. The cognitive-sensory paradigm provides an effective rehabilitation approach for long-lasting improvements in hand motor function. In addition to motivating patients, this approach improves the motor learning process to execute hand movements for performing tasks and supports cognitive engagement for undertaking these tasks. This innovative approach to gamification enhances neuroplasticity, which subsequently causes long-lasting improvements in motor function. Moreover, enhanced visual training in VR games involves activities that enhance spatial perception. Many daily tasks require this skill, and it may be particularly important for people undergoing hand rehabilitation.

Most previous studies [43,46-49] applied machine learning algorithms to EMG signals for pattern recognition and gesture control–based applications such as prosthetic control. However, longitudinal rehabilitation assessments lack the use of machine learning algorithms and clinical measures in correlation with EMG signals. Although most previous research [13,16,19,20,44] assessed rehabilitation before and after interventions, there was a neglect of longitudinal assessments to effectively monitor and evaluate enduring improvements in patients with subacute stroke. Therefore, it is considered necessary to develop VR-based hand rehabilitation games for achieving enduring improvements in hand motor function and to conduct longitudinal assessments for monitoring movement performance. A previous study investigated the subjective assessment of the effectiveness of developed VR-based games based on clinical outcome measures (FMA-UE, ARAT, and BBT) [26], and the results showed significant improvements in hand motor function.

XSL•FO
RenderX

Improvements in Clinical Outcome Measures and the MCID

According to the results in Table 2, the mean difference in the FMA-UE score between the experimental and control groups exceeded the FMA-UE MCID (9.46 > $MCID_{FMA-UE}$). This finding reveals that the experimental group had clinically meaningful improvements in their motor recovery compared to the control group. Thus, the fully immersive VR-based game intervention with enhanced visual training had a positive effect on motor recovery with substantial improvements in the experimental group.

Moreover, the mean difference in the ARAT score between the 2 groups was greater than the ARAT MCID (8.96 > $MCID_{ARAT}$). This finding implies that the experimental group had clinically substantial improvements in their motor recovery due to the intervention involving VR games with enhanced visual training for promoting functional abilities.

Furthermore, the mean difference in the BBT score between the 2 groups exceeded the BBT MCID (20.46 > $MCID_{BBT}$). This finding indicates that the experimental group experienced clinically substantial improvements in hand dexterity due to the task-specific and repetitive VR-based pinch game with enhanced visual training.

Correlations Between EMG Signal Features and Clinical Outcome Measures

At baseline, there was a weak and nonsignificant (P>.05) relationship between muscle activation and hand motor performance in both groups. This finding implies that the muscles had limited strength before the intervention and were not effective in performing hand functional tasks due to motor control impairments. In the early phases of stroke recovery, patients may have ineffective activation of muscle patterns due to neurological and motor control impairments. Weak correlations between EMG features and clinical outcome measures may result from ineffective muscle activation, which may not accurately indicate the functional capacity of the damaged limb.

From Figures 6-11, at week 6, both the experimental and control groups showed positive and significant (P<.05) correlations between EMG signal features and clinical outcome measures. However, the experimental group showed strong positive correlations, while the control group showed moderate positive correlations. The strong positive correlations in the experimental group indicate significant improvements in hand motor function, coordination, hand dexterity, and fine grip strength. These improvements reflect higher FMA, ARAT, and BBT scores with increased EMG signal features. The strong correlations in the experimental group illustrate the effectiveness of cognitive engagement within visual feedback incorporated into VR-based games that significantly improved hand motor function. Furthermore, the VR-based games were repetitive and engaging, enhancing hand motor function recovery in the experimental group. At week 6, the control group demonstrated moderate positive correlations. This implies that improvements in the control group were less substantial than those in the experimental group. Thus, conventional physical therapy alone was beneficial but was not as effective as the approach involving VR-based game therapy used in the experimental group, which enhanced the motor recovery of hand function.

During the follow-up at week 9, there were significant (P<.05) strong and moderate positive correlations in the experimental and control groups, respectively. These results imply that both groups experienced improvements in motor recovery, but the experimental group showed better sustained motor recovery than the control group, indicating a strong association between EMG signal features and improvement in hand motor function.

Results between baseline and follow-up showed notably substantial improvements, with the experimental group showing a change from nonsignificant (P>.05) weak positive correlations to significant (P<.05) strong positive correlations and the control group showing a change from nonsignificant (P>.05) weak positive correlations to significant (P<.05) moderate positive correlations. These findings indicate that the intervention of VR-based games in the experimental group effectively sustained hand motor function, coordination, hand dexterity, and fine grip strength among patients with subacute stroke. VR games that target hand movements can strengthen relevant muscles, and these benefits promote functional gains in hand motor tasks over time. The longitudinal assessment of correlations provides insights into muscle activation and clinical outcome measures in VR game rehabilitation for better evaluation of hand motor function over time to achieve enduring improvements.

Impact of EMG-Based Progression Analysis of Movement Performance

EMG-based classification accuracies were used to evaluate movement performance in assessment weeks for the effectiveness of VR-based hand rehabilitation games. This provided information on the relationship between changes in EMG characteristics and improvements in motor skills. The results in Table 3 revealed significant differences between the experimental and control groups in all classifiers (all P<.05). At baseline, the mean movement classification accuracies in the experimental and control groups were 37.24% (SD 2.6%) and 37.49% (SD 2.4%), respectively. These low accuracies were due to decreased muscle activity patterns in the EMG signals associated with different hand movements.

Figure 12 shows that movement accuracy increased as rehabilitation progressed in subsequent weeks, including weeks 4, 6, and 9 (follow-up). This suggests that advances in hand movements and motor activities were being recognized by EMG features, which were becoming increasingly distinct and reliable. The experimental group exhibited higher classification accuracies in weeks 4, 6, and 9 (follow-up) compared to the control group, as illustrated in Figure 12. This shows that due to challenging levels and cognitive engagement within visual feedback, the experimental group adapted to more functional engagement in VR-based games in subsequent weeks. This increased muscle activity patterns in EMG signals, which improved the motor function of the hand in the experimental group. The control group also showed higher movement classification accuracy over time, as shown in Figure 13. However, the accuracy was lower than that in the experimental group.

XSL•FO
RenderX

In the experimental group, the KNN, RF, and SVM models achieved high movement accuracies, with rates of 84.68%, 83.52%, and 82.56%, respectively. This indicates that patients' motor function improvements were significant owing to the VR-based game intervention, and the approach effectively enhanced muscle control. The greater accuracy in the experimental group highlights the long-term advantages of VR-based rehabilitation. This implies that EMG features are reliable indicators of motor function improvements. The combined accuracies of all classifiers at follow-up, as presented in Figure 14, show that the experimental group had a higher mean accuracy (83.59%, SD 1.1%) than that in the control group (79.20%, SD 0.8%). This reveals that the VR-based hand games efficiently promoted repeated muscle activations, resulting in improved hand motor function, which was sustained in the experimental group. At baseline, there were no significant differences between the groups (P>.05), indicating equivalent starting points. The difference in machine learning prediction accuracy between the 2 groups was statistically significant (P<.05) after week 2, with the experimental group demonstrating consistently superior performance by weeks 4 and 6. The notable differences between the groups demonstrate that the VR-based training not only expedited motor recovery but also generated more distinct, stable, and classifiable EMG activation patterns, which enhanced machine learning prediction reliability. This suggests that immersive VR is more effective in improving the neurophysiological foundations of motor control compared to traditional therapy alone.

VR training improved the quality of EMG signals, as evidenced by the higher accuracies of all 3 classifiers in the experimental group at weeks 4, 6, and 9, by decreasing intraclass variability (making signals for the same movement more consistent) and increasing interclass separation (making different movements more distinct). KNN performed better than RF and SVM, indicating that VR training produced compact, well-clustered EMG feature distributions, where a distance-based approach was adequate for accurate classification. Following VR training, RF and SVM exhibited slightly weaker performance, suggesting that complicated decision boundaries were no longer necessary to achieve high separability in the EMG data. The fact that KNN performed the best indicates that the EMG signals of the patients had become naturally separable, which suggests that their motor recovery proceeded toward more predictable and distinct muscle commands. Conventional therapy only slightly improved the control group's EMG data, which continued to exhibit greater variability and class overlap. The absence of differences among KNN, RF, and SVM classifiers was not able to completely overcome the signal inconsistency, indicating that patients still lacked precise neuromuscular recovery. Patients in the control group did not achieve the same degree of consistent, distinct EMG activation patterns, as noted in the accuracy values. Their motor intent was less predictable because their muscle activity continued to overlap.

Within VR gaming, visual feedback helps maximize the performance of patients with stroke, ultimately contributing to their motion adaptation and improvement. Using a VR headset with hand tracking, patients were able to actively participate in the games while simultaneously viewing their damaged hands. Owing to this advantage, patients could easily move and control the affected hand.

Implications for Stroke Rehabilitation

This study offers several implications for stroke rehabilitation that may inform accessible practice and research on stroke rehabilitation support, suggesting a new VR-based game approach. The integration of EMG features as a quantitative method facilitates the objective monitoring of patient development, potentially reducing dependence on subjective scales alone. The gamified VR tasks enhanced functional engagement and motivation, which have been recognized as obstacles in traditional stroke recovery [3]. The integration of gameplay challenges and increased visual feedback in immersive VR rehabilitation games enhanced compliance and sustained long-term improvements in hand motor function. Furthermore, through the involvement of interactive and task-oriented training, the VR games led to improved functional performance. This signifies that these VR rehabilitation games can serve as valuable complements to conventional care, particularly for patients with reduced motivation or those who have restricted access to supervised therapy sessions. Moreover, a framework has been established for objective recovery monitoring by merging EMG with machine learning.

The findings of this research enable the integration of innovative sensory-cognitive engagement and functionally challenging activities with targeted movements. This contributes to effective improvements in hand motor function and enhances neuroplasticity in the brain. Moreover, in patients with subacute stroke, a gamified setup with sensory-cognitive engagement provides intrinsic motivation to engage actively during VR sessions. From easy to difficult levels, patients adapt to the challenges of motor task activities, which can enhance their involvement and help promote effective recovery of hand motor function. The outcomes obtained from this research underscore the correlations between EMG signals and clinical outcomes over time and the efficacy of a machine learning–based approach involving weekly progression for evaluating hand motor function.

Future Directions

In the future, a complete pipeline will be developed in Unity or C# that can facilitate seamless integration with a VR device. Moreover, the modular architecture offers a scalable framework that can be integrated into a singular system environment in future versions. The VR rehabilitation games implemented in this study have the potential to be more broadly applied across impairment profiles, including activities that require fine-motor dexterity and bilateral coordination.

Limitations

This study has some limitations. First, the VR games were confined to a limited number of exercises. Some individual stroke survivors could require specific exercises for their conditions. Second, blinding in this study was not feasible as the patients and therapists were informed of the intervention. To minimize potential outcome bias, standardized intervention protocols and consistent instructions were used across the groups. Furthermore, the incorporation of objective EMG-based

XSL•FO
RenderX

outcome measures and machine learning–derived movement performance classification accuracies facilitated an unbiased evaluation of motor recovery, thus reducing the possible influence of this constraint. Third, the evaluators in this study were not blinded. Therefore, further objective assessment may be necessary for validation. Fourth, supplementary clinical outcome measures were not included. It may be necessary to consider specific supplementary clinical outcome measures to further evaluate VR-based hand activities.

Conclusion

This study evaluated the effectiveness of immersive VR-based hand games in patients with subacute stroke, using clinical outcome measures (FMA-UE, ARAT, and BBT) and EMG-based features. Significant differences were observed in all clinical outcome measures (all $P<.05$) between the experimental and control groups. The experimental group experienced clinically meaningful improvements, with mean differences in FMA-UE, ARAT, and BBT scores being greater than their MCIDs. Furthermore, the results demonstrate the effectiveness of a machine learning approach involving weekly progression for performance assessment. The better follow-up correlation results (positive, strong, and significant) between EMG features and clinical outcome measures and the higher movement performance accuracies in the experimental group than in the control group indicate that the VR-based hand games were effective and helped sustain motor function in patients with subacute stroke. The findings of this study reveal that conventional therapy combined with immersive VR-based rehabilitation games involving enhanced visual training feedback can enhance motor recovery and movement performance more effectively than conventional therapy alone.

Funding

This work was supported by Abu Dhabi University (grant number: 19300980).

Data Availability

Virtual reality datasets generated during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

Conceptualization: FA
Funding acquisition: UHS
Investigation: FA, AW, MJK, UHS
Methodology: FA, AW, HN
Software: SOG
Supervision: AW, MAI, SOG
Validation: FA, AW, MJK, MAI, HN, SOG, FH, UHS
Visualization: FA, MJK, MAI, FH
Writing – original draft: FA
Writing – review & editing: AW, MJK, MAI, HN, SOG, FH, UHS

Conflicts of Interest

FA and AW own the games developed in Unity3D software. The other authors have no conflicts to declare.

Editorial Notice

This randomized study was retrospectively registered, which was explained by the authors as a misunderstanding of registration requirements. The editor has granted an exception from ICMJE rules mandating prospective registration of randomized trials, because the authors were able to provide confirmation from their institutional review board that the registered study was identical to the study approved by the institutional review board. However, readers are advised to carefully assess the validity of any potential explicit or implicit claims related to primary outcomes or effectiveness.

Multimedia Appendix 1
Visuals of the virtual reality–based rehabilitation games.
[PDF File, 82 KB - games_v13i1e74314_app1.pdf]

Checklist 1
CONSORT-EHEALTH checklist.
[PDF File, 1191 KB - games_v13i1e74314_app2.pdf]

References

XSL•FO
RenderX

1. Wu WX, Zhou CY, Wang ZW, et al. Effect of early and intensive rehabilitation after ischemic stroke on functional recovery of the lower limbs: a pilot, randomized trial. J Stroke Cerebrovasc Dis 2020 May;29(5):104649. [doi: 10.1016/j.jstrokecerebrovasdis.2020.104649] [Medline: 32115341]
2. Chen L, Chen Y, Fu WB, Huang DF, Lo WLA. The effect of virtual reality on motor anticipation and hand function in patients with subacute stroke: a randomized trial on movement-related potential. Neural Plast 2022 Jan 24;2022:1-14. [doi: 10.1155/2022/7399995] [Medline: 35111219]
3. Amin F, Waris A, Iqbal J, et al. Maximizing stroke recovery with advanced technologies: a comprehensive assessment of robot-assisted, EMG-controlled robotics, virtual reality, and mirror therapy interventions. Results in Engineering 2024 Mar;21:101725. [doi: 10.1016/j.rineng.2023.101725]
4. Albert SJ, Kesselring J. Neurorehabilitation of stroke. J Neurol 2012 May;259(5):817-832. [doi: 10.1007/s00415-011-6247-y] [Medline: 21964750]
5. Ying GL, Elamvazuthi I, Khan M, Parasuraman S. Development of EMG based upper arm exoskeleton. Presented at: 2018 International Conference on Recent Trends in Electrical, Control and Communication (RTECC); Mar 20-22, 2018. [doi: 10.1109/RTECC.2018.8625671]
6. Li Y, Lian Y, Chen X, et al. Effect of task-oriented training assisted by force feedback hand rehabilitation robot on finger grasping function in stroke patients with hemiplegia: a randomised controlled trial. J NeuroEngineering Rehabil 2024 May 14;21(1):77. [doi: 10.1186/s12984-024-01372-3]
7. Peng QC, Yin L, Cao Y. Effectiveness of virtual reality in the rehabilitation of motor function of patients with subacute stroke: a meta-analysis. Front Neurol 2021;12:639535. [doi: 10.3389/fneur.2021.639535]
8. Fregna G, Schincaglia N, Baroni A, Straudi S, Casile A. A novel immersive virtual reality environment for the motor rehabilitation of stroke patients: a feasibility study. Front Robot AI 2022;9:906424. [doi: 10.3389/frobt.2022.906424]
9. Park S, Lee D, Hong S, Cho K, Lee G. Feasibility of training using full immersion virtual reality video game in young stroke survivor: a case report. NeuroRehabilitation 2021;48(1):1-8. [doi: 10.3233/NRE-201501] [Medline: 33361614]
10. Yu B, Zhang X, Cheng Y, et al. The effects of the biceps brachii and brachioradialis on elbow flexor muscle strength and spasticity in stroke patients. Neural Plast 2022;2022:1295908. [doi: 10.1155/2022/1295908] [Medline: 35283993]
11. Waris A, Niazi IK, Jamil M, Englehart K, Jensen W, Kamavuako EN. Multiday evaluation of techniques for EMG-based classification of hand motions. IEEE J Biomed Health Inform 2019 Jul;23(4):1526-1534. [doi: 10.1109/JBHI.2018.2864335] [Medline: 30106701]
12. Zhang X, Zhou P. High-density myoelectric pattern recognition toward improved stroke rehabilitation. IEEE Trans Biomed Eng 2012 Jun;59(6):1649-1657. [doi: 10.1109/TBME.2012.2191551] [Medline: 22453603]
13. Bae S, Park HS. Development of immersive virtual reality-based hand rehabilitation system using a gesture-controlled rhythm game with vibrotactile feedback: an fNIRS pilot study. IEEE Trans Neural Syst Rehabil Eng 2023;31:3732-3743. [doi: 10.1109/TNSRE.2023.3312336] [Medline: 37669214]
14. Brambilla E, Stendal K, Sundling V, Calogiuri G. Nature engagement outcomes of viewing nature through a 360° video or a tablet screen: randomized trial. JMIR Serious Games 2025 Jun 16;13:e63424. [doi: 10.2196/63424] [Medline: 40523269]
15. Rodríguez-Hernández M, Polonio-López B, Corregidor-Sánchez AI, Martín-Conty JL, Mohedano-Moriano A, Criado-Álvarez JJ. Effects of specific virtual reality-based therapy for the rehabilitation of the upper limb motor function post-ictus: randomized controlled trial. Brain Sci 2021 Apr 28;11(5):555. [doi: 10.3390/brainsci11050555] [Medline: 33924767]
16. Shahmoradi L, Almasi S, Ahmadi H, et al. Virtual reality games for rehabilitation of upper extremities in stroke patients. J Bodyw Mov Ther 2021 Apr;26:113-122. [doi: 10.1016/j.jbmt.2020.10.006] [Medline: 33992230]
17. Rey B, Oliver A, Monzo JM, Riquelme I. Development and testing of a portable virtual reality-based mirror visual feedback system with behavioral measures monitoring. Int J Environ Res Public Health 2022 Feb 17;19(4):2276. [doi: 10.3390/ijerph19042276] [Medline: 35206464]
18. Gámez AB, Hernandez Morante JJ, Martínez Gil JL, Esparza F, Martínez CM. The effect of surface electromyography biofeedback on the activity of extensor and dorsiflexor muscles in elderly adults: a randomized trial. Sci Rep 2019 Sep 11;9(1):13153. [doi: 10.1038/s41598-019-49720-x] [Medline: 31511629]
19. Aguilera-Rubio Á, Cuesta-Gómez A, Mallo-López A, Jardón-Huete A, Oña-Simbaña ED, Alguacil-Diego IM. Feasibility and efficacy of a virtual reality game-based upper extremity motor function rehabilitation therapy in patients with chronic stroke: a pilot study. Int J Environ Res Public Health 2022 Mar 13;19(6):3381. [doi: 10.3390/ijerph19063381] [Medline: 35329069]
20. Kuo FL, Lee HC, Kuo TY, et al. Effects of a wearable sensor–based virtual reality game on upper-extremity function in patients with stroke. Clin Biomech (Bristol, Avon) 2023 Apr;104:105944. [doi: 10.1016/j.clinbiomech.2023.105944]
21. Arya KN, Verma R, Garg RK. Estimating the minimal clinically important difference of an upper extremity recovery measure in subacute stroke patients. Top Stroke Rehabil 2011 Oct;18 Suppl 1:599-610. [doi: 10.1310/tsr18s01-599] [Medline: 22120029]
22. Gladstone DJ, Danells CJ, Black SE. The Fugl-Meyer assessment of motor recovery after stroke: a critical review of its measurement properties. Neurorehabil Neural Repair 2002 Sep;16(3):232-240. [doi: 10.1177/154596802401105171] [Medline: 12234086]

XSL•FO
RenderX

23. Meng L, Zhang T, Zhao X, et al. A quantitative lower limb function assessment method based on fusion of surface EMG and inertial data in stroke patients during cycling task. Biomed Signal Process Control 2023 Aug;85:104880. [doi: 10.1016/j.bspc.2023.104880]
24. Song S, Wang S, Zeng S, et al. Exploring the time to onset and early predictors of poststroke spasticity combined with surface electromyography: protocol for a nested case-control study. JMIR Res Protoc 2025 Aug 5;14:e65829. [doi: 10.2196/65829] [Medline: 40762971]
25. Zhuang JY, Ding L, Shu BB, Chen D, Jia J. Associated mirror therapy enhances motor recovery of the upper extremity and daily function after stroke: a randomized control study. Neural Plast 2021;2021:7266263. [doi: 10.1155/2021/7266263] [Medline: 34630560]
26. Amin F, Waris A, Syed S, et al. Effectiveness of immersive vrtual reality-based hand rehabilitation games for improving hand motor functions in subacute stroke patients. IEEE Trans Neural Syst Rehabil Eng 2024;32:2060-2069. [doi: 10.1109/TNSRE.2024.3405852]
27. Meyer S, Verheyden G, Kempeneers K, Michielsen M. Arm-hand boost therapy during inpatient stroke rehabilitation: a pilot randomized controlled trial. Front Neurol 2021;12:652042. [doi: 10.3389/fneur.2021.652042] [Medline: 33716948]
28. Mancuso M, Tondo SD, Costantini E, Damora A, Sale P, Abbruzzese L. Action observation therapy for upper limb recovery in patients with stroke: a randomized controlled pilot study. Brain Sci 2021 Feb 26;11(3):290. [doi: 10.3390/brainsci11030290] [Medline: 33652680]
29. Yin CW, Sien NY, Ying LA, Chung S, Tan May Leng D. Virtual reality for upper extremity rehabilitation in early stroke: a pilot randomized controlled trial. Clin Rehabil 2014 Nov;28(11):1107-1114. [doi: 10.1177/0269215514532851] [Medline: 24803644]
30. Narayan Arya K, Verma R, Garg RK, Sharma VP, Agarwal M, Aggarwal GG. Meaningful task-specific training (MTST) for stroke rehabilitation: a randomized controlled trial. Top Stroke Rehabil 2012 May;19(3):193-211. [doi: 10.1310/tsr1903-193]
31. Kim H, Shin JH. Assessment of upper extremity function in people with stroke based on the framework of the ICF: a narrative review. Brain Neurorehabil 2022 Jul;15(2):e16. [doi: 10.12786/bn.2022.15.e16] [Medline: 36743205]
32. Connell LA, Tyson SF. Clinical reality of measuring upper-limb ability in neurologic conditions: a systematic review. Arch Phys Med Rehabil 2012 Feb;93(2):221-228. [doi: 10.1016/j.apmr.2011.09.015] [Medline: 22289230]
33. Hmaied Assadi S, Feige Gross-Nevo R, Dudkiewicz I, Barel H, Rand D. Improvement of the upper extremity at the subacute stage poststroke: does hand dominance play a role? Neurorehabil Neural Repair 2020 Nov;34(11):1030-1037. [doi: 10.1177/1545968320962502] [Medline: 33016204]
34. Wahid MF, Tafreshi R, Langari R. A multi-window majority voting strategy to improve hand gesture recognition accuracies using electromyography signal. IEEE Trans Neural Syst Rehabil Eng 2020 Feb;28(2):427-436. [doi: 10.1109/TNSRE.2019.2961706] [Medline: 31870989]
35. Waris A, Niazi IK, Jamil M, et al. The effect of time on EMG classification of hand motions in able-bodied and transradial amputees. J Electromyogr Kinesiol 2018 Jun;40:72-80. [doi: 10.1016/j.jelekin.2018.04.004] [Medline: 29689443]
36. Imran U, Waris A, Gilani SO, et al. Patient-specific movement regime: investigating the potential of upper-extremity motions for intelligent myoelectric prosthetic control. IEEE Access 2024;12:35663-35682. [doi: 10.1109/ACCESS.2024.3365639]
37. Yin G, Zhang X, Chen D, et al. Processing surface EMG signals for exoskeleton motion control. Front Neurorobot 2020;14:40. [doi: 10.3389/fnbot.2020.00040] [Medline: 32765250]
38. Phukpattaranont P, Thongpanja S, Anam K, Al-Jumaily A, Limsakul C. Evaluation of feature extraction techniques and classifiers for finger movement recognition using surface electromyography signal. Med Biol Eng Comput 2018 Dec;56(12):2259-2271. [doi: 10.1007/s11517-018-1857-5] [Medline: 29911250]
39. Sultana A, Ahmed F, Alam MS. A systematic review on surface electromyography-based classification system for identifying hand and finger movements. Healthcare Analytics 2023 Nov;3:100126. [doi: 10.1016/j.health.2022.100126]
40. Chen YW, Chiang WC, Chang CL, Lo SM, Wu CY. Comparative effects of EMG-driven robot-assisted therapy versus task-oriented training on motor and daily function in patients with stroke: a randomized cross-over trial. J NeuroEngineering Rehabil 2022 Dec;19(1):6. [doi: 10.1186/s12984-021-00961-w]
41. Jung JY, Youn PS, Kim DH. Effects of mirror therapy combined with EMG-triggered functional electrical stimulation to improve on upper extremity function in patient with chronic stroke. Physikalische Medizin, Rehabilitationsmedizin, Kurortmedizin 2021 Apr;31(2):127-135. [doi: 10.1055/a-1210-2930]
42. Lee SH, Cui J, Liu L, Su MC, Zheng L, Yeh SC. An evidence-based intelligent method for upper-limb motor assessment via a VR training system on stroke rehabilitation. IEEE Access 2021;9:65871-65881. [doi: 10.1109/ACCESS.2021.3075778]
43. Zhang X, Zhang M. Study on the methods of feature extraction based on electromyographic signal classification. Med Biol Eng Comput 2023 Jul;61(7):1773-1781. [doi: 10.1007/s11517-023-02812-3] [Medline: 36894795]
44. Song X, van de Ven SS, Chen S, et al. Proposal of a wearable multimodal sensing-based serious games approach for hand movement training after stroke. Front Physiol 2022;13:811950. [doi: 10.3389/fphys.2022.811950]

XSL•FO
RenderX

45. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126. [doi: 10.2196/jmir.1923] [Medline: 22209829]
46. Saeed B, Zia-ur-Rehman M, Gilani SO, et al. Leveraging ANN and LDA classifiers for characterizing different hand movements using EMG signals. Arab J Sci Eng 2021 Feb;46(2):1761-1769. [doi: 10.1007/s13369-020-05044-x]
47. Inam S, Amin F, ur Rehman MZ. Comparative study of flexor and extensor muscles EMG for upper limb prosthesis. Presented at: 2021 15th International Conference on Open Source Systems and Technologies (ICOSST); Dec 15-16, 2021. [doi: 10.1109/ICOSST53930.2021.9683956]
48. Kim DH, Jang SH. Effects of mirror therapy combined with EMG-triggered functional electrical stimulation to improve on standing balance and gait ability in patient with chronic stroke. Int J Environ Res Public Health 2021 Apr 2;18(7):3721. [doi: 10.3390/ijerph18073721] [Medline: 33918288]
49. Castiblanco JC, Ortmann S, Mondragon IF, Alvarado-Rojas C, Jöbges M, Colorado JD. Myoelectric pattern recognition of hand motions for stroke rehabilitation. Biomed Signal Process Control 2020 Mar;57:101737. [doi: 10.1016/j.bspc.2019.101737]

Abbreviations

ARAT: Action Research Arm Test
BBT: Box and Block Test
CONSORT: Consolidated Standards of Reporting Trials
EMG: electromyography
FMA-UE: Fugl-Meyer Assessment-upper extremity
KNN: k-nearest neighbor
MCID: minimal clinically meaningful difference
RF: random forest
ROM: range of motion
SDK: software development kit
SVM: support vector machine
VR: virtual reality

Edited by A Coristine; submitted 21.03.25; peer-reviewed by A Gazis, Q Qiu, V Wattanasoontorn; revised version received 27.09.25; accepted 27.09.25; published 25.11.25.

Please cite as:
Amin F, Waris A, Khan MJ, Ijaz MA, Nazeer H, Gilani SO, Hazzazi F, Shah UH
Evaluating the Effectiveness of Immersive Virtual Reality Rehabilitation Games With Enhanced Visual Training for Hand Motor Function Improvement Using Electromyography: Randomized Controlled Trial
JMIR Serious Games 2025;13:e74314
URL: https://games.jmir.org/2025/1/e74314
doi:10.2196/74314

© Faisal Amin, Asim Waris, Muhammad Jawad Khan, Muhammad Adeel Ijaz, Hammad Nazeer, Syed Omer Gilani, Fawwaz Hazzazi, Umer Hameed Shah. Originally published in JMIR Serious Games (https://games.jmir.org), 25.11.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Influence of Virtual Reality Illusions on Balance Performance and Immersive User Experience in Young Adults: A Within-Subject Experimental Study

R Achintha M Abayasiri[1], MSc, MPhil; Antonio Padilha Lanari Bo[1,2], PhD; Taylor J M Dick[3], PhD; Nilufar Baghaei[1], PhD

[1]School of Electrical Engineering and Computer Science, Faculty of Engineering, Architecture and Information Technology, The University of Queensland, Brisbane, Australia

[2]Department of Mechanical, Electrical and Chemical Engineering, Faculty of Technology, Art and Design, OsloMet – Oslo Metropolitan University, Pilestredet 35, Oslo, Norway

[3]School of Biomedical Sciences, Faculty of Health, Medicine and Behavioural Sciences, The University of Queensland, Brisbane, Australia

Corresponding Author:
Antonio Padilha Lanari Bo, PhD
School of Electrical Engineering and Computer Science, Faculty of Engineering, Architecture and Information Technology, The University of Queensland, Brisbane, Australia

Abstract

Background: Rehabilitation interventions to improve standing balance are often tedious and complex, limiting user engagement and increasing the burden of the clinicians delivering them. Virtual reality (VR) has been incorporated into such practices as a solution and VR illusions have emerged as a method for perturbing balance within interventions. However, the influence of VR illusions on balance performance, such as center of pressure (CoP), and user experience metrics remain under explored.

Objective: This study aimed to evaluate the impact of the VR illusions on standing balance and immersive user experience in young adults.

Methods: Young healthy adults (N=15, aged 18 - 35 years) played a VR table tennis game while standing on a force plate and were provided with eight directional and magnitude-based VR illusions scaled according to participants' heights. VR illusions were generated by offsetting the position of the playing hand in VR and were provided through 8 trials for each participant. Each VR illusion was delivered throughout final 50 seconds of each 70-second trial. Absolute CoP displacements, directional tendency of CoP displacement, and game performance were analyzed to evaluate the impact of the VR illusions. Responses to the User Experience Questionnaire, Slater-Usoh-Steed Presence Questionnaire, NASA Task Load Index, and Virtual Reality Sickness Questionnaire were analyzed to assess the immersive user experience.

Results: Both the magnitude of VR illusion and changes in VR illusion direction led to significantly greater CoP displacements, with high illusion magnitudes, and anterior and posterior directional illusions associated with higher CoP displacements. Conversely, those illusion magnitudes and directions were associated with low game performance. The directional tendency of the CoP displacements varied across the illusion directions but showed a significant association with the illusion directions. Questionnaire responses showed that participants had moderate to high immersive user experience within the VR illusion paradigm.

Conclusions: This study provides a novel approach for the future development of more effective VR-based balance rehabilitation interventions. The results provide inspiration for the development of future VR-based exergames that can perturbate CoP direction and magnitude. By adjusting the difficulty level through directional and magnitude changes in VR illusions, exergames could provide a personalized rehabilitation experience.

(JMIR Serious Games 2025;13:e70376) doi:10.2196/70376

KEYWORDS

VR Illusions; center of pressure; user experience; balance; virtual reality

Introduction

Standing balance deficits pose significant challenges to activities of daily living. Balance rehabilitation paradigms have been developed as an effective approach to address balance impairments. A fundamental aspect of standing balance rehabilitation is the use of perturbations, which challenge postural control and promote improvement in balance performance [1-4]. Center of pressure (CoP) displacement is widely recognized as a key measure of balance performance [5,6], with minimal displacement in response to perturbations indicating improved postural control [7,8]. However, these

XSL•FO
RenderX

perturbations are often delivered physically, and thus, demand substantial therapist involvement as well as constant supervision. Yet, the repetitive and monotonous nature of such activities can lead to decreased patient engagement over time [9,10].

Virtual reality (VR)–based standing balance rehabilitation interventions have emerged as a promising alternative to reduce therapist workload while enhancing patient engagement. By creating interactive and immersive environments, VR systems offer a more engaging approach than traditional methods, potentially increasing patient motivation and adherence to rehabilitation programs [9]. In addition, VR systems allow for precise control over rehabilitation exercises, enabling therapists to tailor interventions to individual patient needs and remotely monitor progress [10,11]. In this context, incorporating VR illusions to balance rehabilitation interventions to create perturbation effects, has gained attention as a method for challenging postural control during rehabilitation [12,13]. These VR perturbations are typically generated by manipulating the visual field, such as rotating the virtual environment's field of view, which can induce transient effects on balance [14-18]. Another common VR technique involves the manipulation of environmental optic flow, where adjustments in virtual motion patterns affect postural control [19-23]. While optic flow speed has been shown to induce temporary postural instability, Ketterer et al [19] cautioned that prolonged exposure to such VR illusions may have negative consequences for postural stability.

Despite these advancements, maintaining participant adherence to VR-based perturbation interventions remains a challenge. Enhancing the perceived realism of VR illusions is crucial to ensuring their effectiveness in inducing the desired postural adjustments in participants. Immersive user experience plays a critical role in enhancing realism and acceptability of VR illusions for balance rehabilitation interventions [10,24]. Key components of immersive user experience include immersion [25,26], comfort [27], sickness [28], and satisfaction [29,30], all of which contribute to the overall success of a rehabilitation program [24,31-33]. Furthermore, studies have shown that gamified interventions not only enhance patient adherence through increased motivation but also improve balance outcomes [34-40]. However, the existing VR illusion–based balance rehabilitation interventions have largely overlooked the impact on user experience, an essential factor for long-term adherence to rehabilitation programs [10,24]. Some studies have noted that certain populations, particularly young, healthy adults, often disregard the VR illusions, undermining the intended perturbation effects [41].

It is apparent that the perturbation effect and immersive user experience should go hand in hand for a successful VR illusion–based balance rehabilitation intervention. Thus, in this study, we aimed to advance the understanding of how VR illusions impact standing balance and immersive user experience. By leveraging the brain's ability to adapt to discrepancies between sensory inputs and motor outputs, known as sensorimotor adaptation [34-37], we generated VR illusions by offsetting the virtual hand position in a commercially available VR table tennis game. The choice of this game was influenced by several factors, including its real-life relatability [38,39], the ability to be played in a standing posture [32], simple game-mechanics, and easily adjustable parameters. In addition, the game's haptic and auditory feedback had the potential to enhance immersive user experience, further supporting our decision to use it [24,33]. Participants played the game standing on a force plate while being prompted to recalibrate their movements to account for the hand offsets. Additionally, we examined the impact of this intervention on immersive user experience through questionnaire responses. Finally, by conducting a multiscale analysis that incorporates both biomechanics and human-computer Interaction metrics, we sought to draw inferences for future exergame designs and VR-based balance rehabilitation interventions.

Methods

Participants

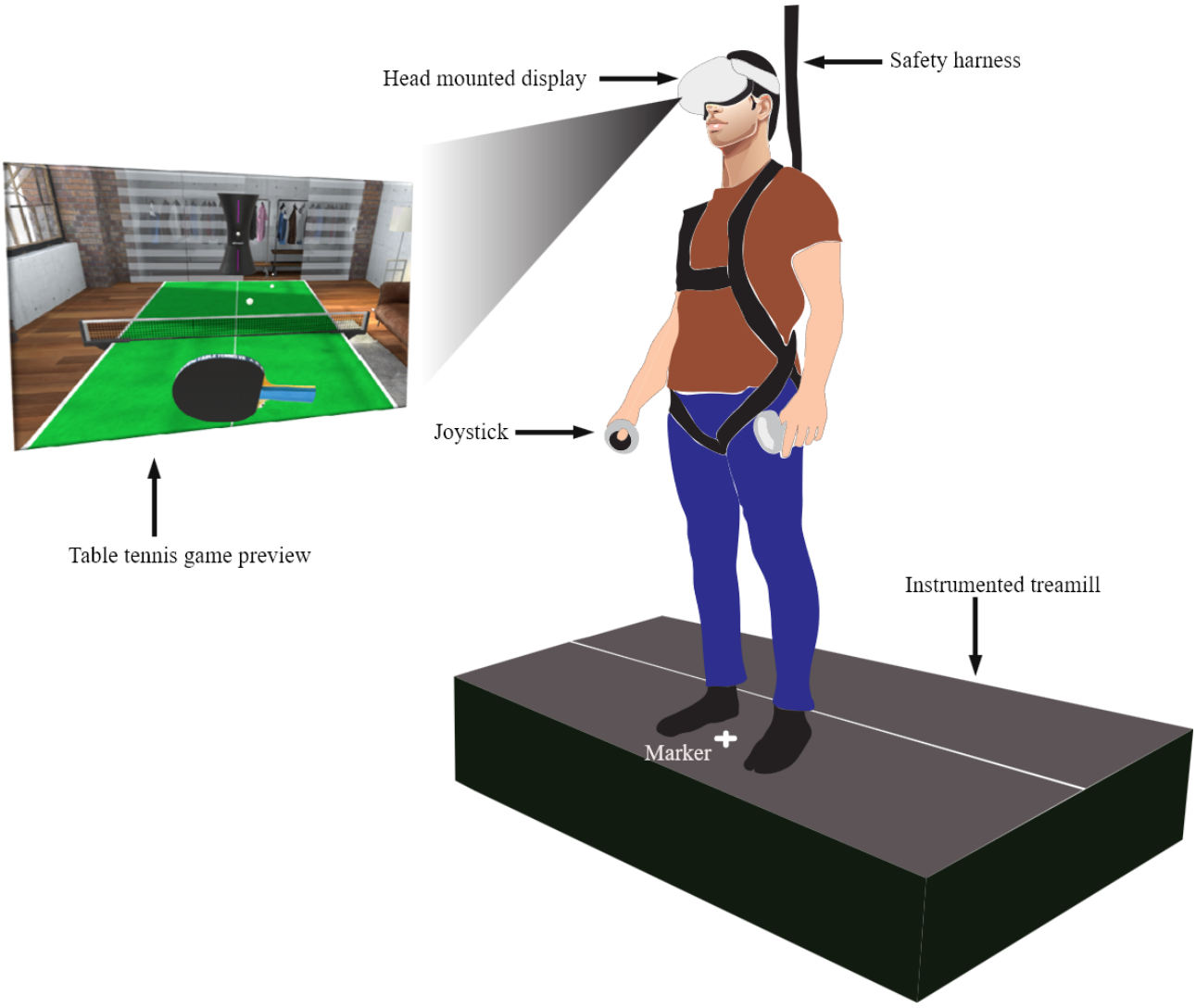
A total of 15 healthy young adults (8 male and 7 female participants; mean age 27.7, SD 3.8 years; mean height 1.7, SD 0.1 m; mean weight 73.62, SD 16.9 kg) were recruited through advertisements posted on university notice boards, social media, and by word of mouth. Once participants expressed their interest, the research team contacted them via email to provide detailed information about the study and confirm their participation. All the recruited participants were reported to be right-hand dominant.

Experimental Setup

The experimental setup was built inside the Neuromuscular Biomechanics laboratory of The University of Queensland (Figure 1). It consisted of immersive VR system based on a Head Mounted Display (HMD) (Quest 2; Meta, formerly Oculus), instrumented split belt treadmill (FIT5; Bertec), a safety harness, a personal computer, and a mobile phone camera (iPhone 14; Apple Inc) to record the participants' gameplay on the monitor of the personal computer. Furthermore, the treadmill has force plates to measure ground reaction forces. During the experiment, treadmill was held static, and the force plate data were acquired at a frequency of 120 Hz and were processed in Qualisys Track Manager system (version 2022.2; Qualisys AB) to extract CoP data. CoP data were further filtered with second-order low-pass Butterworth filters at 6 Hz in MATLAB software (Release 2024a; The MathWorks Inc). Furthermore, a marker was placed on one of the belts of the treadmill to maintain the consistency of the feet placements of the participants.

XSL•FO
RenderX

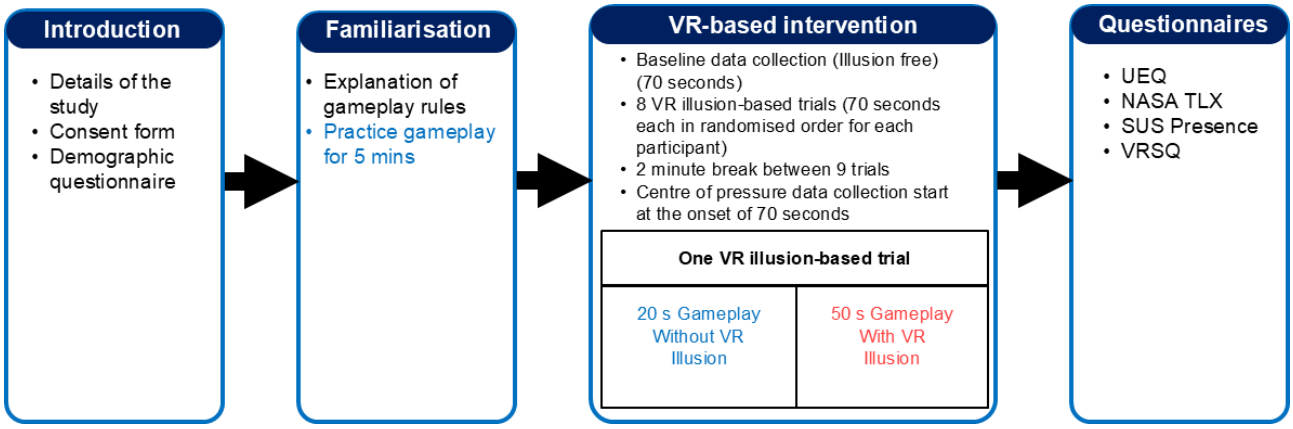
Figure 1. Experimental setup, in which participants stood on a static instrumented treadmill to play a virtual reality table tennis game. They were given instructions to keep their feet apart so that the marker on the treadmill would be in between their feet. They were asked to use the joystick in their dominant hand to hit the ball while keeping their feet steady on the treadmill.



Experimental Protocol

Experimental protocol consists of four stages: introduction, familiarization, data collection during VR-based intervention, and post data collection (Figure 2).

Figure 2. Overview of the experimental protocol. Experiment consists of 4 stages: introduction, familiarization, virtual-reality–based intervention, and post-gameplay data collection. NASA TLX: NASA task load index; SUS: Slater-Usoh-Steed; UEQ: User experience questionnaire; VRSQ: Virtual reality sickness questinnaire.



XSL•FO
RenderX

VR Game Configuration

For this study, the Ball Machine mode of the Eleven Table Tennis VR game (version 1.0; for Fun Labs) was used. The ball feeder mode was set to "default" to feed the ball through the centerline of the table at all times. The time scale was set to "1×" to feed 1 ball every 2 seconds. An independent application programming interface was used together with the game to enable offsetting the virtual representation of the participants' hands with respect to the virtual body. Both the VR game and the external application programming interface were launched simultaneously on the SteamVR (version 2.6; Valve Corporation) platform.

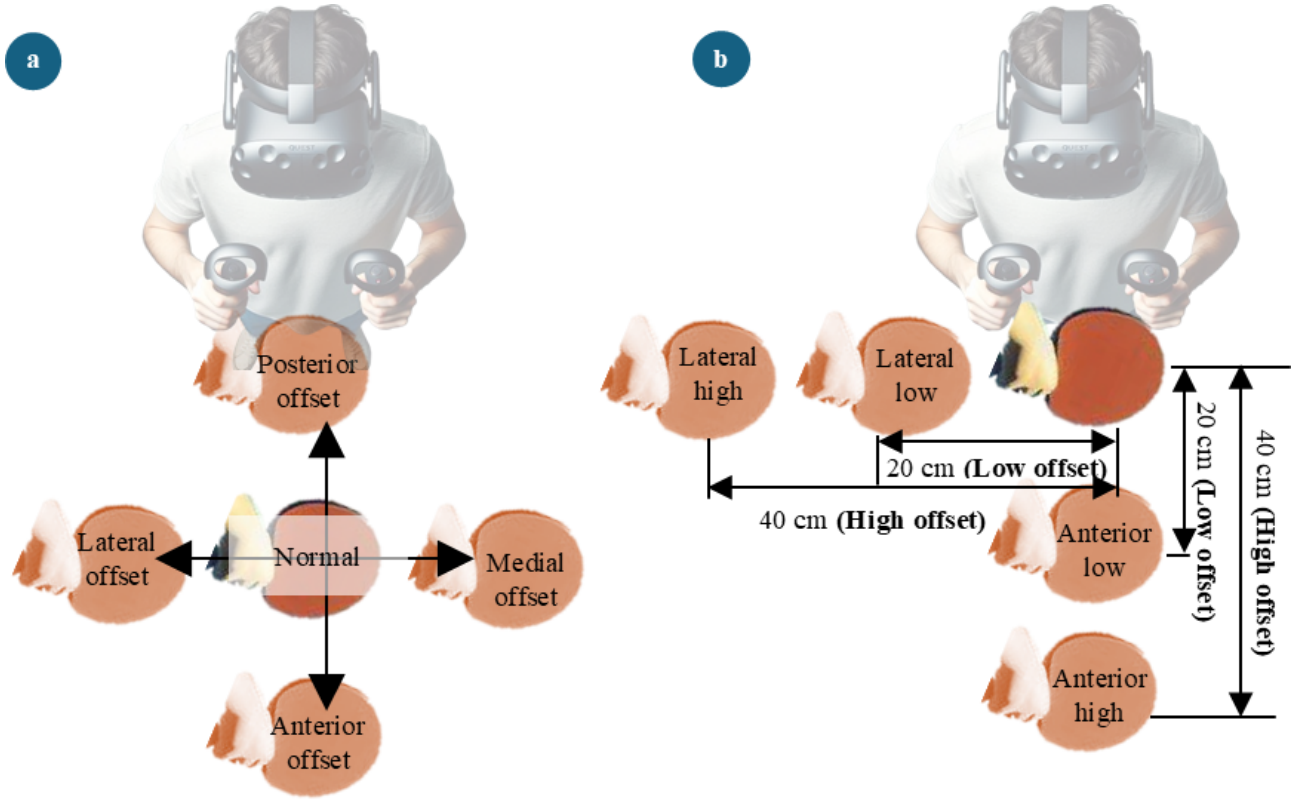
Protocol

Participants were attached to the safety harness when they got onto the treadmill. They were instructed that they could not move their feet while playing the game. They kept their feet shoulder-width apart (approximately 40 cm), centered on the cross mark, and wore the HMD while staring straight ahead to align with the table in the game. Then, they were instructed to return the ball according to the general table tennis rules and to play without moving the feet from their starting position. In addition, participants were instructed that they could stop playing the game if they felt any discomfort, such as dizziness or nausea.

All participants played the table tennis game for 5 minutes initially to become comfortable with the game environment and the rules of the gameplay (familiarization stage). Then, participants were asked to remove the HMD and rest for 2 minutes. After that, gameplay with VR illusions was started. In total, eight types of VR illusions were provided for the participants during this stage (Figure 3). These illusions were created by offsetting the virtual hand used to hit the ball in the VR table tennis game, in 4 directions, namely anterior, posterior, medial, and lateral. In addition, two levels of magnitude (low and high) were incorporated into the illusions. Furthermore, these magnitudes were scaled according to the height of the participants to ensure the uniformity of the applied illusions. Before the onset of this gameplay stage, participants were given further instructions about the gameplay for this stage. They were informed that they would encounter challenges to the usual gameplay but were still required to continue returning the ball to the other side of the table. Each participant played nine 70-second long trials of gameplay. Out of the nine trials, the first trial was conducted to collect the baseline data. Rest of the trials were based on the eight types of virtual illusions. Each virtual illusion–based trial used one type of virtual illusion and the order of applied virtual illusions for trials was randomized for each participant.

Before starting data collection for a trial, the experimenter waited until the participant completed 5 successful consecutive ball returns. Once this was observed, a timer and the collection of data from the force plate of the static treadmill, CoP were started for the trial. For the baseline trial, force data were recorded for 70 seconds. During the virtual illusion–based trials, the experimenter applied the virtual illusion throughout the last 50 seconds of the 70-seconds trial.

Figure 3. Types of virtual reality illusions. (**A**) Directions of the illusions and (**B**) construction of the Illusions combining both magnitudes and directions. Illusion magnitudes were scaled according to the participant's height. (ie, if participant height is 180 cm, low and high values were equal to 20 cm and 40 cm, respectively). Here, examples are depicted for the construction of anterior-low, anterior-high, lateral-low and lateral-high type illusions.



XSL•FO
RenderX

Post Gameplay Data Collection

After finishing the 9 VR gameplay trials including the baseline line trial, participants were asked to answer 4 questionnaires. Those questionnaires were Virtual Reality Sickness Questionnaire (VRSQ) [40], User Experience Questionnaire (UEQ) [42], NASA Task Load Index (NASA-TLX) [43], and Slater-Usoh-Steed (SUS) presence questionnaire [44].

Statistical Analysis

All statistical analyses were performed using R software (version 4.4.1; R Foundation for Statistical Computing). Paired *t* tests were used to evaluate the significance of differences in absolute CoP displacement measures before and after illusion stages for different illusion types. In addition, paired *t* tests were used to assess the statistical significance of mean differences in absolute CoP displacement measures and the number of successful ball returns (gameplay score) across different illusion magnitude levels.

To examine variations in absolute CoP measures and gameplay scores based on the factors of illusion magnitude and direction, linear mixed-effect (LMEs) models were specified, including participants as a random factor, and interaction terms. Tukey post hoc tests were conducted to further analyze the differences between the levels of main effects. In addition, paired *t* tests were conducted to compare absolute CoP displacements between successful and unsuccessful ball returns. Significance was determined at the $P<.05$ level for all the aforementioned statistical tests. We analyzed absolute CoP displacement to capture CoP excursions [45], as considering directional CoP values would only reflect the tendency of movement direction. Furthermore, both mediolateral and anteroposterior directional absolute CoP displacements were considered in the analyses.

To investigate the association between directional tendency and illusion directions, the Fisher exact test, with simulated *P* values based on 10^6 replicates generated through the Monte Carlo simulation was used. The Chi-square test was applied to assess the statistical significance of the relationship between illusion direction and gameplay status (successful or unsuccessful). In both cases, associations were considered significant at $P<.05$.

Ethical Considerations

This study was conducted in the Neuromuscular Biomechanics Laboratory at The University of Queensland, Australia. All study procedures were approved by The University of Queensland Human Research Ethics Committee (approval number 2023/HE001569). Written informed consent was obtained from all participants before their involvement in the study. Participants had no history of leg injuries or fractures within the past 6 months and no neurological disorders affecting visual, cognitive, or vestibular function. Participation was voluntary and no compensation was provided. All recorded data were deidentified using a coding system to ensure confidentiality.

Results

Mean Absolute Center of Pressure Displacement

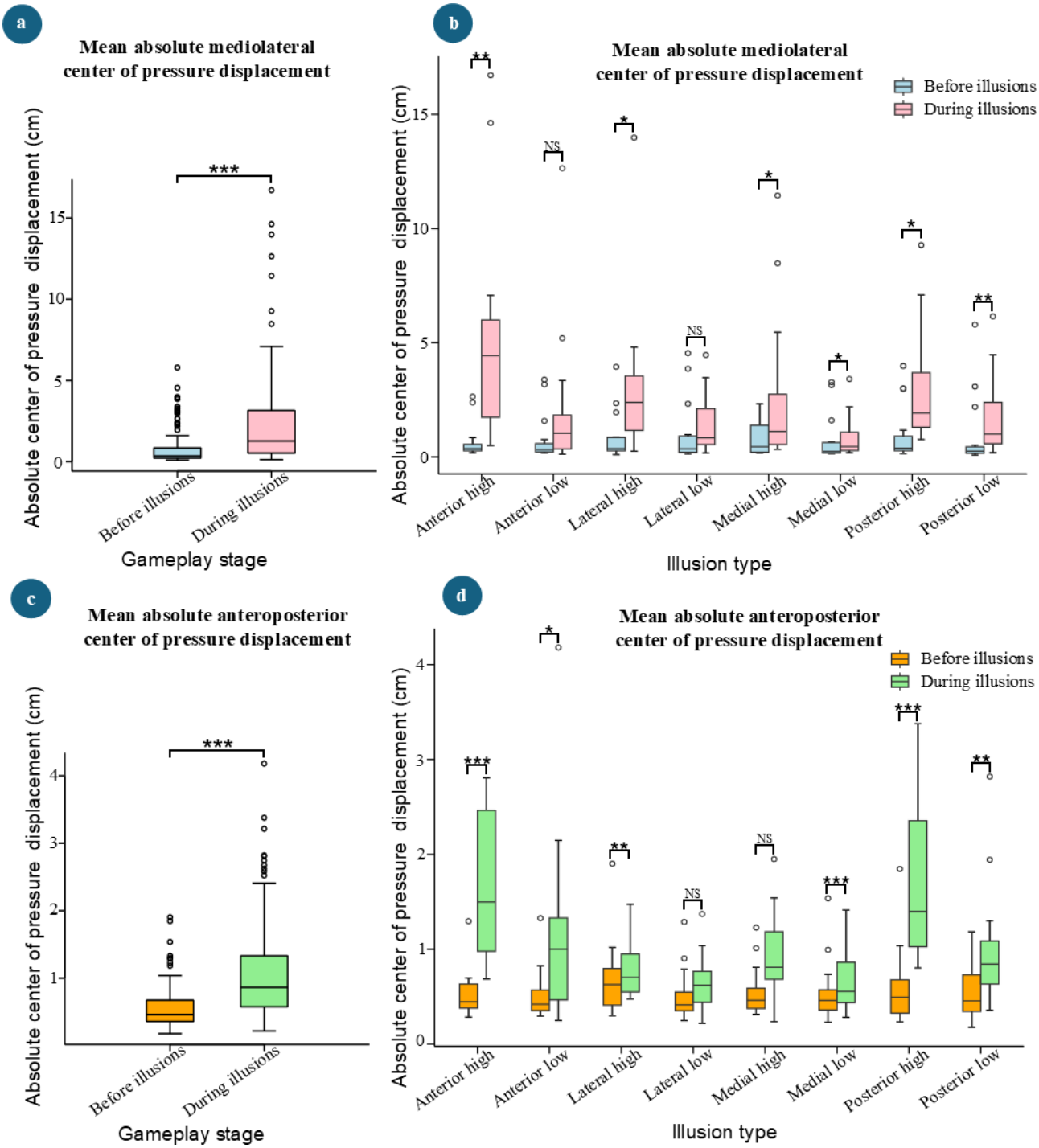
Figure 4 illustrates the comparisons of mean absolute CoP displacement in both the mediolateral and anteroposterior directions. There were significant increases in mean absolute CoP displacements during illusions compared with before the illusion stages for both mediolateral ($P<.001$, mean difference 1.62 cm) and anteroposterior ($P<.001$, mean difference 0.5 cm; Figure 4A and C) directions implying that the VR illusions were able to affect the mediolateral and anteroposterior stability of the participants.

Further comparisons of mean absolute mediolateral CoP displacement showed significant increases during illusions compared with before the illusions for several types of illusions: anterior high ($P=.002$, mean difference 4.4 cm), lateral high ($P=.048$, mean difference 1.8 cm), medial high ($P=.025$, mean difference 2.1 cm), medial low ($P=.045$, mean difference 0.4 cm), posterior high ($P=.015$, mean difference 2 cm), and posterior low ($P=.03$, mean difference 0.8 cm; Figure 4B). In contrast, no significant differences were observed for mean absolute mediolateral CoP displacement for anterior low ($P=.093$, mean difference 1.25 cm) and lateral low ($P=.450$, mean difference 0.1 cm) illusion types implying that the participants' balance was not significantly impacted by these illusion types.

For mean absolute anteroposterior CoP displacement, significant increases were observed during the illusion stages compared with before the illusion stages for the following types: anterior high ($P<.001$, mean difference 1.1 cm), anterior low ($P=.012$, mean difference 0.6 cm), lateral high ($P<.001$, mean difference 0.3 cm), medial low ($P<.001$, mean difference 0.4 cm), posterior high ($P<.001$, mean difference 1.2 cm), and posterior low ($P=.007$, mean difference 0.5 cm; Figure 4D). However, no significant differences were observed for lateral low ($P=.073$, mean difference 0.1 cm) and medial high ($P=.174$, mean difference 0.1 cm) illusion types.

XSL•FO
RenderX

Figure 4. Comparison of absolute mean center of pressure displacement before and during illusions. (**A**) Absolute mean mediolateral center of pressure displacement comparison in gameplay stages. (**B**) Absolute mean mediolateral center of pressure displacement comparison across different types of virtual illusions. (**C**) Absolute mean anteroposterior center of pressure displacement comparison in gameplay stages. (**D**) Absolute mean anteroposterior center of pressure displacement comparison across different types of virtual illusions (NS: $P > 0.05$, *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$).



Maximum Absolute Center of Pressure Displacement

Figure 5 shows the comparisons of maximum absolute CoP displacement. Similar to the behavior of mean absolute CoP displacement, there were significant increases in maximum absolute CoP displacement during illusions compared with before the illusion stages in both the mediolateral (P<.001, mean difference 4 cm) and anteroposterior (P<.001, mean difference 1.9 cm) directions (Figure 5A and C).

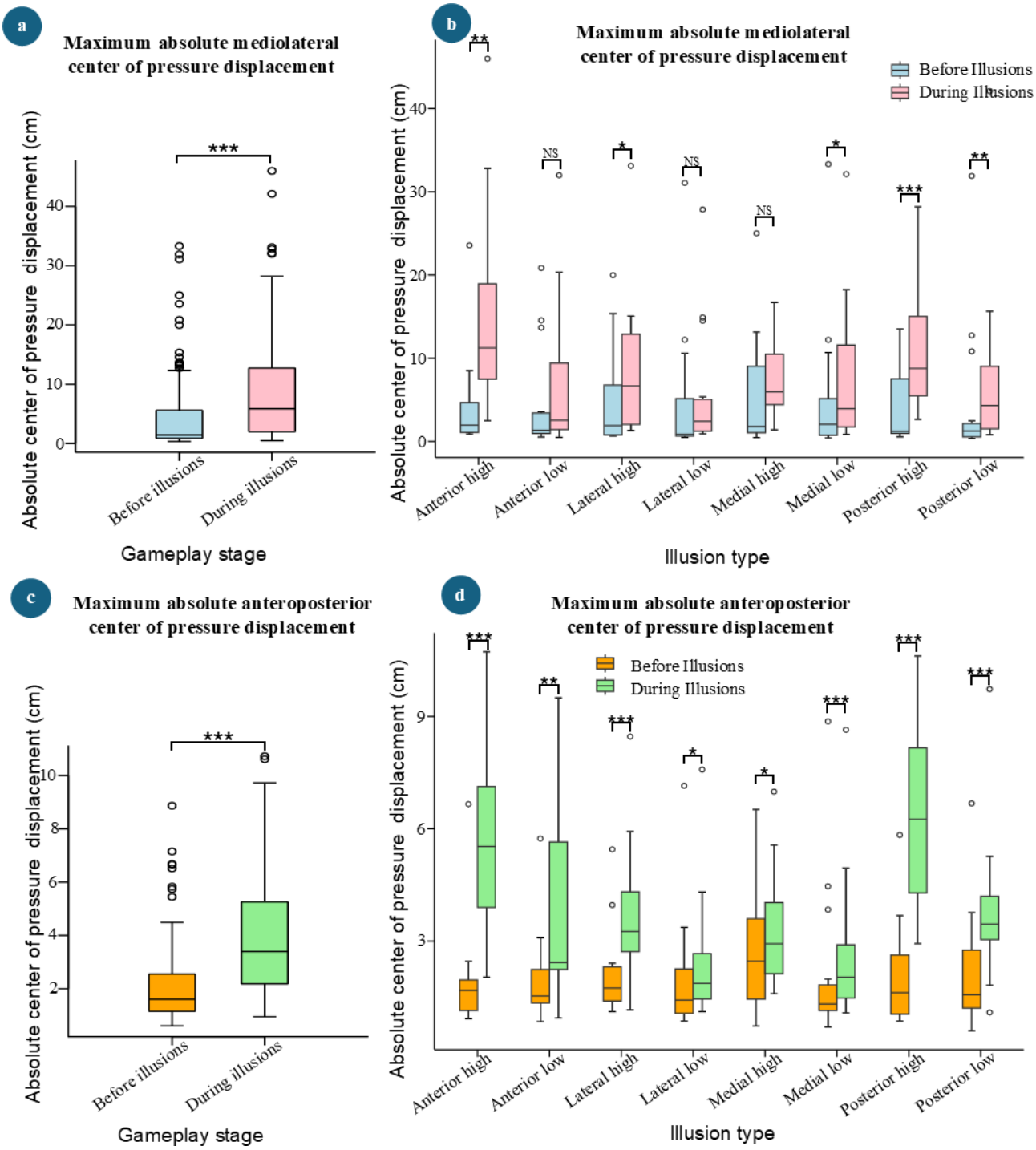
Further analysis of maximum absolute mediolateral CoP displacement before and during the illusion stages for different illusion types revealed significant increases for the following illusion types: anterior high (P=.001, mean difference 10.7 cm); lateral high (P=.041, mean difference 3.8 cm); medial low (P=.048, mean difference 2.5 cm); posterior high (P<.001, mean difference 6.1 cm); and posterior low (P=.003, mean difference 3.5 cm; Figure 5B). Similar to the mean absolute mediolateral CoP comparisons, there were no significant differences in maximum absolute mediolateral CoP displacement for anterior

XSL•FO
RenderX

low (*P*=.078, mean difference 2.7 cm) and lateral low (*P*=.496, mean difference 0.8 cm) illusion types. In addition, no statistically significant increase was observed for the medial high (*P*=.074, mean difference 2.3 cm) illusion type.

In contrast to the behavior of maximum absolute mediolateral CoP displacement, significant increases were observed in maximum absolute anteroposterior CoP displacement during illusion stages compared with before the illusion stages for all types of illusions: anterior high (*P*<.001, mean difference 3.8 cm); anterior low (*P*=.004, mean difference 2 cm); lateral high (*P*<.001, mean difference=1.6 cm); lateral low (*P*=.024, mean difference 0.5 cm); medial high (*P*=.014, mean difference 0.7 cm); medial low (*P*<.001, mean difference 0.6 cm); posterior high (*P*<.001, mean difference 4.2 cm), and posterior low (*P*<.001, mean difference 1.6 cm; Figure 5D).

Figure 5. Comparison of absolute mean center of pressure displacement before and during illusions. (**A**) Absolute mean mediolateral center of pressure displacement comparison in gameplay stages. (**B**) Absolute mean mediolateral center of pressure displacement comparison across different types of virtual illusions. (**C**) Absolute mean anteroposterior center of pressure displacement comparison in gameplay stages. (**D**) Absolute mean anteroposterior center of pressure displacement comparison across different types of virtual illusions (NS: $P > 0.05$, $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$).



XSL•FO
RenderX

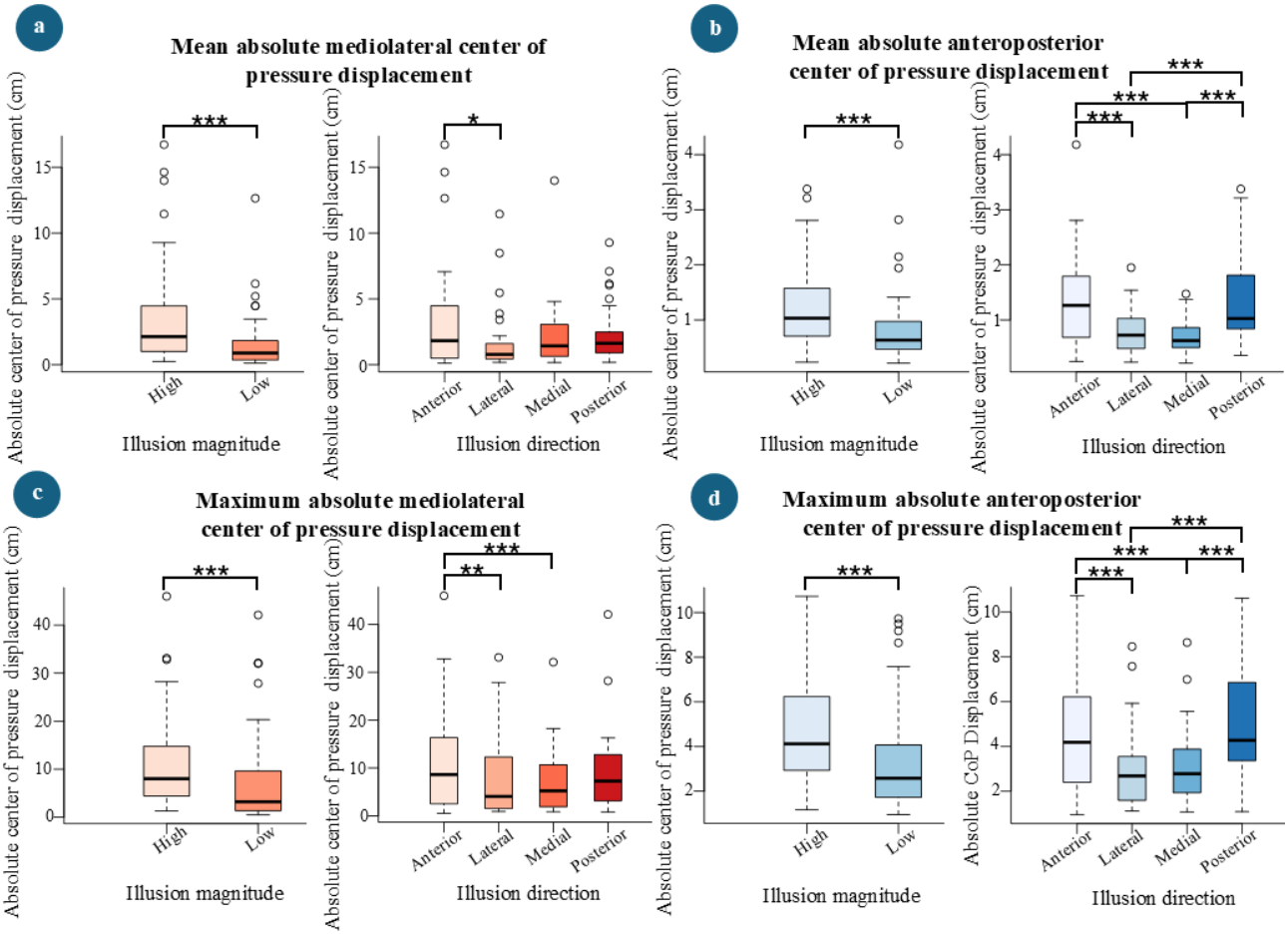
Linear Mixed Effects Models for Absolute Center of Pressure Displacements

ANOVA analyses on the LMEs found no significant interaction effect of illusion direction and magnitude for either mean absolute mediolateral CoP displacement ($F_{3,98}$=0.81, P=.492) or mean absolute anteroposterior CoP displacement ($F_{3,98}$=1.79, P=.155). However, illusion magnitude was identified as a significant main effect, contributing to increased mean absolute mediolateral ($F_{1,98}$=18.24, P<.001) and anteroposterior ($F_{1,98}$=19.30, P<.001) CoP displacements. In addition, illusion direction was found to significantly increase mean absolute CoP displacement in both the mediolateral ($F_{3,98}$=3.12, P=.029) and anteroposterior ($F_{3,98}$=14.04, P<.001) directions. Similarly, illusion magnitude was a significant main effect in increasing maximum absolute mediolateral($F_{1,98}$=12.25, P<.001) and anteroposterior ($F_{1,98}$=28.08, P<.001) CoP displacements. Additionally, illusion direction significantly increased maximum absolute CoP displacement in both mediolateral ($F_{3,98}$=3.13, P=.029) and anteroposterior ($F_{3,98}$=14.51, P<.001) directions. A significant interaction effect was observed for maximum absolute mediolateral CoP displacement ($F_{3,98}$=2.77, P=.046), but no such effect was found for maximum absolute anteroposterior CoP displacement ($F_{3,98}$=1.77, P=.158). Comparison between the high and low illusion magnitude levels revealed that there were significant increases in both mean absolute mediolateral CoP displacement (P<.001, mean difference 1.9 cm) and mean absolute anteroposterior CoP displacement (P<.001, mean difference 0.4 cm) compared with the low magnitude level. High illusion magnitude levels also resulted in significant increases in maximum absolute mediolateral (P<.001, mean difference 3.4 cm) and maximum absolute anteroposterior (P<.001, mean difference 1.5 cm) CoP displacements compared with the low illusion magnitude levels (Figure 6A and B).

Pairwise comparisons using Tukey post hoc tests for illusion directions revealed a significant difference for the anteriorlateral pair (P=.015, estimate 1.8 cm) in mean mediolateral CoP displacement, while no other pairwise comparisons for mean mediolateral CoP displacement showed significant differences. For mean absolute anteroposterior CoP displacement, significant differences were observed for the following illusion direction pairs: anterior-lateral (P<.001, estimate 0.6 cm), anterior-medial (P<.001, estimate 0.7 cm), posterior-lateral (P<.001, estimate 0.6 cm), and posterior-medial (P<.001, estimate 0.7 cm). For maximum mediolateral CoP displacement, significant differences were found for the anterior-lateral (P=.005, estimate 6.2 cm) and anterior-medial (P<.001, estimate 7 cm) pairs, for the maximum absolute anteroposterior CoP displacement, the results were consistent with those for mean absolute anteroposterior displacement, revealing significant differences for the anterior-lateral (P<.001, estimate 1.8 cm), anterior-medial (P<.001, estimate 1.8 cm), posterior-lateral (P<.001, estimate 1.9 cm), and posterior-medial (P<.001, estimate 2 cm) pairs.

XSL•FO
RenderX

Figure 6. Absolute center of pressure displacement across identified main effects from the linear mixed-effects models. (**A**) Mean absolute mediolateral center of pressure displacement across different levels of main effects: illusion magnitude and illusion direction. (**B**) Mean absolute anteroposterior center of pressure displacement across different levels of main effects: illusion magnitude and illusion direction. (**C**) Maximum absolute mediolateral center of pressure displacement across different levels of main effects: illusion magnitude and illusion direction. (**D**) Maximum absolute anteroposterior center of pressure displacement across different levels of main effects: illusion magnitude and illusion direction. (*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$)

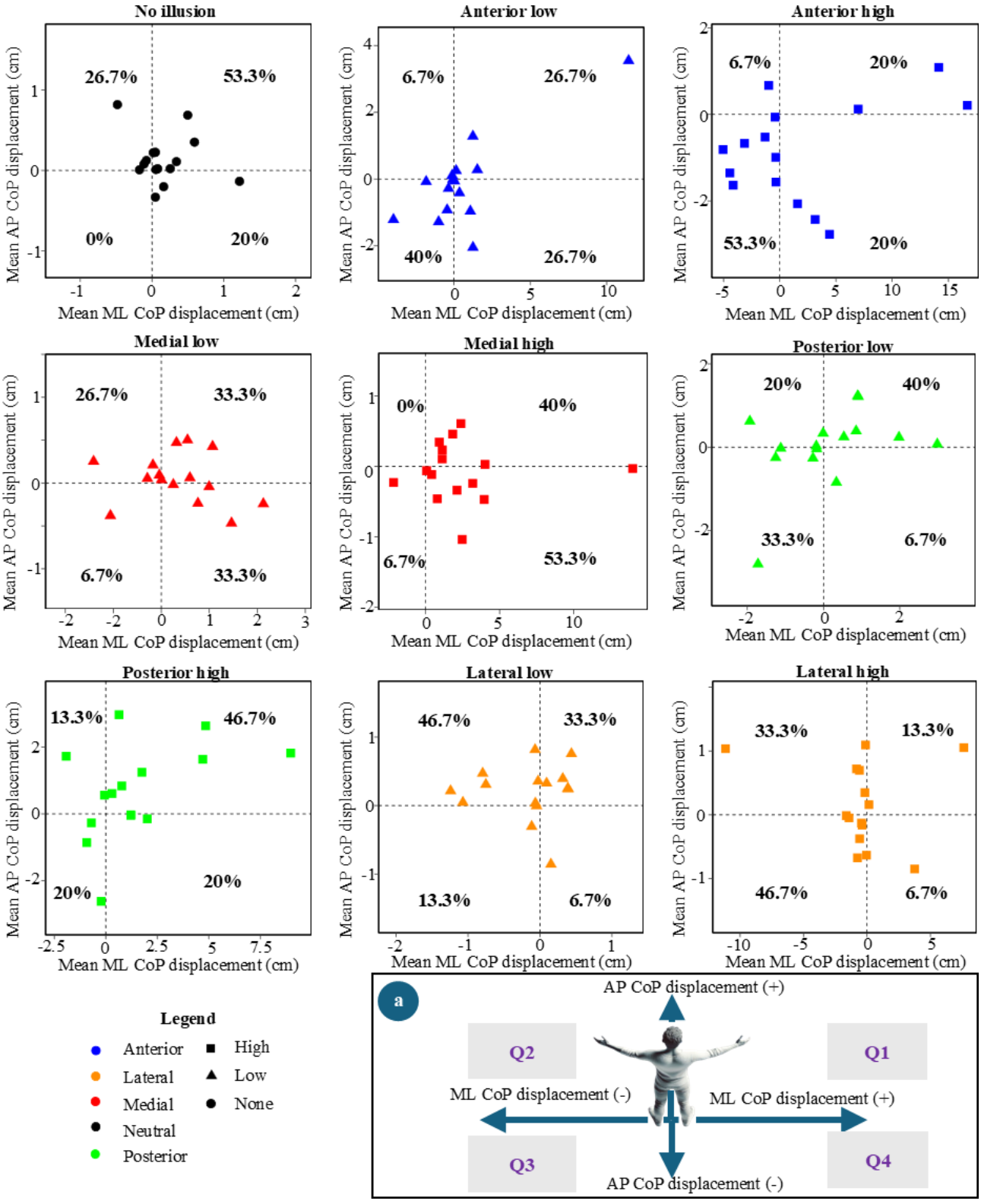


Directional Tendency of Center of Pressure Displacement

Figure 7 illustrates the directional tendency of CoP displacement, where the the mean values of CoP displacement were considered to reveal the direction of the CoP movement. Depending on the direction of the CoP displacement, quadrants were defined (Figure 7A).

XSL•FO
RenderX

Figure 7. Directional tendency of mean center of pressure displacement across all illusion types (including no illusions) and the percentage distribution of center of pressure within each quadrant. (**A**) Definition of the quadrants. Q1-4: quadrants 1-4. AP: Anteroposterior; CoP: Center of pressure; ML: Mediolateral.



For the no illusion stage (53.3%), posterior low illusion (40%), and posterior high illusion (46.7%), the majority of participants' mean CoP displacements were observed in quadrant 1. Notably, the lateral low illusion (46.7%) showed the highest percentage of directional tendency in quadrant 2. For the anterior low illusion (40%), anterior high illusion (53.3%), and lateral high illusion (46.7%), the highest percentage of directional tendency occurred in quadrant 3. The medial high illusion (53.3%) revealed quadrant 4 as having the highest percentage for directional tendency. During the medial low illusion stage, participants exhibited a similar directional tendency in both quadrant 1 and quadrant 3, with 33.3% of the participants' mean CoP displacements falling in each quadrant.

XSL•FO
RenderX

The directional tendency of the mean CoP displacement was further categorized by illusion directions (Table 1). The analysis revealed that quadrant 1 exhibited a higher directional tendency when the illusion direction was posterior (43.3%) and when there was no illusion (53.3%), compared with other illusion directions. Quadrant 2 showed the highest directional tendency during lateral directional illusions (40%), while quadrant 3 had the highest tendency for anterior directional illusions (46.7%). Finally, quadrant 4 ranked highest in directional tendency during medial directional illusions (43.3%). It was found that there is a significant association (P<.001) between illusion direction and the directional tendency of the mean CoP displacement (quadrant).

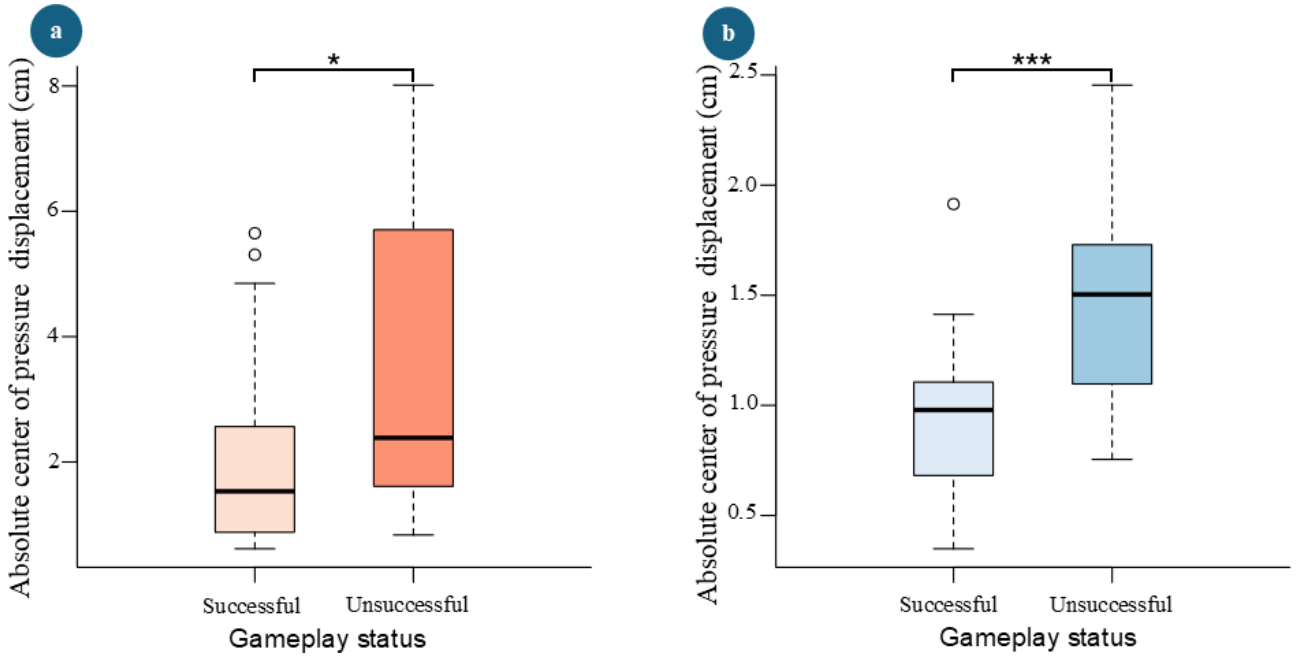
Table . Percentage of data points for mean center of pressure displacement by illusion direction and quadrant (quadrants were defined in Figure 7A).

Illusion direction	Quadrant 1 (%)	Quadrant 2 (%)	Quadrant 3 (%)	Quadrant 4 (%)
No illusion	53.3	26.7	0	20
Anterior	23.3	6.7	46.7	23.3
Posterior	43.3	16.7	26.7	13.3
Medial	36.6	13.3	6.7	43.3
Lateral	23.3	40	30	6.7

Gameplay Performance versus Center of Pressure Displacement

Figure 8 illustrates how mean absolute CoP displacement varied based on whether participants successfully returned the ball to the other side of the table or not.

Figure 8. Mean absolute center of pressure displacement comparison across gameplay statuses. (**A**) In the mediolateral direction. (**B**) In the anteroposterior direction. (*P < 0.05, **P < 0.01, ***P < 0.001)



Mean absolute CoP displacements in both the mediolateral (2.3, SD 1.8) cm and anteroposterior directions (0.9, SD 0.4) cm were lower when participants were successful, compared with unsuccessful attempts, where mean absolute CoP displacements in the mediolateral direction measured 3.5 (SD 2.6) cm and in the anteroposterior direction 1.4 (SD 0.5) cm. Furthermore, paired t test results indicated that the difference in absolute CoP displacements between successful and unsuccessful attempts was statistically significant in both the mediolateral (P=.031) and anteroposterior (P<.001) directions.

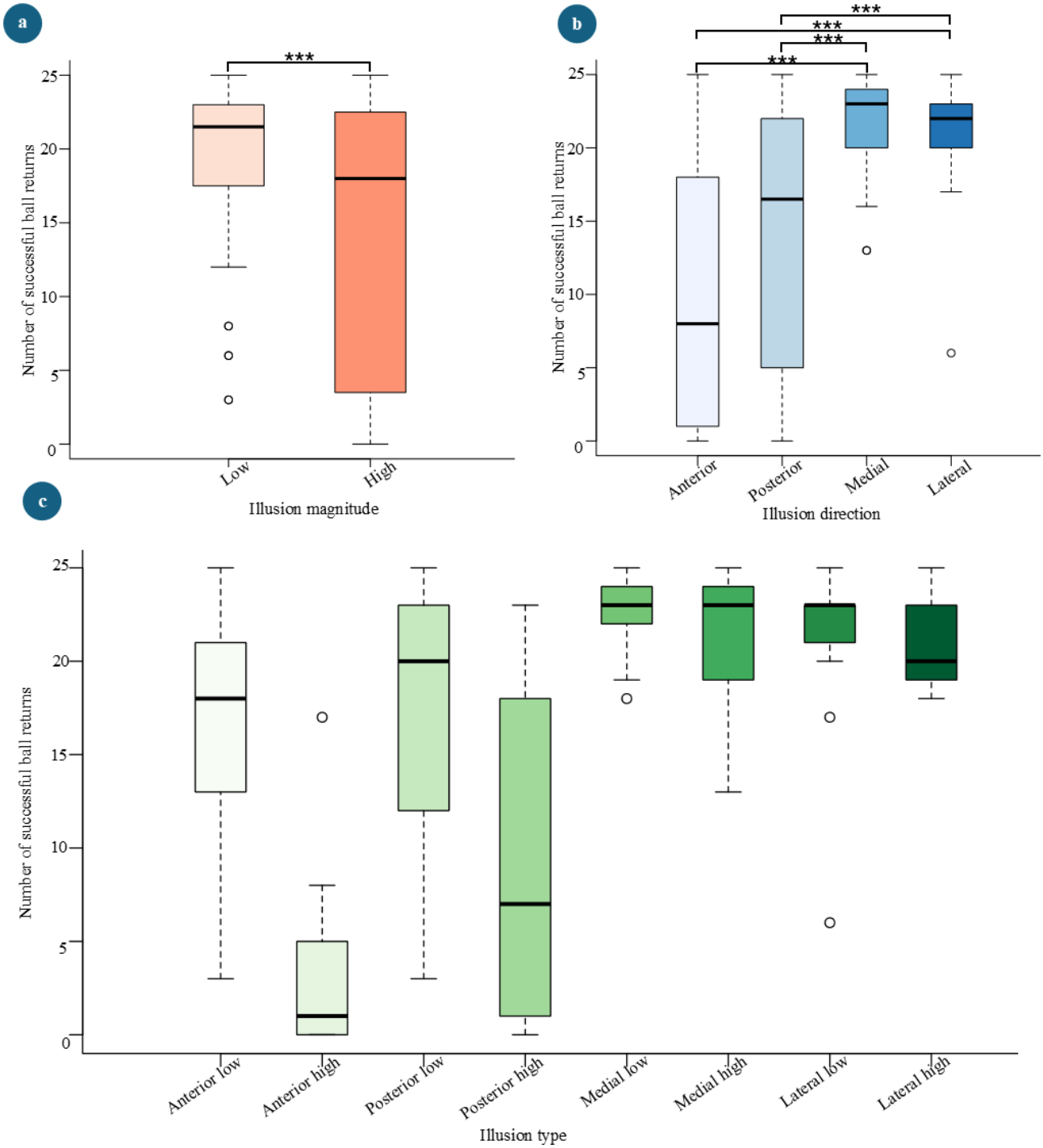
ANOVA analysis conducted on the LME built for the gameplay score revealed that illusion magnitude ($F_{1,84}$=28.64, P<.001) and illusion direction ($F_{3,84}$=31.99, P<.001) were significant main effects in decreasing the score. In addition, a significant interaction effect between illusion magnitude and direction ($F_{3,84}$=8.07, P<.001) was found, further contributing to a decrease in score. A significant difference between the scores during high and low illusion magnitude levels (P<.001) was reported, with an average decrease of 6 successful returns for the high magnitude condition (Figure 9A). Furthermore, significant differences were found in the following pairs of illusion directions: lateral-anterior (P<.001, estimate 11 hits),

XSL•FO
RenderX

medial-anterior (P<.001, estimate 12 hits), lateral-posterior (P<.001, estimate 7 hits), and medial-posterior (P<.001, estimate 7 hits; Figure 9B).

A significant association between illusion direction and gameplay success (P<.001). Gameplay success was further categorized across the different illusion directions (Table 2) to quantify participants' performance. Participants' success percentages were lower during both anterior direction illusions (38.3%) and posterior direction illusions (55.4%) compared with the other 2 directions.

Figure 9. Number of successful returns of the balls. (**A**) Across illusion magnitude levels. (**B**) Across illusion direction levels. (**C**) Across illusion types. (*P < 0.05, **P < 0.01, ***P < 0.001)



XSL•FO
RenderX

Table . Percentage of data points for gameplay status by illusion direction.

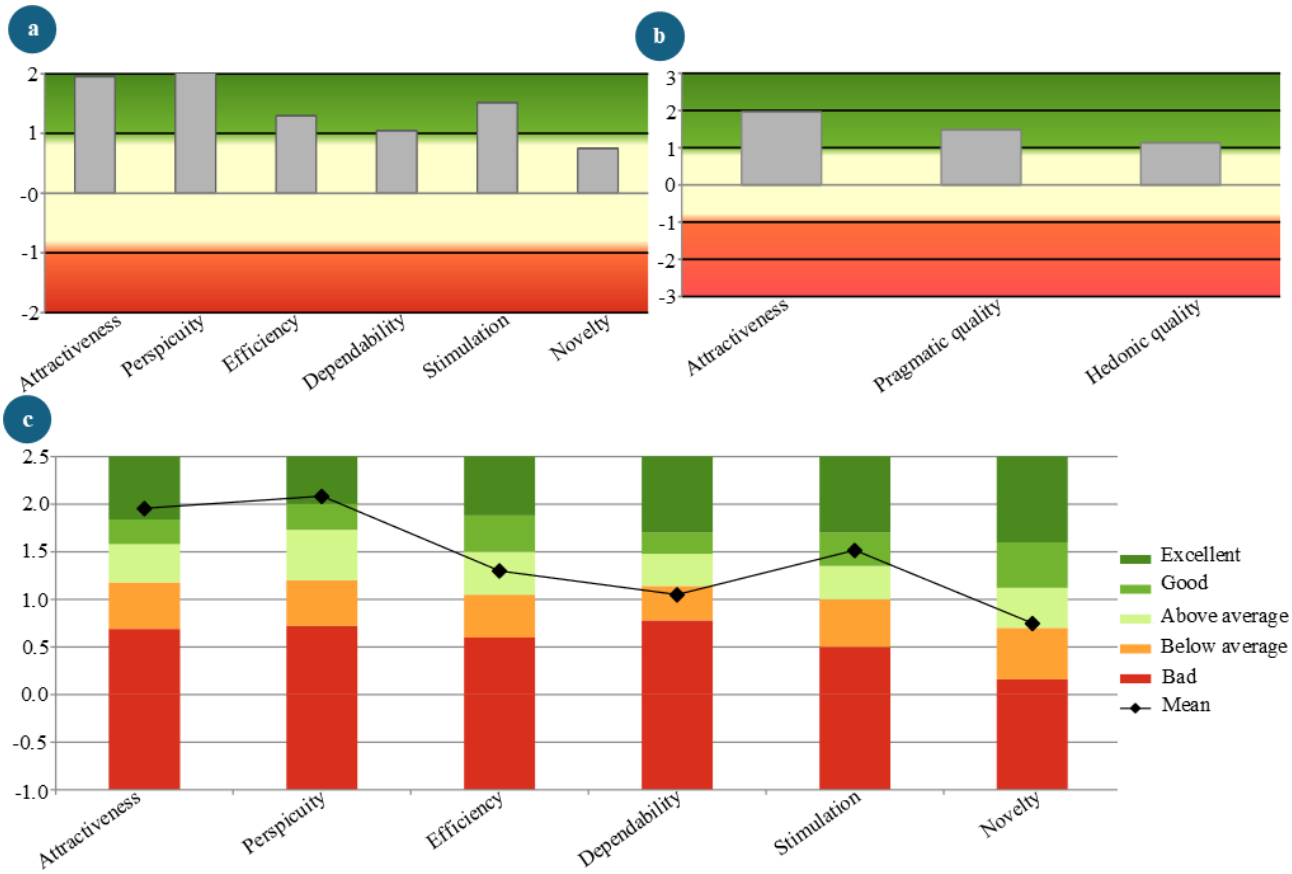
Illusion direction	Successful (%)	Unsuccessful (%)
Anterior	38.3	61.7
Posterior	55.4	44.6
Medial	86.3	13.7
Lateral	83.8	16.2

Questionnaire Responses

User Experience Questionnaire

The UEQ Data Analysis Tool (version 12) was used to analyze the UEQ responses [46]. The results indicate generally positive user perceptions (Figure 10). Attractiveness scored 1.96 (SD 0.42), suggesting that participants found the experiment appealing, while perspicuity, which reflects ease of use and learning, received the highest score of 2.08 (SD 0.46). Efficiency, measuring task performance ease, scored 1.30 (SD 0.72), and dependability scored 1.05 (SD 0.69). Stimulation, which reflects how engaging participants found the experiment, received a score of 1.52 (SD 0.54), while novelty, indicating perceived innovation, had the lowest score at 0.75 (SD 0.89; Figure 10A). Overall, pragmatic quality, which assesses task support, scored 1.48, and hedonic quality, reflecting enjoyment, was 1.13 (Figure 10B).

Figure 10. Distribution of user experience questionnaire responses. (**A**) Across user experience questionnaire scales. (**B**) Across attractiveness, pragmatic quality, and hedonic quality. (**C**) User experience questionnaire results in comparison with benchmark data.



The UEQ benchmark results show strong performance in several areas (Figure 10C). Attractiveness and perspicuity received "Excellent" ratings, placing them within the top 10% of results. Efficiency was rated "Above Average," with 25% of results performing better and 50% performing worse. Dependability was rated "Below Average," with 50% of results performing better and 25% performing worse. Stimulation achieved a "Good" rating, with 10% of results outperforming it and 75% performing worse. Novelty was rated "Above Average," with 25% of results better and 50% worse.

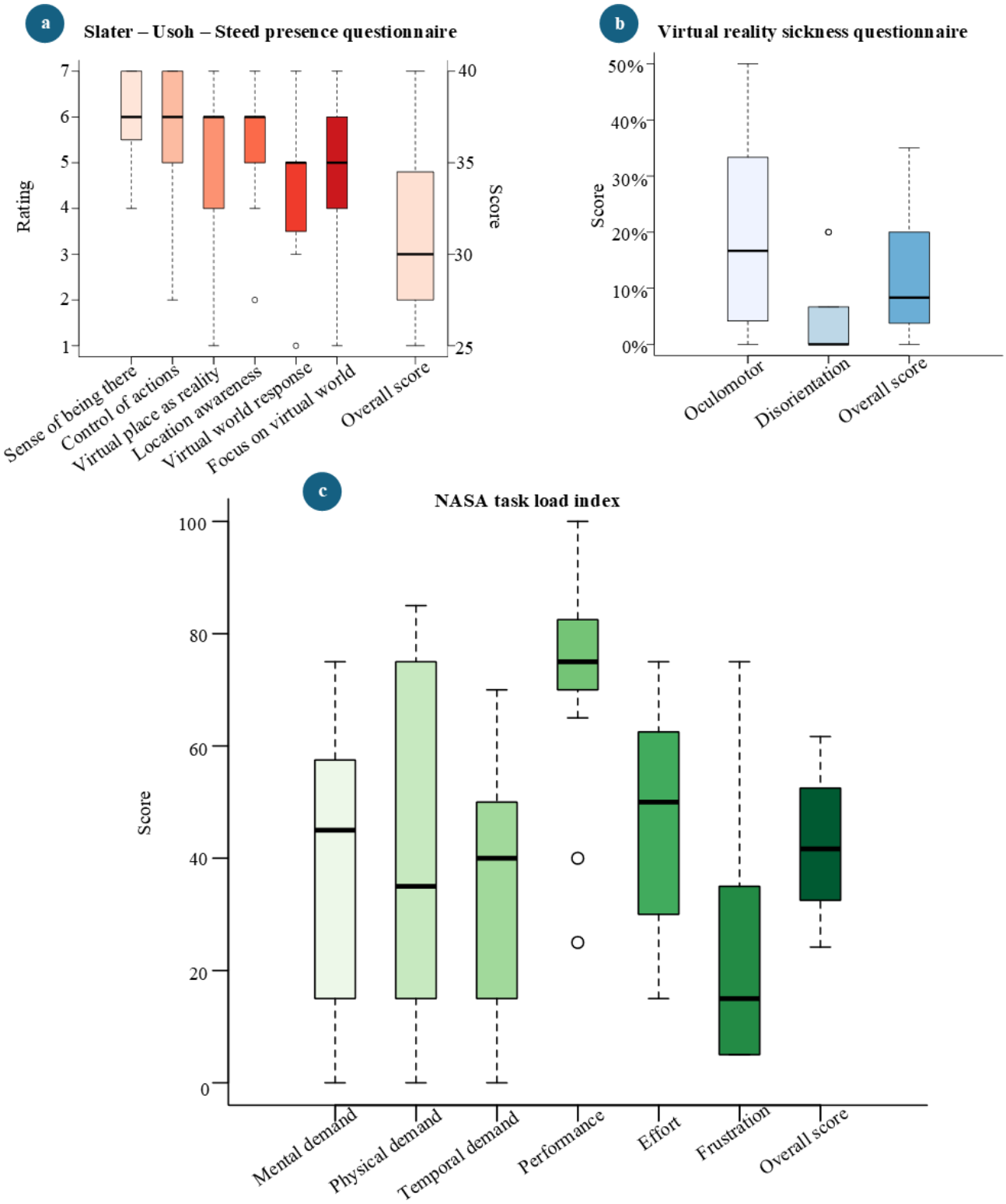
Slater-Usoh-Steed Presence Questionnaire

Participants' responses to the SUS Questionnaire indicated varying levels of perceived presence across the six key aspects (Figure 11A). The question regarding whether the virtual environment felt like reality received a mean score of 6 (SD 0.93), suggesting a high sense of presence. For the question asking whether the environment felt more like a place visited than images seen, the mean was 5.73 (SD 1.49). Awareness of the real-world setting during the experience was rated at 4.93 (SD 1.71). Consistency between the virtual and real-world experiences was scored at 5.33 (SD 1.23). The question about

XSL•FO
RenderX

whether the virtual environment overwhelmed participants' senses had a mean score of 4.4 (SD 1.5), and for the question regarding whether the environment felt more like a place visited than observed, the mean was 4.8 (SD 1.7). Overall, the mean presence score across all 6 questions was 31.2 (SD 4.62), reflecting a moderate to high level of perceived presence.

Figure 11. Distribution of answers for remaining questionnaires. (**A**) Slater-Usoh-Steed Questionnaire. (**B**) Virtual Reality Sickness Questionnaire. (**C**) NASA Task Load Index.



Virtual Reality Sickness Questionnaire

Participants' responses to the VRSQ indicated varying levels of symptoms (Figure 11B). The mean score for oculomotor-related symptoms was 17.78 (SD 16.63), suggesting moderate discomfort related to eye strain or visual issues. For disorientation symptoms, the mean score was lower, at 4.89 (SD 6.89), indicating relatively mild disorientation experiences. The overall mean score across all symptoms was 11.33 (SD

XSL•FO
RenderX

10.76), reflecting a low level of virtual reality sickness symptoms experienced by participants.

NASA Task Load Index

Participants reported varying levels of perceived workload across different task dimensions (Figure 11C). The mean score for mental demand was 36.67 (SD 24.90), while physical demand averaged 41.33 (SD 30.56), indicating moderate to high physical effort. Temporal demand, reflecting time pressure, had a mean score of 35.33 (SD 23.03). Participants rated their performance relatively high, with a mean score of 72.67 (SD 19.44). The effort required to complete tasks was rated at 47.33 (SD 19.81), and frustration levels were relatively low, with a mean score of 24.33 (SD 23.29). The overall workload, combining all dimensions, had a mean score of 42.95 (SD 11.96).

Discussion

Principal Findings

In this study, we incorporated VR illusions generated by offsetting the position of the playing hand within the VR table tennis game. Our selection of the VR illusions was primarily influenced by sensorimotor adaptation phenomena and other theories derived from it. In previous literature, it was highlighted that the users are able to adapt according to the modifications made to the embodiment in VR [47,48]. According to the rubber hand illusion theory, humans can be prompted to accept the rubber hand as their own even though there is a proprioceptive drift from the visual feedback [49-52]. Furthermore, humans have the ability rescale the perceptual-motor system to account for the mismatches and changes in the spatial variables [53-55]. Action-perception coupling was also considered when selecting the suitable VR illusion, as it explains brain's ability to recalibrate the motions according to the feedback of the action [56-58]. This could also explain the fact that participants were able to recalibrate their body posture when they hit or missed the ball. Our hypothesis that posture recalibrations due to the encountered VR illusions could impact the CoP displacements was successfully proven through the obtained results.

Impact on Center of Pressure

CoP displacement was more pronounced during VR illusion stages compared with stages without illusions. Further analyses of CoP displacement revealed that the VR illusions had varying effects on absolute CoP displacements in both the mediolateral and anteroposterior directions.

In previous research, it was implied that the difficulty level and complexity of the VR gameplay could affect the balance performance [59,60]. However, this relationship remains insufficiently characterized. Participants exhibited higher absolute CoP displacements in both the mediolateral and anteroposterior directions when they were unsuccessful in returning the ball, compared with when they were successful (Figure 8). Participants were allowed to familiarize themselves with the game rules before the data collection. The experimenter ensured that each participant scored at least 5 consecutive hits before data collection began. As a result, the first 10 hits, representing the before illusion stage, were recorded with an almost 100% success rate across all participants. This suggests that the increased difficulty of gameplay introduced by the VR illusions could be the primary factor driving the differences in absolute CoP displacements before and during the VR illusion stages.

LMEs revealed that VR illusion magnitude is a significant main effect (P<.001) in increasing absolute CoP displacements in both the mediolateral and anteroposterior directions. In addition, the increase in illusion magnitude was significantly associated with a decrease in gameplay scores (P<.001). Figure 9A illustrates that the scores were higher during the low-magnitude level, indicating that participants found the low VR illusion magnitude more adaptable compared with the high magnitude level. Conversely, mean and maximum absolute CoP displacements in both mediolateral and anteroposterior directions were higher during high VR illusion magnitude levels compared with low magnitude levels (Figure 6). This observation further supports the inversely proportional relationship between gameplay performance (score) and absolute CoP displacement. Furthermore, it emphasizes that the difficulty level of the gameplay can be adjusted by changing the magnitude of the illusion. This finding would allow the therapists to control the level of perturbation effect on the patients in this type of interventions.

Post hoc tests revealed that the scores during anterior and posterior illusion directions differed significantly from the scores during medial and lateral directional VR illusions. Table 2 shows that the success rate of participants in returning the ball dropped to 38.3% and 55.4% during anterior and posterior directional VR illusions, respectively. Based on these observations and the relationship discussed earlier, we would expect the absolute CoP displacement during anterior and posterior directional illusions to be higher than during lateral and medial directional illusions. As anticipated, this pattern was observed in post hoc comparisons for both mean and maximum absolute anteroposterior CoP displacement (Figure 6).

However, this trend was only partially observed for mean and maximum absolute mediolateral CoP displacement. Significant differences were found between anterior and lateral directional illusions for both absolute mediolateral CoP displacement measures. In addition to that, significant differences were reported between anterior and medial directional illusions for maximum absolute mediolateral CoP displacement. The pronounced effect of anterior directional VR illusions may be explained by the fact that participants could easily perceive virtual hand offsets in the anterior direction during gameplay. In contrast, posterior directional VR illusions moved the virtual hand closer to the body, potentially causing the hand to be occluded from the participant's field of view. As a result, participants may have shifted their body in the anteroposterior direction to see the hand properly and hit the ball. The intuitiveness of this adaptation might have caused the higher success percentage during the posterior illusions compared to the anterior illusions.

In this study, participants were instructed to keep their feet stationary, which may have hindered their ability to compensate for the lack of movement freedom when hitting the ball.

XSL•FO
RenderX

However, some VR illusion types might have helped them to compensate for the lack of freedom for body movements. This could explain why no significant differences in absolute CoP displacements were observed between the VR illusion stages and the pre-illusion stages in the comparisons of absolute CoP displacement measures. This finding aligns with the higher scores reported for certain illusion types (Figure 9C), where the differences between before and during illusion stages were not statistically significant for the absolute CoP measures.

Higher F-statistics were reported for illusion direction in LMEs for absolute anteroposterior CoP displacements compared with those for absolute mediolateral CoP displacements, suggesting that directional illusions had a stronger impact on postural control in the anteroposterior direction. This difference could also account for the varying behaviors of absolute mediolateral and anteroposterior CoP measures, as depicted in Figures 4 and 5.

The directional tendencies of participants were analyzed to determine their movement preferences. Fisher exact test confirmed that the tendencies observed in Table 1 were not random but followed a statistically significant pattern. In addition, since all participants in this study were right-handed, it could be hypothesized that these directional tendencies would be mirrored in left-handed participants.

Immersive User Experience Considerations

In this study, we use the term “immersive user experience” to encompass the dimensions captured by the UEQ, SUS, NASA-TLX, and VRSQ questionnaires. Analysis of the UEQ responses showed that participants found the intervention attractive, easy to learn, easy to perform, engaging, and innovative. This was further confirmed by the moderate to high scores in both pragmatic and hedonic qualities. However, the dependability scale scored below average in the benchmark results, suggesting that participants may have found the task predictable. This predictability could be attributed to the consistent ball-feeding method used throughout the experiment.

The SUS presence questionnaire yielded moderate to high scores, indicating that participants felt a strong sense of presence within the VR environment. Crucially, they accepted the virtual hand offsets as part of their virtual body representation, recalibrating their behavior accordingly. Furthermore, moderate NASA-TLX scores, coupled with low overall VRSQ ratings, suggest that the intervention was manageable in terms of perceived workload and induced minimal virtual reality sickness symptoms, making the experience both feasible and user-friendly for participants.

In existing literature on VR illusions for balance, the focus has predominantly been on VR sickness as a user experience dimension [18], and the dimensions such as senses, emotions, and cognition of immersive user experience in VR-based balance interventions is often overlooked [17,61]. Consequently, it was uncertain whether participants would willingly engage in such interventions if asked. Our study’s moderate to high immersive user experience ratings suggest that participants are likely to prefer and engage with our intervention for extended periods.

To enhance our current intervention, future improvements could include diversifying ball feeding methods (eg, serve or dash) and varying ball serving directions. These modifications would address the predictable nature of the game, potentially improving scores on the dependability scale of the UEQ and increasing overall user engagement.

Implications for Standing Balance Rehabilitation Interventions

In balance rehabilitation interventions, CoP displacement is a crucial indicator of balance performance [4,62-64]. Effective balance rehabilitation often involves perturbing postural stability to stimulate improvements [65]. This study demonstrates that our VR-based intervention not only impacts CoP displacement but also allows for predictability in the direction of CoP movements based on VR illusion direction for some extent.

Our intervention enables controlled perturbation effects by adjusting the direction and magnitude of VR illusions, offering flexibility that could benefit clinical populations with compromised balance. In addition, the system requires only a safety harness, a VR HMD, and a VR game, making it a cost-effective and efficient alternative to traditional standing balance interventions, which often demand high therapist involvement. The simple setup allows for easy implementation in both clinical and home-based settings. With this flexibility and ease of implementation, therapists can tailor the VR intervention to design personalized rehabilitation paradigms that enhance engagement and address specific patient needs, potentially leading to improved balance outcomes.

Furthermore, throughout all the eight VR illusion based trials participants responded to the provide VR illusions unlike in some research in literature where participants neglected the VR illusions even though they initially responded to the illusions [19]. This sustained responsiveness suggests that our approach may overcome the sensory reweighting phenomenon observed in other studies, where visual dependency decreases over time as participants adapt to visual perturbations [66,67]. The consistency in participant responses also indicates that our VR paradigm effectively maintains sensorimotor engagement throughout the intervention period, potentially due to the ecological validity of our visual stimuli [68]. This finding aligns with research by Slater and colleagues [69], who demonstrated that contextually relevant visual perturbations sustain postural responses more effectively than abstract visual flows. In addition, the immersive nature of modern VR technology likely contributes to this sustained effect, as higher levels of presence have been correlated with stronger postural responses in virtual environments [70,71]. Future research should investigate the long-term sustainability of these responses during extended rehabilitation protocols, as Keshner and Kenyon [72] suggest that maintaining responsiveness to visual perturbations over multiple sessions is crucial for meaningful balance improvements in clinical populations.

The highest mean difference between the before and during illusion stages for mean absolute mediolateral CoP displacement was 4.4 cm, while it was 1.1 cm for mean absolute anteroposterior CoP displacement. These values suggest that individuals with compromised balance might experience

XSL•FO
RenderX

significant postural challenges, potentially pushing their limits [73]. However, in a controlled virtual environment, these perturbations can be carefully managed, allowing users to safely engage with challenging balance tasks. The ability to adjust both the direction and magnitude of CoP displacement through VR illusions provides an opportunity to incrementally challenge a user's postural control while mitigating the risk of overexertion or injury. Thus, our intervention offers a dynamic and adaptable platform for balance training. This adaptability is particularly beneficial for rehabilitation in clinical populations, allowing for gradual improvement in postural stability while minimizing excessive strain.

It is also important to note that some participants used the aid harness to maintain their balance during gameplay, indicating that the VR-induced perturbations posed a greater challenge for certain individuals. While the harness ensured safety, its use highlights the need to tailor the intervention to varying balance abilities. This safety feature allows for controlled difficulty adjustments, making the intervention suitable for a wider range of clinical populations. Over time, the goal would be to reduce reliance on external support as balance improves and confidence increases.

Limitations and Future Directions

This study was conducted with a small but relatively balanced sample of 15 healthy young adults. While the findings may be generalizable within this demographic, caution should be taken when extending them to older adults or clinical populations. While our results provide valuable insights into VR-based balance rehabilitation interventions, future research should include diverse demographic groups, such as older adults and clinical populations, to explore confounding factors unique to these groups.

In addition to including a more diverse sample across different age groups and clinical backgrounds, future studies should investigate other aspects of lower limb rehabilitation. Specifically, the influence of VR illusions during treadmill walking should be explored, as treadmill-based rehabilitation is widely used in clinical settings. Furthermore, the feasibility of implementing home-based therapeutic interventions will be examined in future work.

Conclusions

We proposed a novel approach to standing balance rehabilitation interventions by leveraging VR illusions to influence CoP displacement while maintaining a high level of immersive user experience. By systematically varying the direction and magnitude of the VR illusions, we successfully induced perturbations that challenged participants' balance, leading to notable differences in CoP displacement across various illusion types. This innovative intervention not only impacted balance performance but also provided valuable insights into participants' responsiveness to different VR conditions. Our findings indicate that the intervention was engaging, with participants reporting a moderate perceived workload, suggesting that the tasks were both manageable and stimulating. Minimal virtual reality sickness was observed, further confirming the feasibility of the intervention. In addition, participants experienced a moderate to high sense of presence, which is crucial for the effectiveness of VR-based interventions. Overall, this study supports the use of VR illusions as a dynamic and adaptable tool for standing balance rehabilitation interventions, offering a controlled yet enjoyable environment to progressively challenge and enhance postural stability. Furthermore, our insights into the relationships between game performance and balance performance, as well as the directional tendencies of CoP movements in response to illusion direction, provide a solid foundation for designing effective exergames for balance rehabilitation interventions.

Acknowledgments

The authors would like to thank all participants for their contribution to the success of this research. We are also grateful to Dr James Williamson and Richard Roth for their initial support during the data collection phase.

Data Availability

Data used for the analyses in this paper are available in Multimedia Appendix 1.

Authors' Contributions

RAMA was involved in the conceptualization, methodology design, data collection, data curation, data analysis, writing of original draft, reviewing and editing of paper. APLB was involved in the conceptualization, methodology design, data analysis, reviewing of paper, and supervision. TJMD was involved in the conceptualization, methodology design, data collection, data curation, data analysis, writing of original draft, reviewing of paper, and supervision. NB was involved in the conceptualization, methodology design, reviewing of paper, project administration, and supervision.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Data analyses and related data files.

XSL•FO
RenderX

[ZIP File, 2907 KB - games_v13i1e70376_app1.zip]

References

1. Rogers MW, Mille ML. Balance perturbations. Handb Clin Neurol 2018;159:85-105. [doi: 10.1016/B978-0-444-63916-5.00005-7] [Medline: 30482334]
2. Ivanenko Y, Gurfinkel VS. Human postural control. Front Neurosci 2018;12:171. [doi: 10.3389/fnins.2018.00171] [Medline: 29615859]
3. Gerards MHG, McCrum C, Mansfield A, Meijer K. Perturbation-based balance training for falls reduction among older adults: current evidence and implications for clinical practice. Geriatr Gerontol Int 2017 Dec;17(12):2294-2303. [doi: 10.1111/ggi.13082] [Medline: 28621015]
4. Tokur D, Grimmer M, Seyfarth A. Review of balance recovery in response to external perturbations during daily activities. Hum Mov Sci 2020 Feb;69:102546. [doi: 10.1016/j.humov.2019.102546] [Medline: 31989948]
5. Richmond SB, Fling BW, Lee H, Peterson DS. The assessment of center of mass and center of pressure during quiet stance: current applications and future directions. J Biomech 2021 Jun 23;123:110485. [doi: 10.1016/j.jbiomech.2021.110485] [Medline: 34004395]
6. Mahmud MR, Stewart M, Cordova A, Quarles J. Standing balance improvement using vibrotactile feedback in virtual reality. 2022 Presented at: Proceedings of the 28th ACM Symposium on Virtual Reality Software and Technology; Nov 29 to Dec 1, 2022; Tsukuba Japan p. 1-11. [doi: 10.1145/3562939.3565638]
7. Wachholz F, Tiribello F, Promsri A, Federolf P. Should the minimal intervention principle be considered when investigating dual-tasking effects on postural control? Brain Sci 2020 Jan;10(1):1. [doi: 10.3390/brainsci10010001]
8. Hunt CM, Widener G, Allen DD. Variability in postural control with and without balance-based torso- weighting in people with multiple sclerosis and healthy controls. Phys Ther 2014 Oct;94(10):1489-1498. [doi: 10.2522/ptj.20130288] [Medline: 24903118]
9. Hernan G, Ingale N, Somayaji S, Veerubhotla A. Virtual reality-based interventions to improve balance in patients with traumatic brain injury: a scoping review. Brain Sci 2024 Apr 26;14(5):429. [doi: 10.3390/brainsci14050429] [Medline: 38790408]
10. Ehioghae M, Montoya A, Keshav R, et al. Effectiveness of virtual reality-based rehabilitation interventions in improving postoperative outcomes for orthopedic surgery patients. Curr Pain Headache Rep 2024 Jan;28(1):37-45. [doi: 10.1007/s11916-023-01192-5] [Medline: 38032538]
11. Darekar A, McFadyen BJ, Lamontagne A, Fung J. Efficacy of virtual reality-based intervention on balance and mobility disorders post-stroke: a scoping review. J NeuroEngineering Rehabil 2015 Dec;12(1):1-14. [doi: 10.1186/s12984-015-0035-3]
12. Assländer L, Albrecht M, Diehl M, Missen KJ, Carpenter MG, Streuber S. Estimation of the visual contribution to standing balance using virtual reality. Sci Rep 2023 Feb 14;13(1):2594. [doi: 10.1038/s41598-023-29713-7] [Medline: 36788259]
13. Keshner EA, Lamontagne A. The untapped potential of virtual reality in rehabilitation of balance and gait in neurological disorders. Front Virtual Real 2021 Mar;2:641650. [doi: 10.3389/frvir.2021.641650] [Medline: 33860281]
14. Peterson SM, Rios E, Ferris DP. Transient visual perturbations boost short-term balance learning in virtual reality by modulating electrocortical activity. J Neurophysiol 2018 Oct 1;120(4):1998-2010. [doi: 10.1152/jn.00292.2018] [Medline: 30044183]
15. Mohebbi A, Amiri P, Kearney RE. Identification of human balance control responses to visual inputs using virtual reality. J Neurophysiol 2022 Apr 1;127(4):1159-1170. [doi: 10.1152/jn.00283.2021] [Medline: 35353629]
16. Chiarovano E, de Waele C, MacDougall HG, Rogers SJ, Burgess AM, Curthoys IS. Maintaining balance when looking at a virtual reality three-dimensional display of a field of moving dots or at a virtual reality scene. Front Neurol 2015;6:164. [doi: 10.3389/fneur.2015.00164] [Medline: 26284023]
17. Chiarovano E, Wang W, Rogers SJ, MacDougall HG, Curthoys IS, de Waele C. Balance in virtual reality: effect of age and bilateral vestibular loss. Front Neurol 2017;8:5. [doi: 10.3389/fneur.2017.00005] [Medline: 28163693]
18. Chander H, Kodithuwakku Arachchige SNK, Hill CM, et al. Virtual-reality-induced visual perturbations impact postural control system behavior. Behav Sci (Basel) 2019 Nov 12;9(11):113. [doi: 10.3390/bs9110113] [Medline: 31718105]
19. Ketterer J, Ringhof S, Gehring D, Gollhofer A. Sinusoidal optic flow perturbations reduce transient but not continuous postural stability: a virtual reality-based study. Front Physiol 2022;13:803185. [doi: 10.3389/fphys.2022.803185] [Medline: 35665227]
20. Gago MF, Yelshyna D, Bicho E, et al. Compensatory postural adjustments in an oculus virtual reality environment and the risk of falling in Alzheimer's disease. Dement Geriatr Cogn Dis Extra 2016;6(2):252-267. [doi: 10.1159/000447124] [Medline: 27489559]
21. Canning CG, Allen NE, Nackaerts E, Paul SS, Nieuwboer A, Gilat M. Virtual reality in research and rehabilitation of gait and balance in Parkinson disease. Nat Rev Neurol 2020 Aug;16(8):409-425. [doi: 10.1038/s41582-020-0370-2] [Medline: 32591756]
22. Rao HM, Talkar T, Ciccarelli G, et al. Sensorimotor conflict tests in an immersive virtual environment reveal subclinical impairments in mild traumatic brain injury. Sci Rep 2020 Sep 8;10(1):14773. [doi: 10.1038/s41598-020-71611-9] [Medline: 32901067]

XSL•FO
RenderX

23. Thompson JD, Franz JR. Do kinematic metrics of walking balance adapt to perturbed optical flow? Hum Mov Sci 2017 Aug;54:34-40. [doi: 10.1016/j.humov.2017.03.004] [Medline: 28371662]
24. Madshaven JM, Markseth TF, Jomås DB, et al. Investigating the user experience of virtual reality rehabilitation solution for biomechatronics laboratory and home environment. Front Virtual Real 2021;2:645042. [doi: 10.3389/frvir.2021.645042]
25. Rose T, Nam CS, Chen KB. Immersion of virtual reality for rehabilitation - review. Appl Ergon 2018 May;69:153-161. [doi: 10.1016/j.apergo.2018.01.009] [Medline: 29477323]
26. Winter C, Kern F, Gall D, Latoschik ME, Pauli P, Käthner I. Immersive virtual reality during gait rehabilitation increases walking speed and motivation: a usability evaluation with healthy participants and patients with multiple sclerosis and stroke. J NeuroEngineering Rehabil 2021 Dec;18(1):68. [doi: 10.1186/s12984-021-00848-w]
27. Fregna G, Schincaglia N, Baroni A, Straudi S, Casile A. A novel immersive virtual reality environment for the motor rehabilitation of stroke patients: A feasibility study. Front Robot AI 2022;9:906424. [doi: 10.3389/frobt.2022.906424] [Medline: 36105763]
28. van Gemert T, Nilsson NC, Hirzle T, Bergström J. Sicknificant steps: a systematic review and meta-analysis of VR sickness in walking-based locomotion for virtual reality. 2024 May 11 Presented at: CHI '24; May 11-16, 2024; Honolulu HI USA URL: https://dl.acm.org/doi/proceedings/10.1145/3613904 [doi: 10.1145/3613904.3641974]
29. Clocchiatti A. Self-avatar motion retargeting for virtual reality post-stroke rehabilitation therapy. Presented at: CHI '24: CHI Conference on Human Factors in Computing Systems; May 11-16, 2024; Honolulu HI USA. [doi: 10.1145/3613905.3651131]
30. Clark J, Kelliher A. Understanding the needs and values of rehabilitation therapists in designing and implementing telehealth solutions. Presented at: CHI '21: CHI Conference on Human Factors in Computing Systems; May 8-13, 2021; Yokohama Japan. [doi: 10.1145/3411763.3451704]
31. Doré B, Gaudreault A, Everard G, et al. Acceptability, feasibility, and effectiveness of immersive virtual technologies to promote exercise in older adults: a systematic review and meta-analysis. Sensors (Basel) 2023 Feb 24;23(5):2506. [doi: 10.3390/s23052506] [Medline: 36904709]
32. Li X, Ren X, Suzuki X, Yamaji N, Fung KW, Gondo Y. Designing a multisensory VR game prototype for older adults - the acceptability and design implications. Presented at: CHI '24: CHI Conference on Human Factors in Computing Systems; May 11-16, 2024; Honolulu HI USA. [doi: 10.1145/3613904.3642948]
33. Kim YM, Rhiu I, Yun MH. A systematic review of A virtual reality system from the perspective of user experience. International Journal of Human–Computer Interaction 2020 Jun 14;36(10):893-910. [doi: 10.1080/10447318.2019.1699746]
34. Bastian AJ. Understanding sensorimotor adaptation and learning for rehabilitation. Curr Opin Neurol 2008 Dec;21(6):628-633. [doi: 10.1097/WCO.0b013e328315a293] [Medline: 18989103]
35. Moore RT, Cluff T. Individual differences in sensorimotor adaptation are conserved over time and across force-field tasks. Front Hum Neurosci 2021;15:692181. [doi: 10.3389/fnhum.2021.692181] [Medline: 34916916]
36. Sarlegna FR, Bernier PM. On the link between sensorimotor adaptation and sensory recalibration. J Neurosci 2010 Sep 1;30(35):11555-11557. [doi: 10.1523/JNEUROSCI.3040-10.2010] [Medline: 20810876]
37. Avraham G, Taylor JA, Breska A, Ivry RB, McDougle SD. Contextual effects in sensorimotor adaptation adhere to associative learning rules. Elife 2020;11:e75801.
38. Baker S, Waycott J, Carrasco R, Hoang T, Vetere F. Exploring the design of social VR experiences with older adults. Presented at: DIS '19; Jun 23-28, 2019; San Diego CA USA p. 303-315 URL: https://dl.acm.org/doi/proceedings/10.1145/3322276 [doi: 10.1145/3322276.3322361]
39. Bradwell HL, Cooper L, Baxter R, et al. Implementation of virtual reality motivated physical activity via omnidirectional treadmill in a supported living facility for older adults: a mixed-methods evaluation.: virtual reality to motivate physical activity for older adults. : ACM; 2024 May 11 Presented at: CHI '24; May 11-16, 2024; Honolulu HI USA p. 1-13 URL: https://dl.acm.org/doi/proceedings/10.1145/3613904 [doi: 10.1145/3613904.3642281]
40. Kim HK, Park J, Choi Y, Choe M. Virtual reality sickness questionnaire (VRSQ): Motion sickness measurement index in a virtual reality environment. Appl Ergon 2018 May;69:66-73. [doi: 10.1016/j.apergo.2017.12.016] [Medline: 29477332]
41. Osaba MY, Martelli D, Prado A, Agrawal SK, Lalwani AK. Age-related differences in gait adaptations during overground walking with and without visual perturbations using a virtual reality headset. Sci Rep 2020 Sep 21;10(1):15376. [doi: 10.1038/s41598-020-72408-6] [Medline: 32958807]
42. Schrepp M, Hinderks A, Thomaschewski J. Applying the user experience questionnaire (UEQ) in different evaluation scenarios. In: Design, User Experience, and Usability Theories, Methods, and Tools for Designing the User Experience: Third International Conference, DUXU 2014, Held as Part of HCI International 2014, Heraklion, Crete, Greece, June 22-27, 2014, Proceedings, Part I 3: Springer:383-392.
43. Hart SG. NASA-Task Load Index (NASA-TLX); 20 years later. Proceedings of the Human Factors and Ergonomics Society Annual Meeting 2006 Oct;50(9):904-908. [doi: 10.1177/154193120605000909]
44. Steed A, Frlston S, López MM, Drummond J, Pan Y, Swapp D. An "In the Wild" experiment on presence and embodiment using consumer virtual reality equipment. IEEE Trans Vis Comput Graph 2016 Apr;22(4):1406-1414. [doi: 10.1109/TVCG.2016.2518135] [Medline: 26780804]

XSL•FO
RenderX

45. Borrelli J, Komisar V, Novak AC, Maki BE, King EC. Extending the center of pressure to incorporate handhold forces: derivation and sample application. J Biomech 2020 May 7;104:109727. [doi: 10.1016/j.jbiomech.2020.109727] [Medline: 32173031]
46. Hinderks A, Schrepp M, Thomaschewski J. User Experience Questionnaire (UEQ). 2024. URL: https://www.ueq-online.org/ [accessed 2025-06-11]
47. Dobricki M, de la Rosa S. The structure of conscious bodily self-perception during full-body illusions. PLOS ONE 2013;8(12):e83840. [doi: 10.1371/journal.pone.0083840] [Medline: 24376765]
48. Slater M, Spanlang B, Sanchez-Vives MV, Blanke O. First person experience of body transfer in virtual reality. PLOS ONE 2010 May 12;5(5):e10564. [doi: 10.1371/journal.pone.0010564] [Medline: 20485681]
49. Rohde M, Di Luca M, Ernst MO. The rubber hand illusion: feeling of ownership and proprioceptive drift do not go hand in hand. PLoS ONE 2011;6(6):e21659. [doi: 10.1371/journal.pone.0021659] [Medline: 21738756]
50. Ehrsson HH, Holmes NP, Passingham RE. Touching a rubber hand: feeling of body ownership is associated with activity in multisensory brain areas. J Neurosci 2005 Nov 9;25(45):10564-10573. [doi: 10.1523/JNEUROSCI.0800-05.2005] [Medline: 16280594]
51. Finotti G, Garofalo S, Costantini M, Proffitt DR. Temporal dynamics of the rubber hand illusion. Sci Rep 2023 May 9;13(1):7526. [doi: 10.1038/s41598-023-33747-2] [Medline: 37160924]
52. Svensson P, Malešević N, Wijk U, Björkman A, Antfolk C. The rubber hand illusion evaluated using different stimulation modalities. Front Neurosci 2023;17:1237053. [doi: 10.3389/fnins.2023.1237053] [Medline: 37781250]
53. Brand MT, de Oliveira RF. Recalibration in functional perceptual-motor tasks: a systematic review. Hum Mov Sci 2017 Dec;56(Pt B):54-70. [doi: 10.1016/j.humov.2017.10.020] [Medline: 29102772]
54. Kunz BR, Creem-Regehr SH, Thompson WB. Does perceptual-motor calibration generalize across two different forms of locomotion? Investigations of walking and wheelchairs. PLoS ONE 2013;8(2):e54446. [doi: 10.1371/journal.pone.0054446] [Medline: 23424615]
55. Wilson R, Alder D, Miller-Dicks M, Poolton J. An examination of perceptual-motor recalibration in a 1-vs-1 anticipation test. J Sports Sci 2023 Sep;41(15):1471-1482. [doi: 10.1080/02640414.2023.2275486] [Medline: 37902048]
56. Huesmann K, Loffing F. Perception-action coupling in anticipation research: a classification and its application to racket sports. Front Psychol 2024;15:1396873. [doi: 10.3389/fpsyg.2024.1396873] [Medline: 39108427]
57. Poljac E, van Schie HT, Bekkering H. Understanding the flexibility of action-perception coupling. Psychol Res 2009 Jul;73(4):578-586. [doi: 10.1007/s00426-009-0238-y] [Medline: 19347358]
58. Schoner G, Dijkstra TMH, Jeka JJ. Action-perception patterns emerge from coupling and adaptation. Ecol Psych 1998 Sep 1;10(3):323-346. [doi: 10.1207/s15326969eco103&4_9]
59. Arnold N, Wilson O, Thompson L. Virtual reality training affects center of pressure (COP)-based balance parameters in older individuals. Appl Sci (Basel) 2024 Jan;14(16):7182. [doi: 10.3390/app14167182]
60. Sylcott B, Lin CC, Williams K, Hinderaker M. Investigating the use of virtual reality headsets for postural control assessment: instrument validation study. JMIR Rehabil Assist Technol 2021 Nov 15;8(4):e24950. [doi: 10.2196/24950] [Medline: 34779789]
61. Phu S, Vogrin S, Al Saedi A, Duque G. Balance training using virtual reality improves balance and physical performance in older adults at high risk of falls. Clin Interv Aging 2019;14:1567-1577. [doi: 10.2147/CIA.S220890] [Medline: 31695345]
62. Elvitigala DS, Matthies DJC, David L, Weerasinghe C, Nanayakkara S. GymSoles: improving squats and dead-lifts by visualizing the user's center of pressure. Presented at: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems; May 4-9, 2019; Glasgow Scotland UK p. 1-12. [doi: 10.1145/3290605.3300404]
63. Quijoux F, Nicolaï A, Chairi I, et al. A review of center of pressure (COP) variables to quantify standing balance in elderly people: algorithms and open-access code. Physiol Rep 2021 Nov;9(22):e15067. [doi: 10.14814/phy2.15067] [Medline: 34826208]
64. Noamani A, Vette AH, Rouhani H. Instrumented functional test for objective outcome evaluation of balance rehabilitation in elderly fallers: a clinical study. Gerontology 2022;68(11):1233-1245. [doi: 10.1159/000521001] [Medline: 35093943]
65. Blenkinsop GM, Pain MTG, Hiley MJ. Balance control strategies during perturbed and unperturbed balance in standing and handstand. R Soc Open Sci 2017 Jul;4(7):161018. [doi: 10.1098/rsos.161018] [Medline: 28791131]
66. Peterka RJ. Sensorimotor integration in human postural control. J Neurophysiol 2002 Sep;88(3):1097-1118. [doi: 10.1152/jn.2002.88.3.1097] [Medline: 12205132]
67. Bronstein AM. A conceptual model of the visual control of posture. Prog Brain Res 2019;248:285-302. [doi: 10.1016/bs.pbr.2019.04.023] [Medline: 31239139]
68. Lubetzky AV, Kelly J, Wang Z, et al. Contextual sensory integration training via head mounted display for individuals with vestibular disorders: a feasibility study. Disabil Rehabil Assist Technol 2022 Jan;17(1):74-84. [doi: 10.1080/17483107.2020.1765419] [Medline: 32421374]
69. Slater M, Khanna P, Mortensen J, Yu I. Visual realism enhances realistic response in an immersive virtual environment. IEEE Comput Graph Appl 2009;29(3):76-84. [doi: 10.1109/mcg.2009.55] [Medline: 19642617]
70. Weech S, Kenny S, Barnett-Cowan M. Presence and cybersickness in virtual reality are negatively related: a review. Front Psychol 2019;10:158. [doi: 10.3389/fpsyg.2019.00158] [Medline: 30778320]

XSL•FO
RenderX

71. Tieri G, Morone G, Paolucci S, Iosa M. Virtual reality in cognitive and motor rehabilitation: facts, fiction and fallacies. Expert Rev Med Devices 2018 Feb;15(2):107-117. [doi: 10.1080/17434440.2018.1425613] [Medline: 29313388]
72. Keshner EA, Kenyon RV. Postural and spatial orientation driven by virtual reality. In: Advanced Technologies in Rehabilitation: IOS Press:209-228 URL: https://ebooks.iospress.nl/volumearticle/12269 [accessed 2025-06-09]
73. Boumer TC, Martello SK, Devetak GF, Iucksch DD, Scrok ND, Manffra EF. Influence of feet position on COP measures in post-stroke patients in quiet standing. Gait Posture 2018 Oct;66:58-62. [doi: 10.1016/j.gaitpost.2018.07.182] [Medline: 30165285]

Abbreviations

CoP: center of pressure
HMD: head mounted display
LME: linear mixed-effects model
NASA-TLX: NASA Task Load Index
SUS: Slater-Usoh-Steed
UEQ: User Experience Questionnaire
VR: virtual reality
VRSQ: Virtual Reality Sickness Questionnaire

Edited by A Coristine; submitted 20.12.24; peer-reviewed by IK Singgih, M Bayattork; revised version received 05.04.25; accepted 06.04.25; published 27.06.25.

Please cite as:
Abayasiri RAM, Padilha Lanari Bo A, Dick TJM, Baghaei N
Influence of Virtual Reality Illusions on Balance Performance and Immersive User Experience in Young Adults: A Within-Subject Experimental Study
JMIR Serious Games 2025;13:e70376
URL: https://games.jmir.org/2025/1/e70376
doi:10.2196/70376

© R Achintha M Abayasiri, Antonio Padilha Lanari Bo, Taylor J M Dick, Nilufar Baghaei. Originally published in JMIR Serious Games (https://games.jmir.org), 27.6.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

The Impact of Active Augmented Reality Games on Physical Activity and Cognition Among Older Adults: Feasibility Study

Amy Shirong Lu[1], PhD; Bhagyashree Parkar[1], MS; Shiveshchandra Gupta[1], MS, MD; Sierra Hall[1], MPH, ALM; Dominika M Pindus[2], PhD; Arthur F Kramer[3], PhD

[1]Health Technology Lab, Department of Communication Studies, College of Arts, Media, and Design, Department of Public Health Science and Health Sciences, Bouvé College of Health Sciences, Northeastern University, 360 Huntington Ave 212A Lake Hall, Boston, MA, United States

[2]Department of Kinesiology and Community Health, Institute of Applied Life Sciences, University of Massachusetts Amherst, Amherst, MA, United States

[3]Beckman Institute for Advanced Science and Technology, University of Illinois, Urbana, IL, United States

Corresponding Author:
Amy Shirong Lu, PhD
Health Technology Lab, Department of Communication Studies, College of Arts, Media, and Design, Department of Public Health Science and Health Sciences, Bouvé College of Health Sciences, Northeastern University, 360 Huntington Ave 212A Lake Hall, Boston, MA, United States

Abstract

Background: Physical activity (PA) enhances physical health as well as cognitive and brain health, yet motivating older adults to initiate and sustain PA remains challenging, a difficulty exacerbated by the COVID-19 pandemic. Active augmented reality (AAR) games, integrating digital gameplay with real-world physical movement, provide an enjoyable and accessible means for PA promotion among older adults in independent living environments, mitigating barriers such as poor weather and unfavorable neighborhood environments. However, limited research has explored the feasibility and impact of AAR interventions in this population.

Objective: This feasibility study examines the acceptability, safety, and preliminary effects of AAR games to enhance PA levels and cognitive functions among older adults, as well as their user experiences. We also examined the practicality of home-based AAR gameplay using minimal equipment and constrained physical space.

Methods: Sixteen independent-living older adults aged 65 - 85 years (mean 74.6, SD 3.73) participated in a single-session AAR intervention using the *Active Arcade* game set by playing four 10-minute AAR games. PA levels were assessed using ActiGraph wGT3x-bt accelerometers and Polar H10 heart rate monitors. Cognitive function was evaluated pre- and post-gameplay using NIH Toolbox's visual reasoning test and Flanker inhibitory control and attention tests. Surveys of PA intention and motivation as well as the gaming experience questionnaire, along with semistructured interviews, were conducted afterwards, providing both quantitative and qualitative insights into the feasibility and appeal of AAR gameplay from the target population.

Results: All participants completed the study protocol without adverse events, demonstrating high feasibility and acceptability. Participants engaged in moderate-to-vigorous PA during 20% - 30% of the gameplay, as measured by accelerometers and heart rate monitors. Of the 16 participants, 7 were taking beta blockers. The mean values of average %HRMax suggest that those not on beta blockers generally met the moderate-intensity threshold, whereas those on beta blockers tended to fall slightly below it. Cognitive assessments revealed significant improvements in visual reasoning postintervention, with the effect sustained after adjustment for age and education (P=.03), suggesting potential cognitive benefits from a single bout of AAR gameplay. Survey responses indicated high levels of PA intention (mean 4.15/5, SD 0.59), motivation (mean 5.67/7, SD 1.24), high positive affect (mean 4.35/5, SD 0.80), and low negative affect (mean 1.30/5, SD 0.46) associated with AAR gameplay. Around 75% of gameplay occurred within a 4×4 ft area (mean 29.77/40 min, SD 2.46), indicating suitability for home environments. Thematically analyzed interview feedback emphasized participants' enjoyment, ease of use, desire for progressive difficulty, and the need to cater to diverse physical abilities and individual preferences.

Conclusions: AAR games are a feasible, accessible, and enjoyable alternative for PA and cognitive engagement among older adults. Future research should investigate the long-term effects, sustainability, and broader applicability of AAR interventions to fully realize their potential in aging populations.

(JMIR Serious Games 2025;13:e73221) doi:10.2196/73221

XSL•FO
RenderX

KEYWORDS

active augmented reality; physical activity; active video games; cognitive function; older adults; feasibility study; health technology; AAR; PA; AVG

Introduction

More than one-fifth of the US residents will be over the age of 65 years by 2030 [1]. This significant demographic shift will increase both the relative and absolute number of older adults, presenting challenges across various aspects of health [2]. Physical activity (PA) levels decrease significantly with age [3,4], and low PA is associated with numerous adverse outcomes such as high blood pressure, frailty, increased fall risk, depression, and cognitive decline [5-7]. Maintaining PA levels and cognitive function is pivotal for functional independence, effective communication, and overall quality of life [8-10]. There is a close connection between the cognitive well-being and older adults' ability to manage daily tasks, with cognitive decline significantly impairing their ability for independent living [11-13].

According to the United Nations, one of the sustainable development goals is to ensure healthy lives and promote well-being for all ages [14]. Moderate-to-vigorous PA (MVPA) is one of the most effective strategies to achieving these goals. It has been shown to decrease morbidity, prevent cognitive decline, and reduce the risk of mild cognitive impairment and Alzheimer disease and related dementias [15-18]. In fact, many forms of PA enhance physical, cognitive, and brain health and serve as a protective factor against future cognitive decline [19,20]. PA also improves cardiorespiratory fitness, reduces frailty and fall risk, and increases quality of life [21-23]. Additionally, PA has demonstrated an antidepressant effect in older adults, further contributing to their emotional and mental well-being [24].

The World Health Organization recommends that adults, including older adults, engage in at least 150 minutes of moderate-intensity or 75 minutes of vigorous-intensity PA per week [25]. However, motivating older adults to initiate and sustain PA remains challenging, exacerbated by the COVID-19 pandemic, which has created additional barriers to PA [26]. Even after the pandemic has passed, its negative impact on older adults persists, contributing to declines in PA levels, mobility, and overall well-being [27]. Common obstacles include low motivation, limited access to facilities, fear of injury or falls, and a lack of knowledge about proper exercise techniques [28]. Addressing these challenges is critical to providing older adults with opportunities to increase PA levels, promoting their physical and cognitive health while fostering overall well-being.

The American College of Sports Medicine has identified the integration of technology to achieve PA goals as "the future of fitness," with exercise programs for older adults ranking third on their list of trends for 2025 [29]. Active video games (AVG), or any video game that requires PA as part of gameplay, have the potential to overcome traditional barriers to PA for older adults. Extended reality technologies, including virtual reality (VR) and augmented reality (AR), have shown promise in promoting PA and enhancing cognitive functioning in older adults [30]. Active augmented reality (AAR) games, a subset of AVG, leverage AR technologies to incorporate active gameplay. Unlike VR, AR integrates digital graphics and information into the user's physical environment, creating an immersive and interactive experience that blends the real and virtual worlds [31].

AR technology has experienced rapid growth, with a projected market value exceeding US $635 billion by 2033 [32,33]. In 2022, over 70% of Americans believed that AR would become an integral part of everyday life, with 89% viewing it as beneficial [34]. Notably, the baby boomer generation has shown a surprising openness to exploring extended reality technologies. A 2023 survey revealed that 70% of baby boomers expressed interest in trying AR/VR technology, suggesting its potential for adoption among older adults [35]. Indeed, older adults have demonstrated positive attitudes and high interest in AR [36-39]. Active AR, or AAR interventions, have been successfully implemented in older adults for various health-related purposes, including improving balance [40] and reducing muscle loss [41]. AAR has also been well received by those with Alzheimer disease and related dementias [42] and Parkinson disease [43], indicating its adaptability across different aging-related conditions.

AAR games integrate real-world movements to enhance physical coordination while keeping users engaged through interactive, immersive experiences. By combining physical exercise with cognitive challenges, AR-based game interventions provide a dual benefit, promoting both physical and cognitive health [44]. While these initial findings are promising, there remains a significant gap in research on AAR's role in promoting PA and preventing cognitive decline. Understanding older adult usage of AAR is crucial for maximizing its potential as a health-promoting intervention.

AAR could serve as an effective, enjoyable, and accessible method for promoting PA motivation and adherence among older adults, mitigating traditional PA barriers such as poor weather and unsafe neighborhood environments. While older adults have historically been underrepresented in digital design, which partly contributes to their lower utilization of AR devices [45], they have shown a growing enthusiasm for AR technologies [33]. AAR games are increasingly gaining popularity as a means of engaging older adults in PA [46,47]. Commercially available platforms, such as Nintendo Wii and Microsoft Kinect, have introduced AAR-based games that enable users to engage in interactive exercise experiments from the comfort of their homes. These systems offer games designed to accommodate varying levels of difficulty and skill, making them a versatile option for older adults seeking to integrate PA into their daily routines. Moreover, AR technology extends beyond physical exercise, contributing to the psychological and social well-being of older adults. Research indicates that AR applications create enjoyable and engaging experiences, which can enhance motivation and promote adherence to exercise

XSL•FO
RenderX

routines [30]. Additionally, the immersive nature of AR fosters mental engagement, helping older adults combat loneliness and social isolation [30].

As AR technology continues to evolve, further research is needed to address existing challenges and investigate the long-term benefits of AR interventions on cognitive health and physical abilities in older adults. This feasibility study seeks to bridge this research gap by examining the positive effects of AR games on PA and cognition in older adults. The study employs a mixed-method design with pre- and postintervention assessments, integrating both quantitative and qualitative research methods to evaluate the impact of AAR games on PA levels and cognitive functioning in older adults. The following research questions (RQs) guide this investigation:

RQ1: Can AAR induce MVPA among older adults?

RQ2: Will AAR improve older adults' PA intention?

RQ3: Can AAR improve cognitive outcomes among older adults?

RQ4: What are older adults' cognitive and affective responses to the AAR gaming experience?

Methods

Ethical Considerations

The Northeastern University Institutional Review Board approved the study (#23-09-28). All participants signed written informed consent. Each participant completed the protocol behind a closed door, and the study data were deidentified before being analyzed. Each participant received a US $100 gift card at the end of the study session.

Sample

Participants were recruited through web advertisements and on-campus posters at the first author's institution. Additionally, potential participants were identified from the IGNITE (Investigating Gains in Neurocognition in an Intervention Trial of Exercise) [48] trial database and contacted by phone. The inclusion criteria targeted adults aged 65 - 85 years, ensuring an equal gender distribution, English language proficiency, absence of cognitive impairments, and sufficient physical ability to complete the study protocol (ie, without significant balance or mobility issues). More specifically, all participants completed a comprehensive neuropsychological evaluation, including the Montreal Cognitive Assessment, and all participants passed without being classified as cognitively impaired [49]. Exclusion criteria included any history of acute or chronic physical, cognitive, or mental health conditions that could interfere with participation in exercise. Approval from participants' primary care providers was also required. To control for potential bias, only individuals who had never played the AAR games selected for the study were invited to participate.

Approximately 160 individuals were contacted by phone, with 65 expressing interest in the study. Of these, 39 older adults underwent screening using the Physical Activity Readiness Questionnaire [50] and the Late Life Function and Disability Screening Function and Disability Component [51]. A total of 17 respondents passed the screening, and 16 participants were included in the study.

AAR Game

For this project, we selected *Active Arcade* (Nex, San Jose, CA) as the AAR game set. As one of the newer AR-based AVGs, *Active Arcade* is powered by proprietary computer vision technology used in pro sports and is freely available on iOS [52] and Android [53] stores. *Active Arcade* tracks body movements via a smartphone or tablet camera, requiring no wearables. It offers full-body games playable solo or with others, making it accessible across age and fitness levels. Its simple setup suits home-based use. A 2022 Apple Design Award finalist, the app has logged over 100 million play sessions since its 2021 launch [54,55]. It is compatible with smartphones and tablets on both major operating systems and can be displayed on any TV using an HDMI adapter [55].

General Procedures

Upon arrival, participants were provided with written informed consent before completing an online questionnaire (see Multimedia Appendix 1) to record demographic information. Research assistants (RAs) measured each participant's height using a stadiometer (ShorrBoard, Weight and Measure, LLC, Olney, MD) and weight with a calibrated scale (SECA 813, SECA Inc., Chino, CA). To ensure focus during the session, the RAs instructed participants to set their personal mobile devices to "Do Not Disturb" mode.

An accelerometer (wGT3x-bt; ActiGraph, Pensacola, FL) was attached to each participant's nondominant hip, and a heart rate (HR) band (Polar H10; Polar, Kempele, Finland) was secured to their chest. Baseline resting HR and rate of perceived exertion (RPE) were recorded using the Borg Category-Ratio (CR-10) [56]. Participants then completed two cognitive tests from the NIH Toolbox, the visual reasoning test, and the Flanker inhibitory control and attention test, before receiving gameplay instructions from the RAs. The NIH Toolbox is a validated, standardized battery developed by the National Institutes of Health to measure neurological and behavioral function across the lifespan [57]. For this study, we selected two specific tests because they target cognitive domains (fluid reasoning and executive function) that have shown associations with PA in older adults [58,59]. Visual reasoning assesses problem-solving and visuospatial skills, while the Flanker task measures attention and inhibitory control. Together, they offer a concise but meaningful assessment of the cognitive processes most likely to be influenced by interactive, movement-based AAR gameplay.

Participants engaged in four 10-minute *Active Arcade* AR mini-games, interspersed with rest intervals, while HR was continuously monitored. The games were randomly assigned in blocks with progressively increasing intensities. The *Active Arcade* game includes a variety of mini games; for this study, participants played *Reaction Flow* (Figure 1), *Whack a Mole* (Figure 2), and two modes of *Super Hits* [Catch (Figure 3) and Slash (Figure 4)], totaling 40 minutes of gameplay (10 min per game). We did not modify any part of the game, as the *Active Arcade* game set was designed for all ages and ability levels,

XSL•FO
RenderX

with games accessible to both children and older adults. These four games, as well as their order of presentation, were pretested and selected by two additional older adults meeting the same inclusion criteria and by lab staff experienced in AVG research, using the same PA measurement protocol, to ensure that the level of difficulty increased progressively. The gameplay took place in a square-shaped space divided into inner (4×4 sq ft), middle (6×6 sq ft), and outer (8×8 sq ft) zones. Games began only when participants' HR was within their resting HR range. After gameplay, participants repeated the NIH Toolbox cognitive tests to assess post-gameplay cognitive performance.

Figure 1. Reaction Flow.

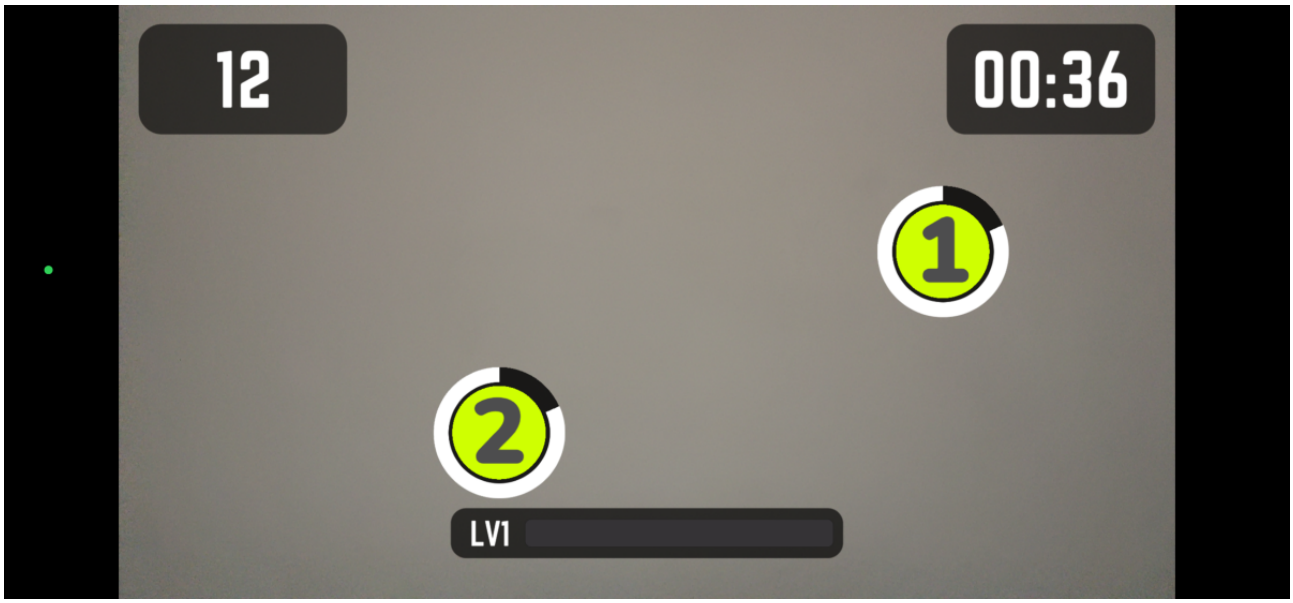
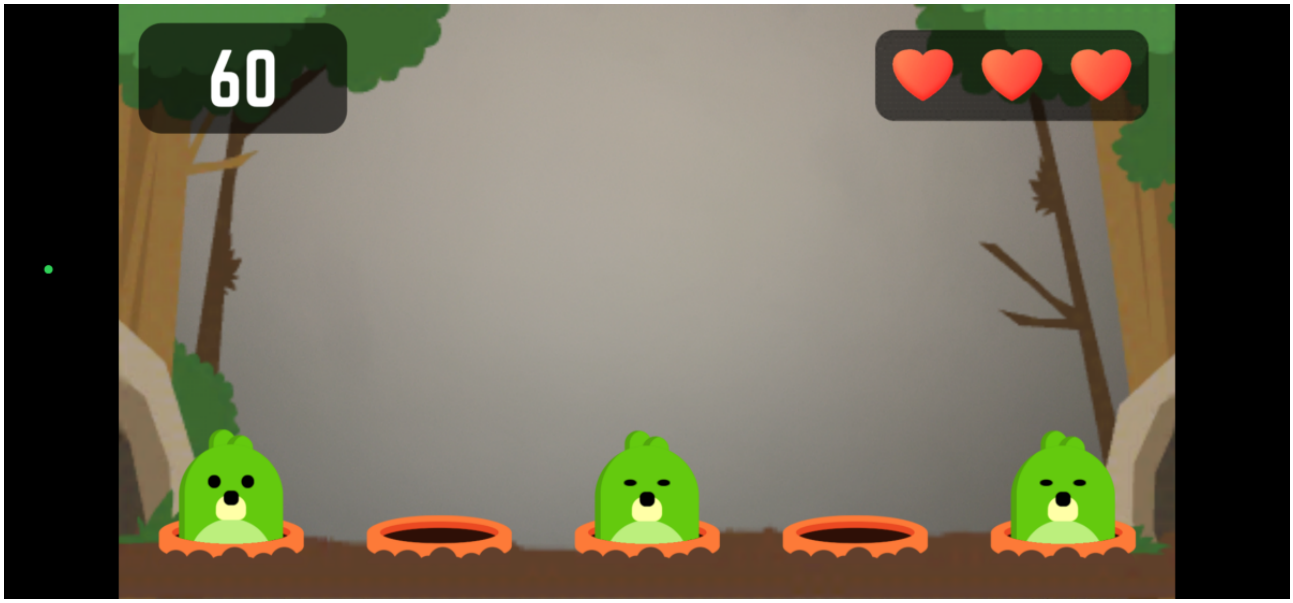


Figure 2. Whack a Mole.



XSL•FO
RenderX

Figure 3. Super Hits—Catch.

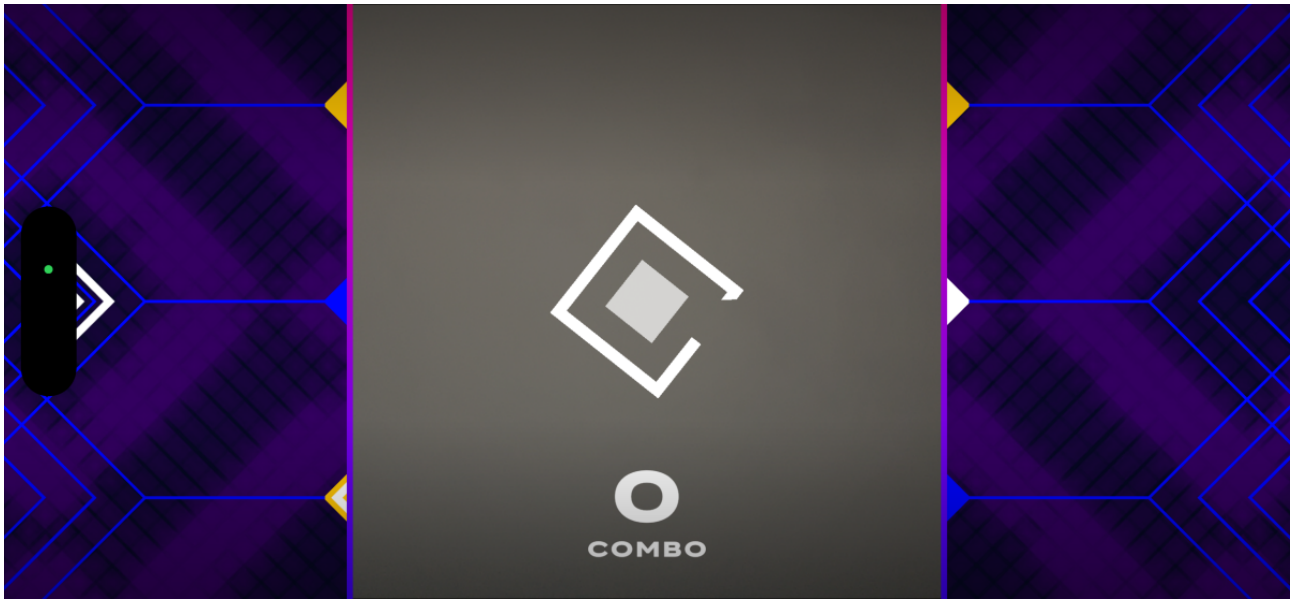
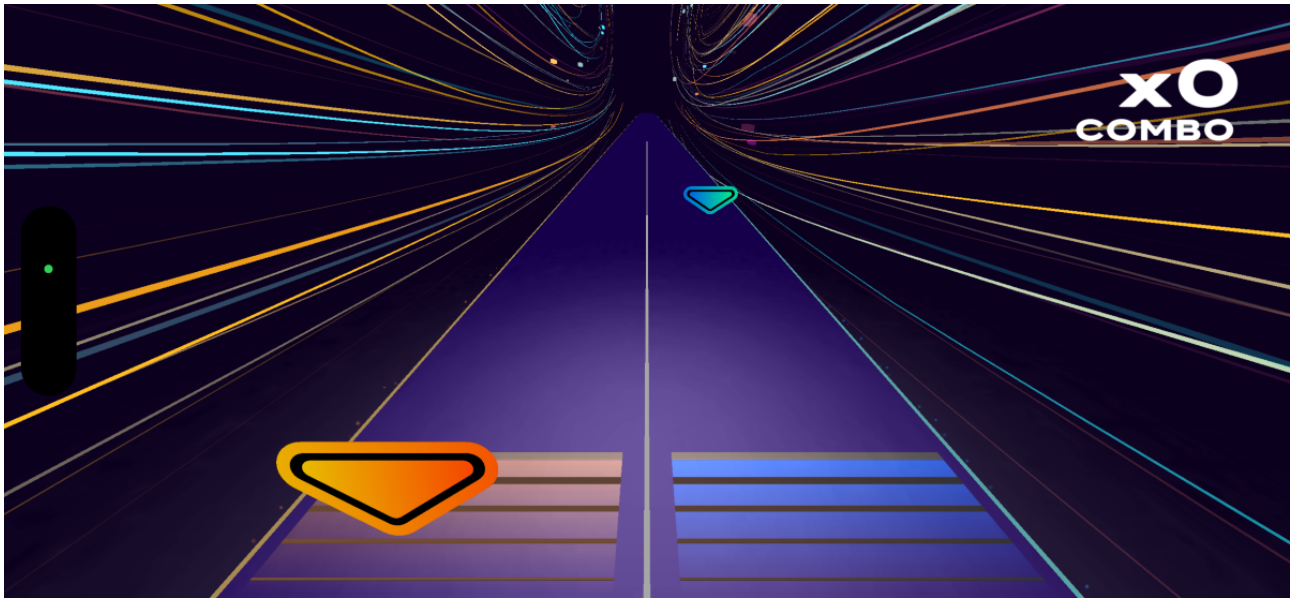


Figure 4. Super Hits—Slash.



After gameplay, participants completed surveys assessing intentions, motivations, and gaming experience. The session concluded with audio- and video-recorded interviews exploring participants' gameplay experience and attitudes toward AR.

Physical Activity Assessments

To assess PA, an ActiGraph wGT3x-bt accelerometer (ActiGraph, Pensacola, FL) was attached to each participant's nondominant hip at a 60-Hz sampling rate, recording three-axis movement (anterior-posterior, vertical, and medial-lateral) [60-62]. The RAs documented the start and end times of each 10-minute gameplay period, using these timestamps to identify PA data for each game session. Data were downloaded via ActiLife software (version 6.13.2) and processed in 1-second epochs, a resolution appropriate for capturing short bursts of PA [63]. Activity counts from all three axes and vector magnitude (VM) were extracted, and validated cut-point sets for older adults [64] were applied to classify PA into light (<2690 VM), moderate (≥2,690 to <6,166 VM), and vigorous (≥6166 VM) (VPA) activity levels [65].

A Polar HR monitor (Polar, Kempele, Finland) was used to continuously track PA intensity relative to each participant's age-predicted maximum HR [66]. PA intensity was calculated using the HR reserve (HRR) formula [HRR=maximum HR (220-age)–resting HR]. Intensity thresholds were defined as light (<40% HRR), moderate (40% - 59% HRR), and vigorous PA (60% - 89% HRR) based on the American College of Sports Medicine guidelines [67]. Individualized cutoff points were applied to determine time spent at each intensity level, with MVPA estimated calculated as the sum of moderate and vigorous PA durations. HR monitors provide a relative estimate of PA intensity while accelerometers will offer an absolute estimation, which is crucial for designing group-based intervention studies due to ease of use in free-living environments [68-71].

XSL•FO
RenderX

Cognitive Assessments

Cognitive function was assessed using two NIH Toolbox (version 3) [72] tests: the visual reasoning test and the Flanker inhibitory control and attention test. The visual reasoning test is an executive function assessment that evaluates nonverbal and visual reasoning, requiring participants to select the response option that is most like or missing from the target images displayed at the top of the screen. The Flanker inhibitory control and attention test measures inhibitory control and attention, where participants must focus on a particular stimulus while ignoring distracting flanking stimuli. See Multimedia Appendix 2 for screenshots of both tests.

These assessments were administered twice—before and after the gameplay—to evaluate cognitive changes. For post-gameplay assessments, RAs ensured that each participant's HR returned to within waited for each participant's HR to drop to within 10 beats per minute (BPM) of their resting HR before initiating the cognitive tests, minimizing potential influences of elevated HR on cognitive performance.

Questionnaires

Three surveys were used to assess participants' PA intentions, motivations, and gaming experience: (1) The AAR/PA Intention Scale [73] evaluates attitudes toward PA, perceived behavioral control, and intention to engage in PA. (2) The AAR/PA Motivation is measured by the PA Enjoyment Scale [74]. (3) The Game Experience Questionnaire (GEQ) [75] is a 33-item scale that evaluates gaming experience across multiple dimensions, including competence, immersion, and emotional responses. These validated instruments provide a comprehensive assessment of participants' engagement with the AAR-based PA and the factors influencing their motivation and experience.

Interviews

Before leaving the experimental session, participants took part in interviews to share their experiences with AAR games, including their overall impressions and interest in continued gameplay. RAs conducted the interviews using standardized questions that had been previously used in similar studies to ensure consistency and comparability across participants [76,77].

Data Analyses

Quantitative data were analyzed using SPSS V.29 (IBM, SPSS Inc., Chicago, IL), with paired-sample *t* tests conducted to assess changes in PA levels and cognitive function. Data are presented as means and standard deviations. All survey questionnaires were scored and analyzed according to their original instructions. The AAR/PA Intention Scale (Cronbach α=0.96) [73] and the PA Enjoyment Scale (Cronbach α=0.86) [74] were calculated by averaging their respective items. All GEQ-derived variables were scored and analyzed per the developer's recommendation [75], with subscales excluded from averaging if a Cronbach α was below 0.6 [78]. Qualitative interview data were examined through thematic analysis to identify central themes and patterns in participants' experiences.

Results

Demographics

The average (SD) age of participants was 74.6 (3.73) years. The majority of participants were right-handed (15/16). Racial distribution was equal, with 8 participants identifying as White and 8 as Black; notably, the majority of White participants were male (5 of 8), whereas the majority of Black participants were female (6 of 8). Education levels varied, with the largest proportion holding a Bachelor's degree (31.25%; Table 1).

Table . Demographics table, expressed as means and standard deviations.

Characteristic	Overall	Female	Male
Number of participants, n (%)	16	9 (56.25)	7 (43.75)
Age (years), mean (SD)	74.6 (3.73)	74.22 (2.58)	75.14 (5.04)
Height (cm), mean (SD)	168.87 (10.12)	164 (6.5)	175.14 (10.88)
Weight (kg), mean (SD)	84.72 (13.14)	81.40 (12.18)	88.98 (14.02)
Dominant hand, n	Right = 15; Left = 1	Right = 9; Left = 0	Right = 6; Left = 1
Right	15	9	6
Left	1	0	1
Race, n			
White	8	3	5
Black	8	6	2
Education level, n (%)			
High school graduate	3 (18.75)	2	1
Professional degree	3 (18.75)	3	0
Bachelor's	5 (31.25)	2	3
Master's	2 (12.5)	2	0
Doctorate	3 (18.75)	0	3

XSL•FO
RenderX

The most prevalent health conditions among participants included high blood pressure and hypertension, with commonly prescribed medications such as amlodipine, losartan, and lisinopril. Statin therapy, including atorvastatin and rosuvastatin, was widely used for cholesterol management. Several participants have histories of significant cardiovascular events, including heart attacks, arrhythmias, and coronary artery stent placements. For instance, one participant experienced a heart attack 6 years ago, while another had stents inserted into the right coronary artery. Other common conditions include osteoarthritis, osteoporosis, and hypothyroidism, along with reports of past cancer diagnoses and respiratory issues. Notably, seven of the 16 participants reported taking beta-blockers, which could lead to underestimation of HR-based PA measurements [79].

Physical Activities and Zone of Exercise

Table 2 presents the time spent in various PA intensities—light, moderate, and vigorous—as recorded by the ActiGraph and Polar devices across four different games [*Reaction Flow*, *Whack A Mole*, *Super Hits* (*Catch*), and *Super Hits* (*Slash*)]. Although, as expected, vigorous PA levels were relatively low across both devices and all games, each game successfully induced, on average, 20% - 30% MVPA among participants.

Table . Time spent in minutes in various PA levels using accelerometer and HR monitors, expressed as means and standard deviations.

	Light	Moderate	Vigorous	MVPA[a]
Reaction Flow				
ActiGraph	3.57 (1.32)	2.73 (1.59)	0.73 (0.88)	3.46 (2.25)
Polar	2.39 (2.11)	2.20 (2.59)	0.62 (1.42)	2.81 (3.45)
Whack A Mole				
ActiGraph	3.76[b] (1.40)	2.21 (1.64)	0.55 (0.87)	2.77 (2.31)
Polar	1.81[b] (1.76)	3.66 (3.49)	0.48 (1.18)	4.14 (3.96)
Super Hits (Catch)				
ActiGraph	3.61 (1.68)	1.42 (1.57)	0.12 (0.17)	1.53 (1.71)
Polar	2.37 (2.00)	1.70 (2.29)	0.73 (1.83)	2.43 (3.41)
Super Hits (Slash)				
ActiGraph	4.03 (1.53)	1.15 (1.57)	0.08 (0.11)	1.23[c] (1.65)
Polar	3.08 (1.84)	2.34 (2.33)	0.75 (1.97)	3.09[c] (3.28)
Overall average				
ActiGraph	3.74[d] (0.21)	1.88 (0.73)	0.37 (0.32)	2.25 (1.05)
Polar	2.41[d] (0.52)	2.47 (0.84)	0.65 (0.12)	3.12 (0.73)

[a]MVPA: moderate-to-vigorous physical activity

[b]*P*=.005

[c]*P*=.02

[d]*P*=.003

Notably, Polar recorded higher MVPA times for most games compared to ActiGraph, except for *Reaction Flow*. ActiGraph consistently recorded higher light PA levels, particularly for *Whack A Mole* ($Light_{ActiGraph}$=3.76 [SD 1.40] versus $Light_{Polar}$=1.81 [SD 1.76]; *P*=.005) and for the overall average ($Light_{ActiGraph}$= 3.74 [SD 0.21] versus $Light_{Polar}$=2.41 [SD 0.52]; *P*=.003). *Whack A Mole* had the highest MVPA per Polar, 4.14 (SD 3.96) minutes, while *Reaction Flow* had the highest MVPA per ActiGraph, 3.46 (SD 2.25) minutes. The difference between the two PA measurement devices was statistically significant only for *Super Hits* (*Slash*) ($MVPA_{ActiGraph}$=1.23 [SD 1.65] versus $MVPA_{Polar}$=3.09 [SD 3.28]; *P*=.02).

Table 3 presents the average time participants spent in different zones within the 8×8 ft space. The majority of gameplay time occurred in Zone 1 (4×4 ft), where participants spent an average of 30.01 (SD 2.14) minutes, accounting for 75% of total gameplay. Zone 2 (6×6 ft) was used moderately, with participants spending 9.15 (SD 1.80) minutes (23% of total gameplay), while Zone 3 (8×8 ft) was rarely used, with only 0.93 (SD 0.35) minutes (2%). This distribution suggests a strong tendency for participants to remain within the smallest zone, highlighting the feasibility of playing these AAR games in constrained spaces, making them particularly suitable for home environments. Additionally, the limited movement across zones suggests that balance, rather than extensive mobility, can be more critical to gameplay, thereby lowering spatial barriers to engaging in PA.

XSL•FO
RenderX

Table . Duration spent in different zones in minutes, expressed as means and standard deviations.

	Reaction Flow	Whack a Mole	Super Hits (Catch)	Super Hits (Slash)	Total duration	% of time spent
Zone 1: 4×4 ft	4.30 (4.14)	8.68 (2.73)	8.39 (3.33)	8.63 (3.15)	30.00 (2.14)	75
Zone 2: 6×6 ft	4.97 (3.66)	1.16 (2.18)	1.64 (3.34)	1.39 (3.15)	9.15 (1.80)	23
Zone 3: 8×8 ft	0.75 (1.63)	0.18 (0.67)	0.00 (0.00)	0.00 (0.00)	0.93 (0.35)	2
Total	10.02 (2.27)	10.02 (4.65)	10.03 (4.45)	10.02 (4.64)	40.00 (0.10)	100

Heart Rate and Rate of Perceived Exertion

All HR values collected during the protocol (in both BPM and %HRmax) reflect averages calculated from second-by-second readings. Before cognitive testing, participants' HR averaged 75.75 (SD 12.85) BPM, with minimal exertion (RPE=0.12, SD 0.5). After the initial test, HR decreased to 72.25 (SD 12.32) BPM when RPE rose modestly to 0.43 (SD 0.72). Prior to gameplay, HR remained stable at 72.31 (SD 12.89) BPM, with a slight RPE increase to 0.62 (SD 0.71).

During gameplay, HR rose significantly to an average of 100 (SD 14.11) BPM, with post-gameplay RPE immediately increasing to 3.25 (SD 1.73), indicating moderate-to-high physical exertion. This converts to an average of 64.1% HR_{Max} during the gameplay, indicating that the gameplay on average elicited moderate PA [67]. The gameplay phase lasted approximately 40 minutes. In addition, participants rested for an average of 8.73 minutes (SD 3.70, range 2 - 25) between each game to allow for HR recovery.

Before the post-cognitive test, HR remained somewhat elevated at 80.31 (SD 14.75) BPM with an RPE of 1.62 (SD 1.74), indicating lingering physical exertion effects. The RAs ensured that participants' HR dropped within 10 BPM of resting HR before initiating the cognitive test. After the test, HR returned closer to baseline at 76 (SD 13.24) BPM and RPE decreased to 1.42 (SD 1.96). See Table 4 for details.

XSL•FO
RenderX

Table . Heart rate and rate of perceived exertion, expressed as means and standard deviations.

Sequence and variables	Value
Pre-cog tests	
Age-predicted maximum HR[a] (HR_{Max})	156.01 (2.63)
Before pre-cog HR (BPM[b])	75.75 (12.85)
Before pre-cog $\%HR_{Max}$	48.55
Before pre-cog RPE[c] (scale)	0.12 (0.5)
After pre-cog HR (BPM)	72.25 (12.32)
After pre-cog $\%HR_{Max}$	46.31
After pre-cog RPE (scale)	0.43 (0.72)
Before gameplay	
HR (BPM)	72.31 (12.89)
Before gameplay $\%HR_{Max}$	46.34
RPE (scale)	0.62 (0.71)
During gameplay	
Duration (mins)	40.00 (0.10)
HR during Game 1 (BPM)	100.00 (14.76)
$\%HR_{Max}$ during Game 1 (%)	64.10
HR during Game 2 (BPM)	100.01 (14.96)
$\%HR_{Max}$ during Game 2 (%)	64.10
HR during Game 3 (BPM)	100.25 (14.35)
$\%HR_{Max}$ during Game 3 (%)	64.26
HR during Game 4 (BPM)	98.69 (14.71)
$\%HR_{Max}$ during Game 4 (%)	63.26
HR per game (BPM)	100.00 (14.11)
$\%HR_{Max}$ during all gameplay (%)	64.10
Resting time for HR recovery (min)	8.73 (3.70)
After gameplay	
HR (BPM)	98.87 (16.82)
After gameplay $\%HR_{Max}$	63.37
RPE (scale)	3.25 (1.73)
Post-cog tests	
Resting time for HR recovery (min)	7.26 (3.77)
Before post-cog HR (BPM)	80.31 (14.75)
Before post-cog $\%HR_{Max}$	51.48
Before post-cog RPE (scale)	1.62 (1.74)
After post-cog HR (BPM)	76 (13.24)
After post-cog $\%HR_{Max}$	48.71
After post-cog RPE (scale)	1.42 (1.96)

[a]HR: heart rate

[b]BPM: beats per minute.

[c]RPE: rate of perceived exertion.

XSL•FO
RenderX

Notably, the beta-blocker group consistently exhibited lower HR throughout the study—approximately 8 BPM less on average—compared to the non-taker group. Given that continuous RPE reporting during gameplay would have been disruptive, RPE was measured immediately after each game, following prior literature [80-82]. Additionally, they reported higher post-gameplay RPE than nontakers, though neither the HR nor RPE differences reached statistical significance due to the small sample sizes (7 takers vs 9 nontakers). Nevertheless, the mean values of average $\%HR_{Max}$ suggest that those not on beta-blockers generally met the moderate-intensity threshold during games, whereas those on beta-blockers tended to fall slightly below the moderate intensity threshold. As such, the data were presented in aggregate below (see Multimedia Appendix 3 for a breakdown between the two groups).

Cognitive Tests

The cognitive assessment, including the Flanker and visual reasoning tests, was conducted pre- and post-AAR gameplay to evaluate potential cognitive benefits. Visual reasoning scores improved significantly post-play (P=.04) in change-sensitive score (CSS), an item response theory–based score metric designed for longitudinal change monitoring. Starting from NIH Toolbox V3, age- and education-adjusted T scores are also provided [83]. The effect remained significant when adjusted by age alone and age and education combined (P's = .03). See Table 5 for details. These results may suggest that visual reasoning abilities benefited from brief single-bout AAR gameplay, possibly due to the interactive, visually stimulating nature of the activities. However, an alternative explanation is that the improvement may be attributed to a practice effect on the task. To distinguish between these two possibilities, future studies should incorporate a control group to isolate the specific cognitive effects of AAR interventions.

Table . Cognitive tests (Flanker and visual reasoning) before and after AAR[a] Play, expressed as means and standard deviations.

Cognitive test	Before	After	T	P
Flanker CSS[b]	503.81 (19.17)	501.81 (15.39)	0.76	.23
Flanker age adjusted standard score	105.69 (21.98)	103.44 (17.60)	0.73	.24
Flanker age edu adjusted standard score	53.50 (14.89)	51.94 (11.81)	0.73	.24
Visual reasoning CSS[b]	499.5 (8.52)	502.69 (10.66)	1.94[c]	.04[c]
Visual reasoning age adjusted standard score	102.88 (17.17)	109.31 (21.89)	1.98[c]	.03[c]
Visual reasoning age edu adjusted standard score	50.44 (10.01)	54.25 (12.08)	1.99[c]	.03[c]

[a]AAR: active augmented reality

[b]CSS: change-sensitive score

[c]P<.05

Questionnaire Results

Table 6 summarizes participants' responses across scales measuring AAR/PA motivation, intention, and game engagement. The AAR/PA intention score was 4.15 (SD 0.59) on a 5-point scale, reflecting a strong intention to continue engaging in AAR activities. Similarly, the AAR/PA motivation scale had a high mean score of 5.67 (SD 1.24) on a 7-point scale, suggesting participants were motivated to continue playing AAR games. In the GEQ subscales (5-point scale), all subscales except flow had an acceptable Cronbach α (≥0.6) and were averaged. Participants reported high positive affect, 4.35 (SD 0.80), and minimal negative affect, 1.30 (SD 0.46), with low tension, 1.25 (SD 0.67), suggesting that gameplay was generally enjoyable and nonstressful. They also reported relatively high competence 3.68 (SD 0.92) and sensory and imaginative immersion 3.64 (SD 0.88), showing that they felt capable and engaged during gameplay. The challenge level was moderate at 2.49 (SD 0.76), reflecting a manageable level of difficulty.

XSL•FO
RenderX

Table . Survey results for the AAR[a]/PA[b] intention, motivation, and Game Experience Questionnaire (GEQ) measures, expressed as means and standard deviations[c].

Survey result	Mean (SD)	Cronbach α
AAR /PA intention	4.15 (0.59)	.86
AAR/PA motivation	5.67 (1.24)	.96
GEQ		
Competence	3.68 (0.92)	.88
Sensory and imaginative immersion	3.64 (0.88)	.81
Flow	—	.43
Tension	1.25 (0.67)	.86
Challenge	2.49 (0.76)	.67
Negative affect	1.30 (0.46)	.83
Positive affect	4.35 (0.80)	.82

[a]AAR: active augmented reality

[b]PA: physical activity

[c]All scales are 5-point scales except for AAR/PA motivation, which is a 7-point scale.

Interview Analysis

We used a thematic analysis approach to examine the qualitative interview data. Two team members independently reviewed the transcripts and conducted open coding to identify initial patterns. Codes were refined through discussion, and themes were inductively developed to reflect key participant perceptions of AAR gameplay, such as enjoyment, accessibility, physical challenge, and suggestions for improvement.

The thematic analysis revealed that older adult participants generally responded positively to the AR games, viewing them as both enjoyable and beneficial for combining physical exercise with mental engagement. Many appreciated the opportunity to play at home, describing it as an accessible and refreshing alternative to traditional exercise settings. While some initially had reservations about the novelty of AR or the perceived cost of setup, these concerns diminished with familiarity. Additionally, participants expressed a desire for more physically exerting games than those provided. Among the games played, *Whack a Mole* emerged as a favorite, with participants appreciating its simplicity and engaging nature. The interactive experience, particularly the combination of music and movement, was highlighted as a factor that fostered focus and enjoyment.

Looking ahead to future game design, participants expressed interest in customizable games that start at an easier level and progressively increase in difficulty, accommodating a range of physical abilities. Many suggested incorporating full-body movements and immersive environments, such as outdoor simulations, to further enhance engagement. Music integration, particularly with familiar tunes, was another highly requested feature. Beyond individual play, participants saw great potential for these games in supporting physical and mental health, particularly for older adults or individuals with limited mobility, ensuring a user-friendly interface for those less familiar with technology and incorporating community or group-based settings to foster social engagement and support.

Discussion

Principal Findings

This feasibility study highlights the potential of AAR games to increase PA and enhance cognitive function among older adults in a user-friendly and safe environment. Participants engaged in MVPA during gameplay, as measured by accelerometers and HR monitors, aligning with the recommended PA levels for older adults. Cognitive assessments revealed significant improvements in visual reasoning, indicating potential cognitive benefits from a single bout of AAR gameplay. Surveys indicated high levels of motivation, enjoyment, and positive emotions associated with AAR gameplay, with participants appreciating its accessibility and suitability for free-living environments. Interview feedback emphasized the need for adaptable game designs to cater to diverse physical abilities and individual preferences. Importantly, despite their various health conditions, all participants successfully completed the study protocol, underscoring the feasibility and acceptability of AAR for the older adult population.

Comparison to Prior Research

Our findings revealed differences in PA measurement between devices, with the HR monitor, Polar, recording slightly higher MVPA levels compared to the accelerometer, ActiGraph. Since HR-based measurements offer a more individualized measurement of PA intensity, MVPA estimates from Polar may provide a more appropriate reflection of exertion levels, whereas ActiGraph measures may be more appropriate on a group but not individual level [84,85]. However, given that 7 out of 16 participants were taking beta-blockers, their actual MVPA level may have been higher than what the Polar has recorded, as beta-blockers can suppress HR responses to physical exertion [79]. These variations in PA measurement emphasize the need for careful selection of appropriate tracking methods in future AAR studies among older adults, particularly when accounting for medication effects on HR-based PA assessments.

XSL•FO
RenderX

Beyond the benefits afforded by the PA engagement, AAR gameplay also improved cognitive function, with significant gains in visual reasoning. This suggests that the AAR games may enhance cognitive processes related to problem-solving and cognitive flexibility. These results align with prior research linking PA to enhanced executive function, indicating that AAR games could stimulate brain regions associated with higher-order cognition and promote neural engagement [59,86,87]. While these effects may be attributed to AAR gameplay, the possibility of a practice effect on cognitive testing cannot be ruled out. Future studies should incorporate control groups to better isolate the specific cognitive effects of AAR interventions.

One of the most notable advantages of AAR gameplay is its accessibility and practicality for home-based exercise. Participants remained engaged and motivated, as reflected by their high AAR/PA intention and motivation scores, along with strong positive affect and immersion ratings. With initial training and a play environment designed to simulate the home environment, participants were able to engage in PA with minimal equipment and space requirements. Notably, the majority of the gameplay occurred within a 4×4 area (Zone 1), highlighting AAR's feasibility in confined spaces and its potential to reduce barriers to PA among those with limited access to outdoor exercise spaces. Additionally, since AAR gameplay relies more on balance than extensive mobility, it offers a viable exercise option for individuals with physical limitations [28].

The qualitative findings further support these conclusions, with participants expressing enthusiasm for AAR games as a viable alternative to traditional exercise. Many appreciated the interactive nature of the games, particularly the integration of music and movement, and welcomed the flexibility of playing in a home-like environment. While some initially faced challenges with gameplay mechanics, their confidence improved over time, and they expressed interest in continued gameplay. Participants also suggested various customization options, such as adjustable difficulty levels, which could enhance long-term engagement and accommodate physical abilities.

Limitations

This study has several limitations. First, the small sample size (n=16), along with the relatively short duration, reduces statistical power and limits the generalizability of the findings, warranting larger-scale studies conducted over a longer period. Future research should aim to recruit a more representative sample. Second, 7 out of 16 participants were taking beta blockers, which may have influenced HR-based PA assessments and potentially led to underestimation of MVPA levels, as reflected in the lower $\%HR_{max}$ observed among takers compared to nontakers. Future research should explore alternative PA measurement methods or incorporate personalized adjustments for HR thresholds in individuals taking beta blockers. Although alternative PA measurement strategies or individualized adjustments to HR thresholds should be explored, to date, no widely accepted method exists for adjusting HR-based PA estimates among beta-blocker users. A larger sample and auxiliary measures such as RPE with a larger sample in both groups might help shed light on the PA intensity levels. Third, the lack of a control group prevents us from drawing causal conclusions regarding cognitive improvements from AAR gameplay. Future studies should include appropriate comparison conditions, such as passive and active control groups, to better isolate the cognitive effects of AAR interventions and strengthen causal inferences.

Future Directions

Given the feasibility-focused nature of the study, we intentionally selected games with more moderate than vigorous PA intensities to prioritize safety. Despite this, several participants expressed a desire for more intense PA options, indicating that older adults may benefit from progressively vigorous AAR gameplay as they grow more comfortable with technology and gradually increase their PA levels. Although appropriate for a feasibility trial, this constrained intensity range may have limited the extent of cognitive changes observed, particularly if higher-intensity PA is a necessary condition for eliciting measurable cognitive benefit, though careful precautions are needed to ensure that high-intensity exercise is not potentially dangerous for older adults who are not regular exercisers. Future research should explore adaptive difficulty settings to allow for gradual progression in PA intensity, accommodating users at different fitness levels. In addition, future studies should examine a wider variety of AAR game types, longer intervention periods, and group-based gameplay to explore potential social and motivational benefits.

Conclusions

Using a mixed-methods one-group pretest-posttest experimental design, the study demonstrated that older adults could engage in MVPA during AAR gameplay and showed significant improvements in visual reasoning. Participants reported high levels of motivation, enjoyment, and positive affect, and qualitative feedback further supported the acceptability and potential benefits of AAR interventions. These findings underscore the need for further exploration of long-term AAR interventions with larger, controlled trials to evaluate the sustained effects and optimize design features. With continued development and refinement, AAR games may serve as a viable and acceptable intervention tool for promoting PA and cognitive engagement among older adults.

Acknowledgments

This study was supported in part by a grant from the National Institute of Diabetes and Digestive and Kidney Diseases (R01DK109316, PI: ASL) and by the College of Arts, Media, and Design and Bouvé College of Health Sciences at Northeastern University, Boston, MA.

XSL•FO
RenderX

Data Availability

The datasets used or analyzed during the current study are available from the corresponding author on reasonable request.

Authors' Contributions

ASL has contributed to Funding acquisition, Conceptualization, Methodology, Resources, Supervision, Project administration, Formal analysis, Writing–original draft, and Writing–review & editing. BP and SG contributed in Investigation, Data curation, and Writing–original draft. SH has contributed to Investigation and Writing–original draft. DP was responsible for Methodology. AFK was responsible for Methodology, Resources, and Writing–review & editing. All authors have read and approved the final version of the manuscript and agree with the order of presentation of the authors.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Online questionnaire.
[DOCX File, 20 KB - games_v13i1e73221_app1.docx]

Multimedia Appendix 2
Screenshots of the visual reasoning (Up) and the Flanker inhibitory control and attention tests (Down).
[DOCX File, 75 KB - games_v13i1e73221_app2.docx]

Multimedia Appendix 3
Heart rate and rate of perceived exertion between beta-blocker nonusers and users, expressed as means and standard deviations.
[DOCX File, 19 KB - games_v13i1e73221_app3.docx]

References

1. Vincent GK, Velkoff VA. The next four decades: the older population in the united states: 2010 to 2050. : US Department of Commerce, Economics and Statistics Administration, US Census Bureau; 2010 URL: https://www.census.gov/content/dam/Census/library/publications/2010/demo/p25-1138.pdf [accessed 2025-10-01]
2. Rowe JW, Kahn RL. Successful aging 2.0: conceptual expansions for the 21st century. J Gerontol B Psychol Sci Soc Sci 2015 Jul;70(4):593-596. [doi: 10.1093/geronb/gbv025] [Medline: 25878054]
3. Milanović Z, Pantelić S, Trajković N, Sporiš G, Kostić R, James N. Age-related decrease in physical activity and functional fitness among elderly men and women. Clin Interv Aging 2013;8:549-556. [doi: 10.2147/CIA.S44112] [Medline: 23723694]
4. Suryadinata RV, Wirjatmadi B, Adriani M, Lorensia A. Effect of age and weight on physical activity. J Public Health Res 2020 Jul 3;9(2):1840. [doi: 10.4081/jphr.2020.1840] [Medline: 32728579]
5. Becofsky K, Baruth M, Wilcox S. Physical activity mediates the relationship between program participation and improved mental health in older adults. Public Health (Fairfax) 2016 Mar;132:64-71. [doi: 10.1016/j.puhe.2015.07.040] [Medline: 26318601]
6. Musich S, Wang SS, Hawkins K, Greame C. The frequency and health benefits of physical activity for older adults. Popul Health Manag 2017 Jun;20(3):199-207. [doi: 10.1089/pop.2016.0071] [Medline: 27623484]
7. Soares-Miranda L, Siscovick DS, Psaty BM, Longstreth WT, Mozaffarian D. Physical activity and risk of coronary heart disease and stroke in older adults: the cardiovascular health study. Circulation 2016 Jan 12;133(2):147-155. [doi: 10.1161/CIRCULATIONAHA.115.018323] [Medline: 26538582]
8. Logsdon RG, Gibbons LE, McCurry SM, Teri L. Assessing quality of life in older adults with cognitive impairment. Psychosom Med 2002;64(3):510-519. [doi: 10.1097/00006842-200205000-00016] [Medline: 12021425]
9. Mosti CB, Rog LA, Fink JW. Differentiating mild cognitive impairment and cognitive changes of normal aging. In: Ravdin LD, Katzen HL, editors. Handbook on the Neuropsychology of Aging and Dementia: Springer International Publishing; 2019:445-463. [doi: 10.1007/978-3-319-93497-6_28]
10. Murman DL. The Impact of Age on Cognition. Semin Hear 2015 Aug;36(3):111-121. [doi: 10.1055/s-0035-1555115] [Medline: 27516712]
11. Hertzog C, Kramer AF, Wilson RS, Lindenberger U. Enrichment effects on adult cognitive development: can the functional capacity of older adults be preserved and enhanced? Psychol Sci Public Interest 2008 Oct;9(1):1-65. [doi: 10.1111/j.1539-6053.2009.01034.x] [Medline: 26162004]
12. Pedone C, Ercolani S, Catani M, et al. Elderly patients with cognitive impairment have a high risk for functional decline during hospitalization: the GIFA study. J Gerontol A Biol Sci Med Sci 2005 Dec;60(12):1576-1580. [doi: 10.1093/gerona/60.12.1576] [Medline: 16424291]

XSL•FO
RenderX

13. Diehl M, Willis SL, Schaie KW. Everyday problem solving in older adults: observational assessment and cognitive correlates. Psychol Aging 1995 Sep;10(3):478-491. [doi: 10.1037//0882-7974.10.3.478] [Medline: 8527068]
14. Transforming our world: the 2030 agenda for sustainable development. United Nations Department of Economic and Social Affairs. 2015. URL: https://sdgs.un.org/2030agenda [accessed 2025-01-28]
15. Blumenthal JA, Smith PJ, Mabe S, et al. Lifestyle and neurocognition in older adults with cognitive impairments. Neurology (ECronicon) 2019 Jan 15;92(3):e212-e223. [doi: 10.1212/WNL.0000000000006784]
16. Brasure M, Desai P, Davila H, et al. Physical activity interventions in preventing cognitive decline and Alzheimer-type dementia: a systematic review. Ann Intern Med 2018 Jan 2;168(1):30-38. [doi: 10.7326/M17-1528] [Medline: 29255839]
17. National Academies of Sciences, Engineering, and Medicine. In: Downey A, Stroud C, Landis S, Leshner AI, editors. Preventing Cognitive Decline and Dementia: A Way Forward: National Academies Press (US); 2017. [doi: 10.17226/24782]
18. Petersen RC. Mild cognitive impairment. Continuum (Minneap Minn) 2016 Apr;22(2 Dementia):404-418. [doi: 10.1212/CON.0000000000000313] [Medline: 27042901]
19. Halloway S, Arfanakis K, Wilbur J, Schoeny ME, Pressler SJ. Accelerometer physical activity is associated with greater gray matter volumes in older adults without dementia or mild cognitive impairment. J Gerontol B Psychol Sci Soc Sci 2019 Sep 15;74(7):1142-1151. [doi: 10.1093/geronb/gby010] [Medline: 29432610]
20. Stubbs B, Eggermont L, Soundy A, Probst M, Vandenbulcke M, Vancampfort D. What are the factors associated with physical activity (PA) participation in community dwelling adults with dementia? A systematic review of PA correlates. Arch Gerontol Geriatr 2014;59(2):195-203. [doi: 10.1016/j.archger.2014.06.006] [Medline: 25034708]
21. Buchner DM, Beresford SA, Larson EB, LaCroix AZ, Wagner EH. Effects of physical activity on health status in older adults. II. Intervention studies. Annu Rev Public Health 1992;13:469-488. [doi: 10.1146/annurev.pu.13.050192.002345] [Medline: 1599599]
22. Pandey A, Garg S, Khunger M, et al. Dose-response relationship between physical activity and risk of heart failure: a meta-analysis. Circulation 2015 Nov 10;132(19):1786-1794. [doi: 10.1161/CIRCULATIONAHA.115.015853] [Medline: 26438781]
23. Ribas TM, Teodori RM, Mescolotto FF, Montebelo M, Baruki SBS, Pazzianotto-Forti EM. Impact of physical activity levels on musculoskeletal symptoms and absenteeism of workers of a metallurgical company. Rev Bras Med Trab 2021 Mar 3;18(4):425-433. [doi: 10.47626/1679-4435-2020-572] [Medline: 33688324]
24. Schuch FB, Vancampfort D, Rosenbaum S, et al. Exercise for depression in older adults: a meta-analysis of randomized controlled trials adjusting for publication bias. Braz J Psychiatry 2016;38(3):247-254. [doi: 10.1590/1516-4446-2016-1915] [Medline: 27611903]
25. Physical activity. World Health Organization. 2022. URL: https://www.who.int/news-room/fact-sheets/detail/physical-activity [accessed 2024-05-12]
26. Suzuki Y, Maeda N, Hirado D, Shirakawa T, Urabe Y. Physical activity changes and its risk factors among community-dwelling Japanese older adults during the COVID-19 epidemic: associations with subjective well-being and health-related quality of life. Int J Environ Res Public Health 2020 Sep 10;17(18):6591. [doi: 10.3390/ijerph17186591] [Medline: 32927829]
27. König M, Gollasch M, Komleva Y. Frailty after covid‐19: the wave after? Aging Med (Milton) 2023 Sep;6(3):307-316. [doi: 10.1002/agm2.12258] [Medline: 37711259]
28. Rúa-Alonso M, Bovolini A, Costa-Brito AR, et al. Exploring perceived barriers to physical activity among older adults living in low-population density regions: gender differences and associations with activity dimensions. Healthcare (Basel) 2023 Nov 11;11(22):2948. [doi: 10.3390/healthcare11222948] [Medline: 37998440]
29. Newsome AM, Batrakoulis A, Camhi SM, et al. 2025 ACSM worldwide fitness trends: future directions of the health and fitness industry. ACSM's Health Fitness J 2024;28(6):11-25. [doi: 10.1249/FIT.0000000000001017]
30. Margrett JA, Ouverson KM, Gilbert SB, Phillips LA, Charness N. Older adults' use of extended reality: a systematic review. Front Virtual Real 2022;2. [doi: 10.3389/frvir.2021.760064]
31. Zhao MY, Ong SK, Nee AYC. An augmented reality-assisted therapeutic healthcare exercise system based on bare-hand interaction. Int J Human Comput Interact 2016 Sep;32(9):708-721. [doi: 10.1080/10447318.2016.1191263]
32. Top augmented reality statistics in 2022. Art labs. 2022. URL: https://artlabs.ai/blog/augmented-reality-statistics-in-2021 [accessed 2023-02-13]
33. Augumented reality market by device type, offering, application, technology, and geography. Marketsandmarkets.com. URL: https://www.marketsandmarkets.com/Market-Reports/augmented-reality-market-82758548.html [accessed 2022-12-30]
34. America's interest in virtual & augmented reality. ARRIS Composites. 2022. URL: https://arriscomposites.com/americas-interest-in-virtual-and-augmented-reality [accessed 2025-09-22]
35. 95% of americans eager to embrace AR/VR technology, but point to price and health concerns as top barriers. AddictiveTips Staff. 2023. URL: https://www.addictivetips.com/virtual-reality-survey [accessed 2025-01-29]
36. Malik SA, Abdullah LM, Mahmud M, Azuddin M. Presented at: 2013 international conference on research and innovation in information systems (ICRIIS); Nov 27-28, 2013; Kuala Lumpur, Malaysia. 2013.[doi: 10.1109/ICRIIS.2013.6716739]. Presented at: . [doi: 10.1109/ICRIIS.2013.6716739]

XSL•FO
RenderX

37. Guerrero Huerta AG, Hernández Rubio E, editors. Interaction Modalities for Augmented Reality in Tablets for Older Adults: Springer International Publishing; 2017. [doi: 10.1007/978-3-319-58753-0_61]
38. Derby JL, Chaparro BS, editors. Use of augmented reality by older adults. In: Human Aspects of IT for the Aged Population Technologies, Design and User Experience: Springer International Publishing; 2020. [doi: 10.1007/978-3-030-50252-2_10]
39. Anderson M, Perrin A. Tech Adoption Climbs Among Older Adults. 2017. URL: https://www.pewresearch.org/internet/2017/05/17/tech-adoption-climbs-among-older-adults [accessed 2024-07-04]
40. Ku J, Kim YJ, Cho S, Lim T, Lee HS, Kang YJ. Three-dimensional augmented reality system for balance and mobility rehabilitation in the elderly: a randomized controlled trial. Cyberpsychol Behav Soc Netw 2019 Feb;22(2):132-141. [doi: 10.1089/cyber.2018.0261] [Medline: 30596530]
41. Jeon S, Kim J. Effects of augmented-reality-based exercise on muscle parameters, physical performance, and exercise self-efficacy for older adults. Int J Environ Res Public Health 2020 May 7;17(9):3260. [doi: 10.3390/ijerph17093260] [Medline: 32392833]
42. Rohrbach N, Gulde P, Armstrong AR, et al. An augmented reality approach for ADL support in Alzheimer's disease: a crossover trial. J Neuroeng Rehabil 2019 Jun 3;16(1):66. [doi: 10.1186/s12984-019-0530-z] [Medline: 31159816]
43. Espay AJ, Baram Y, Dwivedi AK, et al. At-home training with closed-loop augmented-reality cueing device for improving gait in patients with Parkinson disease. J Rehabil Res Dev 2010;47(6):573-581. [doi: 10.1682/jrrd.2009.10.0165] [Medline: 20848370]
44. Goumopoulos C, Drakakis E, Gklavakis D. Feasibility and acceptance of augmented and virtual reality exergames to train motor and cognitive skills of elderly. Computers 2023;12(3):52. [doi: 10.3390/computers12030052]
45. Lee LN, Kim MJ, Hwang WJ. Potential of augmented reality and virtual reality technologies to promote wellbeing in older adults. Appl Sci (Basel) 2019;9(17):3556. [doi: 10.3390/app9173556]
46. Chen M, Tang Q, Xu S, Leng P, Pan Z. Design and evaluation of an augmented reality-based Exergame system to reduce fall risk in the elderly. IJERPH 2020 Oct 1;17(19):7208. [doi: 10.3390/ijerph17197208]
47. Philippe AG, Goncalves A, Korchi K, Deshayes M. Exergaming in augmented reality is tailor-made for aerobic training and enjoyment among healthy young adults. Front Public Health 2024;12:1307382. [doi: 10.3389/fpubh.2024.1307382] [Medline: 38469269]
48. Erickson KI, Grove GA, Burns JM, et al. Investigating gains in neurocognition in an intervention trial of exercise (IGNITE): protocol. Contemp Clin Trials 2019 Oct;85:105832. [doi: 10.1016/j.cct.2019.105832] [Medline: 31465859]
49. Oberlin LE, Wan L, Kang C, et al. Cardiorespiratory fitness is associated with cognitive function in late adulthood: baseline findings from the IGNITE study. Br J Sports Med 2025 Feb 3;59(3):167-176. [doi: 10.1136/bjsports-2024-108257] [Medline: 39658276]
50. Warburton DE, Jamnik VK, Bredin SS, Gledhill N. The physical activity readiness questionnaire for everyone (PAR-Q+) and electronic physical activity readiness medical examination (ePARmed-X+). Health Fitness J Canada 2011;4(2):3-17.
51. Haley SM, Jette AM, Coster WJ, et al. Late life function and disability instrument: II. Development and evaluation of the function component. J Gerontol A Biol Sci Med Sci 2002 Apr;57(4):M217-M222. [doi: 10.1093/gerona/57.4.m217] [Medline: 11909886]
52. Active arcade ios. 2021. URL: https://apps.apple.com/us/app/active-arcade/id1553158383 [accessed 2025-02-12]
53. Active Arcade Android. 2021. URL: https://play.google.com/store/apps/details?id=team.nex.arcadexp&hl=en_US [accessed 2025-02-12]
54. Apple announces winners of the 2022 apple design awards. Apple Inc. 2022. URL: https://www.apple.com/newsroom/2022/06/apple-announces-winners-of-the-2022-apple-design-awards [accessed 2025-07-21]
55. Active arcade. Nex. 2022. URL: https://www.activearcade.ai [accessed 2023-02-17]
56. Morishita S, Tsubaki A, Nakamura M, Nashimoto S, Fu JB, Onishi H. Rating of perceived exertion on resistance training in elderly subjects. Expert Rev Cardiovasc Ther 2019 Feb;17(2):135-142. [doi: 10.1080/14779072.2019.1561278] [Medline: 30569775]
57. Weintraub S, Dikmen SS, Heaton RK, et al. Cognition assessment using the NIH Toolbox. Neurology (ECronicon) 2013 Mar 12;80(11 Suppl 3):S54-S64. [doi: 10.1212/WNL.0b013e3182872ded] [Medline: 23479546]
58. Colcombe S, Kramer AF. Fitness effects on the cognitive function of older adults: a meta-analytic study. Psychol Sci 2003 Mar;14(2):125-130. [doi: 10.1111/1467-9280.t01-1-01430] [Medline: 12661673]
59. Kramer AF, Hahn S, Cohen NJ, et al. Ageing, fitness and neurocognitive function. Nature New Biol 1999 Jul 29;400(6743):418-419. [doi: 10.1038/22682] [Medline: 10440369]
60. Hwang J, Lu AS. Narrative and active video game in separate and additive effects of physical activity and cognitive function among young adults. Sci Rep 2018;8(1):11020. [doi: 10.1038/s41598-018-29274-0]
61. McDonough DJ, Pope ZC, Zeng N, Lee JE, Gao Z. Comparison of college students' energy expenditure, physical activity, and enjoyment during exergaming and traditional exercise. J Clin Med 2018 Nov 10;7(11):433. [doi: 10.3390/jcm7110433] [Medline: 30423805]
62. Strath SJ, Swartz AM, Bassett DR Jr, O'Brien WL, King GA, Ainsworth BE. Evaluation of heart rate as a method for assessing moderate intensity physical activity. Med Sci Sports Exerc 2000 Sep;32(9 Suppl):S465-S470. [doi: 10.1097/00005768-200009001-00005] [Medline: 10993416]

XSL•FO
RenderX

63. Hwang J, Fernandez AM, Lu AS. Application and validation of activity monitors' epoch lengths and placement sites for physical activity assessment in exergaming. J Clin Med 2018 Sep 11;7(9):268. [doi: 10.3390/jcm7090268] [Medline: 30208567]
64. Aguilar-Farias N, Peeters G, Brychta RJ, Chen KY, Brown WJ. Comparing ActiGraph equations for estimating energy expenditure in older adults. J Sports Sci 2019 Jan;37(2):188-195. [doi: 10.1080/02640414.2018.1488437] [Medline: 29912666]
65. Sasaki JE, John D, Freedson PS. Validation and comparison of ActiGraph activity monitors. J Sci Med Sport 2011 Sep;14(5):411-416. [doi: 10.1016/j.jsams.2011.04.003] [Medline: 21616714]
66. Tanaka H, Monahan KD, Seals DR. Age-predicted maximal heart rate revisited. J Am Coll Cardiol 2001 Jan;37(1):153-156. [doi: 10.1016/s0735-1097(00)01054-8] [Medline: 11153730]
67. American College of Sports Medicine. ACSM's Guidelines for Exercise Testing and Prescription: Wolters Kluwer; 2021.
68. Miller NE, Strath SJ, Swartz AM, Cashin SE. Estimating absolute and relative physical activity intensity across age via accelerometry in adults. J Aging Phys Act 2010 Apr;18(2):158-170. [doi: 10.1123/japa.18.2.158] [Medline: 20440028]
69. Troiano RP, Berrigan D, Dodd KW, Mâsse LC, Tilert T, Mcdowell M. Physical activity in the united states measured by accelerometer. Med Sci Sports Exe 2008;40(1):181-188. [doi: 10.1249/mss.0b013e31815a51b3]
70. Puterman E, Lin J, Blackburn E, O'Donovan A, Adler N, Epel E. The power of exercise: buffering the effect of chronic stress on telomere length. PLoS ONE 2010 May 26;5(5):e10837. [doi: 10.1371/journal.pone.0010837] [Medline: 20520771]
71. Chudasama YV, Khunti KK, Zaccardi F, et al. Physical activity, multimorbidity, and life expectancy: a UK Biobank longitudinal study. BMC Med 2019 Jun 12;17(1):108. [doi: 10.1186/s12916-019-1339-0] [Medline: 31186007]
72. Cognition measures. NIH Toolbox. 2023. URL: https://www.healthmeasures.net/explore-measurement-systems/nih-toolbox/intro-to-nih-toolbox/cognition [accessed 2023-02-19]
73. Rhodes RE, Macdonald HM, McKay HA. Predicting physical activity intention and behaviour among children in a longitudinal sample. Soc Sci Med 2006 Jun;62(12):3146-3156. [doi: 10.1016/j.socscimed.2005.11.051] [Medline: 16406632]
74. Kendzierski D, DeCarlo KJ. Physical activity enjoyment scale: two validation studies. J Sport Exerc Psychol 1991;13(1):50-64. [doi: 10.1123/jsep.13.1.50]
75. IJsselsteijn WA, Kort YA, Poels K. The game experience questionnaire. : Eindhoven: Technische Universiteit Eindhoven URL: https://pure.tue.nl/ws/files/21666907/Game_Experience_Questionnaire_English.pdf [accessed 2025-10-01]
76. Lu AS, Buday R, Thompson D, Baranowski T. What type of narrative do children prefer in active video games? an exploratory study of cognitive and emotional responses. In: Tettegah S, Huang WD, editors. Emotions, Technology, and Digital Games: Elsevier Publications; 2016. [doi: 10.1016/B978-0-12-801738-8.00007-5]
77. Lu AS, Green MC, Thompson D. Using narrative game design to increase children's physical activity: exploratory thematic analysis. JMIR Serious Games 2019 Nov 21;7(4):e16031. [doi: 10.2196/16031] [Medline: 31750833]
78. George D, Mallery P. IBM SPSS Statistics 26 Step by Step: A Simple Guide and Reference: Routledge 2019. [doi: 10.4324/9780429056765]
79. Schrack JA, Leroux A, Fleg JL, et al. Using heart rate and accelerometry to define quantity and intensity of physical activity in older adults. J Gerontol A Biol Sci Med Sci 2018 Apr 17;73(5):668-675. [doi: 10.1093/gerona/gly029] [Medline: 29509832]
80. Christen J, Foster C, Porcari JP, Mikat RP. Temporal robustness of the session rating of perceived exertion. Int J Sports Physiol Perform 2016 Nov;11(8):1088-1093. [doi: 10.1123/ijspp.2015-0438] [Medline: 26999454]
81. Ferreira SS, Krinski K, Alves RC, et al. The use of session RPE to monitor the intensity of weight training in older women: acute responses to eccentric, concentric, and dynamic exercises. J Aging Res 2014;2014:749317. [doi: 10.1155/2014/749317] [Medline: 24834354]
82. Foster C, Florhaug JA, Franklin J, et al. A new approach to monitoring exercise training. J Strength Cond Res 2001 Feb;15(1):109-115. [Medline: 11708692]
83. Where are the fully-corrected scores in NIH toolbox V3. NIH Toolbox cognition measures. URL: https://nihtoolbox.zendesk.com/hc/en-us/articles/12647360212884-Where-are-the-fully-corrected-scores-in-NIH-Toolbox-V3 [accessed 2025-01-29]
84. Evenson KR, Buchner DM, Morland KB. Objective measurement of physical activity and sedentary behavior among US adults aged 60 years or older. Prev Chronic Dis 2012;9:E26. [Medline: 22172193]
85. Matthew CE. Calibration of accelerometer output for adults. Med Sci Sports Exerc 2005 Nov;37(11 Suppl):S512-S522. [doi: 10.1249/01.mss.0000185659.11982.3d] [Medline: 16294114]
86. Choi E, Shin SH, Ryu JK, et al. Association of extensive video gaming and cognitive function changes in brain-imaging studies of pro gamers and individuals with gaming disorder: systematic literature review. JMIR Serious Games 2021 Jul 9;9(3):e25793. [doi: 10.2196/25793] [Medline: 34255648]
87. Hötting K, Röder B. Beneficial effects of physical exercise on neuroplasticity and cognition. Neurosci Biobehav Rev 2013 Nov;37(9 Pt B):2243-2257. [doi: 10.1016/j.neubiorev.2013.04.005] [Medline: 23623982]

XSL•FO
RenderX

Abbreviations

AAR: active augmented reality
AR: augmented reality
AVG: active video game
BPM: beats per minute
GEQ: Gaming Experience Questionnaire
HR: heart rate
HRR: heart rate reserve
IGNITE: Investigating Gains in Neurocognition in an Intervention Trial of Exercise
MVPA: moderate-to-vigorous physical activity
PA: physical activity
RA: research assistants
RPE: rate of perceived exertion
RQ: research question
VM: vector magnitude
VR: virtual reality

Edited by A Coristine; submitted 27.02.25; peer-reviewed by P Bhowmick; revised version received 21.08.25; accepted 25.08.25; published 03.10.25.

Please cite as:
Lu AS, Parkar B, Gupta S, Hall S, Pindus DM, Kramer AF
The Impact of Active Augmented Reality Games on Physical Activity and Cognition Among Older Adults: Feasibility Study
JMIR Serious Games 2025;13:e73221
URL: https://games.jmir.org/2025/1/e73221
doi:10.2196/73221

© Amy Shirong Lu, Bhagyashree Parkar, Shiveshchandra Gupta, Sierra Hall, Dominika M Pindus, Arthur F Kramer. Originally published in JMIR Serious Games (https://games.jmir.org), 3.10.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Vision AI-Based Gamified Cognitive Prosthesis for Executive Function: Feasibility and Usability Study

Co Yih Siow[1], MD; Yao-Hua Yang[2], BSc; Cheng-Jui Tsai[3], MD; Wan-Wan Yang[2], MD; Chaur-Jong Hu[4,5], MD; Jia-Ying Sung[4,6], MD, PhD; Jowy Tani[4,6], MD, PhD

[1]Department of Physical Medicine and Rehabilitation, Taipei Medical University Hospital, Taipei Medical University, Taipei, Taiwan

[2]Biomed Innovation Center, Wan Fang Hospital, Taipei Medical University, Taipei, Taiwan

[3]School of Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan

[4]Department of Neurology, School of Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan

[5]Department of Neurology, Shuang Ho Hospital, Taipei Medical University, New Taipei City, Taiwan

[6]Department of Neurology, Wan Fang Hospital, Taipei Medical University, No.111, Sec. 3, Xinglong Rd., Wenshan Dist., Taipei, Taiwan

Corresponding Author:
Jowy Tani, MD, PhD
Department of Neurology, School of Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan

Abstract

Background: Dementia is a progressive neurodegenerative condition marked by cognitive decline and loss of functional independence. Among cognitive domains, executive dysfunction is a critical early contributor to reduced self-care capacity and increased caregiver burden. While cognitive assistive technologies have focused primarily on memory, few tools address executive function in real-time, daily tasks. To fill this gap, we developed a novel gamified cognitive prosthesis that integrates artificial intelligence (AI) and computer vision to guide users step-by-step through a simulated egg-cooking task. This system provides real-time audiovisual feedback to support planning, sequencing, and error correction.

Objective: This study aimed to evaluate whether the AI-based cognitive prosthesis improves task completion time and executive function performance in individuals with mild dementia.

Methods: We conducted a pilot study involving 12 patients with mild dementia and 7 age-matched healthy controls. Participants were asked to complete a 6-step gamified egg boiling task under 2 conditions: with and without guidance. The task was evaluated using a custom “Daily Task Completion Test” and a modified executive function performance test (EFPT) adapted to the cooking activity. Demographic and clinical data (age, sex, education, Mini-Mental State Examination, Clinical Dementia Rating, activities of daily living, instrumental activities of daily living, and Dementia Severity Rating Scale) were recorded. The System Usability Scale (SUS) was also collected postintervention.

Results: In the mild dementia group, AI assistance significantly reduced median task completion time from 134.75 (IQR 92.50 - 134.75) to 92.00 (IQR 65.00 - 92.00; *P*=.03) seconds, and significantly improved the Executive Function Performance Test (EFPT) scores from 4.25 (IQR 1.75 - 4.25) to 1.00 (IQR 0.00 - 1.00; *P*=.005), reflecting a 31.7% improvement in efficiency and a 76.5% reduction in required assistance. No significant changes were observed in the control group. The mean SUS score was 80.53 (SD 24.97), indicating high usability. The AI system achieved a cumulative recognition precision of 0.93 (SD 0.07) and cumulative recall of 0.94 (SD 0.11).

Conclusions: This pilot study provides preliminary evidence that an AI-based cognitive prosthesis can enhance executive function and task performance in individuals with mild dementia. The results support the feasibility of using real-time AI guidance in everyday tasks to promote independence. Given its modular design and promising usability profile, this system may serve as the foundation for future digital therapeutics targeting executive dysfunction. Larger, longitudinal studies are warranted to evaluate sustained cognitive and functional benefits.

(JMIR Serious Games 2025;13:e74157) doi:10.2196/74157

KEYWORDS

artificial intelligence; audiovisual stimulation; cognitive impairment; cognitive prosthesis; digital therapeutics; executive function; mild dementia

XSL•FO
RenderX

Introduction

The rising incidence and prevalence of dementia have led to a substantial global care burden [1,2]. Dementia represents a significant challenge as it gradually diminishes the self-sufficiency of individuals in their everyday routines and results in a notable deterioration in their overall well-being [3]. Consequently, this imposes a substantial burden on both the affected families and society [4-7]. Common causes of dementia include Alzheimer disease, vascular dementia, and frontotemporal dementia, with Alzheimer disease contributing to the largest part of the dementia population [8]. The caregiver burden of patients with Alzheimer disease was reportedly significantly higher compared to the burden borne by caregivers of older adults with psychosis, with spouses bearing the highest burden score [7]. The estimated number of people with dementia exceeded 50 million in 2019, and the cases will surpass 150 million in 2050 globally. Globally, the annual spending on individuals with dementia is over a trillion US dollars, while the forecasted cost for dementia will reach US $1.6 trillion in 2050 [8-10]. The societal burden of dementia is projected to reach epidemic levels across countries in the Asia-Pacific region [11]. Taiwan, similarly, is currently experiencing an increasing burden of dementia. Such statistics underscore the critical need for innovative interventions.

Over the past decade, various assistive technologies have been developed for individuals with dementia, with variable levels of success [12,13]. Memory has been a significant focus for the development of cognitive assistive devices, and previous studies have shown that cognitive prosthesis could be beneficial for patients with impaired facial memory [13,14]. The term cognitive prosthesis refers to a compensatory strategy to modulate the environment for rehabilitation through computer technology focusing on the individual's functional activities [15]; in this study, the term refers to a device that augments the cognitive function of the user. Digital interventions for dementia have also been attempted in previous studies [16-20]. In terms of interventions for Alzheimer disease and the general mild-to-moderate stage of dementia, cognitive stimulation therapy and psychosocial interventions have reportedly been shown to preserve cognitive abilities longer than pharmacological treatments [8]. Cognitive rehabilitation, personalized to each individual's needs, has also been reported to improve the activities of daily living (ADL) score and their cognitive functioning, providing evidence that cognitive assessments through various means, including cognitive assistive technology, may be useful as an intervention for all-cause dementia [21].

However, the global burden of various diseases' impacts is receiving increasing attention based on the results of the global burden of disease study, which the World Health Organization supports [22]. There is a paradigm shift in dementia management toward promoting functional independence, a key factor directly influencing the quality of life and care burden. The role of executive function in dementia care has been increasingly recognized, and artificial intelligence (AI)-based tools are considered valuable for promoting functional independence [23].

Executive function is a multifaceted neuropsychological construct that can be defined as forming, maintaining, and shifting mental sets. It includes basic cognitive processes such as attentional control, inhibition, working memory, cognitive flexibility, judgment, and sequencing. The executive function enables people to formulate goals successfully, plan how to achieve them, and carry out the plans effectively, making it essential for daily life independence [24,25]. It plays a crucial role in everyday functioning and is indispensable for instrumental activities of daily living (IADL) [25]. Numerous studies have pointed out that improvements in IADL and Executive Function Performance Test (EFPT) scores were associated with improved quality of life [26,27], reduced incidents of home accidents [28], decreased caregiver burden [29], and lower cost of care [30]. This highlights the importance of shifting focus to executive functions in dementia assessment and treatment.

Although executive function plays a vital role in daily functioning, few assistive technologies have been developed to provide real-time assistance for executive functioning, helping patients complete daily tasks [31]. There has been computerized cognitive training with the NeuroNation MED app in which the users have shown significant improvement in global cognition in mild cognitive impairment (MCI) patients compared to controls [32]. Another randomized clinical trial featuring Constant Therapy app in over 24 weeks for Alzheimer's patients also showed high adherence and task improvements [33]. Virtual reality and gaming-based therapies also have encouraging cognitive benefits, with moderate-to-large gains in global cognition, memory, and executive skills in MCI and dementia patients [34]. Virtual reality–based shopping programs have also been developed to screen for MCI or as therapy, with promising results for both intended uses [35,36]. However, there have been few assistive technologies using AI to improve executive functions.

In response to this gap, a novel AI-based cognitive prosthesis was introduced, integrating gamification and sensor technology to specifically aid individuals with dementia in the training of their executive functions. The prosthesis simulates everyday tasks—cooking egg using interactive tools—designed to replicate the sequential planning, decision making, and motor coordination required in daily life. The present study investigated the effectiveness of an AI-based cognitive prosthesis on individuals with dementia to improve executive functions.

Methods

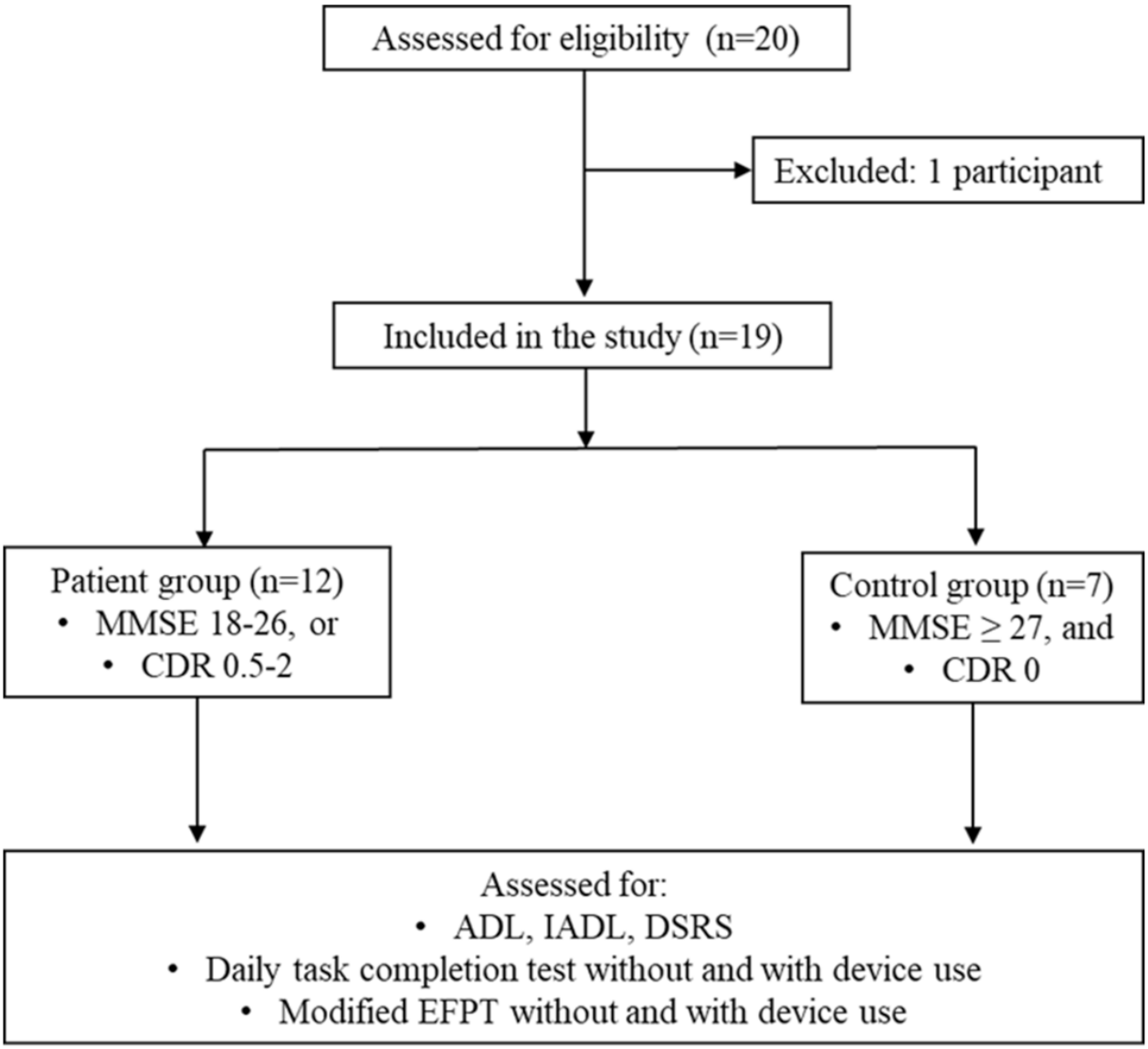
Study Settings and Cohort

Individuals with mild dementia (patient group) and healthy individuals (control group) were recruited from Wan Fang Hospital, Taipei Medical University. Performance differences between the patient and control groups were compared under 2 conditions: with or without the assistance of cognitive guidance devices. A flowchart outlining the experimental process is presented in Figure 1. Mini-Mental State Examination (MMSE) and Clinical Dementia Rating (CDR) tests were done as the participant selection criteria; the age, gender, and

XSL•FO
RenderX

education level of the participants in both groups were also recorded. The participants included in the study were assessed for their ADL, IADL, and Dementia Severity Rating Scale (DSRS) at baseline before being asked to use the cognitive prosthesis. After completing the tasks, the participants were assessed with the modified EFPT and asked to fill out the System Usability Scale (SUS) questionnaire.

Figure 1. Flow chart depicting the experimental process. MMSE: Mini-Mental State Examination; CDR: Clinical Dementia Rating; ADL: activities of daily living; IADL: instrumental activities of daily living; DSRS: Dementia Severity Rating Scale; EFPT: Executive Function Performance Test.



Eligibility Criteria

In total, 12 individuals with mild dementia and 7 healthy individuals were recruited into the study. Individuals with mild dementia were eligible to be enrolled in the study if they met either of the following conditions: (1) MMSE score between 18‐26, or (2) CDR score between 0.5‐1.0 [37]. The age-matched healthy control participants' eligibility criteria were as follows: (1) MMSE score equal to or greater than 27 [38] and (2) CDR score of 0.

Individuals with the following characteristics were excluded from the study: (1) inability to read and understand Chinese, (2) physical impairments, such as upper limb impairments, vision, and hearing impairments, which might hinder the subjects' ability to undergo testing, (3) age less than 60 years, and (4) have not received more than six years of education. All participants were provided written and verbal explanations of the experiment and gave written informed consent.

Ethical Considerations

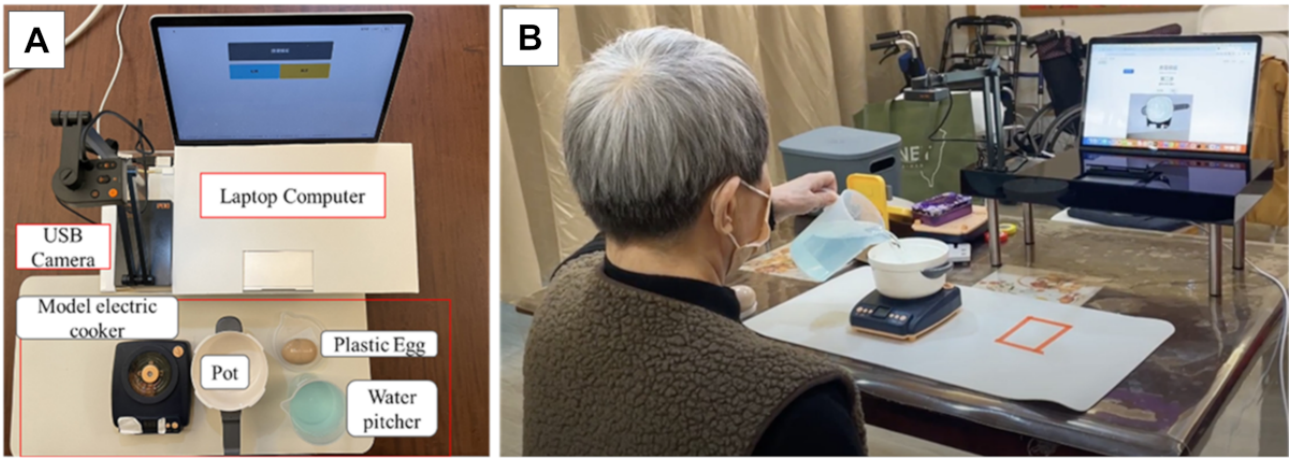
All participants were recruited from Taipei Municipal Wan Fang Hospital, Taipei, Taiwan. The study procedures adhered to the Declaration of Helsinki. The study protocol was approved by the Joint Institution Review Board of Taipei Medical University (Protocol number: N202204004). All participants provided written informed consent after receiving both written and verbal explanations of the study. All participants' identities remained anonymous and undisclosed for privacy and confidentiality protection. No compensation was offered to the participants.

XSL•FO
RenderX

Intervention: Executive Function Cognitive Prosthesis for Simple Cooking

The novel cognitive prosthesis system consisted of a laptop, a universal serial bus camera connected to the laptop computer, and daily task objects. An AI-based executive function cognitive prosthesis software was installed on the laptop computer. Daily task objects—a model electric cooker, a plastic pot, a plastic egg, and a water pitcher—were used in the simulated egg-cooking task. The gamified cognitive prosthesis setup and experimental environment are depicted in Figure 2A and B.

Figure 2. The gamified cognitive prosthesis setup. (**A**) Overall system for the simple cooking task. (**B**) A patient performing the task with the help of the cognitive prosthesis.



In this study, the participants were instructed to perform a simple cooking task—boiling an egg—with or without visual and auditory cues provided by the executive function cognitive prosthesis.

The executive function cognitive prosthesis software was designed to assist users through the necessary steps of the egg boiling task. It analyzed visual data captured by the universal serial bus camera in real-time using a machine learning model to determine the state of each step of the task the participant has done. The machine learning model was trained using the YOLO v5 framework (Ultralytics), and the software was designed to operate on the Web platform. The software was configured to recognize the state of each task step by performing object recognition on the daily task objects—such as an electric cooker, pot, eggs, and water pitcher—by the machine learning model. The software would determine whether each task step is being performed correctly, record the time for each step completed, and then provide visual—both text and graphical—and audio instructions on the next step.

The cognitive prosthesis system incorporates 2 AI modules. The first module uses the Mediapipe hand detection model to monitor the participant's actions, capturing key metrics such as initiation time, task duration, and task completion time [39]. The second module is a custom-trained model based on the YOLOv5 image recognition model [40], which identifies object locations and subsequently assesses task completion status. For custom training, we used 472 images for the training dataset, 93 images for the validation dataset, and 80 images for the test dataset, achieving an accuracy of 95.42%.

To develop the object detection model, we adopted transfer learning, initializing from pretrained weights on the COCO dataset, and fine-tuned the model using our custom dataset of 472 labeled images, covering approximately 20 distinct object classes. The sample images of the training dataset can be found in Figure S1 in Multimedia Appendix 1. Model training was conducted on Google Collab Pro using an NVIDIA Tesla T4 GPU with 16GB of VRAM. We adhered to the standard YOLOv5 training pipeline and employed a comprehensive toolset, including PyTorch for model architecture and training, OpenCV for image preprocessing and augmentation, and NumPy and Pandas for data manipulation. Visualization of model performance was carried out using Matplotlib and Seaborn. In addition, the Ultralytics command-line interface and YAML Ain't Markup Language (YAML)–based configuration files facilitated customizable and reproducible hyperparameter tuning. This integrated setup enabled efficient experimentation while ensuring transparency and reproducibility in our workflow.

The model was trained using the default YOLOv5s input resolution (640×640), with a batch size of 16. Each training cycle comprised 250 epochs, a number selected to ensure model convergence given the relatively small dataset. Training was conducted on Google Colab Pro using a T4 GPU, with each cycle requiring approximately 100 - 130 minutes (1.7 - 2.2 h), depending on runtime conditions and GPU availability. Following each training cycle, we conducted a statistical evaluation that included calculation of mean average precision, class-wise precision, recall, and F_1-scores, as well as confusion matrix generation. This analysis phase added an additional 3 - 5 minutes per cycle. Consequently, each full training-evaluation iteration took approximately 105 - 135 minutes. The process was repeated across 5 - 7 iterations, with hyperparameters adjusted between runs to monitor performance trends and support generalizability of the model.

Camera data were collected every 5 ms and were processed using the aforementioned modules for inference. To ensure recognition stability, a rolling average over 5 frames was used. The recognition results and task progress were integrated continuously to update the system feedback.

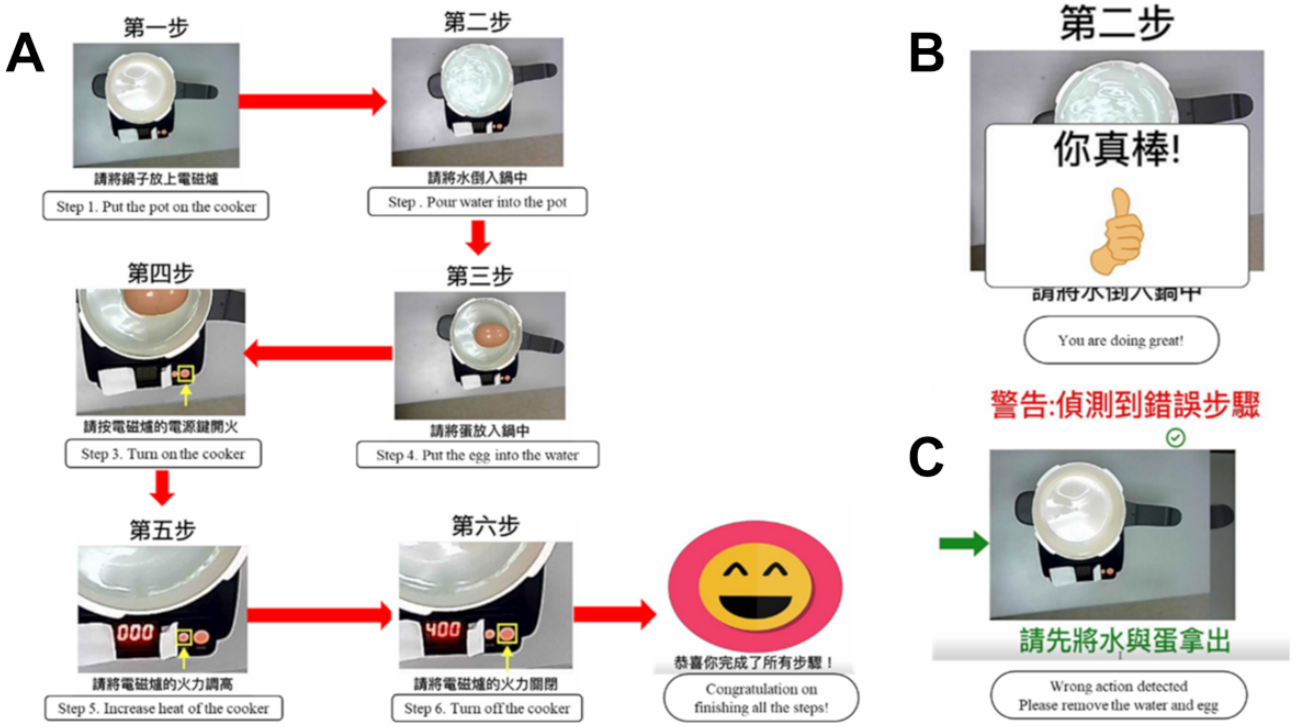
XSL•FO
RenderX

Cognitive Assistance and Clinical Assessments

All participants underwent neuropsychological tests and questionnaires, including ADL, IADL, and DSRS. The participants were then given approximately 2 minutes to read and familiarize themselves with the steps involved in the cooking task. The individuals with dementia were asked to perform the egg boiling task without any assistance from the cognitive prosthesis and assessed for a "daily task completion test" designed for the present study. Each of the 6 task steps was scored on the basis of task performance accuracy and completion time: 2 points for fully correct, 1 for partially incorrect, and 0 for incorrect performance. The perfect score was 12 points, indicating successful completion of all steps.

Subsequently, the participants repeated the egg boiling task with the assistance from the cognitive prosthesis device and were assessed for a "daily task completion test" again, as illustrated in Figure 3A. Guided by the system, the participants would proceed through the steps. Upon correct completion of each step, the system would provide positive reinforcement by displaying the message "great" (Figure 3B), after which the next step was prompted. If a step was performed incorrectly, the system would provide corrective instructions, requiring the participant to repeat the step correctly before proceeding (Figure 3C).

Figure 3. Egg-boiling task of the cognitive prosthesis overview. (**A**) Step-by-step egg cooking process with visual and auditory assistance of a cognitive prosthesis. (**B**) A screenshot showing the correct execution of the first step and transition to the second step, and (**C**) a screenshot showing a mistake in the second step and instructions for correction.



Executive function was evaluated using a modified EFPT that was adapted from the original EFPT of simple cooking task. The scoring ranged from 0 to 25, based on a 5-level cueing system: 0 (no cue required); 1 (verbal assistance); 2 (gestural assistance); 3 (direct verbal assistance); 4 (physical assistance); 5 (task performed by the assessor). A higher score indicates greater executive dysfunction.

Following the experimental task, participants or their caregivers completed the SUS questionnaire, designed to effectively distinguish user-friendly systems from less efficient ones on the basis of ease of use and efficiency.

Object Recognition Algorithm Performance Evaluation

The computing time of the object recognition algorithm was evaluated across the 6 sequential task steps. For each step, the model's processing time was measured and its cumulative precision and recall were evaluated.

Statistical Analysis

The participant clinicodemographic characteristics, modified EFPT scores, task completion time, and completion score were summarized using descriptive statistics. Between-group differences were analyzed using the Mann-Whitney *U* test, whereas the sex distribution differences were analyzed using Fisher exact test. Performance differences before and after assistive device use for each participant were evaluated using the Wilcoxon signed-rank test. User experience data were analyzed descriptively. The classification model performance was assessed using a confusion matrix, with statistical significance set at $P<.05$. Processing delay times for each step of the executive function AI cognitive prosthesis were calculated and are presented as mean (SD) values. The diagram illustrating the statistical analysis done in the study can be found in Figure S2 in Multimedia Appendix 2.

XSL•FO
RenderX

Results

Cohort Characteristics

Clinicodemographic characteristics of the study participants are shown in Table 1. One patient was excluded from the study due to refusal to participate. There is no significant difference in age (*P*=.13) or gender (*P*=1.0) between the control group and the patient group after comparison. No significant differences were observed between the patients and control groups in terms of years of education (10, IQR 10-15 vs 12, IQR 12-16; *P*=.38). However, the patient group had significantly lower MMSE score (24.25, IQR 18.75 - 24.25 vs. 30.00, IQR 30.00 - 30.00; *P*<.001), DSRS scores (18.00, IQR 4.25 - 18.00 vs 0.00, IQR 0.00 - 0.00; *P*<.001), and IADL scores (*P*=.002), as well as significantly higher CDR score (*P*=.008). However, the ADL score difference between the control and patient group was not significant (*P*=.06) (Table 1).

Table . Demographic and clinical characteristics of the patient and control group.

Variables		Patient group (n=12)	Control group (n=7)	*P* value
Demographic profile				
	Age (years), median (IQR)	78.50 (75.25 - 83.75)	72.00 (71.00 - 72.00)	.13
	Sex-(male), n (%)	3 (25)	1 (15)	>.99
	Education (years), median (IQR)	15.00 (10.00 - 15.00)	12.00 (12.00 - 16.00)	.38
Clinical profile, median (IQR)				
	MMSE[a]	24.25 (18.75 - 24.25)	30.00 (30.00 - 30.00)	<.001
	CDR[b]	1.00 (0.50 - 1.00)	0.00 (0.00 - 0.00)	.008
	ADL[c]	100.00 (88.75 - 100.00)	100.00 (100.00 - 100.00)	.06
	IADL[d]	5.50 (3.00 - 5.50)	8.00 (8.00 - 8.00)	.002
	DSRS[e]	18.00 (4.25 - 18.00)	0.00 (0.00 - 1.00)	<.001

[a]MMSE: Mini-Mental State Examination.

[b]CDR: Clinical Dementia Rating.

[c]ADL: activities of daily living.

[d]IADL: instrumental activities of daily living.

[e]DSRS: Dementia Severity Rating Scale.

Comparison of Task Completion Test in Both Groups

Table 2 shows the task performance results for the patient and control groups. At baseline, the patients have longer task completion time (seconds; 134.75, IQR 92.50 - 134.75 vs 34.00, IQR 32.00 - 42.00; *P*=.001) and a worse modified EFPT score (4.25, IQR 1.75 - 4.25 vs 0.00, IQR 0.00 - 0.00; *P*=.001) than the healthy controls.

XSL•FO
RenderX

Table . Comparison of task completion test for the advanced task module in both groups.

		Patient group (n=12)	Control group (n=7)	*P* value between groups
Task completion time (seconds)				
	Without device assistance, median (IQR)	134.75 (92.50 - 134.75)	34.00 (32.00 - 42.00)	.001
	With device assistance, median (IQR)	92.00 (65.00 - 92.00)	39.00 (32.50 - 40.00)	.001
	P value without versus with device	.03	.91	—[a]
Task completion score				
	Without device assistance, median (IQR)	12.00 (10.00 - 12.00)	12.00 (12.00 - 12.00)	.008
	With device assistance, median (IQR)	12.00 (10.75 - 12.00)	12.00 (12.00 - 12.00)	.06
	P value without versus with device	.14	>.99	—
Modified EFPT[b] score				
	Without device assistance, median (IQR)	4.25 (1.75 - 4.25)	0.00 (0.00 - 0.00)	.001
	With device assistance, median (IQR)	1.00 (0.00 - 1.00)	0.00 (0.00 - 0.00)	.10
	P value without versus with device	.005	>.99	—

[a]Not applicable.

[b]EFPT: executive function performance test.

Although the use of assistive device led to no significant improvement in the task completion score, the patient group exhibited significant improvements in the time taken and modified EFPT score after using assistive devices. In the patient group, the time (seconds) required to complete the task significantly (*P*=.03) decreased from 134.75 (IQR 92.50 - 134.75) to 92.00 (IQR 65.00 - 92.00) after the device was used. Also, their modified EFPT score significantly (*P*=.005) decreased from 4.25 (1.75 - 4.25) points before using assistive devices to 1.00 (IQR 0.00 - 1.00) points with use. In contrast, there were no significant differences in the performance of the control group, whether they used assistive devices or not (Table 2).

SUS

The SUS for executive function AI cognitive prosthesis device prototype for both groups is shown in Table 3. The SUS score interpretation is as follows: significant usability issues for <50 score, marginal usability for ≥50 to <68, acceptable usability for ≥68 to 80, and excellent usability for score above 80.

The overall SUS score obtained in this study was 80.53 (SD 24.97); a grade A categorization may be given, denoting acceptable usability. The overall level of the SUS score shows superior performance.

Table . The mean computing time for task completion steps.

	Computing time (ms), mean (SD)	Cumulative precision, mean (SD)	Cumulative recall, mean (SD)
Total Steps	11,720 (2.27)	0.93 (0.07)	0.94 (0.11)
Step 0	1062 (0.44)		
Step 1	1462 (0.99)		
Step 2	3125 (0.89)		
Step 3	1125 (0.81)		
Step 4	1625 (0.81)		
Step 5	2091 (0.83)		
Step 6	1231 (0.73)		

XSL•FO
RenderX

Device Object Recognition Machine Learning Algorithm Performance of the Advanced Task Module

The cumulative precision of the object recognition machine learning model was 0.93 (SD 0.07), while the cumulative recall was 0.94 (SD 0.11). Table 3 shows the mean computing time for task completion steps, with the mean computing time for all steps being 11,720 (SD 2.27 ms; Table 3).

Discussion

Insight From the Cognitive Prosthesis Study

We evaluated the efficacy of an executive function cognitive prosthesis in aiding individuals with mild dementia. The device usage led to 76.5% improvement in the modified EFPT scores and 31.7% reduction in task completion time among patients with mild dementia, indicating improvement in executive function-related tasks. These findings suggest that the cognitive prosthesis has the potential to improve the patients' daily activities.

Differences in Baseline Profile of Patient Group

At baseline, the patients had significantly lower MMSE, IADL, and DSRS scores, as well as higher CDR, compared to the healthy control group; this confirmed the cognitive impairments in the patient group. During the simple cooking task completion, the patients required significantly longer time to complete the steps, along with lower task completion score and worse modified EFPT score; these reflected the deficit in daily task executive abilities in the patient group. These outcomes underscore the need for assistance among dementia patients in performing tasks.

Effects of Cognitive Prosthesis Use on the Patients and Healthy Controls

To address the diverse training requirements of each subject and the different objects familiar to elderly individuals in different regions, we designed this system with a particularly scalable nature from the outset. Each time a new assistive module is designed, the process involves: (1) Designing the task workflow, (2) Collecting correct and incorrect workflow steps as training data for the image recognition model, and (3) Configuring the configuration file to adjust commands, recognition logic, and scoring methods. With a dedicated team of 3 (a designer, an AI engineer, and a system engineer), it takes approximately 2‐4 weeks to complete, demonstrating significant scalability.

The egg boiling task was selected as the initial AI daily task module for patients with mild dementia to assist with executive function. Usage of the cognitive prosthesis device enabled the patient group to complete tasks more efficiently. Notably, the modified EFPT scores showed significant improvement with the use of the device, suggesting that the cognitive prosthesis can enhance independence in performing simple daily tasks.

SUS and Device Performance for the Cognitive Prosthesis

The adequate overall SUS score obtained, together with the excellent machine learning performance, did demonstrate the usability of the device.

Toward the Cognitive Digital Therapeutic System

As the cognitive prosthesis system showed potential to improve EFPT during device use, this implicates the potential of using the similar technology to develop a digital therapeutics system (DTx). DTx should be able to assist the patients to have a longstanding improvement in executive function following regular training. The pilot study results of the integrated task DTx for healthy participants showed improved task completion scores and EFPT. Future interventional study is needed to ascertain whether the effects of cognitive prosthesis on individuals with dementia are sustained, in terms of long-term improvements in IADL and EFPT after extended use.

Limitations

However, this study encountered some difficulties during its execution, and certain limitations should be considered. First and foremost, it is crucial to recognize that this research represents a single experimental intervention for the patients. As a result, it is uncertain whether the observed enhancements in the patients' executive functions will persist or further develop in subsequent evaluations. Second, the number of participants fell short of our initial expectations. Even though the task completion scores of the patient group showed no significant change before and after device assistance, probably due to the limited sample size, the task completion time still significantly decreased by 31.7%.

Future Work and Recommendation

Future studies with extended intervention periods are warranted to obtain more robust results and a thorough grasp of the long-term influence of cognitive prosthesis technology on executive function. Building on the results of this study, future research will expand to larger sample sizes to gather more substantial evidence. We aim to rapidly and efficiently develop AI-based cognitive prosthesis tailored to individual training needs, including hand operations, cognitive processes, and other areas. These developments will consider the cultural backgrounds of users to ensure the assistive technology meets diverse requirements. In the next research phase, we will focus on an integrated task module, specifically the shopping module, using the executive function cognitive prosthesis. We have high expectations for this phase and anticipate recruiting more participants to further demonstrate the efficacy of the executive function cognitive prosthesis.

Conclusion

In conclusion, this study indicates that implementing cognitive prosthesis effectively improved executive function among individuals with mild dementia. The use of visual and auditory cues by the cognitive prosthesis reduced the task completion time in individuals with dementia. The cognitive prosthesis also showed high usability rating, indicating that it can be feasibly used by older adults. While further research is warranted to

XSL•FO
RenderX

explore the potential of this intervention, it is expected that extending the use of cognitive prosthesis holds significant promise in dementia care and rehabilitation interventions. Maximizing their applicability is imperative to enhance the overall quality of care and support those people affected by cognitive impairment.

Acknowledgments

This work was supported by the Higher Education Sprout Project of the Ministry of Education (MOE) in Taiwan, the National Science and Technology Council (NSTC) in Taiwan (Grants no: 107‐2314-B-038‐078-, 108‐2314-B-038‐049-, 110‐2314-B-038-091-, 111‐2314-B-038‐129-, 113-2823-8-038-004-, 114-2314-B-038-063-), Taipei Medical University, Wan Fang Hospital, Taipei, Taiwan (Grants no: TMU111-AE1-B33, 112-wf-IIP-01, 112-wf-f-02, 113-wf-IIP-01, 114TMU-WFH-07).

Data Availability

The datasets analyzed during this study are available from the corresponding authors on reasonable request.

Authors' Contributions

JT was involved in the conceptualization, writing–original draft, methodology, validation, formal analysis, investigation, resources, data curation, visualization, project administration, supervision, funding acquisition, writing–review and editing of this study and article; CYS was involved in the methodology, validation, formal analysis, investigation, resources, data curation, as well as writing–original draft; YHY was involved in the software, visualization, conceptualization, and writing–original draft. CJT was involved in formal analysis, and writing–review and editing; JYS was involved in the data curation and writing–review and editing of the manuscript, as well as supervision of the study; WWY was involved in the formal analysis and writing–review and editing; CJH was involved in the formal analysis and writing–review and editing of the manuscript and study, as well as supervision of the study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Sample images of training dataset.
[DOCX File, 575 KB - games_v13i1e74157_app1.docx]

Multimedia Appendix 2
Statistical analysis flow diagram.
[DOCX File, 106 KB - games_v13i1e74157_app2.docx]

References

1. Nichols E, Steinmetz JD, Vollset SE, et al. Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: an analysis for the Global Burden of Disease Study 2019. Lancet Public Health 2022 Feb;7(2):e105-e125. [doi: 10.1016/S2468-2667(21)00249-8]
2. Nichols E, Szoeke CEI, Vollset SE, et al. Global, regional, and national burden of Alzheimer's disease and other dementias, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet Neurol 2019 Jan;18(1):88-106. [doi: 10.1016/S1474-4422(18)30403-4]
3. Logsdon RG, Gibbons LE, McCurry SM, Teri L. Assessing quality of life in older adults with cognitive impairment. Psychosom Med 2002;64(3):510-519. [doi: 10.1097/00006842-200205000-00016] [Medline: 12021425]
4. Cheng ST. Dementia caregiver burden: a research update and critical analysis. Curr Psychiatry Rep 2017 Aug 10;19(9):64. [doi: 10.1007/s11920-017-0818-2] [Medline: 28795386]
5. Reuben DB, Epstein-Lubow G, Super N. The other dementia breakthrough-comprehensive dementia care. JAMA Neurol 2023 Aug 1;80(8):770-772. [doi: 10.1001/jamaneurol.2023.1252] [Medline: 37213130]
6. Li J, Skinner JS, McGarry K, et al. Declines in wealth among US older adults at risk of dementia. JAMA Neurol 2023 Nov 1;80(11):1250-1252. [doi: 10.1001/jamaneurol.2023.3216] [Medline: 37721762]
7. Sinha P, Desai NG, Prakash O, Kushwaha S, Tripathi CB. Caregiver burden in Alzheimer-type dementia and psychosis: A comparative study from India. Asian J Psychiatr 2017 Apr;26:86-91. [doi: 10.1016/j.ajp.2017.01.002] [Medline: 28483100]
8. Patterson C. World Alzheimer Report 2018. : Alzheimer's Disease International; 2018 URL: https://www.alzint.org/resource/world-alzheimer-report-2018/ [accessed 2025-09-12]

XSL•FO
RenderX

9. Velandia PP, Miller-Petrie MK, Chen C, et al. Global and regional spending on dementia care from 2000-2019 and expected future health spending scenarios from 2020-2050: An economic modelling exercise. EClinicalMedicine 2022 Mar;45:101337. [doi: 10.1016/j.eclinm.2022.101337] [Medline: 35299657]
10. Torabgar M, Figeys M, Esmail S, Stroulia E, Ríos Rincón AM. Machine learning analysis of engagement behaviors in older adults with dementia playing mobile games: exploratory study. JMIR Serious Games 2025 Mar 3;13:e54797. [doi: 10.2196/54797] [Medline: 40029608]
11. Venketasubramanian N, Sahadevan S, Kua EH, Chen CPL, Ng TP. Interethnic differences in dementia epidemiology: global and Asia-pacific perspectives. Dement Geriatr Cogn Disord 2011 Jan 20;30(6):492-498. [doi: 10.1159/000321675]
12. Pappadà A, Chattat R, Chirico I, Valente M, Ottoboni G. Assistive technologies in dementia care: an updated analysis of the literature. Front Psychol 2021;12:644587. [doi: 10.3389/fpsyg.2021.644587] [Medline: 33841281]
13. Van der Roest HG, Wenborn J, Pastink C, Dröes RM, Orrell M. Assistive technology for memory support in dementia. Cochrane Database Syst Rev 2017 Jun 11;6(6):CD009627. [doi: 10.1002/14651858.CD009627.pub2] [Medline: 28602027]
14. Tani J, Yang YH, Chen CM, et al. Domain-specific cognitive prosthesis for face memory and recognition. Diagnostics (Basel) 2022 Sep 16;12(9):2242. [doi: 10.3390/diagnostics12092242] [Medline: 36140643]
15. Cognitive prosthetics: an overview to a method of treatment. NeuroRehabilitation 1999 Feb 1;12(1):39-51. [doi: 10.3233/NRE-1999-12105]
16. Wu S, Ji H, Won J, Jo EA, Kim YS, Park JJ. The effects of exergaming on executive and physical functions in older adults with dementia: randomized controlled trial. J Med Internet Res 2023 Mar 7;25:e39993. [doi: 10.2196/39993] [Medline: 36881445]
17. McCrae CS, Curtis AF, Cottle A, et al. Impact of web-based cognitive behavioral therapy for insomnia on stress, health, mood, cognitive, inflammatory, and neurodegenerative outcomes in rural dementia caregivers: protocol for the NiteCAPP CARES and NiteCAPP SHARES randomized controlled trial. JMIR Res Protoc 2022 Jun 14;11(6):e37874. [doi: 10.2196/37874] [Medline: 35700020]
18. Øksnebjerg L, Woods B, Ruth K, et al. A tablet app supporting self-management for people with dementia: explorative study of adoption and use patterns. JMIR Mhealth Uhealth 2020 Jan 17;8(1):e14694. [doi: 10.2196/14694] [Medline: 31951217]
19. Zhu K, Zhang Q, He B, Huang M, Lin R, Li H. Immersive virtual reality-based cognitive intervention for the improvement of cognitive function, depression, and perceived stress in older adults with mild cognitive impairment and mild dementia: pilot pre-post study. JMIR Serious Games 2022 Feb 21;10(1):e32117. [doi: 10.2196/32117] [Medline: 35188466]
20. Muñoz J, Mehrabi S, Li Y, et al. Immersive virtual reality exergames for persons living with dementia: user-centered design study as a multistakeholder team during the COVID-19 pandemic. JMIR Serious Games 2022 Jan 19;10(1):e29987. [doi: 10.2196/29987] [Medline: 35044320]
21. Dutzi I, Schwenk M, Kirchner M, Bauer JM, Hauer K. Cognitive change in rehabilitation patients with dementia: prevalence and association with rehabilitation success. J Alzheimers Dis 2017;60(3):1171-1182. [doi: 10.3233/JAD-170401] [Medline: 28984597]
22. Mathers CD. History of global burden of disease assessment at the World Health Organization. Arch Public Health 2020;78(1):77. [doi: 10.1186/s13690-020-00458-3] [Medline: 32850124]
23. Patel VL. Cognitive science in the evaluation of medical AI systems. BMJ Health Care Inform 2023 Dec 14;30(1):e100929. [doi: 10.1136/bmjhci-2023-100929] [Medline: 38101808]
24. Cicerone KD, Langenbahn DM, Braden C, et al. Evidence-based cognitive rehabilitation: updated review of the literature from 2003 through 2008. Arch Phys Med Rehabil 2011 Apr;92(4):519-530. [doi: 10.1016/j.apmr.2010.11.015] [Medline: 21440699]
25. Kennedy MRT, Coelho C, Turkstra L, et al. Intervention for executive functions after traumatic brain injury: a systematic review, meta-analysis and clinical recommendations. Neuropsychol Rehabil 2008 Jun;18(3):257-299. [doi: 10.1080/09602010701748644] [Medline: 18569745]
26. Baum CM, Wolf TJ, Wong AWK, et al. Validation and clinical utility of the executive function performance test in persons with traumatic brain injury. Neuropsychol Rehabil 2017 Jul;27(5):603-617. [doi: 10.1080/09602011.2016.1176934] [Medline: 27150506]
27. Sanders D, Scott P. Literature review: technological interventions and their impact on quality of life for people living with dementia. BMJ Health Care Inform 2020 Jan;27(1):e100064. [doi: 10.1136/bmjhci-2019-100064] [Medline: 31948938]
28. Hatice T. Determination of the relationship between the dependence status on daily living activities and home accidents among 65 years of age and older individuals living at home. Turk J Geriatr 2006;9(2):85-93 [FREE Full text]
29. Razani J, Kakos B, Orieta-Barbalace C, et al. Predicting caregiver burden from daily functional abilities of patients with mild dementia. J Am Geriatr Soc 2007 Sep;55(9):1415-1420. [doi: 10.1111/j.1532-5415.2007.01307.x] [Medline: 17767684]
30. van Lier LI, van der Roest HG, Oosten BS, et al. Predictors of societal costs of older care-dependent adults living in the community in 11 european countries. Health Serv Insights 2019;12:1178632918820947. [doi: 10.1177/1178632918820947] [Medline: 30718960]
31. Lee-Cheong S, Amanullah S, Jardine M. New assistive technologies in dementia and mild cognitive impairment care: A PubMed review. Asian J Psychiatr 2022 Jul;73:103135. [doi: 10.1016/j.ajp.2022.103135] [Medline: 35569363]

XSL•FO
RenderX

32. Ferizaj D, Stamm O, Perotti L, et al. Preliminary effects of mobile computerized cognitive training in adults with mild cognitive impairment: interim analysis of a randomized controlled trial. BMC Psychol 2025 Mar 5;13(1):202. [doi: 10.1186/s40359-025-02458-w] [Medline: 40045413]
33. Marin A, DeCaro R, Schiloski K, et al. Home-based electronic cognitive therapy in patients with Alzheimer disease: feasibility randomized controlled trial. JMIR Form Res 2022 Sep 12;6(9):e34450. [doi: 10.2196/34450] [Medline: 36094804]
34. Moulaei K, Sharifi H, Bahaadinbeigy K, Dinari F. Efficacy of virtual reality-based training programs and games on the improvement of cognitive disorders in patients: a systematic review and meta-analysis. BMC Psychiatry 2024 Feb 12;24(1):116. [doi: 10.1186/s12888-024-05563-z] [Medline: 38342912]
35. Park JH. Does the virtual shopping training improve executive function and instrumental activities of daily living of patients with mild cognitive impairment? Asian J Psychiatr 2022 Mar;69:102977. [doi: 10.1016/j.ajp.2021.102977] [Medline: 34998232]
36. Yan M, Yin H, Meng Q, et al. A virtual supermarket program for the screening of mild cognitive impairment in older adults: diagnostic accuracy study. JMIR Serious Games 2021 Dec 3;9(4):e30919. [doi: 10.2196/30919] [Medline: 34870610]
37. Sabbagh M, Sadowsky C, Tousi B, et al. Effects of a combined transcranial magnetic stimulation (TMS) and cognitive training intervention in patients with Alzheimer's disease. Alzheimers Dement 2020 Apr;16(4):641-650. [doi: 10.1016/j.jalz.2019.08.197] [Medline: 31879235]
38. Zhang S, Qiu Q, Qian S, et al. Determining appropriate screening tools and cutoffs for cognitive impairment in the Chinese elderly. Front Psychiatry 2021;12:773281. [doi: 10.3389/fpsyt.2021.773281] [Medline: 34925100]
39. Zhang F, Bazarevsky V, Vakunov A, et al. MediaPipe hands: on-device real-time hand tracking. arXiv. Preprint posted online on Jun 18, 2020. [doi: 10.48550/arXiv.2006.10214]
40. Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: unified, real-time object detection. Presented at: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); Jun 27-30, 2016; Las Vegas, NV, USA. [doi: 10.1109/CVPR.2016.91]

Abbreviations

ADL: activities of daily living
AI: artificial intelligence
CDR: Clinical Dementia Rating
DSRS: Dementia Severity Rating Scale
DTx: digital therapeutics system
EFPT: Executive Function Performance Test
IADL: instrumental activities of daily living
MCI: mild cognitive impairment
MMSE: Mini-Mental State Examination
SUS: System Usability Scale

Edited by A Coristine; submitted 20.03.25; peer-reviewed by A Gazis, P Codyre; revised version received 07.07.25; accepted 09.07.25; published 06.10.25.

Please cite as:
Siow CY, Yang YH, Tsai CJ, Yang WW, Hu CJ, Sung JY, Tani J
Vision AI-Based Gamified Cognitive Prosthesis for Executive Function: Feasibility and Usability Study
JMIR Serious Games 2025;13:e74157
URL: https://games.jmir.org/2025/1/e74157
doi:10.2196/74157

© Co Yih Siow, Yao-Hua Yang, Cheng-Jui Tsai, Wan-Wan Yang, Chaur-Jong Hu, Jia-Ying Sung, Jowy Tani. Originally published in JMIR Serious Games (https://games.jmir.org), 6.10.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

A Resource-Efficient, High-Dose, Gamified Neurorehabilitation Program for Chronic Stroke at Home: Retrospective Real-World Analysis

Spencer A Arbuckle[1*], PhD; Anna Sophie Knill[2,3*], MSc; Michelle H Chan-Cortés[1], PhD; Gabriela Rozanski[4], MSc, PhD; Anastasia Elena Ford[1], MSc; Louis T Derungs[1], MAS; John W Krakauer[1,5,6,7,8], MA, MD; Naveed Ejaz[1], PhD; David Putrino[4], PhD; Jenna Tosto-Mancuso[4*], DPT; Meret Branscheidt[3,9,10*], MD

[1]MindMaze, Lausanne, Switzerland

[10]Department of Health Sciences and Technology, ETH Zurich, Zurich, Switzerland

[2]Rehabilitation Engineering Laboratory, ETH Zurich, Gloriastrasse 39, GLC, Zurich, Switzerland

[3]Lake Lucerne Institute, Vitznau, Switzerland

[4]Abilities Research Center, Icahn School of Medicine at Mount Sinai, New York, NY, United States

[5]Department of Neurology, Johns Hopkins University, Baltimore, MD, United States

[6]Department of Neuroscience, Johns Hopkins University, Baltimore, MD, United States

[7]Department of Physical Medicine and Rehabilitation, Johns Hopkins University, Baltimore, MD, United States

[8]The Santa Fe Institute, Santa Fe, NM, United States

[9]Center for Neurology and Rehabilitation, Cereneo, Vitznau, Switzerland

[*]these authors contributed equally

Corresponding Author:
Anna Sophie Knill, MSc
Rehabilitation Engineering Laboratory, ETH Zurich, Gloriastrasse 39, GLC, Zurich, Switzerland

Abstract

Background: Accumulating evidence and medical guidelines recommend high-dose neurorehabilitation for recovery after stroke. The reality, however, is that most patients receive a fraction of this dose, with therapist availability and costs of delivery being major implementational barriers.

Objective: This study aimed to explore a potential solution by conducting a retrospective analysis of a real-world enhanced clinical service that used gamified self-training technologies at home under remote therapist supervision.

Methods: Data from 17 patients who completed a 12 - 18 week full-body, high-dose neurorehabilitation program entirely at home were analyzed. Program delivery relied primarily on patients training independently (asynchronously) with the MindMotion GO (MindMaze) gamified-therapy solution. Accompanying telerehabilitation training sessions with a therapist occurred weekly while therapists used a web application to continuously monitor and manage the program remotely. Effectiveness of the program was assessed through measured active training time, a measure that more closely reflects delivered dosage as opposed to scheduled dose. Patient recovery was evaluated with standardized impairment and functional clinical measures and patient self-reported outcome measures. Finally, a cost model was computed to evaluate the resource efficiency of the program.

Results: Patients maintained high training adherence throughout the program and reached an average total active training time of 39.7 (SD 21.4) hours, with the majority delivered asynchronously (mean 82.2%, SD 10.8%). Patients improved in both upper-limb (Fugl-Meyer Upper Extremity, mean 6.4, SD 5.1; $P<.001$) and gait and balance measures (Functional Gait Assessment, mean 3.1, SD 2.6; $P<.001$; Berg Balance Scale, mean 6.1, SD 4.4; $P<.001$). Overall, the program was viewed very favorably among patients who completed a post-program survey, with 73.7% (14/19) of respondents being satisfied or very satisfied, while 63.2% (12/19) of respondents reported subjective improvements in physical abilities. Per-patient therapist costs approximated US $338, representing a resource-efficient alternative to delivering the same dose via one-on-one in-person training sessions (US $1903).

Conclusions: This work demonstrates effective high-dose neurorehabilitation delivery via gamified therapy technologies at home. The approach shows that training time can be successfully decoupled from therapist-presence without compromising adherence, outcomes, or patient satisfaction over an extended program period. Given growing concerns over therapist availability and increasing health care costs, this resource-efficient approach can help achieve medical guidelines and complement existing clinic-based approaches.

XSL•FO
RenderX

(JMIR Serious Games 2025;13:e69335) doi:10.2196/69335

KEYWORDS

chronic stroke; neurorehabilitation; telerehabilitation; high-dose therapy; real-world evidence; home-based therapy; gamified technologies; asynchronous training

Introduction

Stroke is a critical global health issue, leaving many individuals with enduring disability [1] and placing an immense economic burden on health care systems. Rehabilitation is pivotal in fostering recovery and restoring functional independence. However, growing evidence suggests that significantly improving clinical outcomes hinges on the provision of high-dose training. Studies indicate that providing at least 20 hours of training over 10 weeks on top of standard therapy is needed to make a clinically meaningful impact [2-6]. Therefore, the importance and clinical benefits of high-dose training are emphasized in existing and emerging medical guidelines for stroke rehabilitation. In the United States, the American Stroke Association guidelines recommend aerobic training of 60 - 300 minutes (over multiple sessions and days) per week in addition to strength and stretching exercises [7], while in the United Kingdom, the recent National Institute for Health and Care Excellence guidelines update recommends at least 3 hours of targeted rehabilitation on 5 days per week [8].

However, despite strong evidence supporting the benefits of high-dose rehabilitation, most patients with stroke fail to receive the recommended amount of therapy dose [6,9-12]. During acute inpatient care, patients were found to be active only 13% of their time and only 5.2% in contact with a therapist [13]. A follow-up review nearly a decade later suggested little had changed [14]. Furthermore, following discharge, patients complete less than 20 minutes of upper extremity training during approximately 45-minute sessions [15,16]. This issue becomes particularly pronounced in the chronic stage, where observational studies in outpatient settings have shown that overall training doses remain low, averaging just one purposeful leg or arm movement per minute [17,18].

Thus, there is a stark gap between the evidence supporting high-dose neurorehabilitation versus what is delivered in traditional standard-of-care. Key implementational barriers that contribute toward this gap are human resource constraints (eg, therapist cost and availability) and poor patient adherence and engagement [2,4,19,20]. Therapist availability in particular is only expected to worsen, with a recent national US survey indicating that 66.6% of therapists intend to leave the profession within the coming year [21].

To address these barriers, numerous studies have tried to dissociate therapist time and outpatient infrastructure from therapy itself, either by incorporating technologies, particularly robotics, into in-clinic training programs [22]; using home-based training with or without the aid of additional technologies [23]; incorporating a training model with a single therapist used to deliver care to multiple patients simultaneously [24,25]; or a combination of approaches [26]. However, none of these approaches have fully overcome implementational barriers: they often fail to free up therapist time, are costly, and have low patient adherence and engagement. To further complicate the matter, studies generally fail to demonstrate that scheduled training time equates to actual delivery of high-dose training—the vast majority do not report on actual therapy time delivered [27].

Among newer solutions, gamified rehabilitation technologies have emerged as promising tools for delivering structured, engaging, and measurable training. These systems typically use motion tracking, game-based tasks, and real-time performance feedback to support motor training and encourage adherence. In stroke rehabilitation, gamified technologies have been reported to produce outcomes comparable to those achieved through conventional therapy [11,28]. A promising feature of these systems is that they offer additional flexibility in how and where training is delivered. Their compatibility with remote care models means that patients can engage in therapy at home, either self-directed according to their own schedule (asynchronously) or under remote therapist supervision (synchronous telerehabilitation), offering a potential path to overcoming key implementational barriers to high-dose rehabilitation. Indeed, a recent clinical trial demonstrated that home-based telerehabilitation delivered via gamified technologies was as effective as in-clinic therapy for patients with stroke; however, its remote implementation still relied on therapist supervision for half of the total scheduled training time [29]. Whether patients can deliver the majority of training independently, using gamified technologies in a minimally supervised asynchronous format, remains relatively unexplored. Such an approach could provide a more scalable pathway to deliver high-dose rehabilitation that aligns with the growing constraints on therapist time and system resources. Demonstrating the feasibility of patients achieving the bulk of their training dose independently is a critical next step.

Thus, in this real-world enhanced clinical service and retrospective analysis, we investigated the extent to which implementational barriers related to delivering high-dose neurorehabilitation can be addressed with a technological approach. Our primary objective was to determine whether gamified training and monitoring technologies can be used to successfully deliver high-dose neurorehabilitation for chronic stroke survivors at home. As part of this aim, we report on therapy adherence and clinical changes experienced by patients. By setting a soft target that 75% of the training dose was to be delivered through patient self-training (asynchronous) with the remaining 25% relying on synchronous telerehabilitation with the therapist, we attempted to dissociate therapeutic dose delivery from therapist availability.

Training was delivered with the aid of a gamified training technology (MindMotion GO) that enabled full-body training of the upper limbs, hand, trunk, and lower limbs. Monitoring components of the gamified technology system were used to

XSL•FO
RenderX

track actual training times and to ensure that therapists were able to consistently oversee patient progress and uphold the quality and success of the training. We performed separate evaluations of patient adherence, clinical effectiveness of training, patient satisfaction, and overall resource and cost efficiency of the program.

Methods

Patients

Patients with stroke were identified from the Mount Sinai Abilities Research Center Clinical Program registry. Patients were informed of the enhanced clinical service and offered the option to participate in high-dose neurorehabilitation, which was offered from October 2022 to August 2023. The program enrolled patients who were ≥22 years of age; had a first stroke episode ≥3 months prior; presented with either arm weakness and/or difficulty with balance; were able to follow multistep commands; and showed willingness and ability to commit to the program length and the weekly training dose accompanied with telerehabilitation sessions. Patients inappropriate for the program included those with plegia of the affected upper limb, severe pain, or severe cognitive or physical challenges that made it difficult for them to participate in neurorehabilitation.

A total of 26 patients with chronic stroke (defined as having experienced their first stroke ≥6 mo prior) were enrolled in the program. Of these, 23 began training (Figure S1 in Multimedia Appendix 1), but 6 either dropped out or were lost to follow-up. Specifically, 1 patient completed the training but declined to attend the second set of clinical evaluations; 3 experienced unrelated medical changes that prevented further participation; and 2 voluntarily withdrew from the program. Data from the 17 patients (8 women; mean age of 54.8 y, SD 14.1 y; time from first stroke episode: mean 5.4 y, SD 4.7 y; Table 1) who completed the training program and underwent both sets of clinical assessments were analyzed retrospectively. As this was a real-world clinical implementation, the sample size for analysis was limited by patient availability, clinical capacity to provide care, and the availability of clinical service data. There were no significant differences regarding chronicity, age, or impairment (tested with the baseline Fugl-Meyer Upper Extremity and Berg Balance Scale measures) between patients who completed the training program and those who did not, and no adverse events were reported during the program.

Table . Baseline characteristics for analyzed patients.

Characteristics	Values (N=17)
Age (years), mean (SD)	54.8 (14.1)
Years post stroke, mean (SD)	5.4 (4.7)
Sex, n (%)	
Male	9 (52.9%)
Female	8 (47.1%)
Stroke type, n (%)	
Ischemic	9 (52.9%)
Hemorrhagic	8 (47.1%)
Affected hemisphere, n (%)	
Left	6 (35.3%)
Right	11 (64.7%)

High-Dose Telerehabilitation Training Program

All patients were assigned to a single-arm treatment group involving a high-dose, home-based training program (Figure 1A–B). For the purposes of this work, we consider a high-dose program to be one that achieves a total dosage above 20 hours scheduled time, as these doses are generally considered to deliver outcomes above the minimum clinically important difference [2,5,29]. Overall, the program goal was to deliver 36 hours of training during the program. The target dose was consistent with a previous large telerehabilitation study that showed significant improvements in upper-limb impairment (42 h [29]). In contrast, training in this program was not limited to one body area [30-32] and could be achieved with multiple effectors (ie, upper-limb, hand, trunk, and lower-limb).

Training dose was delivered via synchronous telerehabilitation and asynchronous training (Table 2). Patients were given the freedom to achieve the 36 hours of training dose over a period of 12-18 weeks and were given a high-level goal to train for at least 120 minutes per week. Ultimately, the exact program duration for each patient was dependent on their availability and desire to continue. Consistent with in-person rehabilitation practice, patients could take short breaks away from the program (these breaks were incorporated in the analysis; see section “Binning Training Data into Weeks” in the Methods). Patients were also free to continue additional regular rehabilitation services.

XSL•FO
RenderX

Figure 1. Design and technologies of the home-based training program. Patients were enrolled into a home-based training program that consisted of self-directed (asynchronous) training sessions performed by the patient each week, complemented by multiple weekly sessions directly supervised by a physical therapist (synchronous telerehabilitation). (**A**) Prior to the program start, patients were evaluated in-clinic by a physical therapist. The therapist would then create a training program for the evaluated patient. Patients set up the training system (MindMotion GO and Izar) at home. (**B**) During the training program, synchronous telerehabilitation sessions were scheduled as per therapist and patient availability, whereas asynchronous training (performed by the patient) or remote monitoring (conducted by the therapist) could be completed at any time. Patients additionally underwent an in-clinic assessment battery prior to the program starting and upon program completion to evaluate the clinical efficacy of the training program (not shown). (**C**) The MindMotion GO telerehabilitation device uses camera-based motion tracking to control gamified training activities displayed on a television screen. (**D**) The Izar, a therapeutic device to train hand function, being grasped in a hand.

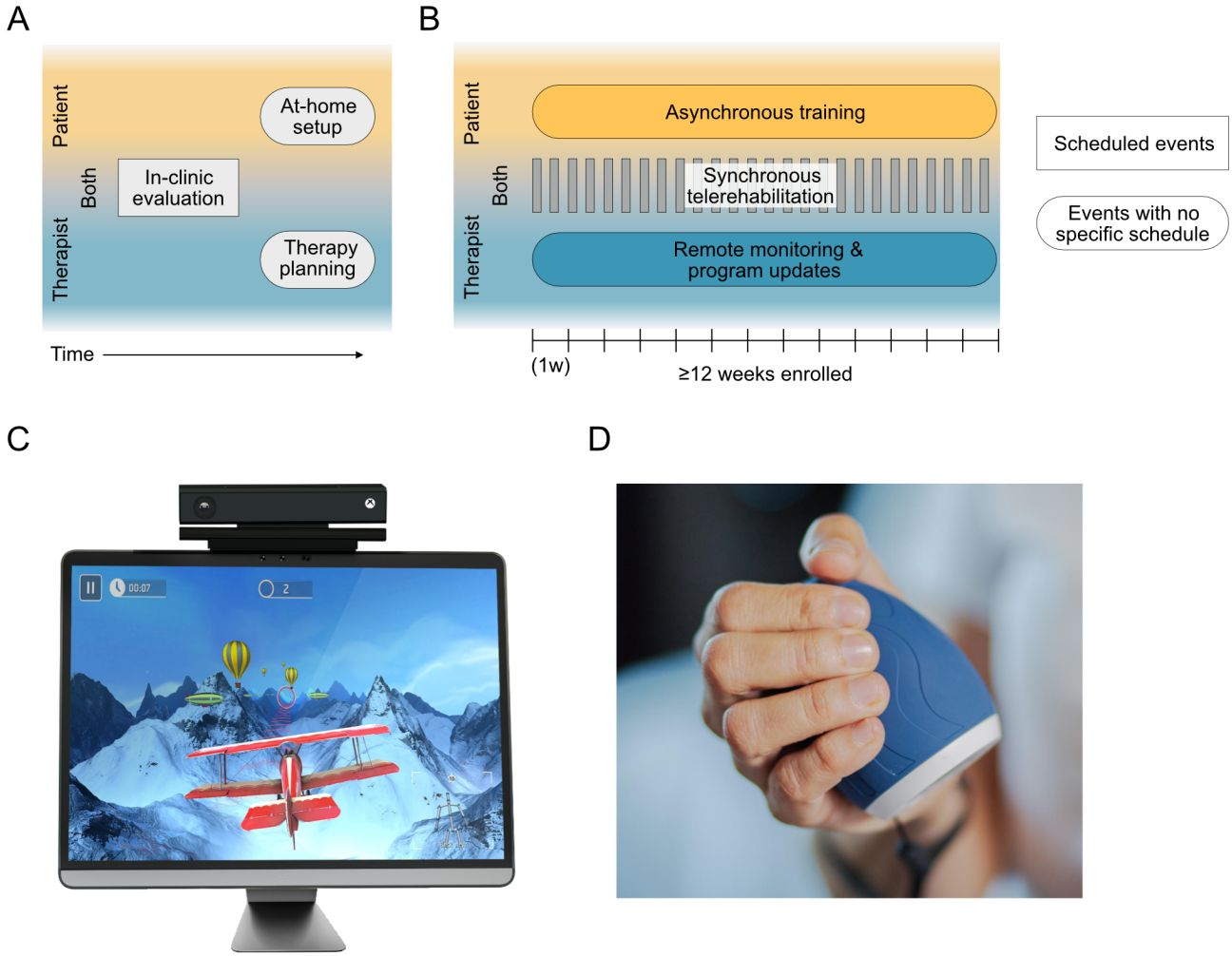


Table . Suggested weekly schedule.

Training type	Planned session duration (min)	Planned sessions per week (#)	Expected training per session (min)	Expected training per week (min)
Asynchronous	at patient discretion	at patient discretion	at patient discretion	≥90
Synchronous	30	1 - 3	15	30 - 45
Total				≥120

Synchronous Sessions

These telerehabilitation sessions were guided directly by a physical therapist who was present via video conferencing. The therapist was free to schedule 30-minute synchronous sessions with a patient at a frequency determined by their clinical judgment. The content of the sessions included gamified training using the MindMotion GO system (see section “Full-Body Gamified System for Asynchronous Training” in the Methods) under visual observation of the therapist and check-in conversations between the patient and the therapist to discuss training goals, training adjustments, feedback, etc.

Asynchronous Sessions

These were self-directed training sessions where patients performed gamified training in a remote environment (ie, at home). In contrast to synchronous training, asynchronous sessions did not include a therapist present. Patients had the ability to independently choose when and for how long to train asynchronously, and the ability to skip or repeat prescribed

XSL•FO
RenderX

activities. At least 75% of a patient's total dose was expected to be delivered asynchronously.

Asynchronous training content for each patient was determined by their supervising therapist, who tailored their training program by selecting gamified activities on the MindMotion GO based on the patient's disease severity, therapeutic goals, preferences, and their physical capacity, with the range of motion calibrations defining the playable range of each activity. This calibration ensured that the gameplay was within the patient's safety envelope (to minimize pain and adverse events such as risk of falls), and was updated as needed. The gamified activities targeted one or more areas of the following: hand, upper limbs, trunk, and/or lower limbs.

Full-Body Gamified System for Asynchronous Training

A commercially available technology (MindMotion GO; Food and Drug Administration Class II) was used to deliver full-body gamified training. The MindMotion GO is a real-time, motion capture–based neurorehabilitation therapy technology (Figure 1C) that (1) enabled therapists to set up and track training, (2) enabled patients to perform asynchronous training sessions, (3) provided teleconsultation capabilities for synchronous training and check-ins, and (4) captured performance measures during gameplay that allowed for the quantification of the dose and nature of training delivered to each patient. At the time of this enhanced clinical service, the MindMotion GO catalog included 36 gamified activities designed to train specific movements relevant to neurorehabilitative therapy (see Table S1 in Multimedia Appendix 1).

Patients' body and limb kinematics were captured by accompanying optical hardware sensors (Kinect and LeapMotion), which were used to drive the mechanics of the gamified activities. An additional peripheral device (Izar, MindMaze; Figure 1D) was used to capture fine-grained forces generated during hand and wrist control and use them for gameplay.

A range of metrics related to a patient's training performance was captured in real-time and uploaded to a web platform (Companion, MindMaze). Therapists used this web platform to remotely set up or adjust therapy plans and to track training progress.

Equipment was shipped to a patient's home, where it was connected to a television. All but one patient successfully installed the system without requiring in-person assistance from the therapy team. Customer and technical support (via telephone or video-call) was provided for the duration of the program.

Quantification of Dose and Nature of Training

We analyzed training data from the MindMotion GO device to assess program feasibility and patient behavior, including training time, duration, and movements trained.

Quantifying Dose Delivered

Active training time (ATT), defined as the time spent engaging in gamified activities (excluding setup, calibration, or pauses), was used as a measure of training dose delivered.

Binning Training Data Into Weeks

Our primary focus was on training program dosage (ie, hours of ATT) rather than program duration (ie, weeks in the program). However, to investigate training behavior over time, training data were organized into "training weeks", defined as consecutive 7-day periods starting on the patient's program start date and ending at their exit assessment. Training weeks could vary across patients and if a patient missed training in a week (ie, zero ATT), it was excluded (5.8% of all total 346 enrolled weeks). For clarity, we use "trained weeks" in figures in the Results section, acknowledging that these may not reflect consecutive weeks.

Assessing Training Consistency

We examined whether patients' asynchronous training behavior changed over time using ordinary least squares regression. A positive or negative beta coefficient indicated whether the proportion of ATT trained asynchronously tended to increase or decrease across weeks, respectively. The magnitude of a coefficient represents the weekly rate of change (eg, a coefficient of 0.001 is +1% increase per week). The beta coefficient for each patient is presented in the Results section.

Analyzing Training Schedules

We tested if patients had idiosyncratic but nevertheless reliable training schedules. To assess patients' training schedules, we calculated the distribution of their total asynchronous ATT across each hour of each day of the calendar week. For each patient, data from even- and odd-numbered training weeks were binned separately. We then computed Pearson correlation between training schedules for even- and odd-numbered weeks within patients to measure schedule reliability. To quantify the uniqueness of each patient's schedule across the cohort, we computed the average Pearson correlations between their even- and odd-week schedules and those of all other patients' schedules. The within-patient and across-patient correlations are presented in the "Results" section.

Capturing Clinical Outcomes and Patient Subjective Experiences

We quantified the effect of training on patients' impairment and function. As training was delivered full-body, we captured a wide range of functional outcome measures and patient reported outcome measures (PROMs) to quantify training-related changes across multiple effectors. A comprehensive list of all outcome measures is available in Multimedia Appendix 1 (Table S2 and Table S5).

Clinical outcome measures and PROMs were captured at the beginning and the end of the program. The clinical outcome measures were assessed in person at the Abilities Research Center by a physical therapist who was not their supervising therapist to retain blindness. PROMs were completed by patients through web surveys captured via Research Electronic Data Capture (REDCap) tool [33,34]. A custom survey was administered to all enrolled patients at the end of the program, regardless of training completion, to gather information on their experiences, satisfaction, motivation, and perceived improvements.

XSL•FO
RenderX

Cost Analysis

We conducted a simplified cost analysis, comparing physical therapist resourcing costs for two scenarios: (1) this program, with training delivered via a combination of asynchronous and synchronous sessions and (2) a scenario where the entire dose is delivered through direct supervision. The median hourly salary for a physical therapist was sourced from the US Bureau of Labor Statistics 2023 [35], which excludes fringe benefits (eg, medical insurance). Therapist costs for both scenarios were calculated by multiplying the total training hours by the median hourly salary (in USD), adjusted for the percentage of training delivered under supervision.

Statistical Analysis

All analyses and statistical tests were performed using R (R Core Team, version 4.3.1 [36-39]). Values presented in the results are the average across patients (plus or minus the SD). For Pearson correlation values, we first Fisher z-transformed the correlations and calculated the mean and SD. The mean and SD were then transformed back to correlations and reported according to convention. To compare if the proportion of asynchronous ATT systematically changed for all patients across the program duration, we used a 2-sided *t* test to compare the beta coefficients to 0 (see section "Assessing Training Consistency" in the Methods). To compare if patients had idiosyncratic asynchronous training schedules, we used a 2-sided paired *t* test to compare the Fisher z-transformed correlations from within and across-patient comparisons of their training schedules (see section "Analyzing Training Schedules" in the Methods). We set the significance level to P=.05 for these statistical tests.

For the analysis of clinical improvements, changes from baseline to program end were tested using paired 2-sided *t* tests. As these tests involved multiple comparisons without any specified a priori hypotheses, we adjusted the significance level by dividing by the number of clinical outcomes examined, which was 12 (ie, Bonferroni correction). For clarity, we report uncorrected P values in the text. Note that not all 17 patients were evaluated on all assessments, and therefore, the degrees of freedom varied across the clinical outcomes.

Results of the exit survey are reported as a proportion of total survey respondents, which differs from patients who completed the program. We chose to present the exit survey results for all respondents to ensure all views were included, irrespective of whether a patient officially completed the program. A total of 19 patients completed the exit survey, of whom 16 completed the program and 3 did not (2 dropped out and 1 was lost to follow-up).

Ethical Considerations

All data were collected as a part of routine clinical care at the Abilities Research Center at the Icahn School of Medicine at Mount Sinai. Institutional review board approval for retrospective analysis of clinical data was obtained from the Mount Sinai Program for Protection of Human Subjects (Icahn School of Medicine at Mount Sinai, New York, United States; STUDY-21 - 00345). This approval includes a consent waiver, allowing for the retrospective analysis of the data for purposes beyond routine clinical care without requiring additional patient consent. Patients received no monetary compensation. Prior to analysis, all data were deidentified by the clinical staff at the Abilities Research Center.

Results

Long-Term, High-Dose Training Can Be Achieved for Patients With Chronic Stroke in the Home

A retrospective analysis was carried out on 17 patients with chronic stroke who completed high-dose neurorehabilitation. Overall, the goal was for patients to receive a minimum of 36 hours of ATT (see section "High-Dose Telerehabilitation Training Program" in the Methods), delivered entirely at home and spread across all major motor functions (ie, upper limbs, hand, trunk, and lower limbs). Training was delivered using a gamified therapy solution (MindMotion GO with Izar), through a combination of the patients following prescribed training plans on their own (self-directed asynchronous training) supplemented by weekly training sessions with a therapist being present via video conferencing (therapist-directed synchronous telerehabilitation). Therapists were able to remotely define and alter training plans and maintain oversight of the program using a web-based application for patient monitoring (Companion).

Training duration and consistency were high. Patients trained an average of 120.51 (SD 52.47) minutes per week for 19.18 (SD 3.24) weeks (Figure 2A), with all patients exceeding the intended minimum duration of 12 weeks. Overall, patients achieved an average cumulative training dose of 39.67 (SD 21.37) hours, slightly surpassing the program goal of 2 hours of ATT per week (Figure 2B). This cumulative training dose was full body, with a slight tendency toward upper limb training (mean 43.20%, SD 9.06% of total ATT) compared to the lower limb (mean 28.40%, SD 8.27%), hand (mean 15.40%, SD 14.09%) and trunk (mean 13.01% SD 4.25%; see Figure 2C). For a breakdown of dosage per each of the 27 possible movements that could be trained, see Figure S2 in Multimedia Appendix 1. Altogether, these results demonstrate the feasibility of hybrid asynchronous and synchronous programs to deliver actual high-dose training distributed across the full body.

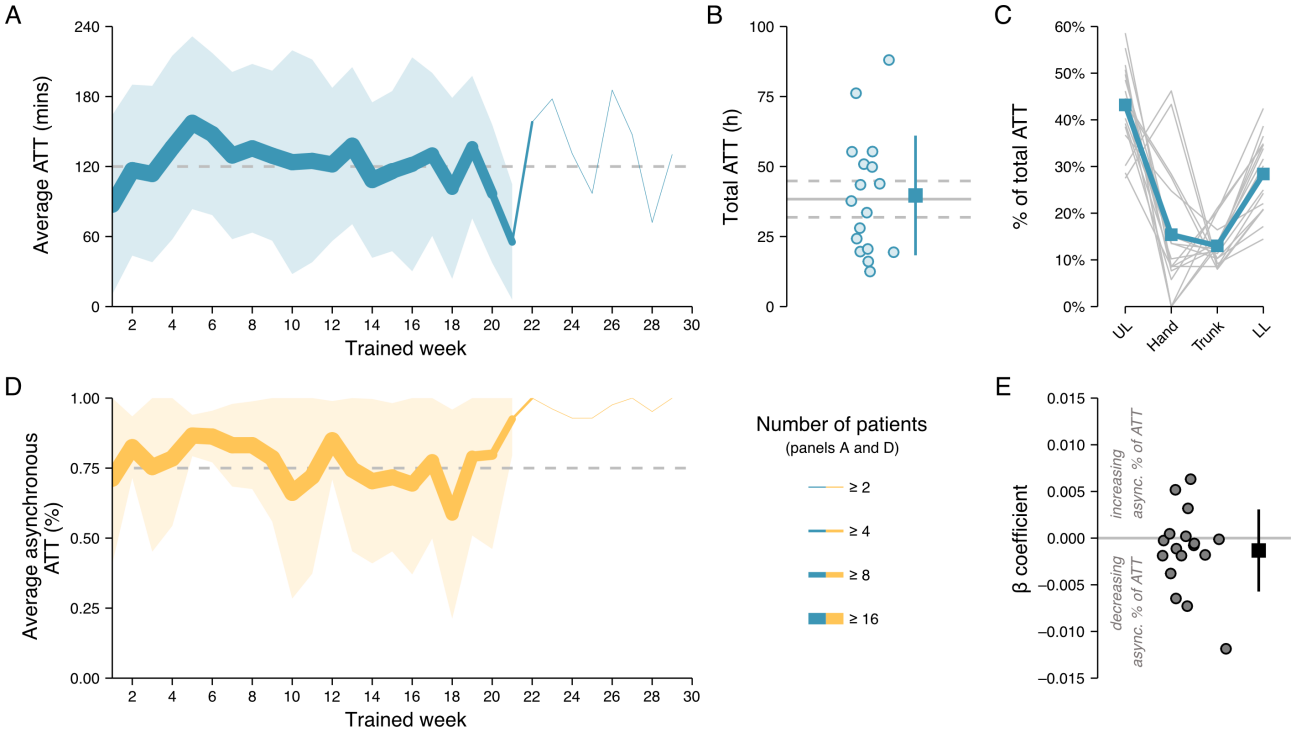
It is important to note that the ATT dose measure used in this analysis is a closer representation of the dose actually delivered in an intervention, as compared to the scheduled dose that is almost exclusively reported in other high-dose neurorehabilitation studies [5,10,29,40,41]. Furthermore, this dose was delivered in addition to any other rehabilitation received by patients. Indeed, 12 out of the 17 patients also reported receiving care outside the program (eg, physiotherapy, occupational therapy, or general exercise), highlighting the program's complementarity to standard-of-care neurorehabilitation. We did not observe any significant differences in program adherence measures between patients who did versus did not report participating in additional therapies (Table S3 in Multimedia Appendix 1).

Despite the extended program duration of approximately 19 weeks, patients reported high levels of satisfaction. In an exit survey, 73.7% of respondents (14/19) indicated they were either

XSL•FO
RenderX

very satisfied or satisfied with the program (Figure S3A in Multimedia Appendix 1). In addition, patients expressed strong motivation to continue training beyond the program length, with 57.9% (11/19) reporting they were very motivated or motivated to continue (Figure S3B in Multimedia Appendix 1). This was particularly apparent in one patient who trained for 29 weeks, far surpassing the official program length.

Figure 2. Training dose. (A) Active training time (ATT) per week averaged across patients. The thickness of the line represents the number of patients who remained in the program by that training week. The shaded area indicates the SD above and below the average (for weeks with n>3), and the dashed line indicates the target weekly dosage of 120 minutes. (B) The total ATT dosage achieved at program end. Each dot corresponds to one patient. The box indicates the average and the vertical line indicates the SD above and below the average. The solid horizontal line indicates the total expected dosage if each patient achieved 120 minutes ATT per week they were training in the program, with the dashed horizontal lines reflecting the SD around the average expected dose (patients completed a varying number of training weeks in the program). (C) Proportion of total ATT that patients spent training their upper limbs (UL), hands, trunk, or lower limbs (LL). Individual patient data are depicted as thin lines, with the average across patients shown as a thicker line. (D) The proportion of weekly ATT that was achieved asynchronously, averaged across patients. Similar formatting as in (A). (E) Quantifying patients' tendency to increase, decrease, or maintain their ratio of asynchronous-to-synchronous ATT across training weeks (see Methods). On average, patients maintained a steady proportion of asynchronously achieved ATT across weeks.



Majority of Program Dose Was Delivered Without the Presence of a Therapist

Therapists scheduled synchronous telerehabilitation sessions with patients as per their clinical discretion to ensure compliance with the overall program. However, the goal was for patients to achieve much of their training dose asynchronously. On average, each patient achieved 34.18 (SD 21.72) hours of training asynchronously, representing 82.19% (SD 10.81%) of their total ATT (Figure 2D). This constituted the majority of their overall dose and exceeded the 75% target.

Patients continued to complete most of their training asynchronously as the program progressed (Figure 2D), with no systematic change in the ratio between synchronous and asynchronous training dose from week to week (Figure 2E; mean β^=−1.3e−3, SD 4.4e-3, 2-sided *t* test versus 0: t_{16}=−1.24, *P*=.23).

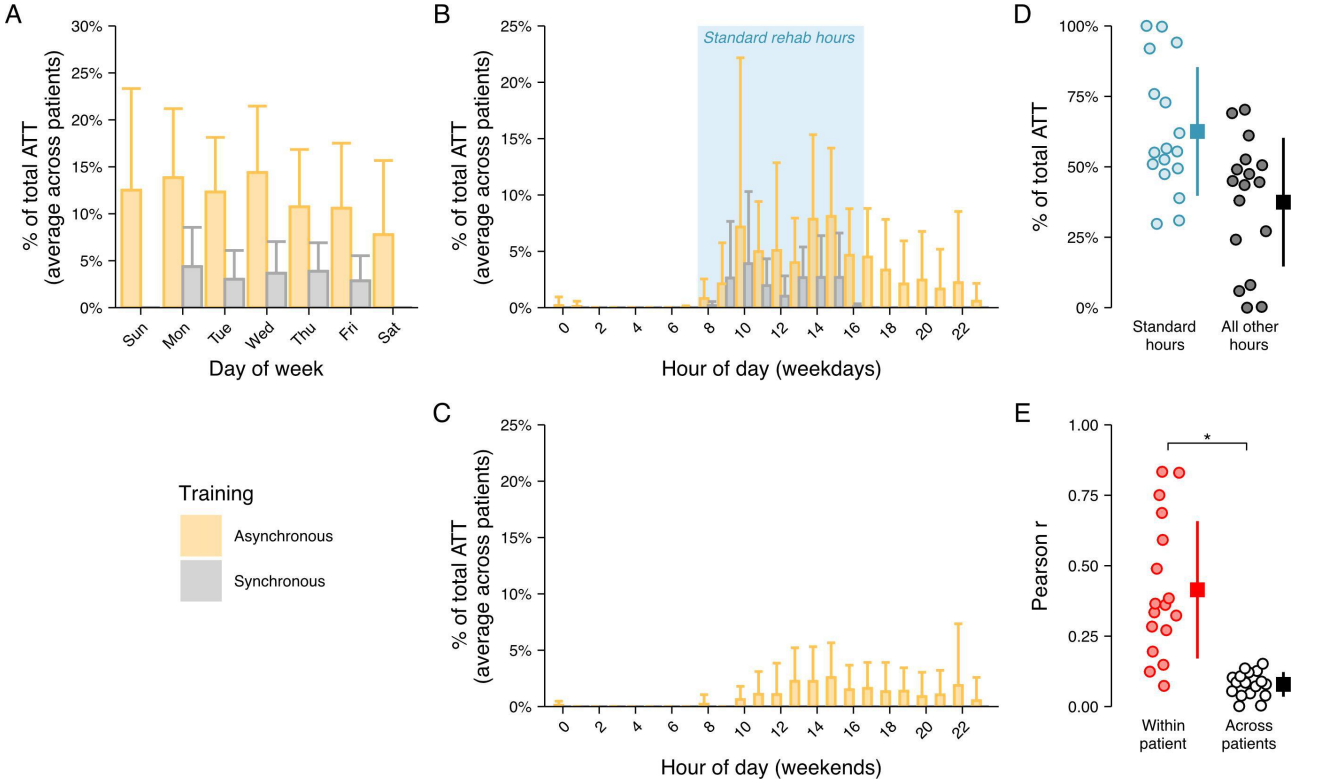
All synchronous sessions were delivered during regular 8 AM-5 PM (Monday-Friday) working hours Figure 3A-C In contrast, asynchronous training offered patients the flexibility of training at any time, decoupled from restrictions of therapist availability. Indeed, a substantial proportion of each patient's total ATT (mean 37.45%, SD 22.84%; mean 15.70, SD 12.29 h) was logged outside of working hours and on weekends (Figure 3D).

Patients' own schedules imposed restrictions on when they were able to train. Interestingly, while the majority of patients remained relatively consistent in their overall training schedules from week to week (mean within-patient Pearson *r*=0.46, SD 0.34, see Figure 3E), the training schedules were highly individual (mean across-patients Pearson *r*=0.08, SD 0.04, see Figure 3E; 2-sided paired *t* test versus within-patient schedule reliability: t_{16}=−4.74, *P*<.001).

Taken together, these results demonstrate that patients remained highly adherent to the home-based training program even if training was largely delivered asynchronously.

XSL•FO
RenderX

Figure 3. Training distribution across days of the week and times of the day. (A) Bar chart depicting the proportion of total ATT (averaged across patients) binned according to the day of the week. Synchronous training times were only logged during the weekdays. Error bars reflect the SD across patients. (B-C) Bar charts depicting the proportion of total ATT (averaged across patients) binned according to the starting hour of the training session for training done on weekdays (B) and weekends (C). Similar formatting as in A. (D) Percent of total ATT that occurred during standard working hours (ie, when the therapist was available for synchronous telerehabilitation sessions; box in B) versus that which occurred during all other hours. Each dot corresponds to one patient. The boxes indicate the average across patients and the vertical lines indicate the SDs above and below the averages. (E) Quantifying the uniqueness of patients' average weekly asynchronous training profiles (see "Methods" section). Each dot in the within-patient grouping reflects a patient's consistency of their weekly training profile, and each dot in the across patients grouping reflects the average correlation between that patient's training profile and the training profiles of all other patients. The boxes and vertical lines are formatted the same as in D. *denotes significant difference between within-patient and across patient-correlations (2-sided paired t test, $P<.01$).



Patients Showed Significant Improvements in Gait, Balance, and Upper-Limb Function

Next, we investigated the effect of the high-dose training on patients' recovery and self-reported well-being. As the delivered training was full-body, we captured a wide range of standardized upper limb, gait and balance, and physiological measures to quantify training-related changes in impairment and function across multiple effectors (see section "Capturing Clinical Outcomes and Patient Subjective Experiences" in the Methods). All measures were obtained at the beginning and end of the training program.

Patients demonstrated positive improvements in three core impairment and functional measures for multiple effectors (see Table 3), namely the Fugl-Meyer Upper Extremity assessment for upper limb impairment, (FM-UE; mean +6.41, SD 5.09, n=17, $P<.001$), the Berg Balance Scale for generalized balance and transfer stability (BBS; mean +6.07, SD 4.43, n=15, $P<.001$), and the Functional Gait Assessment (FGA; mean +3.07, SD 2.55, n=15, $P<.001$) for postural stability and walking. Among these, both the FM-UE and BBS were above the Minimal Detectable Change (MDC) for chronic stroke (MDC=3.2 and 4.66 points, respectively [42,43]), whereas the FGA was not (MDC=4.2 points [44]). These clinical improvements occurred irrespective of whether patients participated in additional therapies (see Table S4 in Multimedia Appendix 1). Although there were positive improvements across most other captured measures, they did not cross thresholds for either statistical significance or minimum detectable change (see Table S2 in Multimedia Appendix 1). These positive improvements in standard clinical measures were complemented by 63.2% of respondents (12/19) self-reporting subjective improvements in their physical abilities and well-being (Figure S3C in Multimedia Appendix 1). Additional patient-reported outcome measures are reported in Table S5 in Multimedia Appendix 1.

Overall, these results demonstrate that patients receiving a high-dose training program at home showed significant improvements in outcome measures related to full-body impairment and function, with these improvements reinforced by patients' self-reported measures.

XSL•FO
RenderX

Table . Core clinical outcomes.

Clinical measure[a]	Number of patients evaluated	Potential range	Evaluation before program start		Evaluation at program end		Δ score (end-baseline scores)		
			mean (SD)	Range	mean (SD)	Range	mean (SD)	*t* (df)[b]	*P* value[c]
FM-UE[d]	17	0 - 66	31.71 (21.22)	8 - 65	38.12 (20.03)	14 - 66	6.41 (5.09)	5.20 (16)	8.8×10^{-5}
ARAT[e]	17	0 - 57	19.00 (21.45)	0 - 57	21.53 (23.29)	1 - 57	2.53 (4.20)	2.48 (16)	2.5×10^{-2}
BBS[f]	15	0 - 56	38.67 (11.18)	14 - 55	44.73 (9.15)	20 - 56	6.07 (4.43)	5.30 (14)	1.1×10^{-4}
FGA[g]	15	0 - 33	9.87 (6.03)	1 - 23	12.93 (6.20)	1 - 28	3.07 (2.55)	4.66 (14)	3.7×10^{-4}
5xStS[h]	15	>0 s	22.00 (10.89) s	8 - 45	17.17 (7.21) s	8 - 34	−4.83 (6.25) s	−3.00 (14)	9.6×10^{-3}
6minWT[i]	14	≥0 ft	579.64 (324.83) ft	120 - 1175	635.00 (371.46) ft	130 - 1420	55.36 (86.77) ft	2.39 (13)	3.3×10^{-2}

[a]Scores of the complete set of 12 clinical outcomes are presented in Table S2 in Multimedia Appendix 1.

[b]The *t* values and associated significance for two-sided paired *t* tests comparing scores at program end to baseline.

[c]Statistical significance is reported as uncorrected values. A Bonferroni correction for 12 comparisons was applied when evaluating significance.

[d]FM-UE: Fugl-Meyer Upper Extremities

[e]ARAT: Action Research Arm Test.

[f]BBS: Berg Balance Scale.

[g]FGA: Functional Gait Assessment.

[h]5xStS: 5 times Stand to Sit.

[i]6minWT: 6-minute Walk Test.

Asynchronous Training Significantly Reduces Costs Associated With Delivery of High-Dose Training

Finally, we computed the personnel costs associated with the delivery of this high-dose training. Based on the current analysis, 39.7 (SD 21.4) hours of cumulative training dose was delivered to each patient on average, with the majority being delivered asynchronously (82.2%, SD 10.8%). Therefore, direct therapist presence was required to deliver only 17.8% of the training dose, and thus the total therapist costs per patient would amount to US $338 (given the median hourly salary for a physical therapist as US $47.94 [35]).

In contrast, the delivery of regular therapy hours relies on direct therapist presence either through in-clinic visits or using teleconferencing tools [45]. Therefore, under a scenario where all training is delivered directly, the costs of delivering 39.7 hours would amount to US $1903—a substantial increase of US $1565 when compared to the current hybrid asynchronous and synchronous program.

Discussion

Principal Findings

While the evidence for high-dose training in stroke rehabilitation is accumulating, the real challenge is in its implementation: how to deliver high-dose training effectively and in a resource-efficient way? Here we demonstrate that a gamified neurorehabilitation program can be used to successfully deliver high-dose full-body training to patients with chronic stroke at home. Using a combination of therapist-directed telerehabilitation sessions (synchronous) and patients' training on their own (asynchronous training), we were able to achieve resource-efficient dose delivery—the program required only a fraction of the therapist's time compared to traditional one-to-one in-person neurorehabilitation sessions. To our knowledge, this work is the first to demonstrate a significant decoupling of active training dose delivery from therapist presence.

Despite this decoupling, patients in the program showed clinical improvements. Unlike most previous approaches, patients showed positive improvements in all measures related to gait, balance, and upper-limb function, thus addressing that stroke typically results in multilimb deficits. These improvements were noticeable by patients, with 63.2% (12/19) reporting higher physical abilities and overall well-being.

The Role of Gamified Technology to Deliver High-Dose Training at Home

A critical issue in most neurorehabilitation studies is the lack of transparent reporting between scheduled therapy and what is truly delivered. A recent review identified the scale of the problem, with approximately only 13% of studies reporting on the actual training dose received [27]. When regular therapy is quantified, the actual time on task is a fraction of that scheduled, and is significantly lower than training doses required for recovery [46].

Using technologies has two advantages here: (1) namely as a means of sensitively and accurately quantifying the amount of

XSL•FO
RenderX

training dose delivered, while (2) also providing a set of training activities that homogenizes the nature of training received by the patient. Such technologies therefore respond to a growing demand for better documentation and transparency of dose delivery in neurorehabilitation [47].

A key element of this program was the significant use of asynchronous training delivery to achieve a higher dose. Our core mitigation strategy against reductions in patient engagement and adherence [48] was to ensure that the therapist was always in the loop, with the synchronous telerehabilitation sessions serving as a means to review and update the plan of care, but also to ensure patient engagement. In addition, gamified activities, a wide variety of exercises, and user-friendly interfaces were identified as important components [49] and were implemented in the technology used. Our results demonstrate the success of these mitigation strategies, with patients able to self-administer over 80% of the total training dose and maintain consistent engagement.

That patients remained engaged in the program suggests that they could realize a principal advantage of training asynchronously: convenience. Patients were able to train anytime without being restricted by therapist availability. Indeed, here we found that when patients were given a high-level training goal, supplied with tailored training programs delivered through gamified technologies, and provided with therapist check-ins, they used both weekdays and weekends to train which is not possible with conventional in-person neurorehabilitation. Furthermore, patients tended to train with highly idiosyncratic schedules that were most likely driven by unique work, life, and social commitments.

Comparing Approaches to Deliver High-Dose Training

In previous studies, a variety of different approaches for delivering high-dose training have been investigated, but few are truly scalable in a resource-constrained health care landscape. High-dose approaches that require one-on-one in-clinic therapy sessions are prohibitively expensive, are often limited to urban facilities, and struggle to scale (ie, treat more patients) without significant staffing [5]. Offering one-to-many training can mitigate the staffing and cost concerns but often imposes restrictions on patient scheduling. Supplementing regular therapy with robotic systems imposes high equipment and maintenance costs [50], and the complexity of these devices constrains them to supervised use in-clinic [51].

In comparison, while home-based programs are an attractive alternative, they are not all equal. Programs that rely solely on passive exercise videos may be inexpensive, but they often struggle with poor therapy adherence rates and a limited ability to customize care plans. And while traditional telerehabilitation benefits from direct therapist and patient contact with respect to adherence, it is again dependent on therapist availability [48].

What remains attractive and potentially scalable, however, are home-based programs that blend primarily asynchronous training with synchronous sessions. Such hybrid programs can be delivered through various commercially available gamified training systems, which integrate motion tracking, performance feedback, and game-based physical exercises to enhance motivation and adherence. These systems have the potential to address geographic, economic, and engagement-related implementational barriers to delivering high-dose training. The MindMotion GO, used in this study, exemplifies this class of technologies. It has previously been implemented in supervised, in-clinic settings for individuals with subacute stroke and multiple sclerosis [52], as well as in outpatient telerehabilitation contexts [53], though detailed clinical outcomes in these cases have not been reported. While these earlier implementations demonstrate usability, they do not clarify whether such systems can enable effective, asynchronous high-dose training in real-world, at-home settings–particularly for individuals with chronic stroke. This study addresses this gap by retrospectively evaluating a hybrid telerehabilitation program in which patients completed the majority of their own training asynchronously using a gamified system. Hybrid programs offer a promising solution to many of the known challenges associated with traditional high-dose rehabilitation delivery, including therapist availability and patient scheduling flexibility.

While recent studies have demonstrated the efficacy of providing high-dose training at home [29,30], the full value of these programs was not realized; there were significant requirements on therapist availability (50:50 synchronous vs asynchronous split) and no accompanying cost analysis. The work here extends these results and demonstrates that technology-enabled asynchronous training allows for high training doses to be delivered with limited demands on therapist availability (82% was delivered asynchronously). Despite reduced therapist presence in the program, patient engagement and satisfaction remained high. We also report costs associated with program delivery, which represent a fraction of the resource costs associated with delivering equivalent high-dose training through scaling up one-to-one neurorehabilitation.

Integrating Asynchronous Training Into Standard-of-Care

For home-based neurorehabilitation programs to gain widespread adoption, it is critical that they integrate into and complement current neurorehabilitation workflows. Here we offer two pragmatic suggestions. For patients that routinely receive in-person outpatient rehabilitation, part of the in-person sessions can be used to review, discuss, and update the plan-of-care for remote asynchronous training. In contrast, patients in more rural settings, who may have limited access to in-person care, could reduce in-person visits with synchronous telerehabilitation sessions, again complemented by training asynchronously. Indeed, following the COVID-19 pandemic, synchronous telerehabilitation in the United States remains a billable service due to public health emergency extensions [54], while new current procedural billing codes have been established by the Center of Medicare & Medicaid Services to provide reimbursement pathways for technologies that enable asynchronous neurorehabilitation services [55,56].

Challenges of Clinical Implementation

Despite the attractiveness of a high-dose neurorehabilitation at home, a few caveats should be considered. In such programs, there is a tangible concern that the importance of therapists is reduced. We unequivocally disagree with this notion. Therapists

XSL•FO
RenderX

are, and will remain, essential for building collaborative relationships with patients that are critical for engagement and recovery [57]. To this end, it is essential that the technologies are designed with therapists' needs in mind. Care should also be taken to mitigate the risk of therapist fatigue and burnout associated with the specific inherent demands of teleconferencing [58].

Home-based asynchronous training offers significant benefits, but it may not be suitable for all patients. Barriers such as limited physical space, privacy concerns, and the need for technical proficiency can hinder widespread adoption of training technology [59]. With respect to this last item, the relatively young average age of our cohort might suggest higher technical literacy and fewer cognitive impairments, which influences generalizability. In addition, deficits in vision, communication, cognition, or a heightened fall risk may also significantly impact the safety and effectiveness of predominantly asynchronous therapy programs [60,61]. Moreover, some patients have historically preferred in-person therapy [28,59]. Taken together, these points only serve to highlight the critical role of therapists in properly selecting and recommending home-based neurorehabilitation only to appropriate patients.

Limitations

Despite promising results, this study has several limitations. First, the absence of a control group prevents us from drawing any definitive causal conclusions about the impact of remote high-dose training on clinical improvements. However, establishing causality was not the primary aim. The goal of this work was to assess if high levels of training dose could be delivered via a real-world clinical service, as emerging consensus from animal models [62] and human trials [2] suggests that it results in improved upper-limb and gait and balance outcomes. Indeed, medical guidelines for stroke advocate high-dose training [7,8], but patients often receive much less in the chronic setting, posing an implementation challenge. For example, in the United States, patients with chronic stroke typically receive very little or no rehabilitation [63-66], leading to a continual decline in functioning across 5 years after ischemic stroke, irrespective of age and stroke severity [67]. Thus, while we cannot make causal claims, improvements observed here contrast with the typical experience of patients with chronic stroke.

Second, the small sample size, constrained by patient availability and clinical capacity, may limit the generalizability of the findings. Future studies should address factors such as patient motivation, comorbidities, socioeconomic status, and caregiver support. Furthermore, of the 23 patients that began training in this clinical service, 6 dropped out (see section "Patients" in the Methods). Although only 2 of the 6 dropouts voluntarily withdrew from care, suggesting a low dissatisfaction rate, we cannot rule out the possibility that patient attrition may have influenced the results.

Third, the cost analysis focused exclusively on therapist time, excluding the cost of equipment. This approach was intentional to ensure applicability across different commercially available gamified telerehabilitation systems, which vary in price. Furthermore, a true comparison of technology costs would require documenting the different technologies routinely used as part of standard-of-care (eg, robotics, virtual reality therapies, antigravity support systems, etc.), which was not captured as part of this study.

Conclusion

This work introduces an asynchronous and gamified technology-driven approach for delivering high-dose training to patients with stroke. We believe this approach has the potential to scale to address therapist shortages economically and effectively. This work, to our knowledge, is the first demonstration of the full potential value of hybrid asynchronous and synchronous high-dose neurorehabilitation programs at home. The program described here is potentially an attractive option to bridge the gap between accumulating scientific evidence and medical guidelines around the need for high-dose training, and the reality of traditional neurorehabilitation today.

Acknowledgments

This study was funded by MindMaze.

Data Availability

De-identified data analyzed in this study are available from the corresponding author on reasonable request.

Authors' Contributions

SAA contributed to data curation, formal analysis, visualization, writing-original draft, writing-review & editing. ASK was responsible for formal analysis, visualization, writing-original draft, writing-review & editing. MHCC was involved in methodology, supervision, writing-review & editing. GR performed investigation. AEF and LTD designed methodology. JWK and NE handled writing-review & editing. DP was involved in conceptualization, resources, project administration. JTM was responsible for conceptualization, methodology, resources, supervision, project administration, writing-review & editing. MB contributed to supervision; writing - original draft; writing - review & editing.

Conflicts of Interest

All authors have completed an ICMJE uniform disclosure form. SAA is an independent researcher contracted by MindMaze; MHCC, AEF, LTD, and NE are or have been previously employed by MindMaze; JWK is a paid consultant for MindMaze; DP,

XSL•FO
RenderX

JTM, and GR have a research service agreement with MindMaze; DP and JTM are board members of Precision Recovery Inc; All other authors declare no other relationships or activities that could appear to have influenced this work.

Multimedia Appendix 1
Supporting figures and tables.
[PDF File, 699 KB - games_v13i1e69335_app1.pdf]

References

1. Steinmetz JD, Seeher KM, Schiess N, et al. Global, regional, and national burden of disorders affecting the nervous system, 1990–2021: a systematic analysis for the Global Burden of Disease Study 2021. Lancet Neurol 2024 Apr;23(4):344-381. [doi: 10.1016/S1474-4422(24)00038-3]
2. Lohse KR, Lang CE, Boyd LA. Is more better? Using metadata to explore dose-response relationships in stroke rehabilitation. Stroke 2014 Jul;45(7):2053-2058. [doi: 10.1161/STROKEAHA.114.004695] [Medline: 24867924]
3. Veerbeek JM, van Wegen E, van Peppen R, et al. What is the evidence for physical therapy poststroke? A systematic review and meta-analysis. PLOS ONE 2014;9(2):e87987. [doi: 10.1371/journal.pone.0087987] [Medline: 24505342]
4. Kwakkel G, van Peppen R, Wagenaar RC, et al. Effects of augmented exercise therapy time after stroke: a meta-analysis. Stroke 2004 Nov;35(11):2529-2539. [doi: 10.1161/01.STR.0000143153.76460.7d] [Medline: 15472114]
5. Ward NS, Brander F, Kelly K. Intensive upper limb neurorehabilitation in chronic stroke: outcomes from the queen square programme. J Neurol Neurosurg Psychiatry 2019 May;90(5):498-506. [doi: 10.1136/jnnp-2018-319954] [Medline: 30770457]
6. Klassen TD, Dukelow SP, Bayley MT, et al. Higher doses improve walking recovery during stroke inpatient rehabilitation. Stroke 2020 Sep;51(9):2639-2648. [doi: 10.1161/STROKEAHA.120.029245] [Medline: 32811378]
7. Billinger SA, Arena R, Bernhardt J, et al. Physical activity and exercise recommendations for stroke survivors: a statement for healthcare professionals from the American heart association/American stroke association. Stroke 2014 Aug;45(8):2532-2553. [doi: 10.1161/STR.0000000000000022] [Medline: 24846875]
8. Bhalla A, Clark L, Fisher R, et al. National clinical guideline for stroke for the United Kingdom and Ireland. Stroke Guidelines. 2023 May 4. URL: https://www.strokeguideline.org/ [accessed 2025-05-21]
9. Pollock A, Farmer SE, Brady MC, et al. Interventions for improving upper limb function after stroke. Cochrane Database Syst Rev 2014 Nov 12;2014(11):CD010820. [doi: 10.1002/14651858.CD010820.pub2] [Medline: 25387001]
10. Daly JJ, McCabe JP, Holcomb J, et al. Long-dose intensive therapy is necessary for strong, clinically significant, upper limb functional gains and retained gains in severe/moderate chronic stroke. Neurorehabil Neural Repair 2019 Jul;33(7):523-537. [doi: 10.1177/1545968319846120] [Medline: 31131743]
11. Krakauer JW, Kitago T, Goldsmith J, et al. Comparing a novel neuroanimation experience to conventional therapy for high-dose intensive upper-limb training in subacute stroke: the SMARTS2 randomized trial. Neurorehabil Neural Repair 2021 May;35(5):393-405. [doi: 10.1177/15459683211000730] [Medline: 33745372]
12. Winstein C, Kim B, Kim S, et al. Dosage matters. Stroke 2019 Jul;50(7):1831-1837. [doi: 10.1161/STROKEAHA.118.023603] [Medline: 31164067]
13. Bernhardt J, Dewey H, Thrift A, et al. Inactive and alone: physical activity within the first 14 days of acute stroke unit care. Stroke 2004 Apr;35(4):1005-1009. [doi: 10.1161/01.STR.0000120727.40792.40] [Medline: 14988574]
14. West T, Bernhardt J. Physical activity patterns of acute stroke patients managed in a rehabilitation focused stroke unit. Biomed Res Int 2013;2013(1):438679. [doi: 10.1155/2013/438679] [Medline: 24024192]
15. Waddell KJ, Birkenmeier RL, Moore JL, et al. Feasibility of high-repetition, task-specific training for individuals with upper-extremity paresis. Am J Occup Ther 2014 Jul 1;68(4):444-453. [doi: 10.5014/ajot.2014.011619] [Medline: 25005508]
16. Hayward KS, Brauer SG. Dose of arm activity training during acute and subacute rehabilitation post stroke: a systematic review of the literature. Clin Rehabil 2015 Dec;29(12):1234-1243. [doi: 10.1177/0269215514565395] [Medline: 25568073]
17. Lang CE, MacDonald JR, Gnip C. Counting repetitions: an observational study of outpatient therapy for people with hemiparesis post-stroke. J Neurol Phys Ther 2007 Mar;31(1):3-10. [doi: 10.1097/01.npt.0000260568.31746.34] [Medline: 17419883]
18. Kimberley TJ, Samargia S, Moore LG, et al. Comparison of amounts and types of practice during rehabilitation for traumatic brain injury and stroke. J Rehabil Res Dev 2010;47(9):851-862. [doi: 10.1682/jrrd.2010.02.0019] [Medline: 21174250]
19. Feigin VL, Owolabi MO, World Stroke Organization–Lancet Neurology Commission Stroke Collaboration Group. Pragmatic solutions to reduce the global burden of stroke: a World Stroke Organization-Lancet Neurology Commission. Lancet Neurol 2023 Dec;22(12):1160-1206. [doi: 10.1016/S1474-4422(23)00277-6] [Medline: 37827183]
20. Lang CE, Lohse KR, Birkenmeier RL. Dose and timing in neurorehabilitation: prescribing motor therapy after stroke. Curr Opin Neurol 2015 Dec;28(6):549-555. [doi: 10.1097/WCO.0000000000000256] [Medline: 26402404]
21. Jannenga DH. The state of rehab therapy. WebPT. 2023. URL: https://www.webpt.com/downloads/the-state-of-rehab-therapy-in-2023 [accessed 2025-06-24]
22. Putrino D, Krakauer JW. Neurotechnology's prospects for bringing about meaningful reductions in neurological impairment. Neurorehabil Neural Repair 2023 Jun;37(6):356-366. [doi: 10.1177/15459683221137341] [Medline: 36384334]

XSL•FO
RenderX

23. Laver KE, Adey-Wakeling Z, Crotty M, et al. Telerehabilitation services for stroke. Cochrane Database Syst Rev 2020 Jan 31;1(1):CD010255. [doi: 10.1002/14651858.CD010255.pub3] [Medline: 32002991]
24. Wuennemann MJ, Mackenzie SW, Lane HP, et al. Dose and staffing comparison study of upper limb device-assisted therapy. NeuroRehabilitation 2020;46(3):287-297. [doi: 10.3233/NRE-192993] [Medline: 32250331]
25. English C, Hillier SL, Lynch EA. Circuit class therapy for improving mobility after stroke. Cochrane Database of Systematic Reviews 2017 Jun;2017(6):S. [doi: 10.1002/14651858.CD007513.pub3]
26. Wolf SL, Sahu K, Bay RC, et al. The HAAPI (Home Arm Assistance Progression Initiative) trial: a novel robotics delivery approach in stroke rehabilitation. Neurorehabil Neural Repair 2015;29(10):958-968. [doi: 10.1177/1545968315575612] [Medline: 25782693]
27. Lohse KR, Pathania A, Wegman R, et al. On the reporting of experimental and control therapies in stroke rehabilitation trials: a systematic review. Arch Phys Med Rehabil 2018 Jul;99(7):1424-1432. [doi: 10.1016/j.apmr.2017.12.024] [Medline: 29412168]
28. Tosto-Mancuso J, Tabacof L, Herrera JE, et al. Gamified neurorehabilitation strategies for post-stroke motor recovery: challenges and advantages. Curr Neurol Neurosci Rep 2022 Mar;22(3):183-195. [doi: 10.1007/s11910-022-01181-y] [Medline: 35278172]
29. Cramer SC, Dodakian L, Le V, et al. Efficacy of home-based telerehabilitation vs in-clinic therapy for adults after stroke: a randomized clinical trial. JAMA Neurol 2019 Sep 1;76(9):1079-1087. [doi: 10.1001/jamaneurol.2019.1604] [Medline: 31233135]
30. Edwards D, Kumar S, Brinkman L, et al. Telerehabilitation initiated early in post-stroke recovery: a feasibility study. Neurorehabil Neural Repair 2023;37(2-3):131-141. [doi: 10.1177/15459683231159660] [Medline: 36876946]
31. Swanson VA, Johnson C, Zondervan DK, et al. Optimized home rehabilitation technology reduces upper extremity impairment compared to a conventional home exercise program: a randomized, controlled, single-blind trial in subacute stroke. Neurorehabil Neural Repair 2023 Jan;37(1):53-65. [doi: 10.1177/15459683221146995] [Medline: 36636751]
32. Lansberg MG, Legault C, MacLellan A, et al. Home‐based virtual reality therapy for hand recovery after stroke. PM R 2022 Mar;14(3):320-328. [doi: 10.1002/pmrj.12598]
33. Harris PA, Taylor R, Thielke R, et al. Research electronic data capture (REDCap)--A metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381. [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
34. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208. [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
35. Occupational outlook handbook“, physical therapist. Bureau of Labor Statistics, US Department of Labor. URL: https://www.bls.gov/ooh/healthcare/physical-therapists.htm [accessed 2024-08-05]
36. R Core Team. R: The R Project for Statistical Computing“, R Foundation for Statistical Computing. URL: https://www.r-project.org/ [accessed 2024-08-05]
37. Grolemund G, Wickham H. Dates and times made easy with lubridate. J Stat Softw 2011 Apr;40:1-20. [doi: 10.18637/jss.v040.i03]
38. Wickham H, François R, Henry L, et al. dplyr: a grammar of data manipulation. 2023 Nov 17. URL: https://cran.r-project.org/web/packages/dplyr/index.html [accessed 2024-08-05]
39. Wickham H, Chang W, Henry L, et al. Ggplot2: Elegant Graphics for Data Analysis: Springer-Verlag New York; 2016. [doi: 10.1007/978-3-319-24277-4]
40. McCabe J, Monkiewicz M, Holcomb J, et al. Comparison of robotics, functional electrical stimulation, and motor learning methods for treatment of persistent upper extremity dysfunction after stroke: a randomized controlled trial. Arch Phys Med Rehabil 2015 Jun;96(6):981-990. [doi: 10.1016/j.apmr.2014.10.022] [Medline: 25461822]
41. Dodakian L, McKenzie AL, Le V, et al. A home-based telerehabilitation program for patients with stroke. Neurorehabil Neural Repair 2017;31(10-11):923-933. [doi: 10.1177/1545968317733818] [Medline: 29072556]
42. See J, Dodakian L, Chou C, et al. A standardized approach to the Fugl-Meyer assessment and its implications for clinical trials. Neurorehabil Neural Repair 2013 Oct;27(8):732-741. [doi: 10.1177/1545968313491000] [Medline: 23774125]
43. Hiengkaew V, Jitaree K, Chaiyawat P. Minimal detectable changes of the Berg balance scale, Fugl-Meyer assessment scale, timed “up & go” test, gait speeds, and 2-minute walk test in individuals with chronic stroke with different degrees of ankle plantarflexor tone. Arch Phys Med Rehabil 2012 Jul;93(7):1201-1208. [doi: 10.1016/j.apmr.2012.01.014] [Medline: 22502805]
44. Lin JH, Hsu MJ, Hsu HW, et al. Psychometric comparisons of 3 functional ambulation measures for patients with stroke. Stroke 2010 Sep;41(9):2021-2025. [doi: 10.1161/STROKEAHA.110.589739] [Medline: 20671244]
45. Centers for Medicare & Medical Servicces. COVID-19 Frequently Asked Questions (FAQs) on Medicare Fee-for-Service (FFS) Billing. 2023 Apr 20. URL: https://www.cms.gov/files/document/03092020-covid-19-faqs-508.pdf#page=173 [accessed 2025-06-25]
46. Lang CE, Macdonald JR, Reisman DS, et al. Observation of amounts of movement practice provided during stroke rehabilitation. Arch Phys Med Rehabil 2009 Oct;90(10):1692-1698. [doi: 10.1016/j.apmr.2009.04.005] [Medline: 19801058]

XSL•FO
RenderX

47. Hayward KS, Churilov L, Dalton EJ, et al. Advancing stroke recovery through improved articulation of nonpharmacological intervention dose. Stroke 2021 Jan;52(2):761-769. [doi: 10.1161/STROKEAHA.120.032496] [Medline: 33430635]
48. Sharififar S, Ghasemi H, Geis C, et al. Telerehabilitation service impact on physical function and adherence compared to face‐to‐face rehabilitation in patients with stroke: A systematic review and meta‐analysis. PM R 2023 Dec;15(12):1654-1672. [doi: 10.1002/pmrj.12988]
49. Threapleton K, Drummond A, Standen P. Virtual rehabilitation: what are the practical barriers for home-based research? Digit Health 2016;2:2055207616641302. [doi: 10.1177/2055207616641302] [Medline: 29942551]
50. Klobucká S, Klobucký R, Valovičová K, et al. Cost-effectiveness analysis of robot-assisted gait training in patients with bilateral spastic cerebral palsy. Cost Eff Resour Alloc 2023 Sep 11;21(1):60. [doi: 10.1186/s12962-023-00475-3] [Medline: 37697377]
51. Klaic M, Fong J, Crocher V, et al. Application of the extended technology acceptance model to explore clinician likelihood to use robotics in rehabilitation. Disabil Rehabil Assist Technol 2024 Jan;19(1):52-59. [doi: 10.1080/17483107.2022.2060356] [Medline: 35400278]
52. Wiskerke E, Kool J, Hilfiker R, et al. Determining the optimal virtual reality exergame approach for balance therapy in persons with neurological disorders using a Rasch Analysis: longitudinal observational study. JMIR Serious Games 2022 Mar 22;10(1):e30366. [doi: 10.2196/30366] [Medline: 35315785]
53. Kim SY, Daley K, Pruski AD, et al. Implementation of a framework for telerehabilitation in clinical care across the continuum during COVID-19 and beyond. Am J Phys Med Rehabil 2022 Jan 1;101(1):53-60. [doi: 10.1097/PHM.0000000000001904] [Medline: 34915545]
54. 117th congress of the united states of america, consolidated appropriations act, 2023. CONGRESSGOV.: Senate and House of Representatives of the United States of America in Congress; 2022 Jan 3. URL: https://www.congress.gov/117/bills/hr2617/BILLS-117hr2617enr.pdf [accessed 2025-06-24]
55. Medical learning network. Telehealth & Remote Patient Monitoring. 2024 Apr. URL: https://www.cms.gov/files/document/mln901705-telehealth-services.pdf [accessed 2024-09-20]
56. Centers for Medicare & Medical Services. PFS federal regulation notices: CMS-1784-F. URL: https://www.cms.gov/medicare/medicare-fee-service-payment/physicianfeesched/pfs-federal-regulation-notices/cms-1784-f [accessed 2024-09-20]
57. Kelly K, Brander F, Strawson A, et al. Pushing the limits of recovery in chronic stroke survivors: a descriptive qualitative study of users perceptions of the queen square upper limb neurorehabilitation programme. BMJ Open 2020 Oct 5;10(10):e036481. [doi: 10.1136/bmjopen-2019-036481] [Medline: 33020085]
58. Laver K, Walker M, Ward N. Telerehabilitation for stroke is here to stay. but at what cost? Neurorehabil Neural Repair 2022 Jun;36(6):331-334. [doi: 10.1177/15459683221100492] [Medline: 35527716]
59. Chen Y, Abel KT, Janecek JT, Chen Y, et al. Home-based technologies for stroke rehabilitation: a systematic review. Int J Med Inform 2019 Mar;123:11-22. [doi: 10.1016/j.ijmedinf.2018.12.001] [Medline: 30654899]
60. Hill AJ, Theodoros D, Russell T, et al. Using telerehabilitation to assess apraxia of speech in adults. Intl J Lang & Comm Disor 2009 Sep 10;44(5):731-747. [doi: 10.1080/13682820802350537]
61. Theodoros DG. Telerehabilitation for service delivery in speech-language pathology. J Telemed Telecare 2008;14(5):221-224. [doi: 10.1258/jtt.2007.007044] [Medline: 18632993]
62. Jeffers MS, Karthikeyan S, Gomez-Smith M, et al. Does stroke rehabilitation really matter? Part B: an algorithm for prescribing an effective intensity of rehabilitation. Neurorehabil Neural Repair 2018 Jan;32(1):73-83. [doi: 10.1177/1545968317753074] [Medline: 29334831]
63. Chan L, Wang H, Terdiman J, et al. Disparities in outpatient and home health service utilization following stroke: results of a 9-year cohort study in northern California. PM R 2009 Nov;1(11):997-1003. [doi: 10.1016/j.pmrj.2009.09.019]
64. Skolarus LE, Feng C, Burke JF. No racial difference in rehabilitation therapy across all post-acute care settings in the year following a stroke. Stroke 2017 Dec;48(12):3329-3335. [doi: 10.1161/STROKEAHA.117.017290] [Medline: 29089456]
65. Ayala C, Fang J, Luncheon C, et al. Use of outpatient rehabilitation among adult stroke survivors - 20 states and the District of Columbia, 2013, and four states, 2015. MMWR Morb Mortal Wkly Rep 2018 May 25;67(20):575-578. [doi: 10.15585/mmwr.mm6720a2] [Medline: 29795076]
66. Young BM, Holman EA, Cramer SC, STRONG Study Investigators. Rehabilitation therapy doses are low after stroke and predicted by clinical factors. Stroke 2023 Mar;54(3):831-839. [doi: 10.1161/STROKEAHA.122.041098] [Medline: 36734234]
67. Dhamoon MS, Moon YP, Paik MC, et al. Long-term functional recovery after first ischemic stroke: the Northern Manhattan Study. Stroke 2009 Aug;40(8):2805-2811. [doi: 10.1161/STROKEAHA.109.549576] [Medline: 19556535]

Abbreviations

5xStS: 5 times Sit to Stand Test
6minWT: 6 Minute Walk Test
ARAT: Action Research Arm Test
ATT: active training time

XSL•FO
RenderX

BBS: Berg Balance Scale
FGA: Functional Gait Assessment
FM-UE: Fugl-Meyer Upper Extremity
MDC: Minimal Detectable Change
PROM: patient reported outcome measure

Edited by A Coristine; submitted 02.12.24; peer-reviewed by J Hao, J Hwang; revised version received 23.04.25; accepted 23.04.25; published 10.07.25.

Please cite as:
Arbuckle SA, Knill AS, Chan-Cortés MH, Rozanski G, Ford AE, Derungs LT, Krakauer JW, Ejaz N, Putrino D, Tosto-Mancuso J, Branscheidt M
A Resource-Efficient, High-Dose, Gamified Neurorehabilitation Program for Chronic Stroke at Home: Retrospective Real-World Analysis
JMIR Serious Games 2025;13:e69335
URL: https://games.jmir.org/2025/1/e69335
doi:10.2196/69335

© Spencer A Arbuckle, Anna Sophie Knill, Michelle H Chan-Cortés, Gabriela Rozanski, Anastasia Elena Ford, Louis T Derungs, John W Krakauer, Naveed Ejaz, David Putrino, Jenna Tosto-Mancuso, Meret Branscheidt. Originally published in JMIR Serious Games (https://games.jmir.org), 10.7.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Exergaming System for Exercise-Based Cardiac Rehabilitation in Patients With Heart Failure: Development and Usability Assessment Study of a Device Prototype

Carles Blasco-Peris[1,2*], MSc; Juan Pedro Alcolea Garrido[3*], MSc; Barbara Seguí[3*], MSc; Rocio Zaragoza[3*], MSc; Vicente Climent-Paya[4*], PhD; Laura Fuertes-Kenneally[4*], MSc; Agustín Manresa-Rocamora[1,5*], PhD; Ana Sanz-Rocher[1,5*], MSc; Sabina Baladzhaeva[1,5*], MSc; José M Sarabia[1,5*], PhD

[1]Instituto de investigación sanitaria y biomédica de Alicante, Alicante, Spain

[2]Department of Physical Education and Sport, Universitat de València, Valencia, Spain

[3]Aiju, Ibi, Spain

[4]Department of Cardiology, Dr Balmis General University Hospital (HGUA), Instituto de investigación sanitaria y biomédica de Alicante, Av. Pintor Baeza, 12, Alicante, Spain

[5]Department of Sport Sciences, Sports Research Centre, Universitat de Miguel Hernández d'Elx, Elche, Spain

[*]all authors contributed equally

Corresponding Author:
Vicente Climent-Paya, PhD
Department of Cardiology, Dr Balmis General University Hospital (HGUA), Instituto de investigación sanitaria y biomédica de Alicante, Av. Pintor Baeza, 12, Alicante, Spain

Abstract

Background: Heart failure (HF) is a growing global health concern, and adherence to early cardiac rehabilitation (CR) remains suboptimal. Exergaming is a promising alternative to conventional exercise programs for patients with HF. However, existing research has limitations, and the integration of exergaming into clinical practice remains challenging. Most notably, current studies often rely on commercially available systems that are not tailored to needs specific to patients with HF, lack long-term adherence strategies, and have limited evaluation in the initial phases of cardiac rehabilitation.

Objective: This study aimed to design, develop, and assess the usability of a novel exergaming prototype (ie, HEFMOB), integrating immersive virtual reality (VR), real-time biometric monitoring, and autonomous session management to support early-phase, exercise-based CR in patients with HF.

Methods: A multidisciplinary team developed HEFMOB through iterative prototyping. The final system included a pedal-based VR cycling game and an upper-limb mobilization minigame, with real-time monitoring of heart rate, blood pressure, and peripheral capillary oxygen saturation. Usability was assessed in two phases: (1) an expert evaluation and refinement phase and (2) a single-session usability phase involving 10 patients with HF (4 female). The sessions were recorded and individually evaluated by 2 researchers using the Serious Game Usability Evaluator tool. After each session, the participants completed the System Usability Scale (SUS) and a subscale of Intrinsic Motivation Inventory (IMI) to rate the usability of the exergaming prototype and enjoyment, respectively. Descriptive statistics were reported.

Results: The participants had a mean age of 64.8 (SD 8.4) years, BMI of 26.7 (SD 4.6) kg/m^2, and left ventricular ejection fraction of 40.5% (SD 7.4%). All participants completed the session without adverse events. The SUS score averaged 71.5, SD 17.8 (indicating good usability) and IMI scores indicated very high enjoyment (mean 25.1, SD 3.5). A total of 136 gameplay-related events were recorded: negative (n=76, mostly confusion), neutral (n=49), and positive (n=11). Interface-related issues (n=61) were most common, followed by design (n=52) and hardware (n=23).

Conclusions: HEFMOB appears to be a promising, engaging, and well-tolerated tool for delivering tailored exergaming interventions in patients with HF. High usability and enjoyment ratings support its acceptability, while structured user experience analysis provided valuable insights for system refinement. This study marks a critical step toward integrating personalized, gamified exercise in inpatient settings, especially where early mobilization is lacking. Building on these findings, future research will assess long-term usability and clinical impact through a multicenter randomized controlled trial.

(JMIR Serious Games 2025;13:e71385) doi:10.2196/71385

KEYWORDS

virtual reality; exercise; game-based rehabilitation; personalized intervention; HEFMOB; early mobilization; early rehabilitation

XSL•FO
RenderX

Introduction

Heart failure (HF) is a growing global health concern in Europe and worldwide, leading to heavy economic costs, frequent hospitalizations, and high mortality [1]. Furthermore, trends suggest that prevalence and health care costs will continue increasing dramatically in the coming years. Despite significant advances in therapy and prevention, mortality and morbidity remain high, and quality of life is poor [2]. Hospitalization for HF is associated with a worse prognosis, reduced quality of life, and high costs [3]. Population aging, the rising prevalence of risk factors like hypertension, diabetes, obesity, and declining cardiovascular prevention and rehabilitation services all contribute to this alarming trend [4,5].

In this context, there is an acute need to optimize therapeutic strategies and administer treatments at the appropriate time. The World Health Organization has emphasized the need to strengthen cardiac rehabilitation (CR) within health care systems to meet the current and future needs of populations [6]. Within CR programs, an important aspect is phase 1 (ie, early mobilization), which is typically initiated within the first 24 to 72 hours of hospital admission once the patient is clinically stable [7]. Scientific evidence has shown that early multidomain physical rehabilitation intervention improves functional capacity and is associated with shorter hospital stays and fewer readmissions in patients with HF [8,9]. However, only 35% of the centers in Spain perform this phase 1 [10], mainly due to the absence of protocols and insufficient personnel, which hinder the application of these strategies. Thus, the low number of patients who do receive phase 1 CR is concerning, given that early CR may be their only opportunity for CR intervention [11]. The absence of financial incentives, space, multidisciplinary teams, and time are the most common barriers to implementing CR programs [12,13]. In light of these challenges, the accelerated pace of technological progress, particularly the convergence of internet of things (IoT) architecture and information and communication technologies, may offer a transformative avenue to address the mounting health care challenges [14-21]. Recent years have seen the emergence of exercise-based CR programs that implement the use of ICT, such as video games and virtual reality (VR), as alternatives to conventional exercise-based CR programs. These approaches represent a new way for patients to learn, participate in rehabilitation processes, receive personalized care, and engage in care interventions [16,22].

In addition to these advantages, immersive VR-based exergaming offers unique benefits: it enables patients to mentally escape the clinical environment, potentially reducing anxiety and enhancing psychological well-being; it has shown high acceptance and tolerance among vulnerable populations, such as older adults; and it effectively combines physical exercise with cognitive stimulation, which may improve mobility and cognitive function [23]. Exergaming has been demonstrated to be a safe and feasible intervention for patients with HF, although its effect on exercise capacity remains unproven in larger trials [24]. Pilot data suggest that exergaming may increase exercise capacity in selected patients [25].

Despite the promising results of exergaming studies, several crucial aspects and limitations have emerged [16,21]. First, the costs of exergaming equipment can be prohibitive [26]. Second, some patients may struggle with adopting new technology, and these commercial solutions may not be sufficiently tailored to individuals' exercise needs [27]. Third, supervision and close monitoring may be necessary for specific patient groups and rehabilitation stages [28]. Finally, maintaining motivation to exercise can be a persistent challenge for patients, and exergaming alone may not be sufficient to ensure long-term adherence [29].

Having said that, few studies have addressed either the development, the testing, or both, of tailor-made exergaming systems in patients with cardiovascular disease [30-38]. Furthermore, there is a dearth of exergaming systems in the early phases of HF hospitalization that effectively tackle the limitations identified. Therefore, the aim of this study was to design, develop, and assess the usability of an exergaming prototype equipped with remote monitoring of biometric parameters and autonomous management functionalities. The prototype was intended to promote exercise-based CR in patients with HF.

Therefore, the aim of this study was to design, develop, and assess the usability of a tailor-made exergaming prototype for patients hospitalized with HF.

Methods

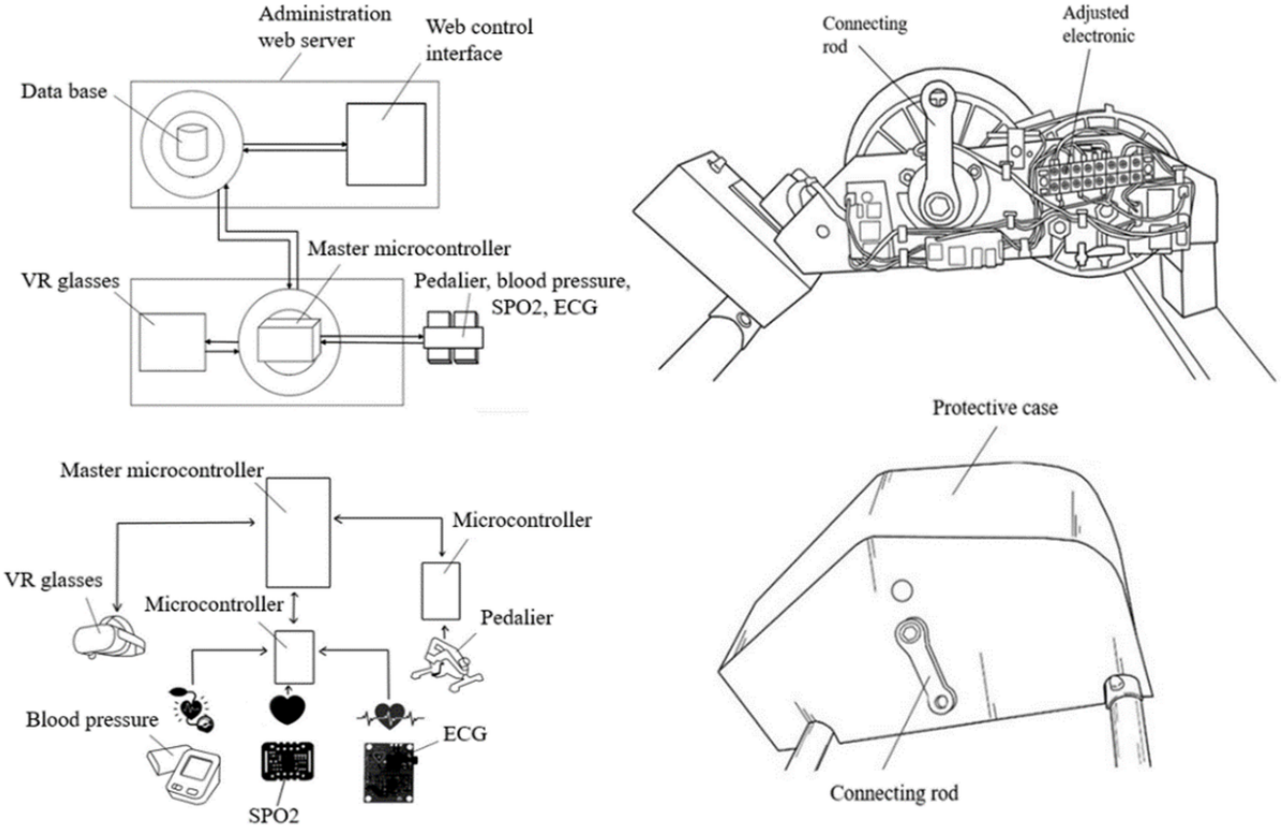
Study Phase 1: Development of the Exergaming Prototype

A multidisciplinary team of engineers, physicians, nurses, physiotherapists, and sports researchers designed and developed an exergaming prototype. The system underwent several development rounds to refine its functionalities and reach unanimous approval from all team members. The final prototype denominated the HF mobilization device (ie, HEFMOB), incorporates real-time monitoring of biometric variables and is specifically tailored for early mobilization in the HF setting.

HEFMOB is a therapeutic rehabilitation assistance device based on physical exercise and gamification. The device consists of the following essential elements: microcontrollers, biometric data measurement devices, an administration web server with a web control interface, data communication technologies, a pedalier, and VR glasses (see Figure 1). The Spanish Agency of Medicines and Health Products approved the prototype to conduct clinical studies as a class IIA clinical product (974/22/EC-R). In addition, its intellectual property rights have been registered with the Spanish Patent and Trademark Office (U202330170).

XSL•FO
RenderX

Figure 1. Prototype scheme diagram and electronic pedalier. VR: virtual reality; ECG: electrocardiogram; SpO_2: peripheral capillary oxygen saturation.



The centerpiece of HEFMOB is the integration of a VR videogame tailored to the Oculus Quest 2 (Meta Platforms), a cutting-edge VR platform renowned for its immersive experiences [39]. The game is integrated into an IoT architecture orchestrated by Raspberry Pi and supported by ESP32 devices that transmit critical physiological parameters [40], such as peripheral capillary oxygen saturation (SpO_2), blood pressure, and heart rate, in real-time through the Message Queuing Telemetry Transport protocol [41].

All collected data are stored in a database and displayed in a dashboard, allowing health care professionals to monitor real-time patient metrics through a web-based control interface during each session. Users engage with the VR game by pedaling through various scenarios while wearing sensors that measure heart rate, SpO_2, blood pressure, and electrical heart activity. An avatar named Cori guides patients throughout the game, offering instructions and information on how to interact with the system autonomously (see Figure 2 left). Pedal resistance is automatically adjusted based on the target heart rate set by the specialist. This target is established by the clinical team according to guidelines for inpatient cardiac rehabilitation in the absence of a stress test and is set approximately 30 beats per minute above the patient's resting heart rate before the session [42]. The system continuously monitors heart rate in real time and adjusts pedaling resistance every 5 seconds to maintain it within ±2 bpm of a target. Based on the average of the last 5 readings, resistance increases or decreases by one unit (range 0 - 8) to match cardiovascular demand, ensuring a safe and adaptive workload. Participants pedal alongside a virtual partner and compete against an opponent. To add excitement, patients collect water bottles to prevent their partner from dehydration. Along the route, participants encounter billboards integrated into the virtual scenario with information relevant to HF pathology. This information, in turn, aims to provide patients with tips on managing their condition better and promote health-conscious behaviors. Once the route phase is over, users engage in a fruit-picking minigame (see Figure 2 right). In this segment, fruits are directed toward the patients, prompting them to carry out upper limb movements corresponding to the fruits' trajectories. The difficulty of the game depends on the speed and frequency at which patients are required to pick the fruit.

XSL•FO
RenderX

Figure 2. Virtual reality–based road cycling tour and fruit-picking minigame.



Following the minigame, the Borg CR-10 scale [42,43] is displayed, and patients are asked to assess their rate of perceived exertion (RPE) after completing both VR segments. These data are also collected by the control system.

Study Phase 2: Usability Assessment

Patients aged 50 to 74 years, diagnosed with HF with mildly reduced (40% to 50%) or reduced (<40%) left ventricular ejection fraction, and without cognitive or physical limitations, were recruited from the HF Unit at Dr Balmis General University Hospital of Alicante. All participants provided written informed consent before being included in the study. Exclusion criteria were unstable angina pectoris, decompensated HF, myocardial infarction in the last 6 months, diabetes mellitus, chronic obstructive pulmonary disease, and complex ventricular arrhythmias. The usability of the prototype was studied in stable patients with HF (ie, chronic heart failure [CHF]) to confirm its safety before using it with decompensated HF. VR experiences were asked before starting usability sessions.

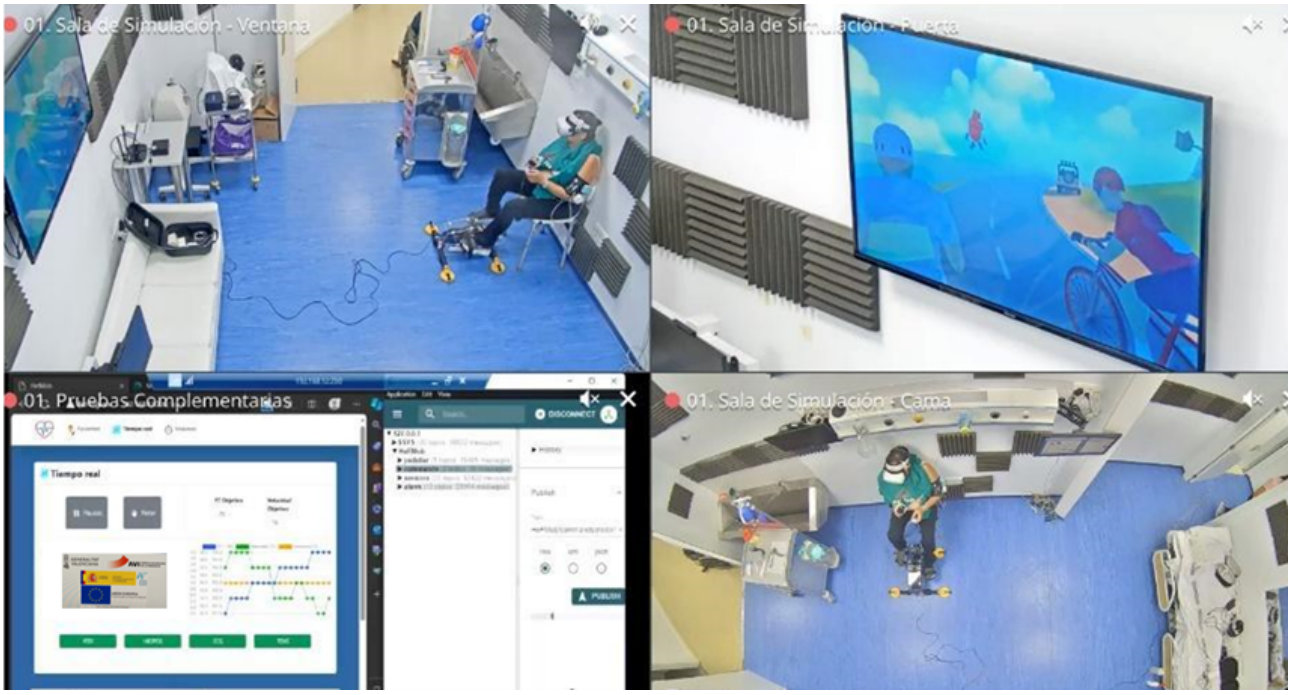
The usability assessment was performed according to evidence-based methods [44], in two stages: (1) expert methods and (2) user methods (ie, patients with CHF). In the first stage, 7 research team members (2 cardiologists, 2 nurses, 2 sport science graduates, and 1 engineer) engaged with the prototype themselves. Based on their insights, necessary modifications to the prototype's functionality were made to improve its usability.

Subsequently, the usability assessment with patients took place. A total of 10 patients with CHF (6 males and 4 females) used the prototype for approximately 20 minutes in a single session in a simulator clinical laboratory. Every session started by placing the hardware elements on the participants. Participants were asked to follow Cori's instructions, while researchers refrained from giving feedback or suggestions during the sessions. Later, Cori provided a brief introduction with a functional explanation and 1 example for familiarization purposes. Thereafter, the cycling tour started, and all patients performed a single 5-minute bout of pedaling at 60% to 70% of their predicted maximum heart rate followed by the fruit-picking minigame, as previously discussed in study phase 1. The session concluded with a question regarding their RPE.

Furthermore, 2 methodological approaches were used to assess familiarization and usability. First, observational analyses were performed while the users interacted with the system. During this stage, patients were instructed to verbalize their thoughts and actions while their sessions were recorded via video and audio using the CAE Learning Space platform (see Figure 3). Subsequently, metrics of game familiarization (ie, percentage of bottles acquired, average time to pick each of the first 5 bottles, and percentage of fruits acquired) were exported from the system database. In addition, 2 research team members independently reviewed the recorded sessions and documented gameplay and usability-related events using the Serious Game Usability Evaluator (SGUE) tool [45].

XSL•FO
RenderX

Figure 3. Example of a recorded participant session.



The SGUE is a structured method designed to evaluate a serious game, focusing not only on the prototype itself but also on the process of interacting with the game and the user's experience. Events recorded with the SGUE are categorized based on their nature (interface, design, and hardware) and the user's reaction (negative, neutral, or positive). This comprehensive approach provides a deeper understanding of how users interact with the game, highlighting areas for improvement in both functionality and user satisfaction [45]. Because the HEFMOB prototype comprises various gadgets, the evaluation also encompassed the interaction with hardware elements. Any disagreements between the two assessors were resolved through consensus. A list of action points to improve system usability was derived from the analysis of the SGUE results. For each action related to a neutral or negative user dimension, we indicated the frequency (number of events that would be solved by this action) and the dispersion (number of users that encountered an event that would be solved by this action). Both numbers were used to determine the priority of each action.

Survey-based methods were also used. Immediately after the session, patients were asked to answer the System Usability Scale (SUS) [46], a 10-item Likert scale with 5 possible levels of agreement, ranging from strongly disagree to strongly agree. The SUS score was interpreted as worst imaginable (≥12.5), awful (≥20.3), poor (≥35.7), ok (≥50.9), good (≥71.4), excellent (≥85.5), and best imaginable (≥90.9) [47]. Finally, patients also responded to the 5-item interest and enjoyment subscale of the Intrinsic Motivation Inventory (IMI) [48]. The IMI score was interpreted as follows: very low level of enjoyment (5 - 9 points), low level of enjoyment (10 - 14 points), moderate level of enjoyment (15 - 19 points), high level of enjoyment (20 to 24 points), and very high level of enjoyment (25 to 35 points) [49]. Mean and SD values are reported.

Ethical Considerations

This study was conducted in accordance with the principles outlined in the Declaration of Helsinki and received approval from the Comité Ético de Investigación con Medicamentos del Departamento de Salud de Alicante – Hospital General (2022/041). All participants provided written informed consent before inclusion. The approved consent form included explicit provisions for the analysis and secondary use of the collected data. To protect privacy and confidentiality, all data were deidentified before analysis, and no personal or identifying information was stored or shared. Participation was entirely voluntary, and no compensation was offered or received. In addition, no individual participant can be identified in any of the images or multimedia appendices included in the manuscript.

Results

Study Phase 2: Usability Assessment

Expert Methods

At this stage, areas for improvement were identified and subsequently addressed (see Multimedia Appendix 1 for the list of needs and improvements in the experts' methods).

Key modifications included adding suction cups to the pedaling mechanism for stability and automating pedaling intensity based on the target heart rate, calculated from the average of the previous 5 heart rate values. This removed the need for a manual intensity test. Usability improvements included removing the password requirement to start sessions, adding portable batteries for the VR glasses, and enabling real-time monitoring and archiving of blood pressure and SpO_2 data, with the option to manually measure blood pressure during adverse events. Minigame updates introduced gameplay explanations, difficulty levels, replay options, and adjustments to reduce dizziness by optimizing collider areas. Additional improvements included

XSL•FO
RenderX

redesigning the pulse oximeter holder for consistent signals, updating blood pressure measurement methods to support the arm, and adding an easily accessible emergency stop button, both physically and via the web interface. These changes enhanced the system's functionality, safety, and user experience. These updates improved the functionality and expert user experience of the system.

User Methods

Table 1 shows the characteristics of the 10 participants with CHF who completed the session. None of the recruited patients had any previous experience with VR.

Table . Participant characteristics (n=10).

Variable	Values
Quantitative variables, mean (SD)	
Age (years)	64.8 (8.4)
Weight (kg)	72.3 (14.1)
Height (cm)	164 (0.1)
BMI (kg/m^2)	26.7 (4.6)
LVEF[a] (%)	40.5 (7.4)
Categorical variables, n (%)	
LVEF	
Mildly reduced	6 (60)
Reduced	4 (40)
Aetiology	
Ischemic	3 (30)
Nonischemic	7 (70)
Sex	
Male	6 (60)
Female	4 (40)
VR[b] experience	
Yes	0 (0)
No	10 (100)

[a]LVEF: left ventricular ejection fraction.

[b]VR: virtual reality.

Regarding familiarization, participants successfully acquired 72 out of 112 bottles during the tandem cycling tour and 179 out of 590 fruits in the fruit-picking minigame. The average time participants took to acquire each of the first 5 bottles show a downward trend along the session (see Figure S1 in Multimedia Appendix 1).

Table 2 shows the number of appearances of each tag and the frequencies for each event type obtained with the SGUE tool. More negative events (76/136, 56%), such as confusion, than neutral (49/136, 36%) and positive events (11/136, 8%) were recorded. In addition, there were more interface events (61/136, 45%) than events related to design (52/136, 38%) and hardware (23/136, 17%).

XSL•FO
RenderX

Table . Tag statistics: the events are categorized into 2 dimensions: the source of the event (interface, design, and hardware) and the reaction of the user.

User-related dimension	System-related dimension						
	Interface, n (%)			Design, n (%)		Hardware, n (%)	Total, n (%)
	Content	Layout and UI	Technical error	Gameflow	Functionality		
Negative							
Annoyed	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Confused	7 (5)	23 (17)	0 (0)	6 (4)	11 (8)	1 (0)	48 (35)
Frustrated	0 (0)	1 (0)	0 (0)	3 (2)	3 (2)	0 (0)	7 (5)
Unable to continue (fatal)	0 (0)	0 (0)	0 (0)	0 (0)	6 (4)	15 (11)	21 (15)
Positive							
Pleasantly frustrated	0 (0)	0 (0)	0 (0)	2 (1)	0 (0)	0 (0)	2 (1)
Reflecting	0 (0)	1 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (0)
Satisfied or excited	0 (0)	0 (0)	0 (0)	1 (0)	0 (0)	0 (0)	1 (0)
Learning	0 (0)	1 (0)	0 (0)	4 (3)	2 (1)	0 (0)	7 (5)
Neutral							
N/A[a]	2 (1)	1 (0)	21 (15)	0 (0)	3 (2)	6 (4)	33 (24)
Suggestion or comment	1 (0)	3 (2)	0 (0)	9 (7)	2 (1)	1 (0)	16 (12)
Total	10 (7)	30 (22)	21 (15)	25 (18)	27 (20)	23 (17)	136 (100)

[a]N/A: not available.

Table 2 summarizes the number of appearances of each tag and the frequencies for different event types, as recorded by the SGUE tool (the dataset is available at [50]). A total of 136 events were categorized. Negative events (n=76, 56%) were the most prevalent, with confusion being the dominant tag (n=48), particularly in interface (n=23) and design-related issues (n=11). For example, users frequently expressed confusion when navigating certain interface features or attempting to understand design elements like layout inconsistencies.

Neutral events (n=49, 36%) primarily involved suggestions or comments (n=16) and technical observations that did not provoke strong reactions, such as noting specific interface layouts or reporting minor glitches.

Positive events (n=11, 8%) were less common, with users occasionally reporting satisfaction (n=2) or instances of learning (n=7). A practical example is a user adapting to a new feature after initial difficulty, ultimately finding the experience rewarding.

From a system-related perspective, interface issues accounted for the largest share of events (n=61, 45%), followed by design-related problems (n=52, 38%) and hardware issues (n=23, 17%). Specific examples of interface challenges included unclear menu structures or misaligned buttons. Hardware-related events, though less frequent, were impactful, such as issues with the positioning or adjustment of sensors, which sometimes affected gameplay.

A list of action points to improve system usability was derived from the analysis of the results in Table 2 (see Table 3). The interface action points provided can be grouped into graphics issues (priority 1) and comprehension issues with the avatar (priorities 5 to 7). The hardware action points shown can be categorized into hardware elements that users interact with (priorities 2 and 8) and hardware elements for user monitoring (priority 9). The design action points presented can be classified as user imprecise execution (priorities 3 and 4).

Table 4 displays the results of the usability assessment. Just 1 participant experienced a light dizziness sensation along the curves in the road cycling tour. All participants expressed interest and enjoyment after the prototype session. The participants who took part in the usability assessment reported very high levels of interest/enjoyment (mean 25.1, SD 3.5) and rated the exergame's usability as good (mean 71.5, SD 17.8).

XSL•FO
RenderX

Table . Excerpt of the prioritized action points list.

Priority	Action	Type	Frequency	Spread
1	Error with an image.	Hardware	20	10
2	Assistance to deliver the second controller.	Interface	11	10
3	Difficulty activating the command.	Design	14	9
4	Technical problem performing the action.	Design	15	8
5	User does not understand Cori's instruction.	Interface	15	8
6	Question regarding Cori's instruction.	Interface	9	6
7	Lack of indication from Cori.	Interface	10	5
8	Issues with the pedal anchor.	Hardware	9	5
9	Maladjustment of one of the devices gathering hemodynamic data.	Hardware	7	4

Table . Usability assessment outcomes.

Participant ID	Dizziness	IMI[a] score	IMI level of interest/enjoyment	SUS[b] score	SUS usability
1	No	24	High	37.5	Poor
2	No	19	Moderate	75	Good
3	No	28	Very high	97.5	Best imaginable
4	No	28	Very high	70	Ok
5	No	25	Very high	45	Poor
6	No	19	Moderate	82.5	Good
7	No	27	Very high	77.5	Good
8	No	28	Very high	75	Good
9	Yes	28	Very high	82.5	Good
10	No	25	Very high	72.5	Good
Mean (SD)	—[c]	25.1 (3.5)	Very high	71.5 (17.8)	Good

[a]IMI: Intrinsic Motivation Inventory.

[b]SUS: System Usability Scale.

[c]Not applicable.

Mean RPE was 4.5 (SD 3.0; see Table S1 in Multimedia Appendix 1). This indicates a moderate perceived effort during the session, with individual values ranging from 0 to 8, reflecting variability in the participants' exertion levels.

Discussion

Principal Findings

This study presents the design, development, and usability evaluation of HEFMOB, an exergaming prototype integrating remote monitoring and autonomous management features to support early-phase, exercise-based CR in patients with HF. The prototype was developed through a multidisciplinary collaboration with technological and health research institutes, hospitals, and academia and tested in a sample of 10 patients, all without previous VR experience. Overall, the usability assessment showed that the system was safe, well tolerated, and highly engaging. Patients reported high levels of interest and enjoyment (IMI mean 25.1, SD 3.5) and rated the system's usability as good (SUS mean 71.5, SD 17.8), with only 1 patient reporting a light sensation of dizziness during a specific activity, which was not classified as an adverse event. Although some negative user feedback was recorded, primarily related to confusion with interface and game design, the system was perceived positively overall. A systematic analysis of 136 user experience events led to a prioritized list of actionable

XSL•FO
RenderX

improvements, which have informed the next version of the prototype. These findings support the feasibility and user acceptability of HEFMOB as a novel solution for implementing early mobilization in hospitalized patients with HF. The analysis revealed a notable prevalence of negative events (n=76) compared to neutral (n=49) and positive ones (n=11), with interface issues (n=61) outweighing design (n=52) and hardware (n=23) concerns. These results are similar to another Serious Game for patients [44], with negative events being the most common, representing approximately 50%, followed by neutral events, and finally, positive events. In software development, it is usual that some functions are obvious for developers and experts but not for real users, leading to negative events such as feelings of confusion or frustration [51].

Action points were formulated based on the events analysis findings, with prioritization determined by the frequency of occurrence and the number of users affected. Addressing interface issues related to graphics emerged as the top action point. Other hardware and design actions were identified to improve interaction with physical elements and enhance game execution. These improvements will be incorporated into the next HEFMOB prototype version.

In the usability assessment, participants largely avoided dizziness, with only 1 participant reporting a mild dizziness sensation during a specific activity. Overall, participants expressed significant interest and enjoyment, rating the exergame's usability positively. Despite the good scores on the SUS, 2 participants rated usability poorly and provided comments about disliking the bicycle exercise and finding the fruit-picking game difficult. Taking into consideration their scores on the IMI (both rated the activity 26, a very high level of enjoyment), we hypothesize that the 2 participants who rated the system usability as poor were influenced by their dislike of exercise and the difficulty of the game rather than system functioning or usability. This could explain the poor usability score. Although patients received a brief introduction and familiarization with the system before each exercise session, age-related factors should also be taken into account. Age and previous gaming experience can influence usability and the occurrence of negative events. Older adults may struggle with digital literacy due to the digital divide, limiting effective interaction with active video games [20]. Proper onboarding is therefore essential to address these challenges, reinforcing the need for tailored familiarization based on users' age and gaming experience [20].

While this feasibility study and its findings are exploratory in nature, they nonetheless offer valuable insights. In a sample of 10 patients with HF, the HEFMOB prototype demonstrated no side effects, good usability, and user enjoyment. Furthermore, in line with Kortum and Peres [52], who established a correlation between SUS scores and satisfaction, the usability assessment outcomes suggest that patients with HF experienced overall satisfaction after a single session with the HEFMOB prototype.

Comparison With Previous Work

Other studies have developed exergaming prototypes for exercise-based CR [30-34,36,38], but few have been used in patient samples that included children and adolescents with congenital heart disease or in patients undergoing nonemergency coronary artery graft surgery [32,35,37]. Among these studies, 2 different exergaming systems stand out: an augmented reality software developed to train patients in physical activities, such as walking around the inpatient ward and climbing stairs [32]; and the MedBike, an innovative rehabilitation home-based device designed to enhance patient adherence, convenience, and engagement [35,37]. MedBike has a 2-tiered interface, enabling clinicians to remotely monitor patients as they exercise on a stationary bicycle while immersed in a VR environment. Advanced sensors continuously track vital signs (ie, electrocardiograph, SpO_2, and blood pressure), providing valuable data for personalized care [30].

HEFMOB and MedBike differ in that HEFMOB has been designed and developed for use in the early phase, in a hospital setting, whereas MedBike is used in an outpatient setting. HEFMOB can be used with any conventional chair, whereas MedBike requires the use of a cycle ergometer. HEFMOB uses immersive VR technology, whereas MedBike does not. In addition, HEFMOB boasts unique functionalities such as autonomous management of sessions guided by an avatar, autonomous regulation of pedaling intensity based on the predetermined target heart rate, inclusion of upper body mobilization exercises, promotion of healthy habits, and management of pathology through in-game advice.

Our results suggest that HEFMOB represents a significant initial step in developing an exergaming prototype customized for patients with HF. It encourages exercise-based CR in a patient population while also incorporating remote biometric monitoring and autonomous management. Based on the findings from this pilot study, we believe that the HEFMOB prototype holds promise for promoting early exercise-based CR at hospital facilities. Its implementation may facilitate the attainment of desirable engagement levels in phase 1 CR programs and compliance with early mobility recommendations in line with clinical guidelines [3,53]. Furthermore, the impact of HEFMOB could be notably amplified given that the early CR phase may be the sole opportunity for CR intervention in a substantial number of patients [11].

Limitations and Research Priorities

The study involved only 10 patients with stable CHF, which may limit the generalizability of the findings to HF inpatients, particularly those who have experienced recent decompensation. A larger sample size would provide more robust evidence regarding the usability of HEFMOB across a broader patient population. In addition, the usability assessment was based on a single session with HEFMOB, which may not fully capture its long-term usability. This phase was necessary before applying HEFMOB in real-world scenarios with patients post decompensation, which will be addressed in subsequent studies. Furthermore, future studies will also evaluate the effectiveness of HEFMOB in improving patient outcomes.

The assessment of usability relied primarily on subjective measures, such as participants' ratings and feedback. While these provide valuable insights, future studies could consider complementing subjective data with additional measures, such as detailed session observations, to provide a more

XSL•FO
RenderX

comprehensive evaluation of HEFMOB's usability. Importantly, this study did not aim to analyze effectiveness, and metrics related to performance or physiological responses were not included, as these will be explored in the planned effectiveness studies.

Conclusions

The findings suggest that HEFMOB is a feasible and well-tolerated tool, demonstrating high user engagement and good usability ratings in a controlled, single-session simulation clinical laboratory environment involving patients with HF in the ambulatory phase. HEFMOB represents a significant initial step in the development of an exergaming prototype tailored to the needs of patients with HF and emerges as a promising and engaging tool for supporting early exercise-based CR.

Future Work

Future research is required to conclusively demonstrate the efficacy and effectiveness of serious exergaming prototypes such as HEFMOB in promoting early mobilization in hospital, particularly through interventions involving multiple sessions. These future analyses should also examine its subsequent impact on duration of hospital stay, readmissions, and behavioral and lifestyle outcomes, including quality of life, improved disease knowledge, functional capacity, and engagement and adherence in physical activity and exercise.

Next steps include the deployment of a new prototype version of the system. A new usability study is planned with a larger sample, multiple sessions, and semistructured interviews with participants who may report poor usability. This approach aims to generate deeper insights into user experience and system performance, guiding further refinements before scaling.

Subsequently, a multicenter randomized controlled trial will be conducted, with sample size calculated based on expected effect sizes. The study will assess outcomes such as functional capacity, muscular strength, HF knowledge, body composition, psychosocial status (anxiety and depression), and user enjoyment. Gender representation will be monitored to address potential biases, and rational exclusion criteria (eg, cognitive impairment, severe comorbidities, and low digital literacy) will ensure participant safety and data quality. Follow-up will include hospital readmissions and all-cause mortality to evaluate the long-term clinical impact of the intervention.

Acknowledgments

We extend our gratitude to Isabel Hernández-Pérez and Noemi Sempere-Ruiz for their invaluable assistance in the recruitment process and session arrangements. Furthermore, we would like to express our thanks to all members of the Simulator and Innovation Laboratory of Alicante (SIMIA-ISABIAL) for their hospitality and cooperation.

This research was funded by the Instituto de Salud Carlos III (ISCIII), co-funded by the European Union (ERDF, "A way to make Europe", grants DTS21/00171 and DTS24/00015); the Institute for Health and Biomedical Research of Alicante (ISABIAL), grant 2022- - 0336; the Valencia Region Innovation Agency (AVI), grant INNVA1/2020/5; and the Valencian Government, grant INVEST/2023/436.

This study adheres to the principles set out in Maastricht University's Policy Framework on Generative Artificial Intelligence (AI), which promotes the responsible, ethical, and transparent use of AI in research. Generative AI tools were used in a limited capacity to support language editing, text summarization, and formatting suggestions during the manuscript preparation phase. All scientific content, data interpretation, and critical analyses were performed by the authors without reliance on AI technologies. No generative AI tools were used for data generation, hypothesis formulation, or statistical analysis. The authors maintained full responsibility for the integrity and accuracy of the scientific content. AI-assisted outputs were critically reviewed and edited by the research team to ensure compliance with academic standards, authorship criteria, and disciplinary norms. The use of AI did not compromise research integrity, participant privacy, or the transparency of methods and findings. Any AI tool used was applied in alignment with existing data protection regulations and institutional ethical guidelines.

Data Availability

The data collected and used for the analysis in this study are openly available in the Zenodo repository at [50,54,55]. All relevant data supporting the findings are either included in the main manuscript or accessible via this public repository. No restrictions apply to data access.

Authors' Contributions

CBP, JPA, BS, RZ, VCP, LF-K, AM-R, and JMS contributed to the conceptualization of the study. CBP, AMR, ASR, and SB performed data curation. CBP, VCP, AM-R, and JMS conducted formal analysis. CBP, VCP, and JMS led the funding acquisition. CBP, VCP, LF-K, AM-R, ASR, SB, and JMS carried out the investigation. AM-R and JMS developed the methodology. CBP, VCP, and JMS managed project administration. JPA, BS, and RZ developed the software. CBP, JPA, BS, RZ, VCP, LF-K, AM-R, and JMS provided resources. VCP and JMS performed supervision. CBP and AM-R performed validation. CBP, AM-R, ASR, and SB contributed to visualization. CBP and AM-R contributed to the writing of the original draft. CBP, VCP, and JMS contributed to writing, review, and editing.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Expert-identified needs and proposed improvements, individual game outcomes with perceived exertion ratings, and average time to task completion
[DOC File, 66 KB - games_v13i1e71385_app1.doc]

References

1. Guha K, McDonagh T. Heart failure epidemiology: European perspective. Curr Cardiol Rev 2013 May;9(2):123-127. [doi: 10.2174/1573403x11309020005] [Medline: 23597298]
2. Rosano GMC, Seferovic P, Savarese G, et al. Impact analysis of heart failure across European countries: an ESC-HFA position paper. ESC Heart Fail 2022 Oct;9(5):2767-2778. [doi: 10.1002/ehf2.14076] [Medline: 35869679]
3. Authors/Task Force Members, McDonagh TA, Metra M, et al. 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: Developed by the Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC). With the special contribution of the Heart Failure Association (HFA) of the ESC. Eur J Heart Fail 2022 Jan;24(1):4-131. [doi: 10.1002/ejhf.2333] [Medline: 35083827]
4. Groenewegen A, Rutten FH, Mosterd A, Hoes AW. Epidemiology of heart failure. Eur J Heart Fail 2020 Aug;22(8):1342-1356. [doi: 10.1002/ejhf.1858] [Medline: 32483830]
5. Mahmood SS, Levy D, Vasan RS, Wang TJ. The Framingham Heart Study and the epidemiology of cardiovascular disease: a historical perspective. Lancet 2014 Mar 15;383(9921):999-1008. [doi: 10.1016/S0140-6736(13)61752-3] [Medline: 24084292]
6. Negrini S. Introduction to the special section “Rehabilitation 2030: a call for action” by the World Health Organization (WHO). Eur J Phys Rehabil Med 2017 Apr;53(2):151-152. [doi: 10.23736/S1973-9087.17.04765-7] [Medline: 28488423]
7. Miao X, Jiang H, Huang X, Zheng Y, Jin S, Wang R. Randomized trial of early exercise rehabilitation and its effects on patients with acute coronary syndrome. Sci Rep 2024;14(1):39487165. [doi: 10.1038/s41598-024-77108-z]
8. Fleming LM, Zhao X, DeVore AD, et al. Early ambulation among hospitalized heart failure patients is associated with reduced length of stay and 30-day readmissions. Circ Heart Fail 2018 Apr;11(4):e004634. [doi: 10.1161/CIRCHEARTFAILURE.117.004634] [Medline: 29643066]
9. Kitzman DW, Whellan DJ, Duncan P, et al. Physical rehabilitation for older patients hospitalized for heart failure. N Engl J Med 2021 Jul 15;385(3):203-216. [doi: 10.1056/NEJMoa2026141] [Medline: 33999544]
10. Pablo-Zarzosa C, Arrate-Estéban V, Castro-Conde A, Ferro-Múgica J, Montiel-Trujillo A. National registry of cardiac rehabilitation units. R-EUReCa [Article in Spanish]. : Spanish Society of Cardiology; 2015 URL: https://secardiologia.es/riesgo/cientifico/6415-registro-nacional-de-unidades-de-rehabilitacion-cardiaca-r-eureka [accessed 2025-07-12]
11. Bjarnason-Wehrens B, McGee H, Zwisler AD, et al. Cardiac rehabilitation in Europe: results from the European Cardiac Rehabilitation Inventory Survey. European Journal of Cardiovascular Prevention & Rehabilitation 2010 Aug;17(4):410-418. [doi: 10.1097/HJR.0b013e328334f42d]
12. Ruivo J, Moholdt T, Abreu A. Overview of cardiac rehabilitation following post-acute myocardial infarction in European Society of Cardiology member countries. Eur J Prev Cardiol 2023 Jul 12;30(9):758-768. [doi: 10.1093/eurjpc/zwad024] [Medline: 36722203]
13. Thomas RJ. Cardiac rehabilitation - challenges, advances, and the road ahead. N Engl J Med 2024 Feb 29;390(9):830-841. [doi: 10.1056/NEJMra2302291] [Medline: 38416431]
14. Aghdam ZN, Rahmani AM, Hosseinzadeh M. The role of the internet of things in healthcare: future trends and challenges. Comput Methods Programs Biomed 2021 Feb;199:105903. [doi: 10.1016/j.cmpb.2020.105903] [Medline: 33348073]
15. Haghi Kashani M, Madanipour M, Nikravan M, Asghari P, Mahdipour E. A systematic review of IoT in healthcare: applications, techniques, and trends. J Network Comp App 2021 Oct;192:103164. [doi: 10.1016/j.jnca.2021.103164]
16. Blasco-Peris C, Fuertes-Kenneally L, Vetrovsky T, Sarabia JM, Climent-Paya V, Manresa-Rocamora A. Effects of exergaming in patients with cardiovascular disease compared to conventional cardiac rehabilitation: a systematic review and meta-analysis. Int J Environ Res Public Health 2022 Mar 15;19(6):3492. [doi: 10.3390/ijerph19063492] [Medline: 35329177]
17. Herren S, Seebacher B, Mildner S, et al. exergame (ExerG)-based physical-cognitive training for rehabilitation in adults with motor and balance impairments: usability study. JMIR Serious Games 2025 Feb 14;13:e66515. [doi: 10.2196/66515] [Medline: 39951650]
18. Li AY, Qiang W, Li JW, Geng Y, Qiang Y, Zhao JJ. Retracted: evaluating the clinical efficacy of an exergame-based training program for enhancing physical and cognitive functions in older adults with mild cognitive impairment and dementia residing in rural long-term care facilities: randomized controlled trial. J Med Internet Res 2025 Feb 19;27:e69109. [doi: 10.2196/69109] [Medline: 39969990]

XSL•FO
RenderX

19. Radhakrishnan K, Baranowski T, Julien C, Thomaz E, Kim M. Role of digital games in self-management of cardiovascular diseases: a scoping review. Games Health J 2019 Apr;8(2):65-73. [doi: 10.1089/g4h.2018.0011] [Medline: 30199275]
20. Sawa R, Saitoh M, Morisawa T, et al. The potential application of commercially available active video games to cardiac rehabilitation: scoping review. JMIR Serious Games 2022 Mar 18;10(1):e31974. [doi: 10.2196/31974] [Medline: 35302503]
21. Schättin A, Häfliger S, Meyer A, et al. Design and evaluation of user-centered exergames for patients with multiple sclerosis: multilevel usability and feasibility studies. JMIR Serious Games 2021 May 7;9(2):e22826. [doi: 10.2196/22826] [Medline: 33960956]
22. Fang Z, Wu T, Lv M, et al. Effect of traditional plus virtual reality rehabilitation on prognosis of stroke survivors: a systematic review and meta-analysis of randomized controlled trials. Am J Phys Med Rehabil 2022 Mar 1;101(3):217-228. [doi: 10.1097/PHM.0000000000001775] [Medline: 33929347]
23. Peng Y, Wang Y, Zhang L, et al. Virtual reality exergames for improving physical function, cognition and depression among older nursing home residents: a systematic review and meta-analysis. Geriatr Nurs (Lond) 2024 May;57:31-44. [doi: 10.1016/j.gerinurse.2024.02.032]
24. Jaarsma T, Klompstra L, Ben Gal T, et al. Effects of exergaming on exercise capacity in patients with heart failure: results of an international multicentre randomized controlled trial. Eur J Heart Fail 2021 Jan;23(1):114-124. [doi: 10.1002/ejhf.1754] [Medline: 32167657]
25. Klompstra L, Jaarsma T, Strömberg A. Exergaming to increase the exercise capacity and daily physical activity in heart failure patients: a pilot study. BMC Geriatr 2014 Nov 18;14:119. [doi: 10.1186/1471-2318-14-119] [Medline: 25407612]
26. Stanmore E, Eost-Telling C, Meekes W, et al. Exergames for falls prevention in sheltered homes: a feasibility study. Front Public Health 2024;12:1344019. [doi: 10.3389/fpubh.2024.1344019] [Medline: 38975352]
27. Krishnan S, Mandala MA, Wolf SL, Howard A, Kesar TM. Perceptions of stroke survivors regarding factors affecting adoption of technology and exergames for rehabilitation. PM R 2023 Nov;15(11):1403-1410. [doi: 10.1002/pmrj.12963]
28. Zhao W. Technology-Enabled Motion Sensing and Activity Tracking for Rehabilitation: Institution of Engineering and Technology; 2022:149-159.
29. Valenzuela T, Okubo Y, Woodbury A, Lord SR, Delbaere K. Adherence to technology-based exercise programs in older adults: a systematic review. J Geriatr Phys Ther 2018;41(1):49-61. [doi: 10.1519/JPT.0000000000000095] [Medline: 27362526]
30. Boulanger P, Pournajib A, Mott W, Schaeffer S. A low-cost virtual reality bike for remote cardiac rehabilitation. Presented at: Virtual Reality and Augmented Reality: 14th EuroVR International Conference, EuroVR 2017; Dec 12-14, 2017; Laval, FRANCE. [doi: 10.1007/978-3-319-72323-5_10]
31. Eichhorn S, Koller V, Schreiber U, Mendoza A, Krane M, Lange R. Development of an Exergame for individual rehabilitation of patients with cardiovascular diseases. Australas Phys Eng Sci Med 2013 Dec;36(4):441-447. [doi: 10.1007/s13246-013-0223-z] [Medline: 24113921]
32. Ghlichi Moghaddam N, Namazinia M, Hajiabadi F, Mazlum SR. The efficacy of phase I cardiac rehabilitation training based on augmented reality on the self-efficacy of patients undergoing coronary artery bypass graft surgery: a randomized clinical trial. BMC Sports Sci Med Rehabil 2023 Nov 17;15(1):156. [doi: 10.1186/s13102-023-00770-9] [Medline: 37978403]
33. Guo S, Grindle GG, Authier EL, et al. Development and qualitative assessment of the GAME(Cycle) exercise system. IEEE Trans Neural Syst Rehabil Eng 2006 Mar;14(1):83-90. [doi: 10.1109/TNSRE.2006.870493] [Medline: 16562635]
34. Hung K, Wan N, Choy SO, Chu C, Chow DHK. Design of an exergaming system with haptic feedback for the investigation of energy expenditure and muscle activities. Presented at: 2016 IEEE 14th International Conference on Industrial Informatics (INDIN); Jul 19-21, 2016; Poitiers, France. [doi: 10.1109/INDIN.2016.7819286]
35. Khoury M, Phillips DB, Wood PW, et al. Cardiac rehabilitation in the paediatric Fontan population: development of a home-based high-intensity interval training programme. Cardiol Young 2020 Oct;30(10):1409-1416. [doi: 10.1017/S1047951120002097] [Medline: 32716280]
36. Soares JC, Vieira A, Gabriel J. Assisted living: home physiotherapy demo. Presented at: 2nd Experiment at International Conference (Exp at); Sep 18-20, 2013; Univ Coimbra, Coimbra, Portugal.
37. Spence C, Khoo N, Mackie A, et al. Exploring the promise of telemedicine exercise interventions in children and adolescents with congenital heart disease. Can J Cardiol 2023 Nov;39(11S):S346-S358. [doi: 10.1016/j.cjca.2023.08.015] [Medline: 37657493]
38. Trombetta M, Bazzanello Henrique PP, Brum MR, Colussi EL, De Marchi ACB, Rieder R. Motion Rehab AVE 3D: A VR-based exergame for post-stroke rehabilitation. Comput Methods Programs Biomed 2017 Nov;151:15-20. [doi: 10.1016/j.cmpb.2017.08.008] [Medline: 28946996]
39. Pereira D, Oliveira V, Vilaça JL, Carvalho V, Duque D. Measuring the precision of the Oculus Quest 2's handheld controllers. Actuators 2023;12(6):257. [doi: 10.3390/act12060257] [Medline: 28946996]
40. Barik L. IoT based temperature and humidity controlling using Arduino and Raspberry Pi. IJACSA 2019;10(9):494-502. [doi: 10.14569/IJACSA.2019.0100966]
41. Kashyap M, Sharma V, Verma G. Implementation and analysis of iot based automation using MQTT. Presented at: 2021 IEEE 4th International Conference on Computing, Power and Communication Technologies (GUCON); Sep 24-26, 2021; Kuala Lumpur, Malaysia p. 1-5. [doi: 10.1109/GUCON50781.2021.9573857]

XSL•FO
RenderX

42. Mytinger M, Nelson RK, Zuhl M. Exercise prescription guidelines for cardiovascular disease patients in the absence of a baseline stress test. J Cardiovasc Dev Dis 2020 Apr 27;7(2):32349219. [doi: 10.3390/jcdd7020015] [Medline: 32349219]
43. Borg G, Linderholm H. Perceived exertion and pulse rate during graded exercise in various age groups. J Intern Med 1967 Jan 12;181(S472):194-206 [FREE Full text] [doi: 10.1111/j.0954-6820.1967.tb12626.x]
44. Borg GAV. Psychophysical bases of perceived exertion. Med Sci Sports Exerc 1982;14(5):377-381. [Medline: 7154893]
45. Moreno-Ger P, Torrente J, Hsieh YG, Lester WT. Usability testing for serious games: making informed design decisions with user data. Advances in Human-Computer Interaction 2012;2012:1-13. [doi: 10.1155/2012/369637]
46. Brooke J. SUS: a “quick and dirty” usability scale. In: Jordan PW, Thomas B, McClelland IL, Weerdmeester B, editors. Usability Evaluation in Industry: London, UK: Taylor & Francis; 1996:189-194.
47. Bango A, Kortum P, Miller J. Determining what individual SUS scores mean: adding an adjective rating scale. J Usability Stud 2009;4(3):114-123 [FREE Full text]
48. McAuley E, Duncan T, Tammen VV. Psychometric properties of the Intrinsic Motivation Inventory in a competitive sport setting: a confirmatory factor analysis. Res Q Exerc Sport 1989 Mar;60(1):48-58. [doi: 10.1080/02701367.1989.10607413] [Medline: 2489825]
49. Ryan RM, Deci EL. Intrinsic and extrinsic motivations: classic definitions and new directions. Contemp Educ Psychol 2000 Jan;25(1):54-67. [doi: 10.1006/ceps.1999.1020] [Medline: 10620381]
50. Full descripition of events identified. Zenodo. URL: https://zenodo.org/records/12161483 [accessed 2025-07-05]
51. Martikainen S, Korpela M, Tiihonen T. User participation in healthcare IT development: a developers’ viewpoint in Finland. Int J Med Inform 2014 Mar;83(3):189-200. [doi: 10.1016/j.ijmedinf.2013.12.003] [Medline: 24382475]
52. Kortum P, Peres SC. The relationship between system effectiveness and subjective usability scores using the system usability scale. Int J Hum Comput Interact 2014 Jul 3;30(7):575-584. [doi: 10.1080/10447318.2014.904177]
53. Pelliccia A, Sharma S, Gati S, et al. 2020 ESC Guidelines on sports cardiology and exercise in patients with cardiovascular disease. Eur Heart J 2021 Jan 1;42(1):17-96. [doi: 10.1093/eurheartj/ehaa605] [Medline: 32860412]
54. Physiological and mechanical parameters collected by the HEFMOB system during patient sessions. Zenodo. URL: https://zenodo.org/records/15322883 [accessed 2025-07-05]
55. Object acquisition during cycling and fruit-picking minigames. Zenodo. URL: https://zenodo.org/records/15323173 [accessed 2025-07-05]

Abbreviations list

CHF: chronic heart failure
CR: cardiac rehabilitation
HF: heart failure
IMI: intrinsic motivation inventory
IoT: internet of things
RPE: rate of perceived exertion
SGUE: Serious Game Usability Evaluator
SpO_2: peripheral capillary oxygen saturation
SUS: System Usability Scale
VR: virtual reality

Edited by A Coristine; submitted 25.01.25; peer-reviewed by C Halliwell, E Kircher, Z Karimi; revised version received 03.06.25; accepted 04.06.25; published 16.07.25.

Please cite as:
Blasco-Peris C, Alcolea Garrido JP, Seguí B, Zaragoza R, Climent-Paya V, Fuertes-Kenneally L, Manresa-Rocamora A, Sanz-Rocher A, Baladzhaeva S, Sarabia JM
Exergaming System for Exercise-Based Cardiac Rehabilitation in Patients With Heart Failure: Development and Usability Assessment Study of a Device Prototype
JMIR Serious Games 2025;13:e71385
URL: https://games.jmir.org/2025/1/e71385
doi:10.2196/71385

© Carles Blasco-Peris, Juan Pedro Alcolea Garrido, Barbara Seguí, Rocio Zaragoza, Vicente Climent-Paya, Laura Fuertes-Kenneally, Agustín Manresa-Rocamora, Ana Sanz-Rocher, Sabina Baladzhaeva, José M Sarabia. Originally published in JMIR Serious Games (https://games.jmir.org), 16.7.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution,

XSL•FO
RenderX

and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Mixed Reality–Based Physical Therapy in Older Adults With Sarcopenia: Preliminary Randomized Controlled Trial

Yeongsang An[1], PT, MS; Seunghwa Min[1], MS; Chanhee Park[2], Prof Dr, PT, PhD

[1]Funrehab Co., Daejeon, Republic of Korea

[2]Department of Physical Therapy, Jeonju University, 303, Cheonjam-ro, Wansan-gu, Jeonbuk-do, Jeonju, Republic of Korea

Corresponding Author:
Chanhee Park, Prof Dr, PT, PhD
Department of Physical Therapy, Jeonju University, 303, Cheonjam-ro, Wansan-gu, Jeonbuk-do, Jeonju, Republic of Korea

Abstract

Background: Sarcopenia in older adults is associated with reduced muscle mass and function, leading to frailty, increased fall risk, and decreased quality of life (QOL). Mixed reality (MR)–based interventions have emerged as promising tools to enhance physical therapy engagement and effectiveness through immersive, interactive environments.

Objective: This study aimed to investigate the effects of a Mixed Reality–Based Physical Therapy (Mr.PT) platform on quadricep muscle thickness, balance confidence, activities of daily living, and QOL in older adults with sarcopenia.

Methods: A preliminary randomized controlled trial was conducted involving 30 older women (mean age 75.3, SD 9.9 y) diagnosed with sarcopenia based on the Asian Working Group for Sarcopenia criteria. Participants were randomly assigned to either the Mr.PT group or a conventional physical activity (CPA) group. Both groups participated in 30-minute exercise sessions, 3 times per week, over 4 weeks. The Mr.PT program used head-mounted MR devices with gamified, interactive training, while the CPA group received standard therapist-led exercises using resistance bands. Outcome measures included ultrasound imaging of quadricep muscle thickness, the Korean version of the Activities of Daily Living scale (Katz Index of Independence in Activities of Daily Living), the Activities-Specific Balance Confidence (ABC) scale, and the 12-Item Short-Form Health Survey. Statistical analysis was performed using repeated-measures ANOVA and Tukey post hoc tests.

Results: The Mr.PT group showed significantly greater improvement in quadricep muscle thickness than the CPA group (P=.001). Within-group improvements in balance confidence (ABC scale) and daily functioning (Katz Index of Independence in Activities of Daily Living) were observed in both groups (P<.05), though between-group differences were not statistically significant. However, the Mr.PT group demonstrated significantly greater gains in QOL as measured by 12-Item Short-Form Health Survey (P=.02). All participants completed the intervention without dropouts or adverse events.

Conclusions: MR-based exercise using the Mr.PT platform appears effective in increasing muscle mass and enhancing QOL among older adults with sarcopenia. Its interactive and adaptive features may improve engagement and motivation, suggesting potential advantages over traditional programs. Further research with larger cohorts and longer follow-up is recommended to confirm these preliminary findings and explore long-term outcomes.

Trial Registration: Clinical Research Information Service KCT0010241; https://cris.nih.go.kr/cris/search/detailSearch.do?seq=29384

(JMIR Serious Games 2025;13:e76357) doi:10.2196/76357

KEYWORDS

sarcopenia; mixed reality rehabilitation; older adults; artificial intelligence; digital health

Introduction

Background

Mixed reality (MR)–based digital therapeutics have emerged as a novel and promising technology for addressing the growing health care demands of aging populations [1]. Sarcopenia, defined as the progressive loss of skeletal muscle mass and strength with advancing age, represents a significant geriatric concern associated with frailty, increased fall risk, impaired activities of daily living (ADL), and diminished quality of life (QOL) [2,3]. Quadricep muscle thickness is a reliable indicator of lower-limb strength, which is essential for mobility and fall prevention. Impairments in balance confidence and ADL are commonly associated with sarcopenia and contribute to loss of independence. Additionally, QOL often deteriorates in this population due to physical limitations [2,3]. This geriatric issue is widespread and progressive and can lead to hospitalization [4-6].

XSL•FO
RenderX

Rationale

Traditional interventions for sarcopenia, such as group-based or resistance exercise programs, offer clinical benefits but are limited by their standardized design, often requiring fixed schedules, group coordination, and in-person delivery [3,7]. These limitations were exacerbated during the COVID-19 pandemic, highlighting the urgent need for flexible, individualized therapeutic strategies [8,9]. Moreover, conventional group exercise typically lacks sufficient personalization, making it difficult to optimally engage and progress participants based on their functional capabilities.

To address these gaps, we developed a Mixed Reality–Based Physical Therapy (Mr.PT) platform designed to deliver personalized, engaging, and accessible exercise interventions. Mr.PT uses commercially available MR devices that integrate digital and real-world elements, enabling users to interact with digital environments while maintaining physical context [9,10]. Mr.PT comprises inexpensive, motivational, fun, and intensive devices [11]. The Mr.PT platform is designed to work with MetaQuest, and the software allows clinicians and clients free access to the platform without spatiotemporal constraints. The movement kinematics when reaching, grasping, transporting, and releasing objects in a digital environment are comparable to those in the physical world [12,13]. Unlike virtual reality (VR), which immerses the user in a fully simulated environment, MR overlays digital elements onto the real world and enables users to interact with both. This creates a more ecologically valid and functionally relevant experience, facilitating engagement and motor learning in rehabilitation settings [14]. Proprioceptive feedback has been suggested to exploit the multimodal aspects of correct active movements. However, clinical research on these systems has mainly focused on neurorehabilitation training rather than geriatric physical therapy.

Study Aim

While MR-based interventions have been explored extensively in the fields of neurorehabilitation and motor recovery, evidence supporting their use in geriatric physical therapy—particularly for sarcopenia management—is scarce. Thus, this study aimed to evaluate and compare the effects of Mr.PT and conventional physical activity (CPA) programs on quadricep muscle thickness, balance confidence, ADL independence, and QOL among older adults with sarcopenia.

Given the accelerating trends of global aging, especially in countries like South Korea and Japan, there is an urgent need to develop accessible, scalable, and individualized rehabilitation interventions. MR-based platforms such as Mr.PT may offer sustainable solutions to enhance functional outcomes and support independent living among older adults.

Methods

Participants

A convenience sample of 30 older adults with sarcopenia (mean age 75.3, SD 9.9 y; all female) was recruited from a geriatric welfare center. Inclusion criteria were as follows: (1) 65 years and older, and (2) diagnosis of sarcopenia by a medical doctor based on established clinical criteria. Sarcopenia was diagnosed according to the 2019 criteria of the Asian Working Group for Sarcopenia, including low skeletal muscle mass (Skeletal Muscle Index<5.7 kg/m² by BIA), low muscle strength (grip strength <18 kg), and poor physical performance [15]. Basic demographic information (age, sex, height, weight, and BMI) was collected during baseline assessment using a standardized intake form completed by participants.

Ethical Considerations

This study was approved by the Institutional Review Board of Jeonju University (Institutional Review Board number jjIRB-250206-HR-2024 - 1203) prior to participant enrollment. All participants provided written informed consent before participation. All personal data were deidentified prior to analysis to ensure confidentiality and privacy. Only study ID numbers were used in the dataset. Participants received a small nonmonetary gift (health supplement worth approximately 30,000 KRW [US $23]) as compensation for their time. No personally identifiable images or data were included in the manuscript or supplementary materials. All ultrasound and evaluation images were anonymized. Although the clinical trial registration (KCT0010241) was finalized after study completion due to administrative delays, the study was retrospectively registered to ensure transparency. Recruitment and data collection were completed before the registration was officially approved.

Randomization and Blinding

Participants were randomized into either the Mr.PT group or the CPA group using a simple coin-flip method. Randomization occurred prior to baseline assessments to minimize bias. Assessments were conducted by blinded evaluators who were unaware of group assignments, ensuring a single-blinded study design.

Experimental Setup

A procedural checklist was used to standardize all experimental procedures. Equipment was calibrated daily before data collection. Baseline assessments included quadricep muscle thickness, the Activities-Specific Balance Confidence (ABC) scale, the Katz Index of Independence in Activities of Daily Living (KIADL), and the 12-Item Short-Form Health Survey (SF-12). Both groups were supervised by licensed physical therapists during exercise sessions. For the Mr.PT group, a computer technician was present during interventions to address potential technical issues but did not interact with participants or influence the sessions.

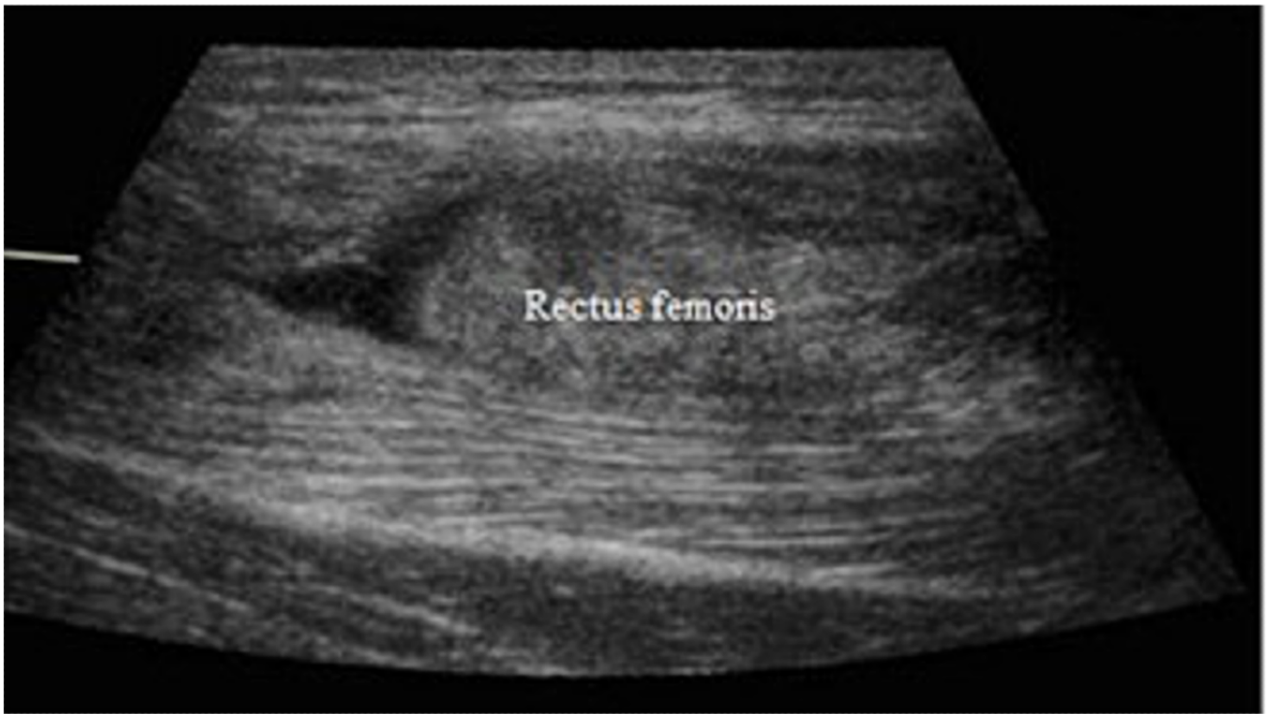
Ultrasound Imaging Measurement of Quadricep Muscle Thickness

An ultrasound probe (Bionet, Seoul, Republic of Korea) was used to obtain images of the rectus femoris muscle. All 2D ultrasound imaging acquisitions were acquired using an ultrasonographic system equipped with a single-sweep volumetric transducer. The apparatus had a linear vascular probe with a bandwidth of 3.3 - 8.0 MHz, an aperture of 2.9 cm, and a maximum scanning depth of 8 cm, in addition to a phased array cardiac probe with a bandwidth of 1.7 - 3.8 MHz and a field of view of 70° [16]. All participants were scanned while

XSL•FO
RenderX

lying comfortably in the supine position. According to a previous study, the examiner positioned the probe on the anterior aspect of the quadriceps, perpendicular to its long axis, halfway between the proximal end of the patella and the anterior superior iliac spine. The examiner identified the rectus femoris, vastus intermedius, femur, and subcutaneous adipose tissues. To reduce artifacts, the skin was treated with excess gel. To examine the inter- and intra-rater reliability of the bilateral limbs, 3 examiners acquired pictures [16]. Among the 3 examiners were 3 physical therapists. The examiners were experts who had used ultrasound for evaluation in a clinical context for a minimum of 3 years. To prevent variations in the measurement and analysis of muscle data, each trial was performed by the 3 examiners separately within 2 hours of the initial assessment [16]. The top border of the rectus femoris was marked with a pointer on the screen of the ultrasound device (Figure 1).

Figure 1. Rectus femoris ultrasound imaging.



This enabled the device to determine the total rectus femoris muscle thickness. Three measurements were made by each examiner to determine intrarater reliability [16,17]. The participants returned to their starting position after every investigation, and any gel or markings were removed from their skin by cleaning it. This minimized the possibility of measurement bias, such as anchoring, and guaranteed independent acquisition of each image and dataset.

Activities-Specific Balance Confidence (ABC) Scale

An indirect indicator of falling risk, the ABC scale, was used to gauge the participants' confidence in their ability to perform a range of ambulatory tasks without feeling unsteady or falling. The 16 ABC scale components have scores ranging from 0 (no confidence) to 100 (full confidence) [18]. The validity and reliability of the measurement tests have been thoroughly proven elsewhere [18].

Katz Index of Independence in Activities of Daily Living

The best tool for determining a client's functional status is a measure of their capacity to perform everyday tasks independently [19]. Age-related changes and health problems frequently result in a decline in the functional status of older adults. This decline often places older adults in a downward health spiral. An effective way to evaluate the health status of older adults is through their functional abilities [19]. An objective assessment that provides objective data helps indicate a decline or improvement in health status, allowing physical therapists to plan and intervene appropriately. The validity and reliability of the measurement tests have been thoroughly proven [19,20].

Geriatric Quality of Life

Twelve items comprise the SF-12, which is a health-related QOL survey. Using this information, we evaluated 8 health areas, including mental and physical health [21]. General health, bodily discomfort, physical roles, and physical functioning were domains associated with physical health. Vitality, social functioning, emotional roles, and mental health are among the scales pertaining to mental health. The validity and reliability of outcome measurement tests have already been demonstrated in other contexts [22].

Intervention

Overview

Both Mr.PT and CPA were assigned to participants in a randomized controlled trial, with a frequency of 30 min, and administered 5 times per week for 4 weeks. All participants were assessed using 2 (pre and post) evaluations.

XSL•FO
RenderX

Mr.PT Group

Participants in the Mr.PT group engaged in an MR-based exercise program using the MetaQuest head-mounted display system (Funrehab, Daejeon, Republic of Korea). The intervention was designed to simulate a variety of functional and cognitive tasks within an interactive digital environment that seamlessly integrated with real-world movement [12,23]. Each session began with participants standing or seated in a safe area, where they wore the MR device and received brief instructions on the day's activities. Initial exercises involved evading digital objects approaching from various directions, requiring participants to shift their weight, adjust their posture, and take small steps to avoid collisions. This activity was intended to enhance dynamic balance, postural control, and lower limb reaction times [10]. The Mr.PT program included interactive MR tasks in which participants reached for and tracked digital objects projected through a head-mounted display. The speed of the digital objects ranged from 1.2 to 2.5 m/s, with target distances between 0.5 and 2.0 meters. The system automatically adjusted difficulty based on task success rates and provided real-time verbal and visual feedback.

Subsequent tasks incorporated upper extremity movements, where participants were asked to "shoot" at digital targets appearing at random locations within their field of vision by performing rapid and precise shoulder abduction or flexion movements. This shooting task aimed to improve upper body coordination, reaction speed, spatial awareness, and proprioception [24]. As participants progressed through the program, combined tasks were introduced, simultaneously requiring both footwork (for dodging) and shoulder movements (for targeting) to complete challenges [25]. These integrated activities were specifically designed to train motor planning, divided attention, and cognitive-motor multitasking, which are often impaired in older adults with sarcopenia. The difficulty level of the digital tasks was automatically adjusted based on participant performance, ensuring an appropriate balance between challenge and skill to maintain motivation and maximize engagement (Figure 2) [26].

Verbal encouragement and visual feedback were provided throughout the sessions by the MR system, further supporting adherence and participant satisfaction. Each MR-based exercise session lasted 30 minutes and was conducted 5 times per week over the 4-week intervention period, under the supervision of a trained physical therapist and a computer technician who managed technical setup but did not influence participant behavior (Figure 3).

Figure 2. Mixed reality platform–based physical exercise therapy.

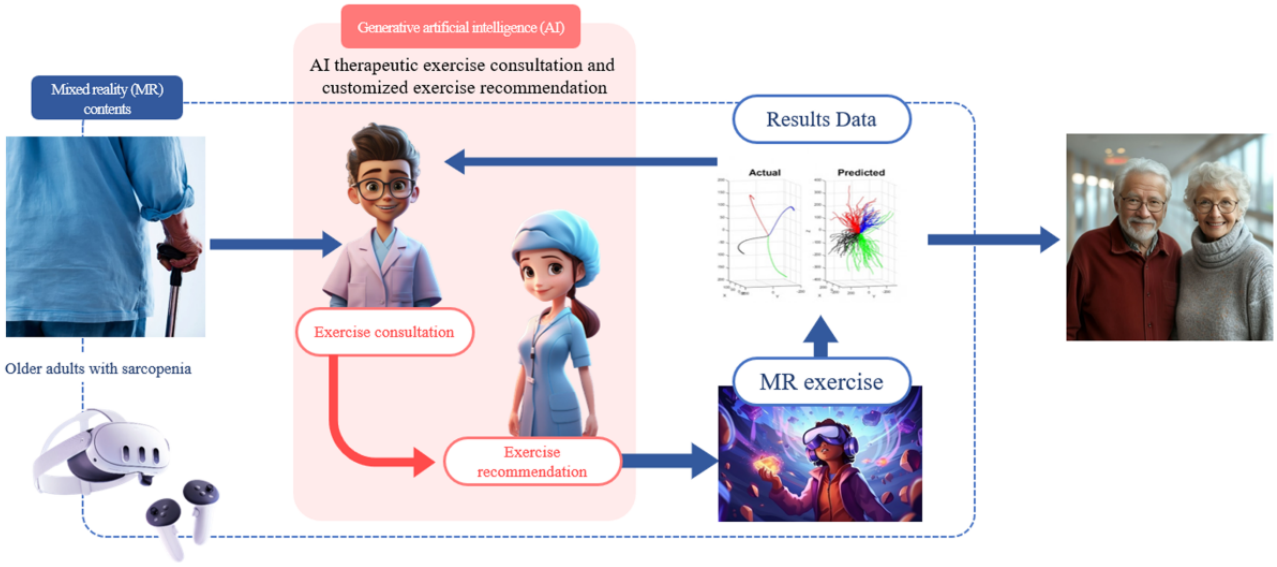
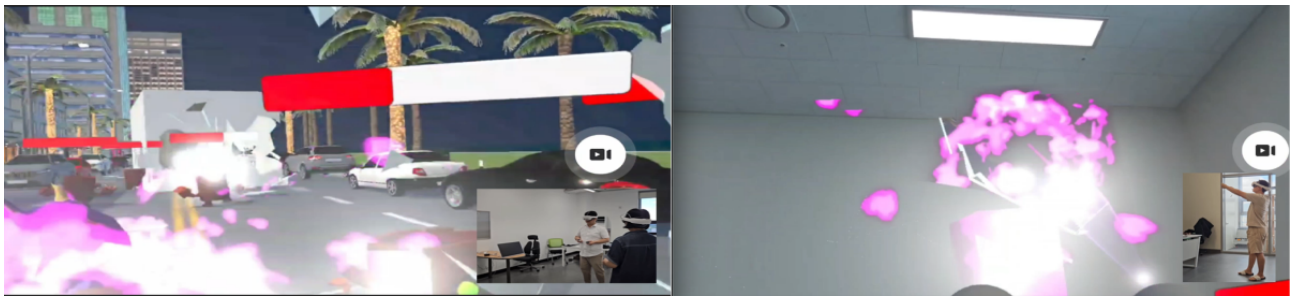


Figure 3. Mixed reality platform.



CPA Group

Participants assigned to the CPA group performed a conventional, therapist-led exercise program focused on maintaining and improving musculoskeletal strength, balance, and mobility. Sessions took place in a standard exercise room equipped with balance mats, chairs, and minimal resistance equipment [27]. Each CPA session consisted of warm-up

XSL•FO
RenderX

activities such as marching in place and dynamic stretching, followed by structured moderate-intensity gait training across flat surfaces. Gait exercises incorporated variations in speed and direction to promote adaptability and coordination. Strength training exercises emphasized the major muscle groups responsible for functional mobility. Participants performed sets of hip flexion, extension, and abduction exercises, as well as knee flexion and extension drills, often using bodyweight resistance or elastic bands to adjust intensity [24]. Balance training involved static and dynamic activities, including one-legged standing with eyes open, tandem standing, and tandem walking along a straight line. Each exercise was modified according to individual ability, with progression achieved by increasing task complexity or duration. Upper extremity exercises targeted shoulder abduction, extension, and flexion using light dumbbells or resistance bands, aiming to maintain upper limb functional strength necessary for daily tasks. The CPA program consisted of therapist-supervised exercises using elastic resistance bands of light to medium strength (1.5 - 3.0 kgf). Exercises included lower-limb strength, balance, and gait training with 10 - 15 repetitions per exercise [14,28]. The Mr.PT group performed individualized MR-based tasks with real-time visual and verbal feedback, adjusting difficulty based on task success. The CPA group received therapist-led physical exercises using elastic resistance bands and balance training. All sessions were conducted individually to control dosage and engagement. Both interventions were conducted 3 times per week for 4 weeks. All 30 participants completed the intervention without dropouts. No adverse effects, such as dizziness or fatigue, were reported during or after the sessions.

Statistical Analysis

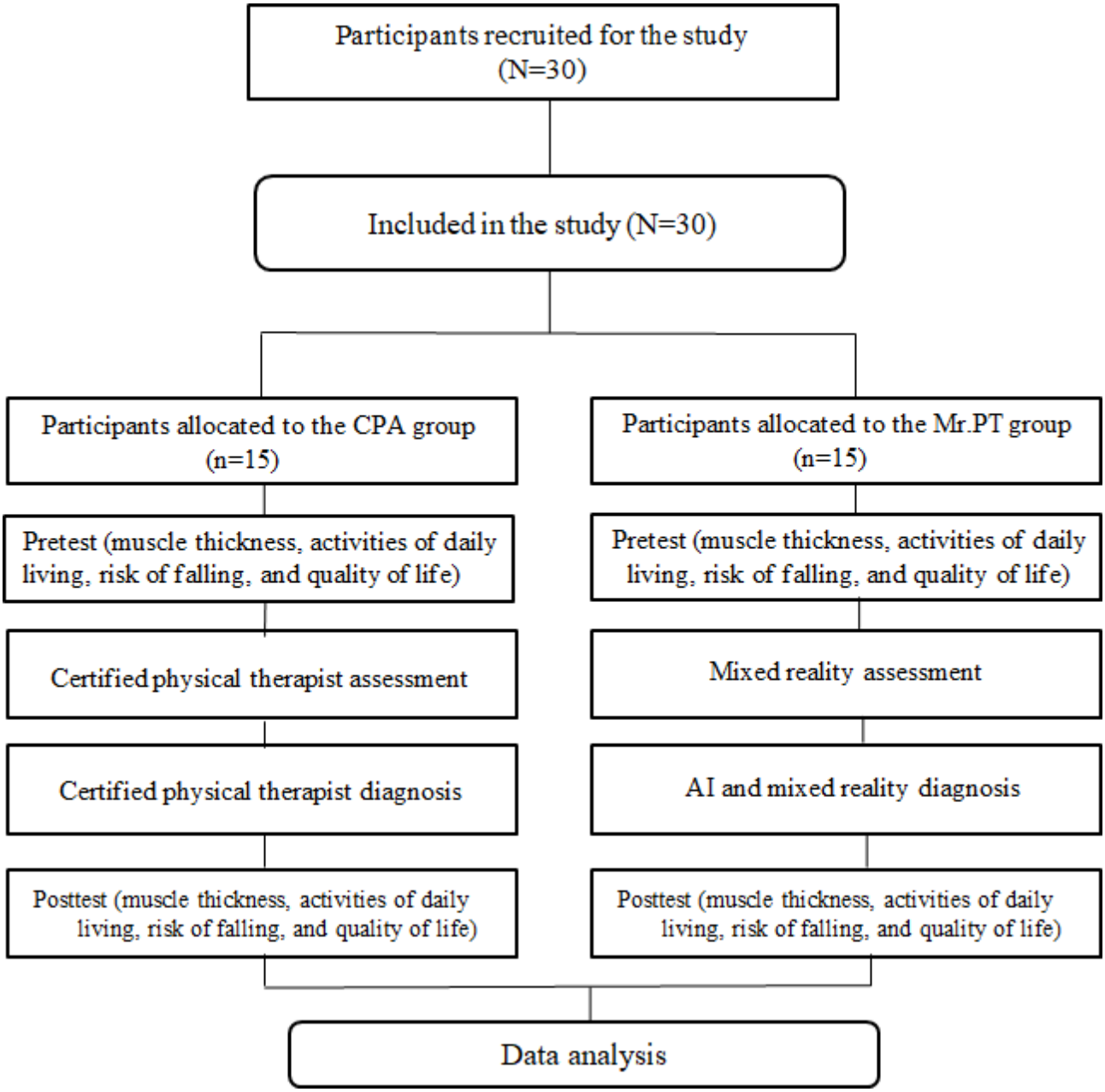
Descriptive statistics (mean, SD) were calculated for all continuous variables. The Kolmogorov-Smirnov test confirmed normality assumptions. Repeated-measures ANOVA was used to compare pre- and postintervention changes within and between groups. Tukey's post hoc tests were applied to examine significant interactions. The chi-square test was used for categorical data (eg, sex). Power analysis using G*Power 3.1.9.7 indicated that a sample size of 27 participants would achieve 80% power with an effect size of 0.5 (α=.05); thus, 30 participants were recruited to allow for potential dropout. Observed power was calculated post hoc for key outcomes. A significance threshold of P<.05 was used for all analyses. Statistical analyses were conducted using SPSS (version 26.0; IBM Corp) was used for the statistical analyses. A P value of .05 was determined.

Results

Figure 4 shows the study flow chart.

XSL•FO
RenderX

Figure 4. CONSORT (Consolidated Standards of Reporting Trials) flow chart. AI: artificial intelligence; CPA: conventional physical activity; Mr.PT: Mixed Reality–Based Physical Therapy.



Demographic Data

Table 1 shows the demographic characteristics of the participants. Each participant completed all experimental tasks and exercises. The CPA and Mr.PT groups did not differ significantly in terms of sex, age, height, weight, or BMI (P>.05).

XSL•FO
RenderX

Table . Participants' demographic characteristics (N=30).

Characteristics	Mr.PT[a] (n=15)	CPA[b] (n=15)	*P* value
Age (years), mean (SD)	76.24 (8.7)	74.88 (9.1)	.68
Sex, n (%)			>.99
Male	0 (0)	0 (0)	
Female	15 (100)	15 (100)	
Height (cm), mean (SD)	153.12 (10.11)	155.88 (8.21)	.31
Weight (kg), mean (SD)	54.13 (6.13)	56.27 (4.88)	.45
BMI (kg/m^2), mean (SD)	20.11 (1.81)	20.83 (0.91)	.23

[a]Mr.PT: Mixed Reality–Based Physical Therapy.

[b]CPA: conventional physical activity.

Quadricep Muscle Thickness and KIADL Measurements

Repeated-measures ANOVA revealed a significant time effect (*P*=.001) for quadricep muscle thickness and KIADL, indicating improved muscle thickness and ADL for participants. No significant difference in KIADL scores was observed between the groups (*P*=.09). The interaction post hoc test revealed a greater improvement in quadricep muscle thickness in Mr.PT than in CPA, suggesting that quadricep muscle thickness is related to sarcopenia after Mr.PT than in CPA (Table 2).

Table . Quadricep muscle thickness and Katz Index of Independence in Activities of Daily Living (KIADL) measurement.

	Mr.PT[a], mean (SD)		CPA[b], mean (SD)		*P* value		
	Pretest	Posttest	Pretest	Posttest	Time effect	Between groups	Time × group
Quadricep muscle thickness (mm)	15.83 (1.83)	17.13 (1.77)	15.71 (2.01)	16.22 (1.38)	.001[c]	.02[c]	.001[c]
KIADL	3.90 (0.32)	4.20 (0.43)	3.80 (0.37)	4.10 (0.41)	.001[c]	.09	.08

[a]Mr.PT: Mixed Reality–Based Physical Therapy.

[b]CPA: conventional physical activity.

[c]*P*<.05.

ABC Scale and SF-12 Measurements

Repeated-measures ANOVA revealed a significant time effect (*P*=.001) for the ABC scale and SF-12, indicating improved fall confidence and QOL for participants. No significant difference in the ABC scale scores was observed between the groups (*P*=.23). The interaction post hoc test revealed a greater improvement in SF-12 in Mr.PT compared with CPA, suggesting that QOL is related to sarcopenia after Mr.PT compared with CPA (Table 3).

Table . Activities-Specific Balance Confidence (ABC) scale and 12-Item Short-Form Health Survey (SF-12).

	Mr.PT[a], mean (SD)		CPA[b], mean (SD)		*P* value		
	Pretest	Posttest	Pretest	Posttest	Time effect	Between groups	Time × group
ABC scale	1223.48 (253.12)	1372.18 (178.55)	1218.68 (188.12)	1349.88 (163.88)	.001[c]	.23	.41
SF-12	60.84 (4.88)	74.20 (7.88)	63.18 (3.78)	74.10 (8.31)	.001[c]	.02[c]	.04[c]

[a]Mr.PT: Mixed Reality–Based Physical Therapy.

[b]CPA: conventional physical activity.

[c]*P*<.05.

Discussion

Principal Findings

The Mr.PT intervention led to a 9.1% increase in quadricep muscle thickness, which aligns with previous studies on immersive VR-based rehabilitation. While functional gains such as gait speed or grip strength were not assessed, literature suggests that each 1 mm increase in quadricep thickness may correspond to approximately a 5% increase in knee extensor force. The improvement in SF-12 (4.68%) exceeds the established minimal clinically important difference of 3 points, supporting its clinical relevance [29,30]. Although Mr.PT

XSL•FO
RenderX

showed superior effects in muscle thickness and QOL, it is important to recognize the practical advantages of the CPA program. CPA involves simple, therapist-guided physical activity using minimal equipment, making it highly scalable and cost-effective for widespread community-based use. The lack of significant between-group differences in ADL and balance confidence (ABC) may reflect the short intervention duration or the limited sensitivity of the assessment tools to detect small changes. Alternatively, it is possible that CPA, despite its simplicity, was equally effective in maintaining these functions in the short term.

Comparison to Prior Work

The quadricep muscle thickness–related sarcopenia analysis demonstrated that the improvement in muscle thickness (9.1%) was significantly different between the groups. This finding is consistent with those of the previous studies. As previously reported, increased muscle strength (10.2%) was observed using VR compared with the control group in 30 participants with sarcopenia [31]. As previously reported, increased muscle parameters (8.8%) were observed using an augmented reality-based exercise program compared with the control group of 30 participants with sarcopenia [32].

VR training has shown promising results in improving physical function and muscle strength in older adults. An 8-week unsupervised VR exercise program significantly enhanced hip muscle strength and balance control in older adults compared with a control group [33]. Similarly, a VR fitness program improved muscular fitness and body composition in young untrained men, although some measures showed less improvement than traditional exercise [34]. VR training also increased lower extremity muscle activation in older participants after an 8-week intervention using a Wii Fit [24]. Previous studies have highlighted the potential of MR technologies to enhance engagement, cognitive-motor integration, and exercise adherence in various populations. As previously demonstrated, muscle thickness assessed by ultrasound can effectively estimate knee extension muscle strength in females [35]. Our findings align with this growing body of evidence and suggest that MR-based exercise interventions may offer meaningful advantages for older adults at risk of sarcopenia.

The activities of daily living analysis demonstrated that the improvement in KIADL (7.7% and 7.9%, respectively) was significantly different pre- and postintervention in both groups. This finding is consistent with those of the previous studies. Liao et al [36] found increased ADL function (7.73%) using VR compared with the control group in 34 older adults with mild cognitive impairment. As previously reported, increased ADL (10.23%) was observed using VR-based applications compared with the control group in older people [37]. As previously found, improved executive function was observed using a VR-based program (22.5%) compared with the control group in 50 community-dwelling older individuals [27]. Gamification platforms increase user loyalty by providing highly immersive, engaging, and entertaining digital environments for various locations and challenges. Consistent with previous analyses, our findings imply that the application of commercial VR systems, including the Xbox Kinect, PlayStation, and Nintendo Wii, may be advantageous and practical for enhancing functional daily living in geriatrics [38,39].

We found that MR was more effective and provided more significant evidence in the field of occupation and ambulation in ADL performance [23,25,26]. VR has emerged as a promising tool for cognitive rehabilitation and training in ADL for patients with neurological disorders. Studies have shown that digital interventions can improve cognitive functioning, attention, memory, and executive function in patients with stroke [5,13]. VR systems allow for personalized progressive training in ADL tasks by providing real-time feedback to patients. These systems offer immersive and realistic environments that simulate everyday activities, making rehabilitation more engaging and motivating for users [40]. Research has also demonstrated the feasibility and usability of VR-based ADL training, with handheld controllers proving to be more intuitive and less demanding than traditional mouse inputs [41]. Overall, VR technology has great potential for enhancing cognitive rehabilitation by offering ecologically valid and personalized training experiences that can effectively supplement conventional therapeutic methods.

The fall confidence analysis demonstrated that the improvement in the ABC scale (12.15% and 10.97%) was significantly different pre and postintervention in both groups. This finding is consistent with those of the previous studies. Using Xbox gaming, as previously found, decreases the risk of falling (7.91%) compared with the control group in 60 older individuals [42]. People find VR games emotionally and physically taxing, and when playing, the players' motions constantly change direction, pace, and speed. Therefore, rapid brain processing, adjustments, and control of body balance are required [26,41]. In addition, older adults can effectively use safe and reasonably priced technologies in their homes [43]. VR technology is promising for fall prevention and improvement in confidence in older adults. VR balance exercise programs can reduce the risk of falling and fear of falling in community-dwelling women [44]. A study using a balance rehabilitation unit found significant improvements in balance parameters, reduced falls, and lower fear of falling in older adults with a history of falls [45]. However, the effectiveness of VR interventions depends on user acceptance. Research on VR headsets for fall prevention has found that the intention to use was positively predicted by perceived usefulness, enjoyment, and ease of use [46].

QOL analysis demonstrated that the improvement in SF-12 (4.68%) was significantly different between the groups. This finding is consistent with those of the previous studies. As previously reported, an improved QOL (15.9%) was observed using VR compared with a control group in 60 older adults [43]. As previously found, an increased QOL (0.37%) was observed using a semi-immersive VR intervention compared with the control group in 100 older adults [47]. VR-based interventions have shown promising results in improving older adults' QOL. Studies have demonstrated that VR can enhance the physical, cognitive, psychological, and social well-being of this population [47]. MR-based exercises have been found to significantly improve overall QOL, mental and physical health, social relationships, and depression symptoms in older adults [13,23]. While VR, augmented, and MR applications have become more

XSL•FO
RenderX

developed in the physical and cognitive domains, research in the areas of psychological and social well-being is still emerging [48]. These findings suggest that digital therapeutic technology has a significant potential to enhance various aspects of older adults' QOL. The superior outcomes in the Mr.PT group may be attributed to enhanced cognitive engagement, task variability, and the real-time feedback provided by the MR environment, which may elicit greater neuromuscular activation compared with the more repetitive nature of CPA.

Limitations

This study has several limitations that should be addressed in future research. First, the sample consisted only of older women, which limits generalizability. Second, ultrasound-based measurements may involve some subjectivity despite being conducted by trained personnel. Third, the 4-week intervention duration may have been insufficient to produce long-term or functional improvements beyond muscle hypertrophy. Although we did not directly assess neurophysiological changes, it is plausible that MR-related improvements are partly mediated by cortical or subcortical adaptations. Confounders such as participant motivation, novelty effects, and attentional demand may also have influenced outcomes. Future studies should consider longer interventions and include more diverse participants and direct measures of functional strength.

Conclusion

Clinical research has indicated that MR platform-facilitated therapeutic exercise is beneficial for improving muscle thickness, ADL, fall confidence, and QOL in older adults with sarcopenia. Our results provide therapeutic evidence that Mr.PT improves the recovery of quadricep muscle thickness and QOL in patients with depression and anxiety. Most importantly, the Mr.PT platform allows autonomous provision of real-time audiovisual feedback and effective training, which could serve as a basis for advanced rehabilitation sciences and medical evidence.

Acknowledgments

This research was supported by the Research Grant of Jeonju University in 2024. This work was supported by the Collabo R&D between Industry, Academy, and Research Institute (RS-2025-02306808) funded by the Ministry of SMEs and Startups (MSS, Korea). The authors did not use generative artificial intelligence (such as ChatGPT or other large language models) to write or generate content for this manuscript.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

YA contributed to the conceptualization, methodology, and investigation of the study, and prepared the original draft of the manuscript. SM was responsible for data curation and formal analysis and participated in reviewing and editing the manuscript. CP provided supervision, managed the project administration, secured funding, and contributed to the review and editing of the manuscript.

Conflicts of Interest

None declared.

Editorial Notice

This randomized study was retrospectively registered on February 28, 2025, as explained by the authors due to administrative delays. Institutional review board approval was granted on February 06, 2025, the date of the first patient enrolment was February 14, 2025, and the last patient data were collected on March 28, 2025. Given this timeline, and because the risk of bias appears low, the editor granted an exception from International Committee of Medical Journal Editors (ICMJE) rules mandating prospective registration of randomized trials. However, readers are advised to carefully assess the validity of any potential explicit or implicit claims related to primary outcomes or effectiveness due to the retrospective registration.

Checklist 1
CONSORT (Consolidated Standards of Reporting Trials) checklist.
[PDF File, 2776 KB - games_v13i1e76357_app1.pdf]

References

1. Mascret N, Delbes L, Voron A, Temprado JJ, Montagne G. Acceptance of a virtual reality headset designed for fall prevention in older adults: questionnaire study. J Med Internet Res 2020 Dec 14;22(12):e20691. [doi: 10.2196/20691] [Medline: 33315019]
2. Cruz-Jentoft AJ, Sayer AA. Sarcopenia. Lancet 2019 Jun 29;393(10191):2636-2646. [doi: 10.1016/S0140-6736(19)31138-9] [Medline: 31171417]

XSL•FO
RenderX

3. Moore SA, Hrisos N, Errington L, et al. Exercise as a treatment for sarcopenia: an umbrella review of systematic review evidence. Physiotherapy 2020 Jun;107:189-201. [doi: 10.1016/j.physio.2019.08.005] [Medline: 32026819]
4. Beaudart C, Demonceau C, Reginster JY, et al. Sarcopenia and health-related quality of life: a systematic review and meta-analysis. J Cachexia Sarcopenia Muscle 2023 Jun;14(3):1228-1243. [doi: 10.1002/jcsm.13243] [Medline: 37139947]
5. Duque G, Boersma D, Loza-Diaz G, et al. Effects of balance training using a virtual-reality system in older fallers. Clin Interv Aging 2013;8:257-263. [doi: 10.2147/CIA.S41453] [Medline: 23467506]
6. Nascimento CM, Ingles M, Salvador-Pascual A, Cominetti MR, Gomez-Cabrera MC, Viña J. Sarcopenia, frailty and their prevention by exercise. Free Radic Biol Med 2019 Feb 20;132:42-49. [doi: 10.1016/j.freeradbiomed.2018.08.035] [Medline: 30176345]
7. Vlietstra L, Hendrickx W, Waters DL. Exercise interventions in healthy older adults with sarcopenia: a systematic review and meta-analysis. Australas J Ageing 2018 Sep;37(3):169-183. [doi: 10.1111/ajag.12521] [Medline: 29638028]
8. Kim JJ, Gu S, Lee JJ, Kim YS, Yoon BC. The effects of virtual reality-based continuous slow exercise on factors for falls in the elderly. J Korean Phys Ther 2012 Apr;24(2):90-97 [FREE Full text]
9. Karnadipa T, Santoso I, Pratama AD, Aulia NN. The effectiveness of virtual reality exercise to reduce risks of fall in elderly with balance disorders: a literature review. Presented at: International Conference on Vocational Education Applied Science and Technology; Oct 26-28, 2022; Teluk Betung, Indonesia. [doi: 10.3390/proceedings2022083002]
10. Tursø-Finnich T, Jensen RO, Jensen LX, Konge L, Thinggaard E. Virtual reality head-mounted displays in medical education: a systematic review. Simul Healthc 2023 Feb 1;18(1):42-50. [doi: 10.1097/SIH.0000000000000636] [Medline: 35136005]
11. Yoo SZ, No MH, Heo JW, et al. Role of exercise in age-related sarcopenia. J Exerc Rehabil 2018 Aug;14(4):551-558. [doi: 10.12965/jer.1836268.134] [Medline: 30276173]
12. Cho KH, Lee KJ, Song CH. Virtual-reality balance training with a video-game system improves dynamic balance in chronic stroke patients. Tohoku J Exp Med 2012;228(1):69-74. [doi: 10.1620/tjem.228.69]
13. Fu Y, Hu Y, Sundstedt V. A systematic literature review of virtual, augmented, and mixed reality game applications in healthcare. ACM Trans Comput Healthcare 2022 Apr 30;3(2):1-27. [doi: 10.1145/3472303]
14. Stone J. Extended reality therapy: the use of virtual, augmented, and mixed reality in mental health treatment. In: The Video Game Debate 2: Routledge; 2020:95-106. [doi: 10.4324/9780429351815-8]
15. Chen LK, Woo J, Assantachai P, et al. Asian working group for sarcopenia: 2019 consensus update on sarcopenia diagnosis and treatment. J Am Med Dir Assoc 2020 Mar;21(3):300-307. [doi: 10.1016/j.jamda.2019.12.012]
16. Pasta G, Nanni G, Molini L, Bianchi S. Sonography of the quadriceps muscle: examination technique, normal anatomy, and traumatic lesions. J Ultrasound 2010 Jun;13(2):76-84. [doi: 10.1016/j.jus.2010.07.004] [Medline: 23396806]
17. Li J, Zhou Y, Lu Y, Zhou G, Wang L, Zheng YP. The sensitive and efficient detection of quadriceps muscle thickness changes in cross-sectional plane using ultrasonography: a feasibility investigation. IEEE J Biomed Health Inform 2014 Mar;18(2):628-635. [doi: 10.1109/JBHI.2013.2275002] [Medline: 24608062]
18. Stasny BM, Newton RA, Viggiano LoCascio L, et al. The ABC scale and fall risk: a systematic review. Phys Occup Ther Geriatr 2011 Aug 22;29(3):233-242. [doi: 10.3109/02703181.2011.572249]
19. Shelkey M, Wallace M. Katz index of independence in activities of daily living. J Gerontol Nurs 1999 Mar;25(3):8-9. [doi: 10.3928/0098-9134-19990301-05] [Medline: 10362969]
20. Cabañero-Martínez MJ, Cabrero-García J, Richart-Martínez M, Muñoz-Mendoza CL. The Spanish versions of the barthel index (BI) and the katz index (KI) of activities of daily living (ADL): a structured review. Arch Gerontol Geriatr 2009;49(1):e77-e84. [doi: 10.1016/j.archger.2008.09.006] [Medline: 18990459]
21. Jakobsson U. Using the 12-item short form health survey (SF-12) to measure quality of life among older people. Aging Clin Exp Res 2007 Dec;19(6):457-464. [doi: 10.1007/BF03324731] [Medline: 18172367]
22. Shou J, Ren L, Wang H, et al. Reliability and validity of 12-item short-form health survey (SF-12) for the health status of Chinese community elderly population in Xujiahui district of Shanghai. Aging Clin Exp Res 2016 Apr;28(2):339-346. [doi: 10.1007/s40520-015-0401-9] [Medline: 26142623]
23. Chignell M, Matulis H, Nejati B. Motivating physical exercise in the elderly with mixed reality experiences. In: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Springer; 2020, Vol. 12203 LNCS:505-519. [doi: 10.1007/978-3-030-50344-4_36]
24. Cho GH, Shin HS. The effect of virtual reality training on lower extremity muscle activation in elderly. J Korean Soc Phys Med 2014 Feb 28;9(1):55-62. [doi: 10.13066/kspm.2014.9.1.55]
25. Papadopoulos T, Evangelidis K, Kaskalis TH, Evangelidis G, Sylaiou S. Interactions in augmented and mixed reality: an overview. Appl Sci (Basel) 2021;11(18):8752. [doi: 10.3390/app11188752]
26. Rokhsaritalemi S, Sadeghi-Niaraki A, Choi SM. A review on mixed reality: current trends, challenges and prospects. Appl Sci (Basel) 2020 ;10(2):636. [doi: 10.3390/app10020636]
27. Gamito P, Oliveira J, Alves C, Santos N, Coelho C, Brito R. Virtual reality-based cognitive stimulation to improve cognitive functioning in community elderly: a controlled study. Cyberpsychol Behav Soc Netw 2020 Mar;23(3):150-156. [doi: 10.1089/cyber.2019.0271] [Medline: 32031888]

XSL•FO
RenderX

28. Corregidor-Sánchez AI, Segura-Fragoso A, Criado-Álvarez JJ, Rodríguez-Hernández M, Mohedano-Moriano A, Polonio-López B. Effectiveness of virtual reality systems to improve the activities of daily life in older people. Int J Environ Res Public Health 2020 Aug 28;17(17):32872313. [doi: 10.3390/ijerph17176283] [Medline: 32872313]
29. Tanimoto Y, Watanabe M, Sun W, et al. Association of sarcopenia with functional decline in community‐dwelling elderly subjects in J apan. Geriatrics Gerontology Int 2013 Oct;13(4):958-963. [doi: 10.1111/ggi.12037]
30. Hays RD, Sherbourne CD, Mazel RM. The RAND 36-item health survey 1.0. Health Econ 1993 Oct;2(3):217-227. [doi: 10.1002/hec.4730020305] [Medline: 8275167]
31. Chen GB, Lin CW, Huang HY, et al. Using virtual reality-based rehabilitation in sarcopenic older adults in rural health care facilities-a quasi-experimental study. J Aging Phys Act 2021 Oct 1;29(5):866-877. [doi: 10.1123/japa.2020-0222] [Medline: 33596540]
32. Jeon S, Kim J. Effects of augmented-reality-based exercise on muscle parameters, physical performance, and exercise self-efficacy for older adults. Int J Environ Res Public Health 2020 May 7;17(9):3260. [doi: 10.3390/ijerph17093260] [Medline: 32392833]
33. Kade D, Akşit K, Ürey H, Özcan O. Head-mounted mixed reality projection display for games production and entertainment. Pers Ubiquit Comput 2015 Jul;19(3-4):509-521. [doi: 10.1007/s00779-015-0847-y]
34. Shimazaki A, Sugiura Y, Mikami D, Kimura T, Sugimoto M. MuscleVR: detecting muscle shape deformation using a full body suit. Presented at: Proceedings of the 8th Augmented Human International Conference; Mar 16-18, 2017; Silicon Valley, California, USA. [doi: 10.1145/3041164.3041184]
35. Muraki S, Fukumoto K, Fukuda O. Prediction of the muscle strength by the muscle thickness and hardness using ultrasound muscle hardness meter. Springerplus 2013;2(1):457. [doi: 10.1186/2193-1801-2-457] [Medline: 24058896]
36. Liao YY, Chen IH, Lin YJ, Chen Y, Hsu WC. Effects of virtual reality-based physical and cognitive training on executive function and dual-task gait performance in older adults with mild cognitive impairment: a randomized control trial. Front Aging Neurosci 2019;11:162. [doi: 10.3389/fnagi.2019.00162] [Medline: 31379553]
37. Arlati S, Colombo V, Spoladore D, et al. A social virtual reality-based application for the physical and cognitive training of the elderly at home. Sensors (Basel) 2019 Jan 10;19(2):261. [doi: 10.3390/s19020261] [Medline: 30634719]
38. Rand D, Kizony R, Weiss PTL. The Sony PlayStation II eyetoy: low-cost virtual reality for use in rehabilitation. J Neurol Phys Ther 2008 Dec;32(4):155-163. [doi: 10.1097/NPT.0b013e31818ee779] [Medline: 19265756]
39. Flynn S, Palma P, Bender A. Feasibility of using the Sony PlayStation 2 gaming platform for an individual poststroke: a case report. J Neurol Phys Ther 2007 Dec;31(4):180-189. [doi: 10.1097/NPT.0b013e31815d00d5] [Medline: 18172415]
40. Yousefi Babadi S, Daneshmandi H. Effects of virtual reality versus conventional balance training on balance of the elderly. Exp Gerontol 2021 Oct 1;153:111498. [doi: 10.1016/j.exger.2021.111498] [Medline: 34311059]
41. Lee JH, Ku J, Cho W, et al. A virtual reality system for the assessment and rehabilitation of the activities of daily living. Cyberpsychol Behav 2003 Aug;6(4):383-388. [doi: 10.1089/109493103322278763]
42. Zahedian-Nasab N, Jaberi A, Shirazi F, Kavousipor S. Effect of virtual reality exercises on balance and fall in elderly people with fall risk: a randomized controlled trial. BMC Geriatr 2021 Sep 25;21(1):509. [doi: 10.1186/s12877-021-02462-w] [Medline: 34563120]
43. Barsasella D, Liu MF, Malwade S, et al. Effects of virtual reality sessions on the quality of life, happiness, and functional fitness among the older people: a randomized controlled trial from Taiwan. Comput Methods Programs Biomed 2021 Mar;200:105892. [doi: 10.1016/j.cmpb.2020.105892]
44. Gerber SM, Muri RM, Mosimann UP, Nef T, Urwyler P. Virtual reality for activities of daily living training in neurorehabilitation: a usability and feasibility study in healthy participants. Presented at: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); Jul 18-21, 2018; Honolulu, HI, USA. [doi: 10.1109/EMBC.2018.8513003]
45. Lin CX, Lee C, Lally D, Coughlin JF. Impact of virtual reality (VR) experience on older adults' well-being. In: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Springer Verlag; 2018, Vol. 10927 LNCS:89-100. [doi: 10.1007/978-3-319-92037-5_8]
46. Faria AL, Andrade A, Soares L, I Badia SB. Benefits of virtual reality based cognitive rehabilitation through simulated activities of daily living: a randomized controlled trial with stroke patients. J Neuroeng Rehabil 2016 Nov 2;13(1):96. [doi: 10.1186/s12984-016-0204-z] [Medline: 27806718]
47. Wang LT. Effects of semi-immersive virtual reality exercise on the quality of life of community-dwelling older adults: three-month follow-up of a randomized controlled trial. Digit Health 2024;10:20552076241237391. [doi: 10.1177/20552076241237391] [Medline: 38449682]
48. Afifi T, Collins N, Rand K, et al. Using virtual reality to improve the quality of life of older adults with cognitive impairments and their family members who live at a distance. Health Commun 2023 Oct;38(9):1904-1915. [doi: 10.1080/10410236.2022.2040170] [Medline: 35253531]

Abbreviations

ABC: Activities-Specific Balance Confidence

XSL•FO
RenderX

ADL: activities of daily living
CPA: conventional physical activity
KIADL: Katz Index of Independence in Activities of Daily Living
MR: mixed reality
Mr.PT: Mixed Reality–Based Physical Therapy
QOL: quality of life
SF-12: 12-Item Short-Form Health Survey
VR: virtual reality

Edited by A Coristine; submitted 22.04.25; peer-reviewed by K Yuenyongchaiwat, SB Guo; revised version received 05.06.25; accepted 05.06.25; published 01.08.25.

Please cite as:
An Y, Min S, Park C
Mixed Reality–Based Physical Therapy in Older Adults With Sarcopenia: Preliminary Randomized Controlled Trial
JMIR Serious Games 2025;13:e76357
URL: https://games.jmir.org/2025/1/e76357
doi:10.2196/76357

© Yeongsang An, Seunghwa Min, Chanhee Park. Originally published in JMIR Serious Games (https://games.jmir.org), 1.8.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Neurotechnology-Based, Intensive, Supplementary Upper-Extremity Training for Inpatients With Subacute Stroke: Feasibility Study

Reut Binyamin-Netser[1,2,3], PhD, OT; Shirley Handelzalts[4], PhD, PT; Noy Goldhamer[2], MBA; Inbar Avni[1,2,3], PhD; Adi Tayer Yeshurun[2], MSci, PT; Yogev Koren[2,4], PhD, PT; Ofri Bibas Levy[5], OT; Shilo Kramer[5], Dr med; Simona Bar Haim[2,4], Prof Dr; Lior Shmuelof[1,2,3], Prof Dr

[1]Department of Cognitive and Brain Sciences, Ben-Gurion University of the Negev, 1 Ben-Gurion Ave, PO Box 653, Beer-Sheva, Israel

[2]Translational Neurorehabilitation Laboratory, Adi Negev Rehabilitation Hospital, Ofakim, Israel

[3]The Zelman Center for Brain Sciences Research, Ben-Gurion University of the Negev, Beer Sheva, Israel

[4]Department of Physical Therapy, Ben-Gurion University of the Negev, Beer Sheva, Israel

[5]Adi Negev Rehabilitation Hospital, Ofakim, Israel

Corresponding Author:
Reut Binyamin-Netser, PhD, OT
Department of Cognitive and Brain Sciences, Ben-Gurion University of the Negev, 1 Ben-Gurion Ave, PO Box 653, Beer-Sheva, Israel

Abstract

Background: Upper-extremity hemiparesis is a common and debilitating impairment after stroke, severely restricting stroke survivors' ability to participate in daily activities and function independently. Alarmingly, only a small percentage of stroke patients fully recover upper extremity function. Animal models indicate that high-dose upper extremity training during the early poststroke phase can significantly enhance motor recovery. However, translating such programs for human patients remains challenging due to resource limitations, patient compliance issues, and administrative constraints.

Objective: This study aimed to assess the feasibility and potential efficacy of an intensive, video game–based upper-extremity training protocol designed to improve movement quality during inpatient stroke rehabilitation. Additionally, it evaluated the resources required for this intervention. Specifically, the protocol provides high-intensity, high-dose training to facilitate motor recovery by engaging patients in targeted interactive exercises.

Methods: Twelve patients with upper-extremity hemiparesis completed a 4-week intensive training program comprising 40 sessions of 60 minutes; the training was conducted for 2 hours per day, 5 days per week. This was delivered in addition to standard care, which included 3 therapeutic sessions daily. Two video game–based platforms were used: one platform (tech 1) targeted proximal movements involving the shoulder and elbow, while the second platform (tech 2) emphasized distal movements of the wrist and fingers. Feasibility was assessed using the measure of time on task and measures of patients' motivation and engagement. Potential effectiveness was assessed using the Fugl-Meyer Assessment of the upper extremity (FMA-UE) scale, Action Research Arm Test (ARAT), and Stroke Impact Scale (SIS).

Results: Of the 12 patients, 8 completed the full protocol, 3 completed 34 - 38 sessions, and 1 completed 27 sessions. On average, patients actively engaged in exercises for 35 (SD 4) minutes per hour on the proximal platform (tech 1) and 37 (SD 2) minutes on the distal platform (tech 2). Patients reported high motivation and enjoyment throughout the sessions, with an Intrinsic Motivation Inventory enjoyment score of 6.49 (SD 0.66) out of 7. Pain levels were minimal, with a visual analogue scale (VAS) mean score of 2.00 (SD 2.32). Significant improvements were observed in motor function assessments: the mean improvement in FMA-UE score was 16.5 (SD 10.2) points, ARAT scores increased by 22.9 (SD 13.1) points, and the SIS Hand Function and Recovery score showed a mean delta of 1.23 (SD 0.80) points and a 23.33% (SD 21.5%) improvement, respectively.

Conclusions: These findings demonstrate that a high-dose, high-intensity, video game–based training protocol is feasible and can be successfully integrated into subacute stroke rehabilitation. Additionally, preliminary evidence suggests that this supplementary intervention may be effective in enhancing motor recovery. This approach holds promise for future stroke rehabilitation protocols by offering an engaging, high-dose, and high-intensity program during early recovery.

Trial Registration: Clinicaltrials.gov NCT04737395, https://clinicaltrials.gov/study/NCT04737395

(JMIR Serious Games 2025;13:e56397) doi:10.2196/56397

XSL•FO
RenderX

KEYWORDS

stroke; rehabilitation; neurotechnology

Introduction

Stroke is the second leading cause of death and the third leading cause of disability-adjusted life-years [1]. One year after a stroke, patients have restricted participation [2]. It is estimated that one- to two-thirds of stroke survivors worldwide require some assistance or are fully dependent on caregivers for activities of daily living [3,4]. The most common impairment after stroke is hemiparesis [5]. Six months after stroke, only 12% of patients regain full functional recovery of the upper extremity (UEs) [6,7]. A poor recovery from stroke significantly impacts the lives of not only the individuals who experience the stroke, but also their families and caregivers.

Effective neurorehabilitation is essential for enhancing stroke recovery [8]. Nevertheless, at 3 months after stroke, patients achieve only 70% of their maximal potential recovery of the UEs [9-12]. One potential way to improve recovery is to increase the dosage of treatment. Increasing treatment dosage beyond standard care has been shown to be effective for the UEs [13-17]. Furthermore, the timing of treatment delivery is also important, as both nonhuman stroke models and studies in humans demonstrate a time-sensitive period of plasticity [18,19]. For example, adding 20 hours of training during the acute and subacute phases was shown to be effective in improving UE motor recovery, whereas adding the same number of training hours during the chronic phase was not [19,20]. Importantly, most studies investigating the effect of supplementary treatments were conducted with chronic stroke patients (>6 months after stroke onset) [13,14,21]. While the effect of treatment dosage on subacute recovery has been examined before in small-scale studies [15-17], its feasibility and potential efficacy needs to be established. Increased treatment dosage and intensity in the subacute phase should be defined with respect to the standard care that the patients receive. We define an increase in dosage as a supplementary treatment of more than 1 hour a day for at least 4 weeks.

The intensity and content of the training program may also affect recovery. It has been reported that patients make functional movement in only half of the therapy sessions, with 32 functional repetitions per session [22]. Furthermore, in clinical practice, rehabilitation therapies focus on the level of function (ie, task-oriented training, focusing on completing a task), rather than the level of motor impairment and quality of movement, which focus on the typicality of the performed movements [22]. Indeed, previous trials with low doses of task-oriented therapies show limited effects [23-26]. Considering these results, it was recently suggested to emphasize the quality of movement, training intensity, and an enriched environment in subacute stroke rehabilitation [27]. One way to achieve these goals is by integrating engaging video games in therapy. Adding these treatments to conventional therapy was shown to be effective (compared to conventional therapy alone [16,28]) and feasible [29].

Applying these recommendations in the subacute stroke population has several challenges. A major difficulty is that stroke patients are often withdrawn and passive, and spend extended time being sedentary [30]. Supplementary treatments should therefore be highly motivating to address the potential low compliance of the patients. Another challenge is related to the administrative complexity of adding substantial amounts of treatment to the schedule of the patients without compromising their care. A third challenge is recruiting enough personnel for supplementary treatments. Lastly, reliance on video game technologies that provide patients with online feedback on their performance requires highly reliable technologies that can be controlled by nontechnical staff. All these aspects should be developed in the context of feasibility and implementation studies.

The aim of this study was to implement and assess the feasibility of adding 40 hours of UE training based on engaging video games that emphasize the quality of movement and intensity during the early subacute phase after a stroke. We specifically aimed to (1) evaluate the feasibility of adding 2 hours a day of intensive video game–based rehabilitation during a period of 4 weeks, (2) evaluate the resources and ability to manage and implement the study and intervention, and (3) conduct a preliminary evaluation of patients' responses to the intervention.

Methods

Patients

Twelve patients in the subacute phase of stroke were recruited for the study. The patients were recruited at the Adi Negev Nahalat Eran Rehabilitation Village (Israel). Inclusion criteria were as follows: age ≥18 years; ischemic or hemorrhagic stroke (hemispheric or brainstem), confirmed by computed tomography or magnetic resonance imaging; first-ever stroke or previous stroke with no UE weakness before the second incident; ≤1 week to ≤6 weeks after stroke onset, active shoulder flexion of at least 20° and partial wrist and/or finger active movement (due to a limitation of the technologies used in the intervention); Fugl-Meyer Assessment of the upper extremity (FMA-UE) score <58; and the ability to provide informed consent. Patients were excluded if they had a painful shoulder limiting an active forward reach, severe spasticity, non-neural loss of range of motion, cognitive or communication impairments as determined by the clinical team, or an unstable medical condition.

All patients were included in the final analyses. All enrolled patients had experienced a first stroke (ie, there were no cases of recurrent stroke in the population).

Intervention

Overview

The aim of the program was to increase the amount of UE practice using customized video game–based platforms that included immersive, challenging, and rewarding virtual environments. In each session, patients had to achieve an explicit

XSL•FO
RenderX

goal while receiving online feedback and rewards for goal attainment. The intervention was composed of 120 minutes of UE training a day (divided into 2 therapy sessions of 60 minutes), and was given for 4 weeks, for a total of 40 hours. Each session was staffed with a 1:1 physical therapist (PT)/occupational therapist (OT) to patient ratio.

We used two platforms: one for proximal arm training (shoulder and elbow; tech 1) and another for distal arm training (wrist and fingers; tech 2).

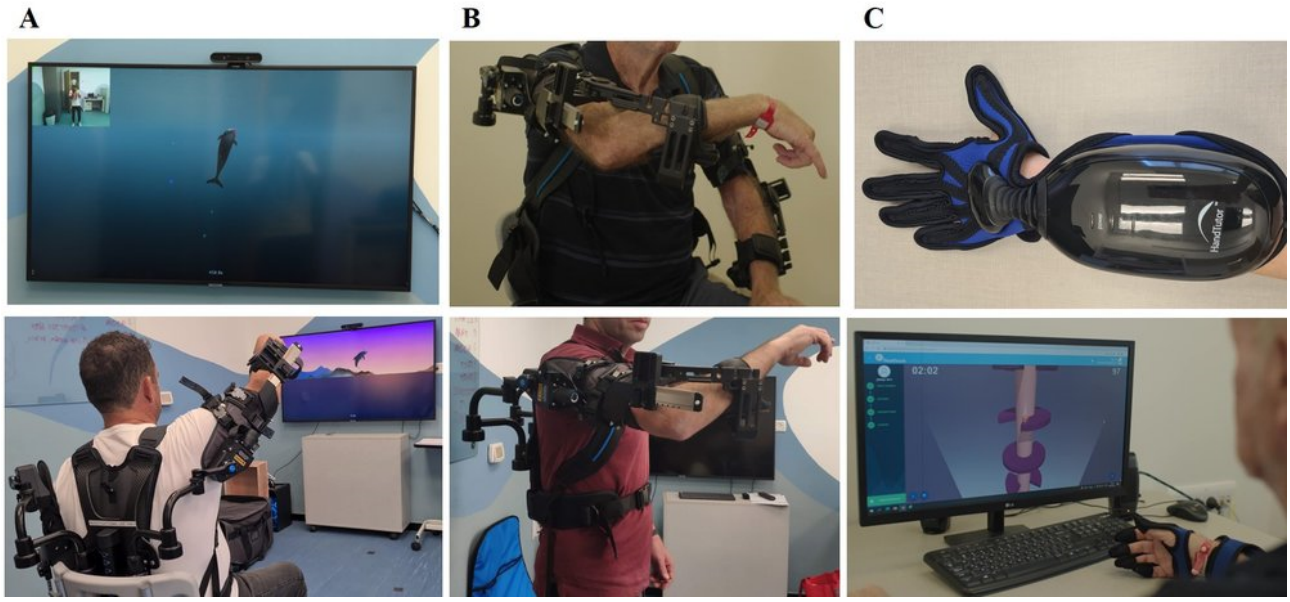
Training on both platforms was conducted while the patients were sitting on a plastic bath chair with back support and no arm support. The legs of the chair had an antislip rubber coating. On both platforms, a PT or OT was present throughout each session and provided verbal and tactile feedback to ensure movement quality and lack of compensations, and to encourage exploration of the full workspace (in tech 1). Each session was dedicated to a single platform. At least 3 sessions per week were dedicated to each platform.

Patients continued with their regular rehabilitative routine, which included daily physical and occupational therapy sessions and speech therapy if needed, as well as group work, hydrotherapy, and gym workouts during weekdays (Sunday to Friday). The regular rehabilitation protocol included at least 3 treatments a day. Patients were also encouraged to work on cardiovascular fitness during the program, as well as to use their paretic UE in daily activities. Outcome measures were collected throughout the intervention period on a session-by-session basis. To evaluate the potential effectiveness of the intervention, motor function was assessed before the intervention, immediately after the intervention, and at 12 (±14 days) and 24 (±14 days) weeks after stroke. Follow-up assessments were scheduled according to the stroke time and not the intervention to assess the potential effectiveness of the treatment at the end of the early and late subacute stages.

MindPod Dolphin (Tech 1) for Proximal Arm Training

The patients used a custom-designed, immersive, animation-based experience named MindPod Dolphin, which is based on patented technology licensed from KATA at Johns Hopkins University to MSquare Healthcare Inc, a MindMaze Group company (Figure 1A) [16]. In this game, patients control the swimming of a virtual dolphin by moving their shoulders and elbow joints and performing complex exploratory movements. The dolphin swims through different ocean scenes with multiple goals such as chasing and eating fish, escaping a shark, and jumping out of the water. This game was designed to promote the movement quality of the shoulder and elbow in all planes (3D movements; this included shoulder abduction, adduction, flexion, and extension and elbow flexion and extension) throughout the active ranges of motion. The game uses a real-time motion capture system. The paretic arm was supported by a passive mechanical exoskeleton vest (Ekso UE; Ekso Bionics; Figure 1B). The degree of support was provided by the exoskeleton using springs with different tension levels. The amount of spring tension needed was adjusted for each patient to allow active movement of the shoulder to at least 90° of flexion. The support of the vest allowed the patients to play with their paretic arm in an extended active range in all directions. The arm weight support level was titrated as the patient progressed through the sessions by reducing the level of support or removing the vest. No active assistance was given by the therapist during the intervention. The game was designed in a way that every level was more difficult than the one before. Each level had a goal (eg, to eat a specific number of fish), and when the goal was reached, the following level began.

Figure 1. (A) Tech 1: MindPod. (B) Ekso UE vest for arm weight support. (C) Tech 2: HandTutor, an ergonomic, wearable glove.



HandTutor (Tech 2) for Distal Arm Training

The HandTutor system (MediTouch; Figure 1C) consists of an ergonomic wearable glove and a dedicated video game that allow practice of active wrist movements, grip control, and finger individuation in a virtual environment. Patients progress through games while adjusting the range of motion that is being practiced according to their abilities. Patients perform flexion

XSL•FO
RenderX

and extension movements of their fingers or wrist separately or simultaneously. In each session, the therapist chooses the trained movements (emphasizing grip abilities and finger individuation). In one game, for example, patients controlled a fishing rod up and down by flexing and extending an individual finger to catch swimming fish. The finger range of movement was set before the game started. If the task was too easy, the therapist increased the difficulty level by making the fish move faster or by adding another control dimension, such as adding movement of the wrist on top of the fingers in a manner that the fingers controlled the fishing rod movement (up or down) and the wrist controlled the boat movement (left or right).

Outcome Measures of Feasibility

Adherence rates to study procedures were documented by the therapists for each session. Time on task (ToT) was measured in each session (in minutes) using a stopwatch. The distance that the arm reached on the task (in meters) was measured in tech 1. Game levels and the amount of weight support were documented by the therapist in each session, as were rehabilitation sessions outside of the intervention (ie, usual care) during the study period. Pain level was monitored using the visual analogue scale (VAS) at the end of each session and before and after the entire intervention. The VAS is a self-report measure consisting of a 10-cm line with a statement at each end representing one extreme of pain intensity (“no pain” and “pain as bad as it could possibly be”). The patient marks the line with a pen at a point corresponding to their present pain level [31]. Exercise intensity was measured using the rating of perceived exertion (RPE), where patients subjectively rate their level of exertion during exercise (1=“did not put in an effort at all” to 10=“put in an extreme effort”) [32]. Participation was measured using the Pittsburgh Rehabilitation Participation Scale (PRPS), a clinician-rated instrument to assess a patient’s participation in therapy using effort and motivation estimates [33]. This measure was taken at the end of each practice session. Motivation was measured using the Intrinsic Motivation Inventory (IMI; Multimedia Appendix 1), a multidimensional measurement intended to assess patients’ subjective experience related to a target activity in laboratory experiments. The instrument assesses patients’ interest/enjoyment, perceived competence, effort, value/usefulness, felt pressure and tension, and perceived choice while performing a given activity. We used the Task Evaluation Questionnaire, a specific version of the IMI, to assess only 4 subscales: interest/enjoyment, perceived competence, pressure/tension, and perceived choice. Motivation was assessed through rating the agreement with proposed statements using a Likert scale (1=“not at all true” to 7=“very true”) [34]. Motivation was assessed at the end of the intervention period.

Adverse events and problems with the operation of the interventions were documented by the therapists throughout the study period.

We found that 4.2% of data were missing from the RPE reports, 5% from the PRPS reports, and 9% from the VAS reports. The data were missing due to the failure of therapists to fill in the forms. Statistical analyses were conducted without the missing values.

Outcome Measures of Potential Efficacy

The FMA-UE assessment was used to assess motor impairment. The FMA-UE is scored on an ordinal 3-point scale. The maximum score for the FMA-UE is 66 for each arm, with a higher score indicating better arm motor status [35]. The FMA-UE has shown good reliability, validity, and sensitivity to poststroke motor changes [36]. The Action Research Arm Test (ARAT) was used to assess motor activity using 19 tests of arm motor function, both distally and proximally (grasp, grip, pinch, and gross movement). Each test is given an ordinal score of 4 values, with higher values indicating better arm motor status. The total ARAT score is the sum of the 19 tests, and thus the maximum score is 57 [37]. The ARAT demonstrates high reliability and good validity, as well as sensitivity to spontaneous and therapy-related gains after a stroke [38] The Stroke Impact Scale (SIS) hand domain (section 7; Multimedia Appendix 2), version 2.0, was used to assess disability and health-related quality of life using a questionnaire. It includes 64 self-report items, divided into 8 areas: strength (physical difficulties), memory and thought, emotion, communication, activities of daily living, mobility, manual function, and social participation. The rating of the items was carried out on an ordinal scale of 5 levels. A separate score was given to each field according to an algorithm. A final score was given in the range between 0 and 100, where lower scores indicate a negative effect in that area on health and perception of quality of life. In addition, overall recovery from the event (section 9) was assessed by a percentage rating on a VAS between 0 (no recovery) and 100 (full recovery) [39,40].

Statistical Analyses

All analyses and statistical calculations were performed using SPSS (version 21; IBM Corp). Means and SDs were calculated for all outcome measures. Paired-sample 2-tailed *t* tests were performed to compare the motor outcome measures before and after the intervention.

Ethical Considerations

The study was approved by the Regional Ethical Review Board at Sheba Medical Center, Israel (6218 - 19-SMC), and registered with ClinicalTrials.gov (NCT04737395) prior to the start of enrollment. All patients signed a consent form before participating in the experiment. The original informed consent form allowed secondary analyses without additional consent. All study data are anonymous. The enrolled patients volunteered for this study.

Results

Overview

Twelve patients participated in the study, aged 44 to 71 years (mean age 61.67, SD 8.80 years; n=3 women; mean time after stroke 32.42, SD 16.45 days). Patients varied in the amount of formal education they had received (0-15 years), and their medical conditions (these included obesity, schizophrenic disorder, hyperlipidemia, hypothyroidism, essential hypertension, and depression) (Table 1).

XSL•FO
RenderX

Table . Patients' characteristics (N=12).

Patient	Age (years)	Sex	Education (years)	Time from stroke to intervention (days)	Type of stroke	Dominant hand	Impaired side	Stroke location	MoCA[a] score
1	49	M[b]	12	38	H[c]	L[d]	R[e]	Thalamus and basal ganglia (L)	20
2	64	M	12	27	I[f]	R	R	Thalamus (L)	25
3	44	M	14	69	H	R	R	Basal ganglia (L)	19
4	66	M	8	49	I	L	R	MCA[g] (L)	18
5	56	F[h]	12	25	H	R	R	Basal ganglia (L)	25
6	68	F	10	44	I	R	L	Basal ganglia (R)	14
7	71	M	0	34	I	R	R	Occipital (L)	8
8	55	F	12	25	I	L	L	MCA (R)	21
9	61	M	15	15	I	R	L	Periventricular (R)	20
10	69	M	10	10	I	R	L	Periventricula (R)	19
11	70	M	12	17	I	R	R	PCA[i] (L)	22
12	67	M	12	36	I	R	R	Periventricular (L)	27
Overall, mean (SD)	61.7 (8.8)	—[j]	10.7 (3.8)	32.4 (16.4)	—	—	—	—	19.8 (4.9)

[a]MoCA: Montreal Cognitive Assessment score (maximum score=30); MoCA score was converted from Mini Mental State Examination score.

[b]M: male.

[c]H: hemorrhagic stroke.

[d]L: left.

[e]R: right.

[f]I: ischemic stroke.

[g]MCA: middle cerebral artery.

[h]F: female.

[i]PCA: posterior crebral artery.

[j]Not applicable.

Training Adherence and ToT

The mean number of delivered sessions was 38 (SD 4). A total of 8 of our 12 (67%) patients completed the 40 intervention sessions. Patients 3 and 11 did not complete the protocol due to COVID-19, patient 4 due to poor motivation at the end of the intervention, and patient 12 due to a medical condition (Table 2). All patients completed pre- and postintervention assessments. A total of 9 (75%) patients completed the 12-week postintervention assessment, and 7 (58%) completed the 24-week postintervention assessment (the missing data is because the patients failed to attend). The mean ToT was 35 (SD 2.44) minutes per 1-hour session (33, SD 4.35 minutes for tech 1 and 37, SD 2.09 minutes for tech 2; Figure 2). For tech 1, total arm distance was also measured, reaching a mean distance of 246 (SD 87.35) meters (Figure 3). In addition to the intervention, all patients received an average of 2.7 (SD 0.36) standard care rehabilitation hours a day. In total, patients received an average of 54 (SD 8.05) hours of standard care rehabilitation during the intervention period.

XSL•FO
RenderX

Table . Training attendance and time on task (ToT).

Patient	Total intervention sessions	Mean ToT tech 1 (minutes), mean (SD)	Mean ToT tech 2 (minutes), mean (SD)
1	40	37 (4)	39 (2)
2	40	41 (2)	38 (3)
3	27	35 (4)	33 (4)
4	36	37 (8)	36 (4)
5	41	32 (6)	34 (7)
6	40	36 (5)	37 (4)
7	40	28 (8)	37 (3)
8	40	31 (7)	37 (6)
9	40	31 (8)	41 (3)
10	40	29 (7)	35 (7)
11	34	27 (6)	37 (4)
12	38	31 (8)	37 (5)
Overall, mean (SD)	38 (4)	33 (4)	37 (2)

The effort level that was reported by the patients was high (RPE; see Methods section). The mean effort patients reported was 7.14 (SD 2.18, on a scale of 10) points (for tech 1: 7.38, SD 2.21 points; for tech 2: 6.90, SD 2.21 points). This suggests that patients were motivated to make an effort in each training session. Participation assessment by the therapists was also high (PRPS; see Methods section). The mean participation of patients was 5.42 (SD 0.49) points (for tech 1: 5.45, SD 0.49 points; for tech 2: 5.40, SD 0.51 points). Pain was measured using the VAS (see Methods section). This measure was taken after each session to determine if the training was associated with any pain. The mean VAS patients reported was 2.00 (SD 2.32) points (for tech 1: 2.39, SD 2.73 points; for tech 2: 1.61, SD 1.86 points). This suggests that pain was generally low.

Motivation was assessed at the end of the intervention (IMI; see Methods section). The mean score in each category was as follows: interest/enjoyment was 6.49 (SD 0.66), perceived competence was 6.18 (SD 0.80), perceived choice was 5.42 (SD 1.39), and pressure/tension was 1.75 (SD 0.97) points. This suggests that patients enjoyed the intervention, felt that they could do it, and felt that joining the intervention was their choice. They did not feel under pressure during the intervention sessions.

No adverse events were recorded during the intervention sessions.

XSL•FO
RenderX

Figure 2. Total time on task. (A) Tech 1 (shoulder and elbow movement practice). (B) Tech 2 (wrist and finger movement practice). The data represent each session. All 12 patients are represented. Each patient is represented by a different line. The average of all patients is represented by the black lines.

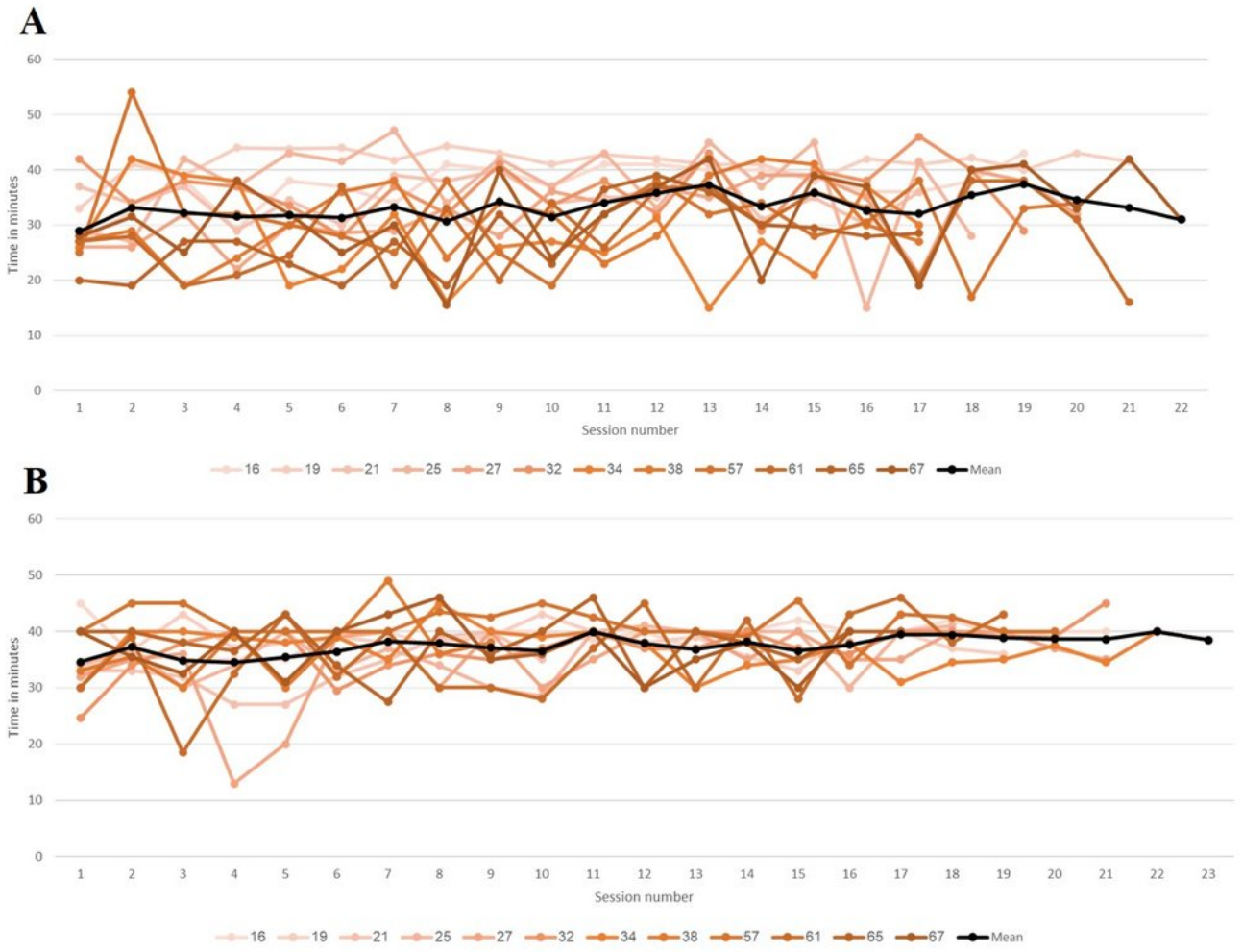
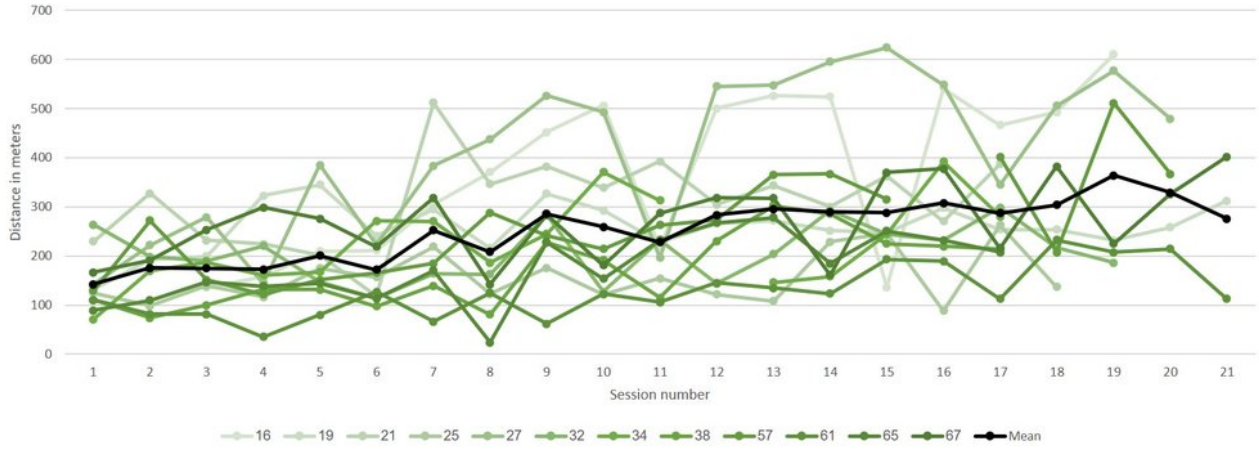


Figure 3. Total arm distance for tech 1 (shoulder and elbow movement practice) for each of the sessions. All 12 patients are represented. Each patient is represented by a different line. The average of all patients is represented by the black line.



Feasibility of Operation

The intervention hours were scheduled before and after the regular rehabilitation hours, from 7 to 8 AM and from 3 to 4 PM, to minimize the effect of the intervention on standard care (3 treatments per day) and to secure therapists for the intervention. This decision led to extension of the treatment day for the enrolled patients, and was dependent on good cooperation between the ward management staff, who had to make sure the patients were prepared to start the rehabilitation day earlier. Another challenge raised by this decision was that even though the interventions were conducted on a single patient at a time, 7 trained therapists were required to fill the intervention schedule.

XSL•FO
RenderX

Motor Outcomes

The scores of the outcome measures before and after intervention are shown for each patient in Table 3. The average FMA-UE recovery was 16.50 (SD 10.2) points. This difference is significant (t_{11}=−5.58; P<.001) and higher than the minimal clinically important difference (MCID; 9 - 10 points) [41]. The average ARAT recovery was 22.92 (SD 13.1) points. This difference is also significant (t_{11}=−6.05; P<.001) and higher than the MCID (12 - 17 points) [42]. The average SIS Function and SIS Recovery changes were 1.23 (SD 0.8) points and 23.33% (SD 21.5%), respectively. These differences were significant (t_{11}=−5.01; P<.001 and t_{11}=−3.77; P=.003, respectively). The outcome measures that were obtained at the end of the intervention were largely stable at 12 and 24 weeks after stroke: the average FMA-UE was 48.78 (SD 12.08) and 48.57 (SD 11.31) points, respectively (in comparison to the average of 49.75, SD 10.87 points after the intervention). The average ARAT scores were 39.67 (SD 20.12) and 35.86 (SD 18.08) points, respectively (in comparison to the average of 39.0, SD 16.89 points after the intervention). The average SIS Function scores were 3.44 (SD 0.73) and 3.11 points (SD 1.26), respectively (in comparison to the average of 2.98, SD 1.01 points after the intervention). The average SIS Recovery scores were 69.44% (SD 17.04%) and 72.14% (SD 14.39%), respectively (in comparison to the average of 71.25%, SD 13.84% after the intervention). Generally, patients maintained their level of motor function 12 and 24 weeks after stroke.

Table . Stroke patients' measurement scores.

Patient	Preintervention FMA-UE[a] motor score	FMA-UE motor change[b]	Preintervention ARAT[c] score	ARAT score change	Preintervention SIS[d] Function score	SIS Function change	Preintervention SIS Recovery (%)	SIS Recovery change
1	41	21	24	32	2.6	1	30	40
2	27	9	2	10	1	0.8	20	60
3	25	3	5	0	2.4	0.2	85	-5
4	18	26	3	29	1	0.2	45	25
5	25	28	0	49	1.2	1.2	10	40
6	40	12	36	18	2	2.2	50	10
7	33	10	26	14	1.4	1	30	20
8	38	8	12	19	2.6	1.8	80	20
9	36	27	30	27	1.4	1.4	15	55
10	30	32	11	36	1.2	3.2	80	0
11	54	3	36	14	2.4	0.6	60	15
12	32	19	8	27	1.8	1.2	70	0
Overall, mean (SD)	33.3 (9.5)	16.5 (10.2)	16.1 (13.5)	22.9 (13.1)	1.7 (0.6)	1.2 (0.8)	47.9 (27.0)	23.3 (21.5)

[a]FMA-UE: Fugl-Meyer Assessment of the upper extremity motor score (maximum score=66).

[b]Change indicates the delta between the scores before and after the intervention.

[c]ARAT: Action Research Arm Test (maximum score=57).

[d]SIS: Stroke Impact Scale (this scale has 2 subscales: Function, maximum score=5, and Recovery [in %], maximum=100%).

Discussion

Principal Findings

This study demonstrates the feasibility of an intensive supplementary subacute training protocol of 40 sessions. In most cases, the protocol was implemented successfully, and incomplete implementation was due to reasons that were unrelated to the intervention protocol. The proportion of ToT to total time was above 60%, and participants found the intervention to be motivating, enjoyable, and engaging.

Intensive UE Training in Patients With Subacute Stroke Is Feasible

We found that it is possible to add 2 hours of intensive UE training per day during hospitalization. This finding has 2 aspects: patients in the subacute stage could endure 2 additional hours of intensive training a day and the medical system could deliver the training. The following elements are recommended for program success: (1) administrative support—this program required the full cooperation of the rehabilitation ward staff to prepare the patients on time (prioritizing them) and to organize their treatment schedule in a way that considered factors such as meals, rest periods, and the mandatory number of usual care treatments per day; (2) engagement of therapists and patients—multiple therapists should be trained to administer the intervention to ensure a smooth flow of treatments (without cancelations; coordination of the clinical staff is also important for information exchange regarding the participating patients); (3) reliable and attractive technologies—the technologies we chose to use in this intervention were reliable, highly motivating, enjoyable, and challenging.

XSL•FO
RenderX

Feasibility Outcome Measures

We expected to reach two-thirds of the total time dedicated for each training session in ToT (ie, 40 minutes ToT per 60-minute session), which would have been similar to the expected ratio in a standard therapy session. However, we reached only 35 - 37 minutes per hour due to extended preparation time, technology maintenance, and patient fatigue. Integrating the intervention within the standard care schedule may increase the proportion of ToT. Notably, compared to other studies that investigated intervention protocols in the subacute phase after a stroke, our protocol is very intensive [19,20,23-26,43]. Importantly, in addition to the potential benefits of early intervention [19,20,29], the ability to provide an intensive intervention protocol in the subacute phase during hospitalization is highly cost-effective since it does not require inpatient or outpatient programs, such as those that are used in the chronic phase after a stroke [13,14].

The Motor Benefits of the Intervention

Substantial improvement was reported in all motor outcome measures. This improvement was larger compared to other reported recovery measures [13,14,44]. A new study showed that adding 30 hours of UE training resulted in better motor improvement than the standard of care alone [16]. That study showed an improvement of 12.5 - 13.4 points in the FMA-UE (in comparison to 16.5 points of improvement in our study) and 13.4 - 14.7 points in the ARAT (in comparison to 22.9 points of improvement in our study). Based on the apparent increased recovery following 40 hours of training, we recommend studying the efficacy of a supplementary intervention protocol of at least 40 hours. Furthermore, the effect of supplemental treatment on recovery may not be independent of standard care treatments. We therefore call for reporting the standard care content and dosage in future intervention studies.

Implementation Recommendations

Implementation of the intensive treatment could be improved in different ways: the efficacy and feasibility of the intervention could be enhanced by increasing the number of patients per therapist (OT or PT), so that multiple patients can practice at the same time. Practice could also be administered by a PT or OT assistant. Moreover, it may be possible to develop an implementation model where treatment is given during the rehabilitation day and not off hours. Other suggested avenues for improvement are to expand this protocol in time and to include a telerehabilitation component. Although this type of rehabilitation is relatively new and has not been extensively studied [45], it has been shown to be feasible [44]. Therefore, we suggest further studies regarding the implementation of high-dosage and high-intensity treatment with a such a component. Ultimately, the efficacy of this intervention compared to the standard of care and to an increased dosage of the standard of care should be studied in randomized controlled trials.

Limitations

Since this is a feasibility study, the reported recovery measures cannot be assigned to the intervention. A randomized controlled trial should be conducted to establish the efficacy of the intervention. Another limitation is the lack of knowledge regarding the content and intensity of standard UE training that our patients received. We can estimate that out of the 3 hours of standard care treatment they received a day, about 1.5 hours were dedicated to UE training and the other 1.5 were dedicated to other areas, such as cognition, the lower extremities, and psychological aspects. We recommend that future studies monitor the content of standard care. Last, as treatment intensity and dosage increases, fatigue, a prominent effect of stroke [46], may become a limiting factor of training adherence. We suggest that future studies assess fatigue using a valid and reliable tool such as the Fatigue Severity Scale [47].

Conclusion

Intensive UE training is practical and well received by patients with subacute stroke. Successful implementation depends on the clinical setting, technologies, and resources. We call for studying the implementation of this protocol with multiple patients at a time and during clinical days and for randomized control trials to study the efficacy of the intervention. Our preliminary results suggest that subacute intensive rehabilitation may improve motor recovery after a stroke.

Acknowledgments

We thank Dr Oren Barzel for his clinical assistance. We thank the reviewers for their constructive comments. This work was supported by the Israeli Science Foundation (grant 1244/22 to LS).

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Intrinsic Motivation Inventory.
[PDF File, 85 KB - games_v13i1e56397_app1.pdf]

Multimedia Appendix 2

XSL•FO
RenderX

Stroke Impact Scale.
[PDF File, 134 KB - games_v13i1e56397_app2.pdf]

References

1. Feigin VL, Forouzanfar MH, Krishnamurthi R, et al. Global and regional burden of stroke during 1990-2010: findings from the Global Burden of Disease Study 2010. Lancet 2014 Jan 18;383(9913):245-254. [doi: 10.1016/s0140-6736(13)61953-4] [Medline: 24449944]
2. Hartman-Maeir A, Soroker N, Ring H, Avni N, Katz N. Activities, participation and satisfaction one-year post stroke. Disabil Rehabil 2009;29(7):559-566. [doi: 10.1080/09638280600924996] [Medline: 17453976]
3. Miller EL, Murray L, Richards L, et al. Comprehensive overview of nursing and interdisciplinary rehabilitation care of the stroke patient: a scientific statement from the American Heart Association. Stroke 2010 Oct;41(10):2402-2448. [doi: 10.1161/STR.0b013e3181e7512b] [Medline: 20813995]
4. Feigin VL, Norrving B, Mensah GA. Global burden of stroke. Circ Res 2017 Feb 3;120(3):439-448. [doi: 10.1161/CIRCRESAHA.116.308413] [Medline: 28154096]
5. Langhorne P, Coupar F, Pollock A. Motor recovery after stroke: a systematic review. Lancet Neurol 2009 Aug;8(8):741-754. [doi: 10.1016/S1474-4422(09)70150-4] [Medline: 19608100]
6. Kwakkel G, Kollen BJ, van der Grond J, Prevo AJH. Probability of regaining dexterity in the flaccid upper limb: impact of severity of paresis and time since onset in acute stroke. Stroke 2003 Sep;34(9):2181-2186. [doi: 10.1161/01.STR.0000087172.16305.CD] [Medline: 12907818]
7. Nijland RHM, van Wegen EEH, Harmeling-van der Wel BC, Kwakkel G, EPOS Investigators. Presence of finger extension and shoulder abduction within 72 hours after stroke predicts functional recovery: early prediction of functional outcome after stroke: the EPOS Cohort Study. Stroke 2010 Apr;41(4):745-750. [doi: 10.1161/STROKEAHA.109.572065] [Medline: 20167916]
8. Pollock A, Farmer SE, Brady MC, et al. Interventions for improving upper limb function after stroke. Cochrane Database Syst Rev 2014 Nov 12;2014(11):CD010820. [doi: 10.1002/14651858.CD010820.pub2] [Medline: 25387001]
9. Prabhakaran S, Zarahn E, Riley C, et al. Inter-individual variability in the capacity for motor recovery after ischemic stroke. Neurorehabil Neural Repair 2008;22(1):64-71. [doi: 10.1177/1545968307305302] [Medline: 17687024]
10. Winters JM. Telerehabilitation research: emerging opportunities. Annu Rev Biomed Eng 2002;4:287-320. [doi: 10.1146/annurev.bioeng.4.112801.121923] [Medline: 12117760]
11. Buch ER, Rizk S, Nicolo P, Cohen LG, Schnider A, Guggisberg AG. Predicting motor improvement after stroke with clinical assessment and diffusion tensor imaging. Neurology (ECronicon) 2016 May 17;86(20):1924-1925. [doi: 10.1212/WNL.0000000000002675] [Medline: 27164664]
12. Goldsmith J, Kitago T, Garcia de la Garza A, et al. Arguments for the biological and predictive relevance of the proportional recovery rule. Elife 2022 Oct 18;11:e80458. [doi: 10.7554/eLife.80458] [Medline: 36255057]
13. Ward NS, Brander F, Kelly K. Intensive upper limb neurorehabilitation in chronic stroke: outcomes from the Queen Square programme. J Neurol Neurosurg Psychiatry 2019 May;90(5):498-506. [doi: 10.1136/jnnp-2018-319954] [Medline: 30770457]
14. McCabe J, Monkiewicz M, Holcomb J, Pundik S, Daly JJ. Comparison of robotics, functional electrical stimulation, and motor learning methods for treatment of persistent upper extremity dysfunction after stroke: a randomized controlled trial. Arch Phys Med Rehabil 2015 Jun;96(6):981-990. [doi: 10.1016/j.apmr.2014.10.022] [Medline: 25461822]
15. Waddell KJ, Birkenmeier RL, Moore JL, Hornby TG, Lang CE. Feasibility of high-repetition, task-specific training for individuals with upper-extremity paresis. Am J Occup Ther 2014;68(4):444-453. [doi: 10.5014/ajot.2014.011619] [Medline: 25005508]
16. Krakauer JW, Kitago T, Goldsmith J, et al. Comparing a novel neuroanimation experience to conventional therapy for high-dose intensive upper-limb training in subacute stroke: the SMARTS2 randomized trial. Neurorehabil Neural Repair 2021 May;35(5):393-405. [doi: 10.1177/15459683211000730] [Medline: 33745372]
17. Birkenmeier RL, Prager EM, Lang CE. Translating animal doses of task-specific training to people with chronic stroke in 1-hour therapy sessions: a proof-of-concept study. Neurorehabil Neural Repair 2010 Sep;24(7):620-635. [doi: 10.1177/1545968310361957] [Medline: 20424192]
18. Zeiler SR, Hubbard R, Gibson EM, et al. Paradoxical motor recovery from a first stroke after induction of a second stroke: reopening a postischemic sensitive period. Neurorehabil Neural Repair 2016 Sep;30(8):794-800. [doi: 10.1177/1545968315624783] [Medline: 26721868]
19. Dromerick AW, Geed S, Barth J, et al. Critical Period After Stroke Study (CPASS): a phase II clinical trial testing an optimal time for motor recovery after stroke in humans. Proc Natl Acad Sci U S A 2021 Sep 28;118(39):e2026676118. [doi: 10.1073/pnas.2026676118] [Medline: 34544853]
20. Dromerick AW, Edwardson MA, Edwards DF, et al. Critical Periods After Stroke Study: translating animal stroke recovery experiments into a clinical trial. Front Hum Neurosci 2015;9:231. [doi: 10.3389/fnhum.2015.00231] [Medline: 25972803]
21. Kerr A, Grealy MA, Kuschmann A, Rutherford R, Rowe P. An intensive exercise program using a technology-enriched rehabilitation gym for the recovery of function in people with chronic stroke: usability study. JMIR Rehabil Assist Technol 2023;10:e46619. [doi: 10.3389/fresc.2021.820929] [Medline: 37477954]

XSL•FO
RenderX

22. Lang CE, Macdonald JR, Reisman DS, et al. Observation of amounts of movement practice provided during stroke rehabilitation. Arch Phys Med Rehabil 2009 Oct;90(10):1692-1698. [doi: 10.1016/j.apmr.2009.04.005] [Medline: 19801058]
23. Winstein CJ, Wolf SL, Dromerick AW, et al. Effect of a task-oriented rehabilitation program on upper extremity recovery following motor stroke: the ICARE randomized clinical trial. JAMA 2016 Feb 9;315(6):571-581. [doi: 10.1001/jama.2016.0276] [Medline: 26864411]
24. Winstein CJ, Wolf SL, Dromerick AW, et al. Interdisciplinary Comprehensive Arm Rehabilitation Evaluation (ICARE): a randomized controlled trial protocol. BMC Neurol 2013 Jan 11;13:5. [doi: 10.1186/1471-2377-13-5] [Medline: 23311856]
25. Kwakkel G, Winters C, van Wegen EEH, et al. Effects of unilateral upper limb training in two distinct prognostic groups early after stroke: the EXPLICIT-Stroke randomized clinical trial. Neurorehabil Neural Repair 2016 Oct;30(9):804-816. [doi: 10.1177/1545968315624784] [Medline: 26747128]
26. Kwakkel G, Meskers CGM, van Wegen EE, et al. Impact of early applied upper limb stimulation: the EXPLICIT-Stroke programme design. BMC Neurol 2008 Dec 17;8:49. [doi: 10.1186/1471-2377-8-49] [Medline: 19091088]
27. Krakauer JW, Cortés JC. A non-task-oriented approach based on high-dose playful movement exploration for rehabilitation of the upper limb early after stroke: a proposal. NeuroRehabilitation 2018;43(1):31-40. [doi: 10.3233/NRE-172411] [Medline: 30056438]
28. Anwar N, Karimi H, Ahmad A, et al. Virtual reality training using Nintendo Wii games for patients with stroke: randomized controlled trial. JMIR Serious Games 2022 Jun 13;10(2):e29830. [doi: 10.2196/29830] [Medline: 35699989]
29. Broderick M, O'Shea R, Burridge J, Demain S, Johnson L, Bentley P. Examining usability, acceptability, and adoption of a self-directed, technology-based intervention for upper limb rehabilitation after stroke: cohort study. JMIR Rehabil Assist Technol 2023 Aug 21;10:e45993. [doi: 10.2196/45993] [Medline: 37603405]
30. Bernhardt J, Dewey H, Thrift A, Donnan G. Inactive and alone: physical activity within the first 14 days of acute stroke unit care. Stroke 2004 Apr;35(4):1005-1009. [doi: 10.1161/01.STR.0000120727.40792.40] [Medline: 14988574]
31. Huskisson EC. Measurement of pain. J Rheumatol 1982;9(5):768-769. [Medline: 6184474]
32. American College of Sports Medicine. ACSM's Guidelines for Exercise Testing and Prescription, 9th edition: Lippincott Williams & Wilkins; 2013.
33. Lenze EJ, Munin MC, Quear T, et al. The Pittsburgh Rehabilitation Participation Scale: reliability and validity of a clinician-rated measure of participation in acute rehabilitation. Arch Phys Med Rehabil 2004 Mar;85(3):380-384. [doi: 10.1016/j.apmr.2003.06.001] [Medline: 15031821]
34. Nunes C, Darin T. Cross-cultural adaptation of the Intrinsic Motivation Inventory Task Evaluation Questionnaire into Brazilian Portuguese. In: IHC '23: Proceedings of the XXII Brazilian Symposium on Human Factors in Computing Systems: Association for Computing Machinery; 2023. [doi: 10.1145/3638067.3638083]
35. Fugl-Meyer AR, Jääskö L, Leyman I, Olsson S, Steglind S. The post-stroke hemiplegic patient. 1. A method for evaluation of physical performance. Scand J Rehabil Med 1975;7(1):13-31. [Medline: 1135616]
36. Duncan PW, Propst M, Nelson SG. Reliability of the Fugl-Meyer Assessment of sensorimotor recovery following cerebrovascular accident. Phys Ther 1983 Oct;63(10):1606-1610. [doi: 10.1093/ptj/63.10.1606] [Medline: 6622535]
37. Yozbatiran N, Der-Yeghiaian L, Cramer SC. A standardized approach to performing the Action Research Arm Test. Neurorehabil Neural Repair 2008;22(1):78-90. [doi: 10.1177/1545968307305353] [Medline: 17704352]
38. Platz T, Pinkowski C, van Wijck F, Kim IH, di Bella P, Johnson G. Reliability and validity of arm function assessment with standardized guidelines for the Fugl-Meyer Test, Action Research Arm Test and Box and Block Test: a multicentre study. Clin Rehabil 2005 Jun;19(4):404-411. [doi: 10.1191/0269215505cr832oa] [Medline: 15929509]
39. Duncan PW, Wallace D, Lai SM, Johnson D, Embretson S, Laster LJ. The Stroke Impact Scale Version 2.0. Stroke 1999 Oct;30(10):2131-2140. [doi: 10.1161/01.STR.30.10.2131] [Medline: 10512918]
40. Edwards B, O'Connell B. Internal consistency and validity of the Stroke Impact Scale 2.0 (SIS 2.0) and SIS-16 in an Australian sample. Qual Life Res 2003 Dec;12(8):1127-1135. [doi: 10.1023/a:1026109920478] [Medline: 14651430]
41. Arya KN, Verma R, Garg RK. Estimating the minimal clinically important difference of an upper extremity recovery measure in subacute stroke patients. Top Stroke Rehabil 2011 Oct;18 Suppl 1:599-610. [doi: 10.1310/tsr18s01-599] [Medline: 22120029]
42. Lang CE, Edwards DF, Birkenmeier RL, Dromerick AW. Estimating minimal clinically important differences of upper-extremity measures early after stroke. Arch Phys Med Rehabil 2008 Sep;89(9):1693-1700. [doi: 10.1016/j.apmr.2008.02.022] [Medline: 18760153]
43. Dromerick AW, Lang CE, Birkenmeier RL, et al. Very early constraint-induced movement during stroke rehabilitation (VECTORS): a single-center RCT. Neurology (ECronicon) 2009 Jul 21;73(3):195-201. [doi: 10.1212/WNL.0b013e3181ab2b27] [Medline: 19458319]
44. Edwards D, Kumar S, Brinkman L, et al. Telerehabilitation initiated early in post-stroke recovery: a feasibility study. Neurorehabil Neural Repair 2023;37(2-3):131-141. [doi: 10.1177/15459683231159660] [Medline: 36876946]
45. Bashir A. Stroke and telerehabilitation: a brief communication. JMIR Rehabil Assist Technol 2020 Jul 17;7(2):e18919. [doi: 10.2196/18919] [Medline: 32706739]
46. Cumming TB, Packer M, Kramer SF, English C. The prevalence of fatigue after stroke: a systematic review and meta-analysis. Int J Stroke 2016 Dec;11(9):968-977. [doi: 10.1177/1747493016669861] [Medline: 27703065]

XSL•FO
RenderX

47. Lerdal A, Gay CL. Fatigue in the acute phase after first stroke predicts poorer physical health 18 months later. Neurology (ECronicon) 2013 Oct 29;81(18):1581-1587. [doi: 10.1212/WNL.0b013e3182a9f471] [Medline: 24078734]

Abbreviations

ARAT: Action Research Arm Test
FMA-UE: Fugl-Meyer Assessment of the upper extremity
IMI: Intrinsic Motivation Inventory
MCID: minimal clinically important difference
OT: occupational therapist
PRPS: Pittsburgh Rehabilitation Participation Scale
PT: physical therapist
RPE: rating of perceived exertion
SIS: Stroke Impact Scale
ToT: time on task
UE: upper extremity
VAS: visual analog scale

Edited by A Coristine, C Prahm; submitted 24.01.24; peer-reviewed by E Vallefuoco, M Michielsen; revised version received 08.11.24; accepted 11.11.24; published 13.02.25.

Please cite as:
Binyamin-Netser R, Handelzalts S, Goldhamer N, Avni I, Tayer Yeshurun A, Koren Y, Bibas Levy O, Kramer S, Bar Haim S, Shmuelof L
Neurotechnology-Based, Intensive, Supplementary Upper-Extremity Training for Inpatients With Subacute Stroke: Feasibility Study
JMIR Serious Games 2025;13:e56397
URL: https://games.jmir.org/2025/1/e56397
doi:10.2196/56397

© Reut Binyamin-Netser, Shirley Handelzalts, Noy Goldhamer, Inbar Avni, Adi Tayer Yeshurun, Yogev Koren, Ofri Bibas Levy, Shilo Kramer, Simona Bar Haim, Lior Shmuelof. Originally published in JMIR Serious Games (https://games.jmir.org), 13.2.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Reward Feedback Mechanism in Virtual Reality Serious Games in Interventions for Children With Attention Deficits: Pre- and Posttest Experimental Control Group Study

Hao Fang[1], PhD; Changqing Fang[1], BE; Yan Che[2], PhD; Xinyuan Peng[3], MA; Xiaofan Zhang[4], PhD; Di Lin[2], MEng

[1]School of Art & Design, Wuhan Institute of Technology, Wuhan, China

[2]Engineering Research Center for Big Data Application in Private Health Medicine of Fujian Universities, Putian University, Putian, China

[3]School of Arts and Communication, China University of Geosciences, Wuhan, China

[4]Department of Psychiatry, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

Corresponding Author:
Di Lin, MEng
Engineering Research Center for Big Data Application in Private Health Medicine of Fujian Universities
Putian University
No. 1133, Xueyuan Middle Street
Chengxiang District
Putian, 351100
China
Phone: 86 15959234669
Email: lindi1015@163.com

Abstract

Background: Virtual reality (VR) serious games, due to their high level of freedom and realism, influence the rehabilitation training of inhibitory control abilities in children with attention-deficit/hyperactivity disorder (ADHD). Although reward feedback has a motivating effect on improving inhibitory control, the effectiveness and differences between various forms of rewards lack empirical research.

Objective: This study aimed to investigate the effectiveness of different forms of reward feedback on the inhibitory control abilities of children with attention deficits in a VR serious game environment.

Methods: This study focuses on children who meet the diagnostic criteria for ADHD tendencies, using a 2 (material rewards: coin reward and token reward) × 2 (psychological rewards: verbal encouragement and badge reward) factorial between-subject design (N=84), with a control group (n=15) for pre- and posttest experiments. The experimental group received VR feedback reinforcement training, while the control group underwent conventional VR training without feedback. The training period lasted 0.5 months, with each intervention session lasting 25 minutes, occurring twice daily with an interval of at least 5 hours for 28 sessions. Before and after training, the Swanson, Nolan, and Pelham, Version IV Scale (SNAP-IV) Scale, stop signal task, inhibition conflict task, and Simon task were administered to assess the hyperactivity index and the 3 components of inhibitory control ability. The pretest included the SNAP-IV Scale and 3 task tests to obtain baseline data; the posttest involved repeating the above tests after completing all training. Data were entered and analyzed using SPSS (IBM) software. Independent sample *t* tests were performed on the experimental and control groups' pre- and posttest task results to determine whether significant differences existed between group means. Paired sample *t* tests were also conducted on the SNAP-IV Scale's pre- and posttest results to assess the intervention effect's significance.

Results: Reward feedback was more effective than no reward feedback in improving behaviors related to attention deficits in children. Material rewards showed significant effects in the Stop-Signal Task (F_1=13.04, P=.001), Inhibition Conflict Task (F_1=7.34, P=.008), and SNAP-IV test (F_1=69.23, P<.001); mental rewards showed significant effects in the Stop-Signal Task (F_1=38.54, P<.001) and SNAP-IV test (F_1=70.78, P<.001); the interaction between the 2 showed significant effects in the Stop-Signal Task (F_1=4.47, P=.04) and SNAP-IV test (F_1=23.85, P<.001).

Conclusions: Combining material and psychological rewards within a VR platform can effectively improve attention-deficit behaviors in children with ADHD, enhancing their inhibitory control abilities. Among these, coin rewards are more effective

XSL•FO
RenderX

than token rewards, and verbal encouragement outperforms badge rewards. The combined feedback of coin rewards and verbal encouragement yields the most significant improvement in inhibitory control abilities.

(JMIR Serious Games 2025;13:e67338) doi:10.2196/67338

KEYWORDS

serious games; virtual reality; attention deficit; inhibitory control ability; reward feedback

Introduction

Attention-deficit/hyperactivity disorder (ADHD) is a common neurodevelopmental disorder in children, with primary clinical symptoms including inattention and behaviors related to attention deficits [1]. Current research indicates that attention deficits in patients with ADHD are primarily caused by damage or deficiencies in the inhibitory control centers. This damage leads to impairments in executive function, which is a major cause of attention-deficit behaviors in children [2]. With the development of digital health care, virtual reality (VR) technology, characterized by immersion, interactivity, and user engagement with the environment and narrative, has emerged as a promising tool for ADHD rehabilitation training in children [3-6]. Previous studies have demonstrated the feasibility of using VR in treating children with ADHD [7-12]. However, these studies remain in the experimental and exploratory phases, particularly regarding the psychological mechanisms related to inhibitory control [13,14]. Therefore, it is crucial to study the effects of different mechanisms on children's inhibitory control abilities within a VR environment.

Studies have found that inhibitory control interventions typically use reinforcers during training. Using rewards as reinforcers can normalize the inhibitory control of children and adolescents with ADHD to a manageable baseline level. The reward mechanism serves as a supplementary intervention for ADHD, offering a potential approach to improving inhibitory control in affected children [15]. Feedback is also considered a part of reinforcement, as it is real-time and continuous, helping players continuouly improve their behavior. Rewards, on the other hand, are provided after a summary of performance and serve as an incentive to reinforce long-term participation and progress. Specifically, reward feedback involves motivating children toward desired goals through external targets outside of the task [16], such as parental recognition, badges, or praise. By using external rewards to achieve objectives, correct behaviors are repeatedly practiced and reinforced, guiding and strengthening proper awareness to achieve the goal of behavioral training. This method has proven to be highly effective in the rehabilitation training of children with ADHD [17,18]. Existing studies on the impact of reward feedback on children's cognitive behavior choices focus largely on factors such as the form, timing, content, and conditions of the feedback [19]. The most frequently examined aspect is the form of reward, whether material rewards, social rewards, activity-based rewards, or token rewards, on cognitive behavior choices of children with ADHD. In multimedia learning environments, the introduction of various reward forms such as electronic badges, points, coins, verbal praise, and animated expressions has enriched the feedback system for children [20-24]. However, few studies have further refined the comparison of different reward feedback characteristics to determine whether there are differences in their impact on inhibitory control.

In VR training environments, the forms of reward feedback are also highly diverse. Covington and Mueller [25] have noted the use of virtual currencies, electronic badges, points, visual feedback, auditory feedback, and combined visual-auditory feedback (eg, graphics, animations, sound effects, and graphic-sound combinations) in training environments. Among these, point-based feedback is considered the most representative form of material reward feedback. Points can quantitatively reflect students' cumulative learning behaviors and outcomes. On various platforms, points often appear in the form of coins, diamonds, stars, or small red flowers. Mental reward feedback is primarily provided through evaluations from parents, teachers, or psychological experts, with feedback presented as verbal assessments of overall status or positive reinforcement [26]. In addition, badges are seen as symbols or markers of achievement or skill, reflecting the holder's training habits and serving as a tool for motivation and personal habit development. Therefore, further exploration of feedback optimization and innovation in VR environments, especially regarding physiological and dynamic feedback, remains of significant importance for more effective ADHD intervention training [27].

Based on the aforementioned studies, researchers have identified deficits in self-control abilities among children with ADHD, specifically in maintained response inhibition, dominant response inhibition, and interference response inhibition [28]. The high level of freedom and complexity in VR platforms may exacerbate these issues and increase the difficulty of task selection and judgment. The realism and diversity in virtual environments cause children with ADHD to overly focus on stimuli that attracts them, making it difficult to concentrate on targets, with unclear reward feedback guidance [13,15,16,23]. At the same time, current research focuses on the presence of reward feedback and the differential effects of material versus mental feedback on inhibitory control [29,30]. However, few studies have examined the differences between various forms of reward feedback in VR environments. This study aims to fill this research gap by conducting a content analysis of existing reward feedback intervention forms, designing different levels of independent variables, and mapping them into the VR environment to explore the effects of various reward feedback forms on the inhibitory control abilities of children with attention deficits.

XSL•FO
RenderX

Methods

Participants

The experiment initially contacted the parents of 150 children to complete the Swanson, Nolan, and Pelham, Version IV Scale (SNAP-IV) Scale (with scores ranging from 0 to 3; a score greater than 1.6 indicates a diagnosis of ADHD tendencies). Based on the scale, 112 children were selected, with 5- to 6-year-old children showing ADHD tendencies chosen as the study participants (Multimedia Appendix 1). The experiment was conducted in person, with the instructor introducing the experimental procedure and requirements to the participants. After signing the informed consent form, the instructor accompanied the participants throughout the entire experiment. Before the experiment, none of the participants had undergone any training, so there was no consideration of excluding participants with higher familiarity.

In this experiment, the independent variables were the forms of material and mental reward feedback in the VR platform. A 2 (material rewards: coins and tokens) × 2 (mental rewards: badges and verbal expressions) factorial design was used, with a control group and pre- and posttests. The experimental group (N=84) participated in VR feedback–enhanced training, while the control group (n=15) received regular VR training without feedback (correct or incorrect). Before and after the training, participants completed the SNAP-IV Scale, Stop-Signal Task, Inhibition Conflict Task, and Simon Task tests. The total training duration was 0.5 months, with each intervention session lasting 25 minutes, conducted twice per day with an interval of more than 5 hours between sessions, for a total of 28 sessions. Outcome variables included the error rate as an indicator of sustained response inhibition deficits, error rate as an indicator of dominant response inhibition deficits, and error rate effect size (ie, the difference between error rates in inconsistent and consistent trials) as an indicator of interference response inhibition deficits. In addition, the ADHD index and task completion scores were evaluated.

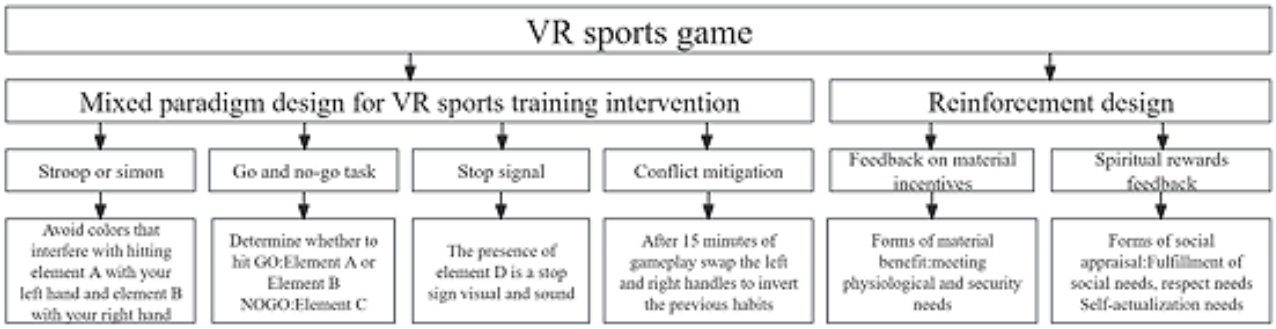
Experimental Materials

The VR training game integrates cognitive and physical training, simulating a VR-based small ball game. In the game, players use a controller to hit 3 types of dynamic objects (Elements A, B, and C) launched from 4 random positions in front of them, adjusting their movements according to the launch angle [31] (Figure 1).

This study categorized VR reward feedback designs by analyzing 33 products across 9 children's VR platforms. Results showed that material rewards in VR are primarily based on a points system, with coins accounting for 61%. While coins function similarly to other tokens, their symbolic meaning as currency requires further exploration. Thus, material rewards were divided into coin rewards and other token rewards for comparison. Verbal encouragement and badge rewards are the primary forms of psychological rewards presented by existing products. Trophies and honorary titles are similar to badges in practical terms, both serving as symbols of honor, while verbal encouragement falls under visual and auditory feedback. Therefore, this study categorizes psychological rewards into verbal encouragement and badge rewards. The study focused on the combined effects of mental and material rewards (Table 1).

Based on children's preference for bright colors, correct answers were associated with a yellow-dominant theme. The experimental group materials were designed as illustrated: from left to right and top to bottom, they represent Coin Reward and Verbal Encouragement, Token Reward and Verbal Encouragement, Coin Reward and Badge, and Token Reward and Badge (Figure 2).

Figure 1. Diagram of the logic and interactive postures combining cognitive and motor training in virtual reality (VR) serious games.



XSL•FO
RenderX

Table 1. Classification and distribution of reward feedback on virtual reality children's training platforms.

Reward feedback and level of incentive feedback	N[a]/(S[b] %)
Material rewards	
Coin reward	20 (61)
Token rewards	13 (39)
Mental rewards	
Verbal rewards	19 (58)
Emoji rewards	14 (42)
Badge rewards	17 (51)
Trophy rewards	3 (9)
Honorary title rewards	2 (6)

[a]N: the number of rewards for feedback in 33 products across 9 VR educational resource platforms.

[b]S: the 33 products of the 9 VR educational resource platforms studied.

Figure 2. Screenshot of feedback types in serious games for the experimental group.



Procedure and Outcome Measurement

In the pretest, the SNAP-IV Scale was first used to assess the participants' hyperactivity index, followed by the stop signal task, Simon task, and inhibition conflict task to test their response inhibition abilities. The posttest involved repeating the same testing procedure after the completion of the VR rehabilitation training.

Stop-Signal Task

In this task, participants needed to complete 2 objectives. First, when symbols appeared to the left or right of a blue star, they responded by pressing the corresponding arrow key on the keyboard based on the symbol's position relative to the star. Second, participants were instructed not to respond when a yellow sun appeared. There was a total of 120 trials, 30 of which were stop trials, accounting for 25% of the total. The error rate served as a reference indicator for maintained response inhibition deficits.

Inhibition Conflict Task

The task was divided into 2 parts: first, participants pressed the "B" key (blue) when a blue square appeared and pressed the "G" key (green) when a green square appeared, regardless of the spatial location of the squares. This part included 80 trials. Second, the task reversed, participants pressed the "G" key when a blue square appeared and the "B" key when a green square appeared, with 80 trials in total. The error rate in the second part served as an indicator of dominant response inhibition deficits.

Simon Task

Participants were instructed to ignore the spatial location of the arrows and press the key corresponding to the arrow's direction,

XSL•FO
RenderX

even if their response was incorrect, and continue to the next trial. The trials in which the spatial location and arrow direction matched were called “consistent trials,” while those in which they did not match were called “inconsistent trials.” The error rate effect size was calculated as the difference between the error rates in inconsistent and consistent trials. There were 80 inconsistent trials and 80 consistent trials, with each trial separated by a 1500-millisecond interval. The error rate effect size was used as an indicator of interference inhibition deficits.

Statistical Analyses

To ensure the power of the experiment, we set the statistical power at 80% and the significance level at α=.05, using G*Power (Franz Faul, Edgar Erdfelder, Albert-Georg Lang, and Axel Buchnere) software to estimate the sample size, which required approximately 15 participants per group [32]. In this study, the collected data will be entered and analyzed using SPSS software. In data processing, outliers were identified using the SD method: data points deviating from the mean by more than ±2 SD were considered outliers and removed. For missing data, variables with a missing ratio exceeding 30% were deleted. Individual records with missing values were directly removed to ensure the reliability and consistency of the data analysis. To assess the effects of different reward feedback on hyperactivity indices and inhibitory control abilities in children with ADHD, a series of scientific statistical methods were applied and systematically analyzed. First, a one-way ANOVA was used to compare the pretest hyperactivity index results across different groups, verifying whether the baseline data of each group were comparable before the experiment. Subsequently, paired-sample *t* tests were used to analyze the changes in error rates before and after the intervention in both the experimental and control groups to evaluate the intervention effects of VR rehabilitation training. To further explore the effects of different reward feedback on intervention outcomes, the researchers used independent-sample *t* tests to compare the training results between the experimental groups. Simultaneously, a 2-way ANOVA was used to analyze the main effects and interactions of material and psychological rewards on the error rates of the Stop Signal Task, Inhibition Conflict Task, Simon Task, and the Snap-Iv Hyperactivity Index. The researchers further analyzed the detailed relationships between specific variables using simple effects tests for significant interactions. In addition, ANOVA was conducted to compare the mean differences between the material reward and psychological reward groups, to clarify the specific effects of different reward types.

Ethical Considerations

The study has been reviewed and approved by the Research Ethics Committee of Wuhan University of Engineering (approval number WIT2024-001), and permission has been granted to analyze previously collected deterministic data. As the study participants are minors, their guardians signed an informed consent form after being informed about the study’s objectives, methods, and potential risks (Multimedia Appendix 2). The research data have been anonymized or deidentified and are used solely for research analysis, with strict protection of participants’ privacy and information security.

Results

Overview

Based on the pretest and posttest data, extreme values and invalid data were removed, and the number of participants in each group was balanced. A total of 13 participants were excluded, resulting in 99 valid data sets, with 21 participants in the Coin Reward and Badge Reward Group, 21 participants in the Token Reward and Badge Reward Group, 21 participants in the Token Reward and Verbal Encouragement Group, 21 participants in the Coin Reward and Verbal Encouragement Group, and 15 participants in the no Reward Feedback Group.

Comparison of Material and Mental Rewards on the Hyperactivity Index of Children With ADHD

The error rates of the experimental and control groups before and after training in different tasks (Stop-Signal Task, Inhibition Conflict Task, and Simon Task) were compared. A paired-sample *t* test was conducted to compare pretest and posttest results. The results are shown in Table 2.

Before the intervention, there were no statistically significant differences between the experimental and control groups in SNAP-IV scores, Stop-Signal Task error rates, Inhibition Conflict Task error rates, and Simon Task error rates, indicating good comparability between the 2 groups. After the intervention, paired-sample *t* tests showed significant statistical differences between the pretest and posttest results of both groups (P<.001), indicating intervention effects. However, postintervention comparisons showed that the experimental group had significantly lower scores or error rates across all indicators compared to the control group (P<.001). This result suggests that interventions with reward feedback in VR training are more effective than those without reward feedback in improving ADHD-related behavioral deficits in children.

XSL•FO
RenderX

Table 2. Comparison of error rates in different response inhibition tasks before and after intervention between the experimental and control groups (based on paired-samples *t* test).

	Pretest, mean (SD)	Posttest, mean (SD)	*t* test (*df*)
SNAP-IV[a] score			
Experimental group (N=84)	1.36 (0.07)	1.09 (0.09)	24.94 (83)
Control participants (n=15)	1.35 (0.02)	1.31 (0.02)	8.35 (14)
Stop signal task error rate			
Experimental group (N=84)	0.21 (0.01)	0.10 (0.02)	77.78 (83)
Control participants (n=15)	0.21 (0.01)	0.14 (0.01)	29.40 (14)
Suppressing conflict task error rates			
Experimental group (N=84)	0.27 (0.01)	0.16 (0.01)	108.27 (83)
Control participants (n=15)	0.27 (0.001)	0.19 (0.01)	28.77 (14)
Simon mission error rate			
Experimental group (N=84)	0.21 (0.02)	0.09 (0.01)	82.03 (83)
Control participants (n=15)	0.20 (0.01)	0.12 (0.003)	32.60 (14)

[a]SNAP-IV: Swanson, Nolan, and Pelham-IV rating scales.

Comparison of Inhibitory Control Improvement in Children With ADHD Across Different Reward Feedback Levels

Independent-sample *t* tests were conducted on pre- and posttest results of sustained response inhibition training (Stop-Signal Task), dominant response inhibition deficit training (Inhibition Conflict Task), and interference response inhibition deficit training (Simon Task) for 6 groups that received different reward feedback combinations. The results are mentioned in Tables 3 and 4.

XSL•FO
RenderX

Table 3. Results of 2-way ANOVA on the effects of reward types on different response Inhibition Tasks and the SNAP-IV[a] Task.

Source of variation	Square sum	*df*	Mean square	*F* test (*df*)	*P* value	*Partial Eta square*
Stop signal task (R^2=0.41)						
Intercept	0.83	1	0.83	4837.32 (1)	<.001	0.98
Material	0.002	1	0.002	13.04 (1)	.001	0.14
Mental	0.01	1	0.01	38.54 (1)	<.001	0.33
Material and mental	0.001	1	0.001	4.47 (1)	.04	0.05
Inaccuracies	0.01	80	0	N/A[b]	N/A	N/A
Conflict suppression task (R^2=0.24)						
Intercept	2.08	1	2.08	18,821.42 (1)	<.001	1.00
Material	0.001	1	0.001	7.34 (1)	.008	0.08
Mental	0	1	0	3.23 (1)	.08	0.04
Material and mental	0	1	0	0.46 (1)	.50	0.01
Inaccuracies	0.01	80	0	N/A	N/A	N/A
Simon mission (R^2=0.08)						
Intercept	0.63	1	0.63	3,752.46 (1)	<.001	0.98
Material	0	1	.00	2.42 (1)	.12	0.03
Mental	0	1	.00	1.17 (1)	.28	0.01
Material and mental	0	1	.00	1.83 (1)	.18	0.02
Inaccuracies	0.01	80	.00	N/A	N/A	N/A
SNAP-IV task (R^2=0.672)						
Intercept	101.90	1	101.90	37,599.07 (1)	.00	1.00
Material	0.19	1	0.19	69.23 (1)	.00	0.46
Mental	0.19	1	0.19	70.78 (1)	.00	0.47
Material and mental	0.07	1	0.07	23.85 (1)	.00	0.23
Inaccuracies	0.22	80	0.003	N/A	N/A	N/A

[a]SNAP-IV: Swanson, Nolan, and Pelham-IV rating scales.

[b]N/A: not available.

XSL•FO
RenderX

Table 4. Results of one-way ANOVA on the effects of reward types on task error rates (within the experimental group: material vs psychological rewards).

	Square sum	*df*	Mean square	*F* test	*P* value
Stop signal task					
Material					
Comparison	0.002	1	0.002	13.04 (1)	.001
Inaccuracies	0.01	80	0	N/A[a]	N/A
Mental					
Comparison	0.01	1	0.01	38.54 (1)	<.001
Inaccuracies	0.01	80	0	N/A	N/A
Conflict suppression task					
Material					
Comparison	0.001	1	0.001	7.34 (1)	.008
Inaccuracies	0.01	80	0	N/A	N/A
Mental					
Comparison	0	1	0	3.23 (1)	.08
Inaccuracies	0.01	80	0	N/A	N/A
Simon mission					
Material					
Comparison	0.001	1.00	0	3.51 (1)	.07
Inaccuracies	0.02	80	0	N/A	N/A
Mental					
Comparison	0.000	1	0	2.04 (1)	.16
Inaccuracies	0.02	80	0	N/A	N/A
SNAP-IV[b] task					
Material					
Comparison	0.19	1.00	0.19	69.23 (1)	<.001
Inaccuracies	0.22	80	0.003	N/A	N/A
Mental					
Comparison	0.19	1	0.19	70.78 (1)	<.001
Inaccuracies	0.22	80	0.003	N/A	N/A

[a]N/A: not available.

[b]SNAP-IV: Swanson, Nolan, and Pelham-IV rating scales.

Analysis of Stop-Signal Task Error Rate

The dependent variable was the Stop-Signal Task error rate, and the independent variables were material and mental rewards. A 2-way ANOVA was performed. As shown in Table 3, the 2-way ANOVA revealed that material rewards had a significant effect on Stop-Signal Task error rates (F_1=13.04, P=.001), indicating a main effect of material rewards on error rates. Mental rewards also showed a significant effect (F_1=38.54, P<.001), suggesting a main effect of mental rewards on Stop-Signal Task error rates. Furthermore, the interaction between material and mental rewards was significant (F_1=4.47, P=.04).

A comparison of the error rates between the material reward and psychological reward groups reveals (Multimedia Appendix 3) that in both material and psychological rewards, the verbal encouragement group outperforms the badge group (0.082<0.106 and 0.099<0.111), and the coin group overall outperforms the token group (0.106<0.111 and 0.082<0.099).

From Table 4, ANOVA analysis showed that the differences within the material and mental reward groups were statistically significant. In both the coin and token groups, the mean error rates in the Stop-Signal Task differed significantly between the badge and verbal expression groups. Further simple effects tests show that the coin verbal encouragement group performs better than the coin expression group and the coin badge group (the average scores for the verbal encouragement groups in both

XSL•FO
RenderX

reward conditions were 0.082 and 0.099, respectively), with the badge group showing slightly poorer improvement.

Analysis of Inhibition Conflict Task Error Rate Effect

The dependent variable was the error rate in the Inhibition Conflict Task, and the independent variables were material and mental rewards. A 2-way ANOVA was conducted. As shown in Table 3, the analysis revealed that material rewards had a significant effect (F_1=7.34, P<.001), indicating a main effect of material rewards on the error rate in the Inhibition Conflict Task. Mental rewards did not show a significant effect (F_1=3.23, P=.08), indicating that the main effect of mental rewards on the error rate does not exist. In addition, the interaction between material and mental rewards was also not significant (F_1=0.46, P=.50).

A comparison of the material and psychological rewards reveals (Multimedia Appendix 3) that in both material and psychological rewards, the verbal encouragement group outperforms the badge group (0.153<0.156 and 0.158<0.163). ANOVA results in Table 4 show that the differences within the material reward groups were statistically significant (P=.008), with the coin group (mean 0.15) outperforming the token group (mean 0.16). Although the differences within the mental reward groups were not statistically significant (P=.08), the verbal expression group had a better overall performance than the badge group. Both material and mental rewards were more effective than the control group (mean 0.18), confirming the effectiveness of reward feedback.

Analysis of Simon Task Error Rate Effect Size

The dependent variable was the error rate effect size in the Simon Task, and the independent variables were material and mental rewards. A 2-way ANOVA was conducted. As shown in Table 3, material rewards did not show a significant effect (F_1=2.42, P=.12), indicating that there is no main effect of material rewards on the error rate effect size in the Simon Task. Mental rewards also did not show a significant effect (F_1=1.17, P=.28), indicating that there is no main effect of mental rewards on the error rate effect size. In addition, the interaction between material and mental rewards was not significant either (F_1=1.83, P=.18).

A comparison of material and psychological rewards (Multimedia Appendix 2) shows that, in terms of means, the coin verbal encouragement group (mean 0.08) < coin badge group (mean 0.08) < token verbal encouragement group (mean 0.09) < token badge group (mean 0.10), with the coin group outperforming the token group overall, and the verbal encouragement group outperforming the badge group. The coin group overall outperformed the token group, and the verbal expression group outperformed the badge group. However, ANOVA results showed that the differences within the material and mental reward groups were not statistically significant (as shown in Table 4).

Analysis of SNAP-IV Hyperactivity Index

The dependent variable was the SNAP-IV hyperactivity index, and the independent variables were material and mental rewards. A 2-way ANOVA was conducted to study the relationship between material and mental rewards and the hyperactivity index. As shown in the table, material rewards had a significant effect (F_1=69.23, P<.001), indicating that material rewards have a main effect on the hyperactivity index. Mental rewards also showed a significant effect (F_1=70.78, P<.001), indicating that mental rewards have a main effect on the hyperactivity index. In addition, the interaction between material and mental rewards was significant (F_1=23.85, P<.001).

A comparison between material and mental rewards shows that the verbal expression group outperformed the badge group for both reward types, and the coin group outperformed the token group overall. ANOVA analysis revealed statistically significant differences within the material and mental reward groups, indicating that the mean hyperactivity index in the badge and verbal expression groups differed between the coin and token groups. Further simple effects tests showed that in both material and psychological rewards, the verbal encouragement group significantly outperformed the badge group, with differences being statistically significant (P<.001).

Discussion

Principal Findings

This study found that VR intervention training with reward feedback significantly improved inhibitory control in children with ADHD, with material rewards proving more effective than psychological rewards. Specifically, coin rewards were more effective than token rewards, and verbal encouragement outperformed badge rewards. The combination of material and psychological rewards had the most optimal effect on enhancing inhibitory control, particularly the combination of coin rewards and verbal encouragement, which significantly improved the 3 main components of sustained response inhibition, dominant response inhibition, and interference response inhibition. The experimental results are consistent with the conclusions of Sagvoldon's dynamic developmental theory model of ADHD (2005), which posits that ADHD involves a complex interaction between genetic predisposition and environmental influences [33]. Therefore, given the conflict between the flexibility of VR environments and the developmental characteristics of children with ADHD, incorporating the motivating effects of reward feedback into intervention training is both important and necessary.

Comparison With Previous Work

The results indicate that material rewards had significant main effects in the Stop-Signal Task, Inhibition Conflict Task, and SNAP-IV tests, with coin rewards proving significantly more effective than token rewards. The experimental study found that material rewards increase individuals' focus and effort on tasks, thereby improving inhibitory control. When individuals are aware that they will receive a tangible material reward after completing a task, they are more motivated to suppress impulses and temptations, focusing on achieving the goal. This may be because material rewards are directly related to individuals' physiological needs and desires, which activate the reward system, thus enhancing cognitive control. This finding is consistent with Fosco et al [34], who demonstrated that coin

XSL•FO
RenderX

rewards significantly improve task performance in children with ADHD. In addition, studies using functional magnetic resonance imaging (fMRI) have found that material rewards enhance activity in the prefrontal cortex, particularly in the control networks related to inhibitory control. This suggests that material rewards actively improve individuals' inhibitory control by activating neurobiological mechanisms. *Meyer* [35] further demonstrated that increasing coin rewards more effectively enhances task performance in children with ADHD, particularly by improving attention and inhibitory abilities. However, the difference in effectiveness between coin and token rewards may be due to several factors. Studies have suggested that while both coin and token rewards operate effectively in symbolic economic environments, the actual use of money is less likely. However, the stronger monetary significance of coins makes them more effective in encouraging correct behavior in children. This type of reinforcement has been shown to be effective in children with similar disorders, such as autism spectrum disorder. This finding is consistent with the current study, where the controlling significance of coin rewards prompted children to adjust their learning behavior to obtain rewards [36].

In addition, the study found that mental rewards did not show significant main effects in the Inhibition Conflict Task and Simon Task. Children with ADHD may exhibit differences in their sensitivity to rewards and motivation levels [37], being more responsive to immediate and tangible material rewards, while their response to abstract mental rewards may be weaker. Material rewards typically have clear and immediate effects, such as snacks or toys. Mental rewards, such as praise or recognition, tend to be more indirect and abstract, and may not be as immediately noticeable as material rewards. For children with ADHD, immediacy and visibility are crucial to the effectiveness of rewards, making mental rewards relatively less effective. This may also be related to the characteristics of children with ADHD, such as difficulty with attention, impulsivity, and delayed gratification [38-41]. ADHD is associated with abnormalities in the dopamine system, which plays a key role in reward processing and motivation regulation. Mental rewards typically involve internal satisfaction and motivational activation, but children with ADHD may have dopamine regulation deficits, leading to a diminished response to mental rewards.

The study further examined the interaction between material and mental rewards on inhibitory control abilities. The interaction effect was significant only in the Stop-Signal Task and SNAP-IV test, with the verbal expression group outperforming the badge group. The combination of verbal expression and coin rewards was superior to the combination of badges and tokens, indicating that combining material and mental rewards can enhance the motivational effect for children with ADHD. Material rewards provide direct external motivation, while mental rewards offer intrinsic and social motivation. The 2 complement each other, enhancing the responses of children with ADHD to rewards. Research has shown that compared with token-based material rewards, the monetary nature of coin rewards can stimulate higher training motivation. This aligns with previous studies using fMRI and event-related potentials (ERPs), which demonstrated that monetary rewards more significantly activate brain regions involved in reward feedback processes [42,43]. In the combination of token rewards and verbal encouragement, verbal encouragement dominated the reward effect, contributing to a gradual and lasting improvement in intervention outcomes [44]. The superiority of verbal encouragement over badges is consistent with BERNIS's findings, where smiley face (expression) feedback showed a significant negative correlation with P2, P3, and feedback-related negativity amplitudes, indicating that smiley faces as reinforcers were not particularly effective. In addition, related studies have shown that children with ADHD lack awareness of others' emotions [45,46], similar to children with autism spectrum disorder. Patients with ADHD often have difficulty understanding social cues [47]. While patients with ADHD may have social interest, they often struggle to evaluate social feedback, such as facial expressions [36]. In contrast, verbal encouragement provides more positive informational value, enhancing children's sense of competence and, in turn, increasing their intrinsic motivation [42], which positively impacts task performance. This further explains why verbal encouragement outperforms badges.

Strengths and Limitations

The results of this study not only validate the core principles of the Health Behavior Change Theory and Gamification Theory but also provide important empirical support for their development. The Health Behavior Change Theory emphasizes the key role of external incentives and internal motivation in shaping behavior [48,49]. This study, through the design of material rewards (such as coins) and psychological rewards (such as verbal encouragement) in VR training, provided immediate external incentives, enhanced children's sense of competence and intrinsic motivation, and significantly improved their inhibitory control, thereby validating the positive impact of immediate feedback on behavioral self-regulation. Gamification Theory advocates motivating users to engage and change behavior through game elements such as reward mechanisms and immediate feedback. This study shows that the combination of coins and verbal encouragement had the most effective motivational impact, not only increasing task interest but also enhancing children's sense of achievement, reflecting the core principles of competence, relatedness, and autonomy in Gamification Theory.

In addition, this study is the first to systematically compare different reward forms (such as coins vs tokens and verbal encouragement vs badges) and their combined effects. By integrating existing fMRI and ERP studies, it reveals the neurobiological basis of reward mechanisms in enhancing inhibitory control through the activation of the prefrontal cortex. This further refines the theoretical framework and fills the gap in research on the effects of reward mechanisms in ADHD interventions. The findings suggest that for special populations (such as children with ADHD), it is necessary to adjust the design of social incentives, prioritizing more direct forms of motivation such as verbal encouragement. At the same time, the role of immediate rewards in promoting long-term behavior change should be emphasized, providing theoretical support and design insights for health behavior interventions in these populations [50,51].

XSL•FO
RenderX

By integrating VR technology to provide an immersive intervention environment, multidimensionally assessing children's inhibitory control (such as sustained response inhibition, dominant response inhibition, and interference response inhibition), and exploring the neurobiological mechanisms of reward feedback in cognitive-behavioral improvement, this study offers new perspectives from both practical and theoretical levels. It not only advances the development of Health Behavior Change Theory and Gamification Theory but also provides important references for the design of educational games and VR intervention tools for children.

However, the study should also address several influencing factors. First, the characteristics and environmental backgrounds of children with ADHD may vary, resulting in individual differences in their responses to rewards. We attempted to recruit a diverse sample to address this, but individual variability could still influence the results. Future studies should control for these factors or explore how they mediate reward effectiveness. Second, the reward culture within the family, school, and social environments could affect how children respond to reward feedback. This study did not account for these external influences, which may limit the generalizability of the findings. In future research, it would be helpful to assess these cultural influences and their potential impact on reward responses [52]. Third, the study sample was relatively small and homogeneous regarding age and gender. A more extensive and diverse sample, including a wider age range and varying levels of ADHD severity, could provide a more comprehensive understanding of how rewards affect inhibitory control. In addition, future studies should explore the effects of different reward intensities and their long-term impact on behavior. Fourth, this study did not use precise neural monitoring instruments, such as electroencephalogram or eye-tracking devices, to monitor children's attention control during the intervention. Future studies should incorporate these tools to provide a more accurate assessment of the neurobiological mechanisms underlying the effects of reward feedback.

Conclusion

This empirical study explored the design of reward feedback to enhance inhibitory control in children with attention deficits. It examined the effects of VR interventions with and without reward feedback in ADHD training, as well as the influence of material and mental rewards on 3 main components of ADHD in children (sustained response inhibition deficit, dominant response inhibition deficit, and interference response inhibition deficit). The study further investigated the impact of the levels of independent variables on inhibitory control ability.

The study yielded the following conclusions: (1) a VR intervention platform using both material and mental rewards as feedback can effectively improve sustained response inhibition deficits, dominant response inhibition deficits, and interference response inhibition deficits in children with ADHD, thereby enhancing overall inhibitory control abilities; (2) material reward feedback using coin rewards is more effective in improving inhibitory control abilities in children than token (candy) rewards; (3) mental reward feedback using verbal encouragement is more effective in improving inhibitory control abilities than badge rewards; and (4) the combination of coin rewards and verbal encouragement feedback yields the best results in enhancing children's inhibitory control abilities.

Acknowledgments

We thank the Wuhan Institute of Technology for their support of this research. This work was supported by the General Project of the National Social Science Fund (NSSF) of China in Art (project 23BG133), the National Natural Science Foundation of China (NSFC) General Project (project 82071542), and the Foundation of Engineering Research Center for Big Data Application in Private Health Medicine of Fujian Universities at Putian University (MKF202206).

Data Availability

All data generated or analyzed during this study are included in this published article (and Multimedia Appendices 1-3 information files).

Authors' Contributions

HF, CF, YC, and XP were responsible for the conceptualization and methodology of the study. HF, CF, YC, and XP developed the tools and settings. HF, CF, and XP conducted the data collection and formal analysis. HF, CF, YC, and XP were involved in data curation. HF, CF, YC, XP, XZ, and DL contributed to drafting the manuscript. All authors reviewed and edited the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
SNAP-IV (Swanson, Nolan, and Pelham, Version IV Scale) scale (parent version; 2018).
[DOCX File , 19 KB - games_v13i1e67338_app1.docx]

XSL•FO
RenderX

Multimedia Appendix 2
Informed Consent Form.
[DOCX File , 24 KB - games_v13i1e67338_app2.docx]

Multimedia Appendix 3
Comparison of error rates between material and spiritual rewards.
[DOCX File , 18 KB - games_v13i1e67338_app3.docx]

References

1. Diagnostic and Statistical Manual of Mental Disorders: DSM-5. Washington, DC: American Psychiatric Association; 2013.
2. Peñarrubia M, Navarro-Soria I, Palacios J, Fenollar-Cortés J. ADHD symptomatology, executive function and cognitive performance differences between family foster care and control group in ADHD-diagnosed children. Children (Basel) 2021;8(5):405 [FREE Full text] [doi: 10.3390/children8050405] [Medline: 34067856]
3. Boulton CA, Kent C, Williams HT. Virtual learning environment engagement and learning outcomes at a 'bricks-and-mortar' university. Computers & Education 2018;126:129-142. [doi: 10.1016/j.compedu.2018.06.031]
4. Freina L. A literature review on immersive virtual reality in education: state of the art and perspectives. 2015 Presented at: A literature review on immersive virtual reality in education: state of the art and perspectives. proceedings of the 11th International Scientific Conference e Learning and Software for Education; 2023 April 27-28; Bucharest (Romania) p. 133-141. [doi: 10.12753/2066-026x-15-020]
5. Ng HYH, Wu CW, Hsu H, Huang C, Hsu AL, Chao YP, et al. Neurological evidence of diverse self-help breathing training with virtual reality and biofeedback assistance: extensive exploration study of electroencephalography markers. JMIR Form Res 2024;8:e55478 [FREE Full text] [doi: 10.2196/55478] [Medline: 39642375]
6. Pandria N, Petronikolou V, Lazaridis A, Karapiperis C, Kouloumpris E, Spachos D, et al. Information system for symptom diagnosis and improvement of attention deficit hyperactivity disorder: protocol for a nonrandomized controlled pilot study. JMIR Res Protoc 2022;11(9):e40189 [FREE Full text] [doi: 10.2196/40189] [Medline: 36169998]
7. Yu C, Wang C, Xie Q, Wang C. Effect of virtual reality technology on attention and motor ability in children with attention-deficit/hyperactivity disorder: systematic review and meta-analysis. JMIR Serious Games 2024;12:e56918 [FREE Full text] [doi: 10.2196/56918] [Medline: 39602820]
8. Wong KP, Zhang B, Lai CYY, Xie YJ, Li Y, Li C, et al. Empowering social growth through virtual reality-based intervention for children with attention-deficit/hyperactivity disorder: 3-arm randomized controlled trial. JMIR Serious Games 2024;12:e58963 [FREE Full text] [doi: 10.2196/58963] [Medline: 39467288]
9. Cho Y, Talboys SL. Trends in South Korean medical device development for attention-deficit/hyperactivity disorder and autism spectrum disorder: narrative review. JMIR Biomed Eng 2024;9:e60399 [FREE Full text] [doi: 10.2196/60399] [Medline: 39405518]
10. Cho B, Kim S, Shin DI, Lee JH, Lee SM, Kim IY, et al. Neurofeedback training with virtual reality for inattention and impulsiveness. Cyberpsychol Behav 2004;7(5):519-526. [doi: 10.1089/cpb.2004.7.519] [Medline: 15667046]
11. Cho BH, Ku J, Jang DP, Kim S, Lee YH, Kim IY, et al. The effect of virtual reality cognitive training for attention enhancement. Cyberpsychol Behav 2002;5(2):129-137. [doi: 10.1089/109493102753770516] [Medline: 12025879]
12. Rohani DA, Sorensen HBD, Puthusserypady S. Brain-computer interface using P300 and virtual reality: a gaming approach for treating ADHD[J]. 2014 Presented at: 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; 2014 August 26-30; Chicago, IL, USA p. 3606-3609. [doi: 10.1109/embc.2014.6944403]
13. Ma I, van Duijvenvoorde A, Scheres A. The interaction between reinforcement and inhibitory control in ADHD: a review and research guidelines. Clin Psychol Rev 2016;44:94-111. [doi: 10.1016/j.cpr.2016.01.001] [Medline: 26802874]
14. Luman M, Oosterlaan J, Sergeant JA. The impact of reinforcement contingencies on AD/HD: a review and theoretical appraisal. Clin Psychol Rev 2005;25(2):183-213. [doi: 10.1016/j.cpr.2004.11.001] [Medline: 15642646]
15. Sader A, Walg M, Ferdinand NK. Rewards can foster response execution and response inhibition in young children diagnosed with ADHD. Journal of Cognition and Development 2022;24(1):37-54. [doi: 10.1080/15248372.2022.2110875]
16. Scheres A, Dijkstra M, Ainslie E, Balkan J, Reynolds B, Sonuga-Barke E, et al. Temporal and probabilistic discounting of rewards in children and adolescents: effects of age and ADHD symptoms. Neuropsychologia 2006;44(11):2092-2103. [doi: 10.1016/j.neuropsychologia.2005.10.012] [Medline: 16303152]
17. Wilbertz G, van Elst LT, Delgado MR, Maier S, Feige B, Philipsen A, et al. Orbitofrontal reward sensitivity and impulsivity in adult attention deficit hyperactivity disorder. Neuroimage 2012;60(1):353-361. [doi: 10.1016/j.neuroimage.2011.12.011] [Medline: 22197790]
18. Fu Y. Research on the design of children's ADHD education and rehabilitation game based on virtual reality technology D. Central Academy of Fine Arts. 2019. URL: https://tinyurl.com/mr3rv2zf [accessed 2025-01-12]
19. Cao L, Shen T, Fu Y. Research on the assessment method of attention deficit hyperactivity disorder based on virtual reality technology[J]. Zhejiang Clinical Medicine 2022;2022(002):024. [doi: 10.4324/9780203065563-16]

XSL•FO
RenderX

20. Gao N. A study of classroom rewards for middle school teachers D. Shaanxi Normal University. 2017. URL: https://cdmd.cnki.com.cn/Article/CDMD-10718-1018055309.htm [accessed 2025-01-12]
21. Yu P. An experimental study of the effects of reward modality, goal orientation and material properties on FOK judgmentD. Qufu Normal University. 2006. URL: https://tinyurl.com/2s3ry8vy [accessed 2025-01-12]
22. Yang, Shaobin. Self-delayed gratification ability of 3-5-year-old children and its relationship with parenting styles D. Hebei Normal University. 2012. URL: https://tinyurl.com/2ertw8pz [accessed 2025-01-12]
23. Friedman NP, Miyake A. The relations among inhibition and interference control functions: a latent-variable analysis. J Exp Psychol Gen 2004;133(1):101-135. [doi: 10.1037/0096-3445.133.1.101] [Medline: 14979754]
24. Unsworth N, Miller JD, Lakey CE, Young DL, Meeks JT, Campbell WK, et al. Exploring the relations among executive functions, fluid intelligence, and personality. Journal of Individual Differences 2009;30(4):194-200. [doi: 10.1027/1614-0001.30.4.194]
25. Covington MV, Mueller KJ. Intrinsic versus extrinsic motivation: an approach/avoidance reformulation. Educational psychology 2001;13(2). [doi: 10.4135/9781412963848.n148]
26. Wiguna T, Wigantara NA, Ismail RI, Kaligis F, Minayati K, Bahana R, et al. A four-step method for the development of an ADHD-VR digital game diagnostic tool prototype for children using a DL model. Front Psychiatry 2020;11:829 [FREE Full text] [doi: 10.3389/fpsyt.2020.00829] [Medline: 32973578]
27. Peng C, Wang D, Zhang Y, Xiao J. A visuo-haptic attention training game with dynamic adjustment of difficulty. IEEE Access 2019;7:68878-68891. [doi: 10.1109/access.2019.2918846]
28. Wodka EL, Mahone EM, Blankner JG, Larson JCG, Fotedar S, Denckla MB, et al. Evidence that response inhibition is a primary deficit in ADHD. J Clin Exp Neuropsychol 2007;29(4):345-356. [doi: 10.1080/13803390600678046] [Medline: 17497558]
29. Pliszka SR, Borcherding SH, Spratley K, Leon S, Irick S. Measuring inhibitory control in children. Journal of Developmental & Behavioral Pediatrics 1997;18(4):254-259. [doi: 10.1097/00004703-199708000-00005]
30. Sutcubasi B, Metin B, Tas C, Krzan FK, Sarı BA, Ozcimen B, et al. The relationship between responsiveness to social and monetary rewards and ADHD symptoms. Cogn Affect Behav Neurosci 2018;18(5):857-868. [doi: 10.3758/s13415-018-0609-1] [Medline: 29943173]
31. Fang H, Zhong C, Liao W. A study on the design of a virtual reality intervention training game for children with ADHD[J]. Ergonomics 2023;29(02):1-7. [doi: 10.13837/j.issn.1006-8309.2023.02.0001]
32. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Hillsdale, NJ: Lawrence Erlbaum Associates; 1989.
33. Sagvolden T, Johansen EB, Aase H, Russell VA. A dynamic developmental theory of attention-deficit/hyperactivity disorder (ADHD) predominantly hyperactive/impulsive and combined subtypes. Behav Brain Sci 2005;28(3):397-419; discussion 419. [doi: 10.1017/S0140525X05000075] [Medline: 16209748]
34. Fosco WD, Hawk LW, Rosch KS, Bubnik MG. Evaluating cognitive and motivational accounts of greater reinforcement effects among children with attention-deficit/hyperactivity disorder. Behav Brain Funct 2015;11:20 [FREE Full text] [doi: 10.1186/s12993-015-0065-9] [Medline: 25926127]
35. De Meyer H, Tripp G, Beckers T, van der Oord S. Conditional learning deficits in children with ADHD can be reduced through reward optimization and response-specific reinforcement. Res Child Adolesc Psychopathol 2021;49(9):1165-1178 [FREE Full text] [doi: 10.1007/s10802-021-00781-5] [Medline: 33792820]
36. Albrecht K, Abeler J, Weber B, Falk A. The brain correlates of the effects of monetary and verbal rewards on intrinsic motivation. Front Neurosci 2014;8:303 [FREE Full text] [doi: 10.3389/fnins.2014.00303] [Medline: 25278834]
37. Doidge JL, Flora DB, Toplak ME. A meta-analytic review of sex differences on delay of gratification and temporal discounting tasks in ADHD and typically developing samples. J Atten Disord 2021;25(4):540-561. [doi: 10.1177/1087054718815588] [Medline: 30596297]
38. Crone EA. The role of the medial frontal cortex in the development of cognitive and social-affective performance monitoring. Psychophysiology 2014;51(10):943-950. [doi: 10.1111/psyp.12252] [Medline: 24942498]
39. Shulman EP, Smith AR, Silva K, Icenogle G, Duell N, Chein J, et al. The dual systems model: review, reappraisal, and reaffirmation. Dev Cogn Neurosci 2016;17:103-117. [doi: 10.1016/j.dcn.2015.12.010] [Medline: 26774291]
40. Braver TS, Krug MK, Chiew KS, Kool W, Westbrook JA, Clement NJ, MOMCAI group. Mechanisms of motivation-cognition interaction: challenges and opportunities. Cogn Affect Behav Neurosci 2014;14(2):443-472 [FREE Full text] [doi: 10.3758/s13415-014-0300-0] [Medline: 24920442]
41. Steinberg L. A social neuroscience perspective on adolescent risk-taking. Dev Rev 2008;28(1):78-106 [FREE Full text] [doi: 10.1016/j.dr.2007.08.002] [Medline: 18509515]
42. Bora E, Pantelis C. Meta-analysis of social cognition in attention-deficit/hyperactivity disorder (ADHD): comparison with healthy controls and autistic spectrum disorder. Psychol Med 2016;46(4):699-716. [doi: 10.1017/S0033291715002573] [Medline: 26707895]
43. Demurie E, Roeyers H, Baeyens D, Sonuga-Barke E. Common alterations in sensitivity to type but not amount of reward in ADHD and autism spectrum disorders. J Child Psychol Psychiatry 2011;52(11):1164-1173. [doi: 10.1111/j.1469-7610.2010.02374.x] [Medline: 21223259]

XSL•FO
RenderX

44. Cheo R. Small rewards Or some encouragement? Using an experiment in China to test extrinsic motivation on academic performance. Singapore Econ. Rev 2017;62(04):797-808. [doi: 10.1142/s0217590817400276]
45. Buitelaar JK, van der Wees M, Swaab-Barneveld H, van der Gaag RJ. Theory of mind and emotion-recognition functioning in autistic spectrum disorders and in psychiatric control and normal children. Dev Psychopathol 1999;11(1):39-58. [doi: 10.1017/s0954579499001947] [Medline: 10208355]
46. Yuill N, Lyon J. Selective difficulty in recognising facial expressions of emotion in boys with ADHD. General performance impairments or specific problems in social cognition? Eur Child Adolesc Psychiatry 2007;16(6):398-404. [doi: 10.1007/s00787-007-0612-5] [Medline: 17401608]
47. Nijmeijer JS, Minderaa RB, Buitelaar JK, Mulligan A, Hartman CA, Hoekstra PJ. Attention-deficit/hyperactivity disorder and social dysfunctioning. Clin Psychol Rev 2008;28(4):692-708. [doi: 10.1016/j.cpr.2007.10.003] [Medline: 18036711]
48. Thompson D. Designing serious video games for health behavior change: current status and future directions. J Diabetes Sci Technol 2012;6(4):807-811 [FREE Full text] [doi: 10.1177/193229681200600411] [Medline: 22920806]
49. Garett R, Young SD. Health care gamification: a study of game mechanics and elements. Tech Know Learn 2018;24(3):341-353. [doi: 10.1007/s10758-018-9353-4]
50. Baranowski T, Buday R, Thompson D, Lyons EJ, Lu AS, Baranowski J. Developing games for health behavior change: getting started. Games Health J 2013;2(4):183-190 [FREE Full text] [doi: 10.1089/g4h.2013.0048] [Medline: 24443708]
51. Estévez-López F, Levelt L, van 't Veer J, Hrehovcsik MM, Visch V, Bramer WM, et al. Game mechanics in eHealth interventions promoting self-management in young people with chronic diseases: a protocol for a systematic review and meta-analyses from the eHealth junior consortium. BMJ Open 2022;12(9):e059581 [FREE Full text] [doi: 10.1136/bmjopen-2021-059581] [Medline: 36167372]
52. Sader A, Walg M, Ferdinand NK. Rewards can foster response execution and response inhibition in young children diagnosed with ADHD. Journal of Cognition and Development 2022;24(1):37-54. [doi: 10.1080/15248372.2022.2110875]

Abbreviations

ADHD: attention-deficit/hyperactivity disorder
ERP: event-related potential
fMRI: functional magnetic resonance imaging
SNAP-IV: Swanson, Nolan, and Pelham, Version IV Scale
VR: virtual reality

Edited by A Coristine; submitted 09.10.24; peer-reviewed by X Liu, L Wu; comments to author 19.11.24; revised version received 24.12.24; accepted 27.12.24; published 24.02.25.

Please cite as:
Fang H, Fang C, Che Y, Peng X, Zhang X, Lin D
Reward Feedback Mechanism in Virtual Reality Serious Games in Interventions for Children With Attention Deficits: Pre- and Posttest Experimental Control Group Study
JMIR Serious Games 2025;13:e67338
URL: https://games.jmir.org/2025/1/e67338
doi:10.2196/67338
PMID:

©Hao Fang, Changqing Fang, Yan Che, Xinyuan Peng, Xiaofan Zhang, Di Lin. Originally published in JMIR Serious Games (https://games.jmir.org), 24.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

A Novel Eye Tracking–Based Gamified Assessment of Contrast Sensitivity Function in Children: Prospective Development and Reliability Study

Yunsi He[1], MD; Yijing Zhuang[1], MD; Lei Feng[1], BM; Xuanyu Xu[2], BA; Ying Deng[2], BE; Moruomi Li[2], MA; Yangfei Pang[1], MM; Ying Yao[1], MD; Wentong Yu[1], MM; Zixuan Xu[1], MD; Yusong Zhou[1], MD; Yudan Zhong[1], MM; Qiuying Li[1], MM; Qingqing Ye[1], BM; Junpeng Yuan[1], BS; Yun Wen[1*], MD, PhD; Jinrong Li[1*], MD, PhD

[1]State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-sen University, Guangdong Provincial Key Laboratory of Ophthalmology and Visual Science, Guangdong Provincial Clinical Research Center for Ocular Diseases, 54 Xianlie South Road, Guangzhou, China

[2]Lightspeed Studios, L&Q Oasis PTE. Ltd, Singapore, Singapore

[*]these authors contributed equally

Corresponding Author:
Jinrong Li, MD, PhD
State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-sen University, Guangdong Provincial Key Laboratory of Ophthalmology and Visual Science, Guangdong Provincial Clinical Research Center for Ocular Diseases, 54 Xianlie South Road, Guangzhou, China

Abstract

Background: Reliable assessment of visual function in young children remains a challenge. Contrast sensitivity function (CSF) is a sensitive and fundamental index of visual performance, yet existing pediatric CSF assessments lack objectivity and adaptability. To bridge this methodological gap, we developed a novel eye tracking–based gamified contrast sensitivity function (ETGCSF) tool that integrates gaze-based detection with interactive gameplay to objectively quantify CSF in an engaging and child-centered manner.

Objective: This study aimed to (1) establish the feasibility and test-retest reliability of the ETGCSF tool in preschool-aged children and (2) evaluate whether optimization using adaptive algorithms and enhanced gamification elements could improve test efficiency while maintaining reliability.

Methods: This was a prospective study with 2 sequential cohorts. A total of 80 Chinese children aged 3 to 6 years were pragmatically recruited from Zhongshan Ophthalmic Center between May 2021 and July 2023. On the basis of timing of data collection, 35% (28/80) of the children were included in experiment 1 (mean age 5.24, SD 0.15 years), and 65% (52/80) were included in experiment 2 (mean age 4.76, SD 0.11 years). Children completed 2 runs of ETGCSF test. Experiment 1 used the baseline ETGCSF protocol, and experiment 2 used the optimized protocol. Primary outcomes were test-retest reliability of the area under the log contrast sensitivity function curve (AULCSF) and CSF acuity, reported as intraclass correlation coefficients (ICCs) with 95% CIs.

Results: In experiment 1, the ETGCSF tool showed strong reliability, with ICCs of 0.890 (95% CI 0.741 - 0.951) for AULCSF and 0.890 (95% CI 0.763 - 0.949) for CSF acuity. The median test duration was 482 (IQR 451 - 569) seconds. In experiment 2, the optimized ETGCSF reduced median test duration to 241 (IQR 189 - 296) seconds (P<.001) while maintaining comparable reliability. AULCSF estimates varied by 0.03 log units across 2 runs (95% CI −0.51 to 0.57; t_{51}=0.749; P=.46), with an ICC of 0.851 (95% CI 0.740 - 0.914; P<.001) that was not significantly different from that of experiment 1 (z=0.660; P=.51). Similarly, CSF acuity estimates varied by 0.004 log units (95% CI –0.33 to 0.32; t_{51}=0.192; P=.85), with an ICC of 0.832 (95% CI 0.708 - 0.904; P<.001), also comparable to that of experiment 1 (z=–0.925; P=.36).

Conclusions: This study introduces a paradigm shift in pediatric visual assessment by leveraging objective eye tracking and gamified engagement to transform contrast sensitivity testing into a scalable, child-friendly process. The ETGCSF tool demonstrated strong reliability and markedly improved efficiency in assessing CSF in preschool children aged 3 to 6 years. These findings support ETGCSF as a promising tool for real-world clinical practice, and its modular design holds potential for future adaptations ranging from streamlined rapid screening in very young children to full CSF profiling for research.

(JMIR Serious Games 2025;13:e81082) doi:10.2196/81082

KEYWORDS

pediatric vision assessment; contrast sensitivity function; eye tracking; gamified testing; test–retest reliability

XSL•FO
RenderX

Introduction

The early years of life represent a critical period for visual development during which appropriate visual input is essential to ensure proper maturation of the visual system [1]. Atypical visual experiences during this stage can lead to long-term or even permanent vision impairments if left unaddressed. Therefore, early detection and timely intervention are crucial in mitigating the impact of pathological visual deprivation and promoting optimal visual outcomes. Vision assessment during this sensitive period plays a pivotal role in the prevention and treatment of visual impairments, offering a window of opportunity to prevent lifelong consequences [2].

Despite recent advances in pediatric vision assessment techniques, significant challenges remain in evaluating visual function in young children. Current methods face limitations stemming from 2 primary factors: the cooperation abilities of young children and the inherent constraints of existing assessment tools. Children often struggle to engage with tests that require subjective feedback or behavioral responses, such as verbal descriptions or manual pointing. To address this, techniques such as preferential looking [3] rely on examiners observing subtle indicators, including eye and head movements, facial expressions, and nonverbal gestures, to infer visual responses [4]. More recently, new contrast sensitivity tests such as the Double Happy test, a preferential looking–based contrast sensitivity assessment, have been developed to incorporate fun stimuli to appeal to children [5]. However, these methods are inherently subjective, with outcomes influenced by the examiner's experience and judgment, leading to potential variability and bias. Furthermore, many current assessment tools focus predominantly on high-contrast visual stimuli, such as stripes or patterns, which only assess a narrow range of visual function. This lack of comprehensive evaluation limits the ability to detect subtle but clinically significant deficits in visual performance, particularly in low-contrast environments or other essential aspects of visual development. These challenges underscore the urgent need for innovative, objective, and reliable assessment approaches capable of overcoming these limitations and providing a more accurate assessment of early childhood visual function.

Among various measures of spatial vision, the contrast sensitivity function (CSF) stands out for its ability to provide a comprehensive assessment of spatial vision and detect subtle visual deficits [6]. CSF is particularly valuable in diagnosing ocular diseases in which early visual impairments may not be detectable through standard methods, and it also plays a crucial role in monitoring disease progression and assessing treatment efficacy [7,8]. Widely used in both clinical and research settings, CSF is traditionally measured using chart-based [9,10] or computer-based psychophysical methods [11-13]. Recent advancements in infrared eye tracking technology have further enhanced the potential of CSF assessment by enabling the use of eye movements as objective, nonverbal indicators of visual function. By monitoring different aspects of eye movement patterns, such as fixation [4,14-18], microsaccades [19], saccades [20-22], smooth pursuit [23,24], and optokinetic nystagmus [25,26], the detection or discrimination threshold to the visual stimulus can be estimated objectively and precisely. The literature indicates that eye tracking can quantify pediatric detection thresholds at specific spatial frequencies (SFs), and several studies have further derived full CSF profiles, highlighting its potential for pediatric CSF assessment.

Another major challenge in contrast sensitivity testing for children is the prolonged testing time and the associated risk of attention lapses, particularly during repeated exposure to the same pattern. As a psychophysical assessment, the reliability of contrast sensitivity measurements is highly dependent on children's sustained attention and cooperation. Lapses typically underestimate true visual ability [27,28]. Incorporating gamified elements into the testing process can help mitigate this problem by enhancing test appeal and increasing children's interest and engagement [29]. For example, a tablet-based interactive "pop CSF" game was developed in which children touched the screen to "catch" a Gabor patch; this approach successfully estimated CSF and differentiated visual impairment in children [30,31], underscoring the promise of gamification for pediatric CSF assessment. However, because touch-based paradigms rely on overt motor responses, they are prone to response bias and attentional lapses and may be systematically confounded by eye-hand coordination demands, potentially diluting the validity of the CSF estimate.

Building on this evidence, our preliminary studies have demonstrated the feasibility of integrating gaze position recording with the preferential looking technique to objectively quantify CSF, offering a rapid, automated, and reliable approach to measuring spatial vision in adults [32]. Nonetheless, translating these gains to testing in preschool-aged children requires further optimization. In particular, it is necessary to shorten test time, sustain engagement, and minimize the need for verbal or manual responses given the limited cooperation and attention span of young children [29,30,33]. These considerations motivate this work, which targets an objective, child-centered CSF paradigm designed for efficiency, engagement, and robustness in real-world pediatric settings.

In this study, we introduced a novel eye tracking–based gamified contrast sensitivity function (ETGCSF) assessment designed to mitigate the challenges of pediatric vision assessment. The objective of this study was to develop and validate the ETGCSF test for children by (1) assessing its feasibility and reliability and (2) determining the impact of optimization on its efficiency, reliability, and user engagement. This approach aimed to balance these factors, providing a robust foundation for its application in clinical and real-world pediatric vision assessments.

Methods

Ethical Considerations

This study adhered to the principles of the Declaration of Helsinki and was reviewed and approved by the institutional ethics committee of the Zhongshan Ophthalmic Center, Sun Yat-sen University (approval 2023KYPJ343). Written informed consent was obtained from the parents or legal guardians of all participants before study enrollment. All data were anonymized before analysis, and no personally identifiable information was

XSL•FO
RenderX

retained in the research dataset to ensure participant privacy and confidentiality. Participants did not receive financial compensation. No images or supplementary materials that could reveal the identity of participants are included in this manuscript.

Participants

This was a prospective development and reliability study with 2 sequential cohorts. Participants were recruited via clinic-based pragmatic sampling during routine outpatient visits from the visual rehabilitation clinic at the Zhongshan Ophthalmic Center, a tertiary ophthalmic hospital in Guangzhou, China, between May 2021 and July 2023. The sample size was determined pragmatically based on the number of eligible children available, clinic flow, and staff availability during the study period. Inclusion criteria were age of >3 and <6 years, uncorrected visual acuity (UCVA) better than 1.0 logarithm of minimum angle of resolution (logMAR), and parental or guardian consent to participate. Exclusion criteria were blepharoptosis or nystagmus given their potential to interfere with accurate eye tracking. Participation was voluntary, and refusal to participate did not affect the children's ongoing clinical care.

Distance UCVA was measured using a Tumbling E Early Treatment Diabetic Retinopathy Study chart (Wehen Vision) illuminated to a luminance of 200 cd/m^2 and viewed from 4 meters. Visual acuity was recorded in logMAR units, with each correctly identified letter contributing to the score.

A total of 80 Chinese children aged 3 to 6 years were included in the study. Of the 80 participants, 28 (35%; mean age 5.24, SD 0.15 years; n=12, 43% male) were assigned to experiment 1, and 52 (65%; mean age 4.76, SD 0.11 years; n=24, 46% male) were assigned to experiment 2 based on the timing of data collection. Experiment 1 assessed the feasibility and test-retest reliability of the ETGCSF tool to establish its baseline performance in young children. Experiment 2 built on these findings by incorporating enhanced gamified elements such as cartoon characters, adaptive feedback, and fewer trials to improve test efficiency and participant engagement. Examiners completed protocol training before testing. Each participant underwent 2 consecutive ETGCSF tests on the same eye to evaluate within-session variability. Each measurement was conducted independently, with both examiners and participants masked to previous results.

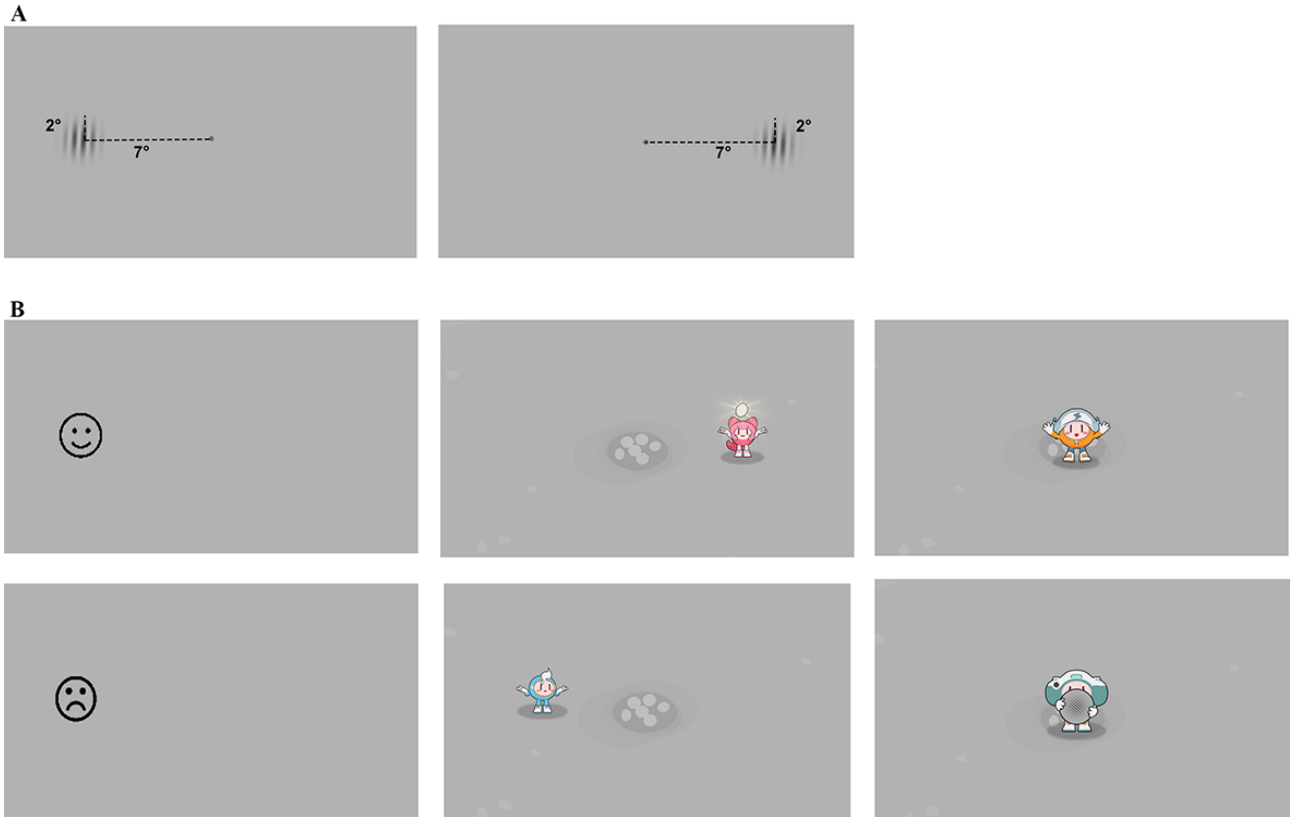
Equipment and Stimuli

Gaze position was recorded using a Tobii Pro Nano remote eye tracker (Tobii AB) with a sampling rate of 65 Hz. Visual stimuli were displayed on a 27-inch Republic of Gamers Swift PG278QR monitor (ASUSTeK Computer Inc) with a resolution of 2560 × 1440, a refresh rate of 165 Hz, and a luminance of 52.1 (SD 1.3) cd/m^2. Real-time gaze monitoring was conducted using a 14-inch Lenovo Xiaoxin Air-14 IIL desktop computer (2020 model).

The primary visual stimulus was a Gabor patch with a radius of 2 degrees of visual angle. To minimize detection cues, the edges were blurred using a half-Gaussian ramp spanning 0.5 degrees of visual angle. The orientation of the patches was randomized between 0 and 180 degrees. Both the contrast and SF were dynamically adjusted using an adaptive algorithm. In each trial, a Gabor patch was randomly displayed on the left or right side of the monitor, positioned 7 degrees from the center (Figure 1A).

XSL•FO
RenderX

Figure 1. Visual stimulus and feedback designs for the eye tracking–based gamified contrast sensitivity function test from a prospective study on its development and reliability. This study was conducted among preschool children (aged 3 - 6 years) at the Zhongshan Ophthalmic Center in Guangzhou, China, between May 2021 and July 2023. (A) Schematic showing the location of Gabor patches, positioned 7 degrees to the left or right of the central fixation point. (B) Evolution of feedback designs: the stick figures used in the initial protocol (experiment 1, left) were replaced with gamified animations featuring the Quantum Family in the optimized protocol (experiment 2, middle and right).



Interactive Feedback Design

To maintain engagement, experiment 1 used feedback stimuli consisting of blinking fruit-shaped stick figures at the center of the screen to capture attention and ensure central fixation. For correct responses, a smiling face was displayed, whereas incorrect responses triggered a crying face (Figure 1B, left panel).

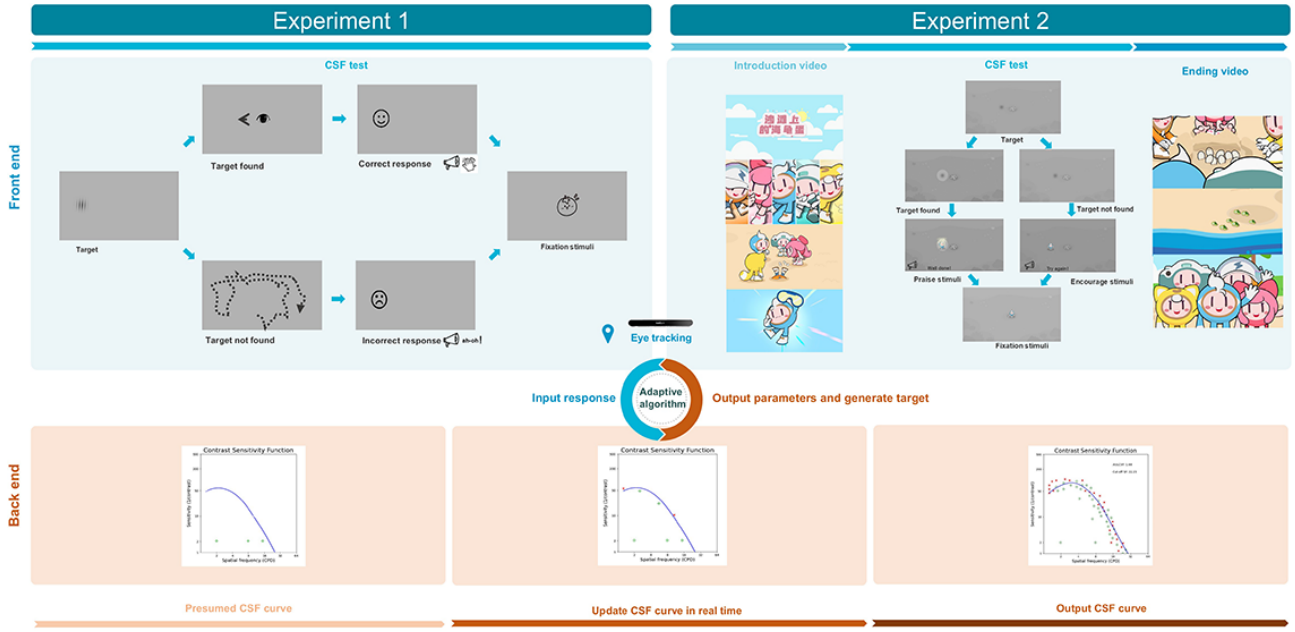
Building on these insights, experiment 2 introduced gamified feedback designed in collaboration with the renowned game company LightSpeed Studios. On the basis of the validation of our feasible method in experiment 1, we further incorporated game elements into the design, with the Quantum Family cartoon characters integrated to make the task more vivid and engaging. At the start of each trial, the character greeted the child with a friendly “Hi” (attention cue). Feedback was tailored to responses: correct answers triggered an animation of the character digging up a turtle egg with praise, whereas incorrect answers prompted encouraging gestures. Additional elements included a bonus cue for 3 consecutive correct responses (eg, enthusiastic greetings) and an instruction cue to re-explain the task after 3 consecutive incorrect responses (Figure 1B, middle and right panels).

CSF Module and Eye Tracking Integration

The test program comprised 2 interconnected modules: the CSF module and the eye tracking module (Figure 2). The CSF module used the quick contrast sensitivity function (qCSF) algorithm, a Bayesian adaptive procedure developed by Lesmes et al [34], to iteratively update 4 key parameters of the CSF curve: peak sensitivity (γ_{max}), peak SF (f_{max}), bandwidth (β), and truncation on the low-frequency side (δ). These parameters modeled the CSF as a truncated log-parabola, capturing individual variations in contrast sensitivity across SFs. By integrating previous knowledge of the CSF’s general shape with trial-by-trial data, the qCSF algorithm selected the optimal combination of contrast and SF at each trial to maximize information gain, improving test efficiency while maintaining reliability and precision. Upon test completion, the algorithm generated a complete CSF curve and computed 2 critical metrics: the area under the log contrast sensitivity function curve (AULCSF) and CSF acuity.

XSL•FO
RenderX

Figure 2. Testing procedure and data processing pipeline for the eye tracking–based gamified contrast sensitivity function (CSF) test. This figure illustrates the workflow for both the baseline (experiment 1) and optimized (experiment 2) protocols from a prospective development and reliability study. This study was conducted among preschool children (aged 3 - 6 years) attending a visual rehabilitation clinic at the Zhongshan Ophthalmic Center, Guangzhou, China (May 2021-July 2023). The system engaged children via a front-end interface, monitored gaze accuracy through an eye tracking module, and dynamically updated stimulus parameters and the CSF curve via a back-end algorithm.



The eye tracking module monitored gaze position in real time to determine response accuracy based on fixation behavior. A correct response was defined as maintaining gaze within a rectangular area centered on the stimulus (4.4 degrees of visual angle) for at least 0.8 seconds, whereas gaze outside this area for more than 4 seconds indicated an incorrect response. After each trial, the eye tracking module communicated the response accuracy to the CSF module, which then adjusted the contrast and SF of the target stimulus for the next trial to maximize information gain. This real-time integration ensured efficient and adaptive testing.

Refinements in Testing Parameters

In experiment 1, the CSF module required 60 trials per session to generate a complete CSF curve, with the SF of the target stimuli ranging from 0.5 to 32 cycles per degree (cpd). Building on these findings, experiment 2 introduced refinements to improve testing efficiency and better accommodate pediatric participants. Enhancements to the Bayesian adaptive algorithm reduced the number of trials to 30 per session while maintaining reliability and precision. Additionally, the SF range was adjusted to 0.5 to 15 cpd to reflect the developmental visual characteristics of young children.

Testing Setup and Experimental Protocols

General Testing Setup and Monitoring

Testing was conducted in a dimly lit room, with children seated either alone or on a parent's lap at a viewing distance of 60 cm. Each eye was tested monocularly, with the other eye patched. Before testing, a 2-point calibration was performed using the Tobii Pro Nano's built-in calibration program to ensure the accuracy of eye tracking. Real-time gaze positioning was monitored throughout the test, enabling the examiner to detect significant deviations in head position or signs of distraction. Adjustments were made as needed, and the test was terminated if necessary to ensure reliable data collection.

Experiment 1: Standard Testing Procedure

In experiment 1, each trial began with a fruit-shaped stick figure displayed at the center of the screen to capture the child's attention, followed by the appearance of the Gabor patch target. Feedback was provided based on the child's response: a smiling face for correct responses or a crying face for incorrect responses, signaling the transition to the next trial. Each session consisted of 60 trials, after which the program generated a complete CSF curve (Figure 2, bottom left panel).

Experiment 2: Gamified Testing Procedure

In experiment 2, enhanced gamification elements were introduced to improve participant engagement. The session began with an introductory video featuring the Quantum Family cartoon characters, who accidentally scattered turtle eggs on a beach. The narrative introduced a magical pair of glasses that transformed the beach into a grayscale image, revealing the striped cloth (representing the Gabor patch) as the hidden location of the eggs. Children were invited to locate the eggs and return them to the nest.

After this, an instructional video explained how to identify the Gabor patch. Each trial began with an attention cue followed by the target presentation. On the basis of the child's response, a reward cue (eg, praise animations) or a motivation cue (eg, encouraging gestures) was displayed to transition to the next trial. Each session consisted of 30 trials. The session concluded with a video showing the Quantum Family returning the eggs to the nest, where they hatched into turtles and swam back to the sea. A CSF curve was generated and displayed at the end of the session (Figure 2, bottom right panel).

XSL•FO
RenderX

The Gabor patch was presented within a predefined target area on a uniform grayscale background, and no other patterns were present within this region, ensuring that gaze responses reflected detection of the Gabor stimulus itself.

Statistical Analysis

The primary CSF metrics included the AULCSF and CSF acuity. AULCSF, calculated by integrating the fitted CSF curve across SFs from 1.5 to 18 cpd, reflected overall spatial vision function. Notably, the fitted curve could extend beyond the maximum stimulus frequency actually presented. CSF acuity was defined as the logMAR-equivalent value corresponding to the SF at a 100% contrast threshold, representing the ability to discern fine details under medium to high contrast conditions. This conversion provided a pragmatic means of situating the psychophysical results within a clinically familiar framework.

Statistical analyses were conducted using SPSS (version 26.0; IBM Corp). Continuous variables with normal distributions were expressed as means and SDs, whereas nonnormally distributed variables were reported as medians with IQRs. The Spearman rank correlation coefficient was used to analyze relationships between test duration and age. Test reliability was assessed using the intraclass correlation coefficient (ICC) [35]; Bland-Altman analysis [36]; and the coefficient of repeatability (CoR), with lower CoR values indicating greater consistency. ICCs were compared across methods using the Fisher *Z* transformation to test for significant differences. Fractional rank precision (FRP) was calculated to assess the test's ability to distinguish participants based on their scores, with higher FRP values (closer to 1.0) indicating greater test-retest consistency [37]. A *P* value of less than .05 was considered statistically significant. Results were reported in compliance with the Guidelines for Reporting Reliability and Agreement Studies criteria [38].

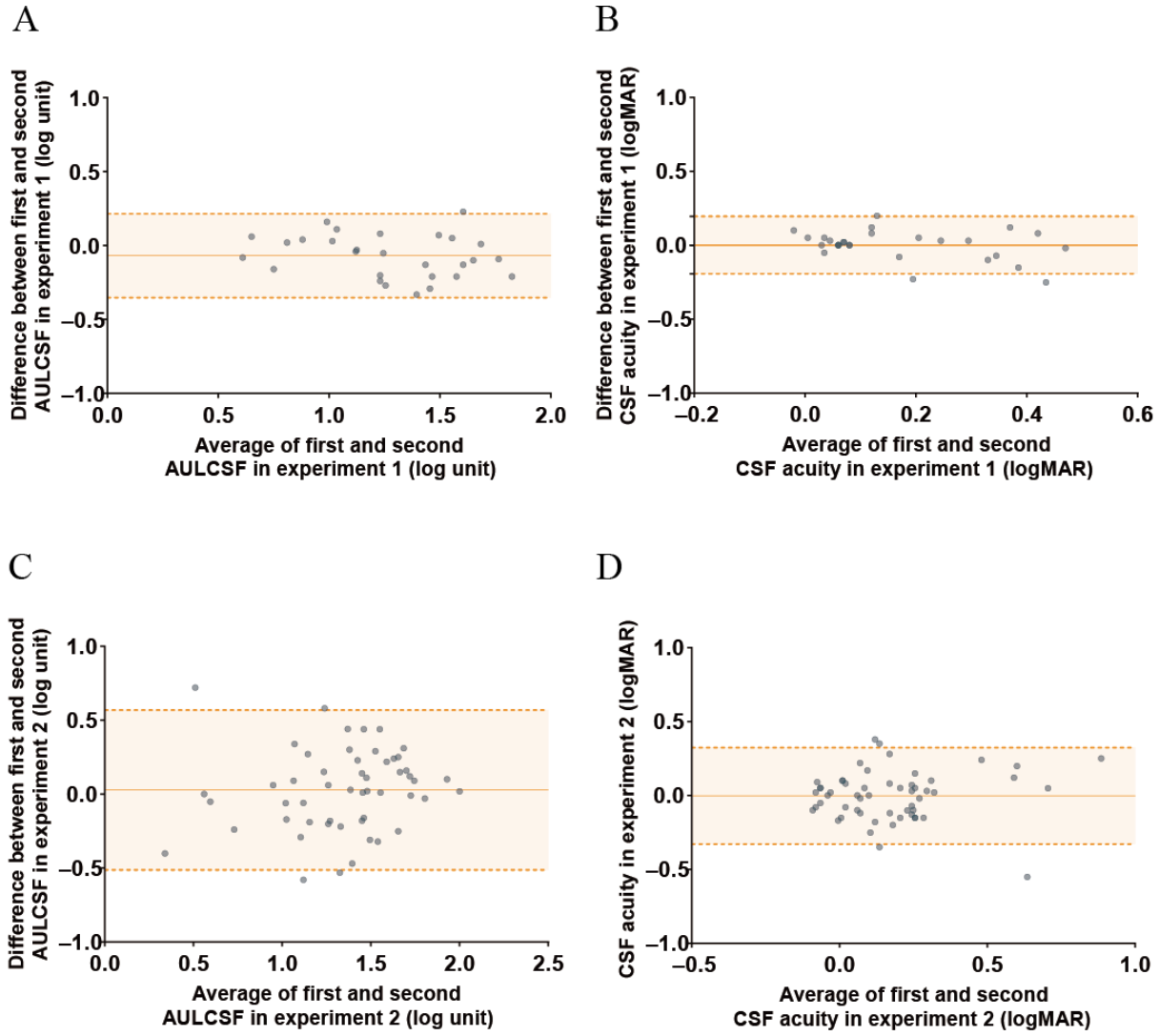
Results

Experiment 1

All 28 children successfully completed 2 rounds of ETGCSF testing in the same eye. The mean UCVA was 0.33 (SD 0.04) logMAR. The median test duration was 482 (IQR 451-569) seconds. Test duration was not correlated with age (Spearman ρ=−0.006; *P*=.98). A mild learning effect was observed in the AULCSF, with estimates varying by a mean of 0.07 log units across 2 runs (95% CI −0.35 to 0.22; t_{27}=2.755; *P*=.01; Figure 3A), with a CoR of 0.319. The ICC for the AULCSF was 0.890 (95% CI 0.741 - 0.951; *P*<.001), and the FRP was 0.80. For CSF acuity, no learning effect was observed, with estimates varying by 0.001 log units across 2 runs (95% CI −0.19 to 0.19; t_{27}=0.057; *P*=.96; not significant; Figure 3B), with a CoR of 0.194. The ICC for CSF acuity was 0.890 (95% CI 0.763 - 0.949; *P*<.001), and the FRP was 0.79.

XSL•FO
RenderX

Figure 3. Test-retest agreement of the eye tracking–based gamified contrast sensitivity function (CSF) measured using Bland-Altman analysis. This analysis was part of a prospective study conducted among preschool children (aged 3 - 6 years) at the Zhongshan Ophthalmic Center, Guangzhou, China (May 2021-July 2023). Plots display the difference between test and retest measurements against their mean for the area under the log contrast sensitivity function (AULCSF) and CSF acuity in the baseline protocol (A and B) and the optimized protocol (C and D). In each plot, the central solid orange line represents the mean difference (bias), and the top and bottom dotted orange lines indicate the upper and lower limits of the 95% agreement interval. Individual data points are shown as semitransparent green circles, with darker shading indicating overlapping values. logMAR: logarithm of minimum angle of resolution.



Experiment 2

All 52 children successfully completed 2 ETGCSF runs in the same eye. The mean UCVA was 0.37 (SD 0.17) logMAR. The median test duration was 241 (IQR 189-296) seconds, significantly shorter than in experiment 1 (z=–7.344; P<.001). Test time was not associated with age (Spearman ρ=0.056; P=.74). No learning effect was observed between initial and retest measurements for either AULCSF or CSF acuity. AULCSF estimates varied by 0.03 log units across 2 runs (95% CI −0.51 to 0.57; t_{51}=0.749; P=.46; not significant; Figure 3C), with a CoR of 0.538. The ICC for the AULCSF was 0.851 (95% CI 0.740 - 0.914; P<.001), closely matching the value from experiment 1 (z=0.660; P=.51). The FRP was 0.82. Similarly, CSF acuity estimates varied by 0.004 log units (95% CI −0.33 to 0.32; t_{51}=0.192; P=.85; not significant; Figure 3D), with a CoR of 0.327. The ICC for CSF acuity was 0.832 (95% CI 0.708 - 0.904; P<.001), which was also comparable to that observed in experiment 1 (z=–0.925; P=.36). The FRP in this case was 0.80.

Discussion

Principal Findings

This study evaluated whether an ETGCSF paradigm functions effectively in preschoolers and whether a more engaging, adaptive version could achieve faster testing without sacrificing reliability. In preschool children, the initial implementation was feasible and reproducible, with high test-retest reliability. The optimized version achieved substantially shorter testing time while maintaining comparable reliability. Overall, these results indicate that the ETGCSF tool provides an objective, child-friendly readout of CSF that balances precision and throughput, supporting its suitability for real-world pediatric use.

XSL•FO
RenderX

With the development of infrared eye tracking technology, eye movements have been adopted as nonverbal cues in psychophysical measurements. Building on this concept, our study incorporated eye tracking technology and further optimized the testing interface and underlying algorithm to create a child-friendly paradigm that preserves objectivity while enhancing efficiency and engagement. Cartoon feedback and gamified settings increased appeal, whereas the adaptive algorithm efficiently selected optimal stimulus parameters to balance test duration and coverage. The resulting ETGCSF paradigm is specifically tailored to the developmental and behavioral characteristics of pediatric populations. Although this study focused primarily on quantitative reliability, informal observations also suggested that the gamified interface in experiment 2 promoted higher engagement. Children appeared more attentive, with fewer interruptions or signs of fatigue during the test than in experiment 1. While these impressions were not formally measured, they support the potential of game-based strategies to improve cooperation in pediatric testing—a hypothesis that may warrant future investigation using standardized behavioral metrics or caregiver-reported engagement scales.

In addition to eye tracking and gamified design, a key design choice concerned the SF range. In experiment 1, we used a broader prior, which was comprehensive but raised occasional concerns about potential aliasing at very high frequencies. Although the adaptive algorithm seldom probed these extremes in preschoolers, this experience informed the next iteration. In response, experiment 2 narrowed the upper range to minimize distortion, shorten testing time, and align more closely with real-world pediatric needs. The finalized window of 0.5 to 15 cpd is intentional: it targets the clinically common and diagnostically challenging mild to moderate deficits, where sensitive detection can influence management. Severe impairment is typically identifiable using simpler, established tools. In this context, ETGCSF is designed to fill the gap between high-contrast acuity tests and comprehensive psychophysics by providing an objective, scalable CSF readout suited to routine pediatric workflows and adaptable to telehealth.

The ETGCSF tool further demonstrated substantial efficiency gains in experiment 2, where reductions in trial numbers shortened the median test duration compared to experiment 1. This improvement was achieved without compromising reliability or validity. The lack of a significant correlation between test duration and age should be interpreted within the context of the study's relatively narrow age range. This finding indicates that test time remained consistent across the specific preschool cohort examined in this study. Future studies encompassing a wider age distribution are warranted to conclusively determine the influence of age on test efficiency.

This study established the reliability and feasibility of the ETGCSF tool for assessing CSF in children. Experiment 1 demonstrated excellent test-retest reliability for both AULCSF and CSF acuity [35] despite individual variability in measurements, likely due to the heterogeneous population. To facilitate clinical interpretation, CSF acuity was expressed in logMAR-equivalent units converted from the cutoff SF, providing a pragmatic bridge between psychophysical measures and conventional acuity metrics. Experiment 2 incorporated a more efficient design with a reduced trial burden, which shortened test durations and mitigated learning effects during retesting. This gain in efficiency involved a trade-off, manifesting as a slight reduction in test precision (evidenced by higher CoR values and wider limits of agreement); however, the preserved high ICCs affirm the method's reliability for group-level analyses. Moreover, the improved rank precision in experiment 2 further underscores its robustness in distinguishing subtle changes in visual performance. Together, these findings indicate that the ETGCSF tool can balance efficiency and reliability in pediatric vision assessments. In a study using the qCSF method for repeated testing in adults with multiple sclerosis, Rosenkranz et al [39] reported a mean difference of –0.049 log units for the AULCSF with a CoR of 0.228 and a mean difference of –0.036 log units for CSF acuity with a CoR of 0.137. Finn et al [40] reported that the qCSF-derived AULCSF and CSF acuity showed excellent reliability, with a CoR of 0.16 and 0.17 log units, respectively. In comparison, the retest variability of our method was larger for both of these metrics, which is potentially attributable to poorer compliance in preschool children compared to adults. Commonly used clinical methods such as the Pelli-Robson chart, the CSV-1000, and the qCSF test typically demonstrate repeatability coefficients ranging from approximately 0.23 to 1.38 log contrast sensitivity across SFs [13,41], and the CoR in our study fell within this range. While the CoR is prone to scaling artifacts, the FRP's strength lies in providing a fair, scale-invariant comparison of reliability between multiple tests administered to the same study population. Although FRP values cannot be directly compared across studies, the high FRP observed in our cohort indicates that our method has excellent discriminative ability and stability, effectively distinguishing individuals with better and worse visual function within the study population.

Limitations

This study has several limitations. First, as participants were primarily recruited from a clinical setting and limited to preschool children aged 3 to 6 years, the findings may not fully reflect the tool's performance in community-based assessment programs or in younger children; future work should include more diverse child populations, including younger children and those with varied ocular and systemic conditions. Second, we did not benchmark the ETGCSF tool against gold-standard methods such as the preferential looking protocol; direct head-to-head comparisons are needed to confirm its clinical equivalence. Third, although the gamified design appeared to improve engagement, we did not quantitatively measure compliance or attention; incorporating objective engagement metrics will clarify the true impact of gamification on test reliability.

Future Work

Looking forward, the modular design of the ETGCSF paradigm allows for adaptation to diverse clinical and research needs. While this study validates the full CSF assessment, the platform could be streamlined for rapid screening in challenging populations (eg, very young children) by focusing on 1 or 2

XSL•FO
RenderX

critical SFs. On the other hand, the full CSF profile remains invaluable for investigating developmental trajectories and nuanced visual deficits in neuro-ophthalmic disorders. Therefore, the ETGCSF tool provides a versatile foundation for both efficient clinical screening and in-depth vision science research.

Conclusions

This study introduces ETGCSF as an objective, efficient, and child-centered assessment of contrast sensitivity for preschool children. Its distinctive contribution lies in coupling eye tracking–derived, nonverbal readouts with gamified task flow, allowing attention to be stabilized and responses to be inferred without speech or pointing while maintaining reliability. Addressing limitations of response-dependent methods, this paradigm functions as a modular framework that can be configured for quick screening or comprehensive CSF profiling and aligned with routine clinics, community screening, and telehealth. Although validation across younger and more diverse populations and direct benchmarks against gold-standard pediatric tests are still required, the findings support practical, near-term translation and point toward broader adoption of objective, digital vision assessments.

Funding

JL is supported by the National Natural Science Foundation of China (82371088) and Guangdong Basic Research Center of Excellence for Major Blinding Eye Diseases Prevention and Treatment (2024-PIZC-040). YW was supported by Science and Technology Projects in Guangzhou (2025A04J3592).

Data Availability

The datasets generated or analyzed during this study are available from the corresponding authors on reasonable request.

Authors' Contributions

YW is the co-corresponding author and can be reached via email at jennywenyun@163.com or phone at 86 020 87330351. Authors Yunsi He, Yijing Zhuang, and Lei Feng are co–first authors.

Conflicts of Interest

None declared.

References

1. Brémond-Gignac D, Copin H, Lapillonne A, Milazzo S, European Network of Study and Research in Eye Development. Visual development in infants: physiological and pathological mechanisms. Curr Opin Ophthalmol 2011 Apr;22 Suppl:S1-S8. [doi: 10.1097/01.icu.0000397180.37316.5d] [Medline: 21478704]
2. US Preventive Services Task Force, Grossman DC, Curry SJ, et al. Vision screening in children aged 6 months to 5 years: US Preventive Services Task Force recommendation statement. JAMA 2017 Sep 5;318(9):836-844. [doi: 10.1001/jama.2017.11260] [Medline: 28873168]
3. Fantz RL. Pattern vision in newborn infants. Science 1963 Apr 19;140(3564):296-297. [doi: 10.1126/science.140.3564.296] [Medline: 17788054]
4. Hathibelagal AR, Leat SJ, Irving EL, Nandakumar K, Eizenman M. Measuring infant visual acuity with gaze tracker monitored visual fixation. Optom Vis Sci 2015 Jul;92(7):823-833. [doi: 10.1097/OPX.0000000000000613] [Medline: 26002001]
5. Mayer DL, Taylor CP, Kran BS. A new contrast sensitivity test for pediatric patients: feasibility and inter-examiner reliability in ocular disorders and cerebral visual impairment. Transl Vis Sci Technol 2020 Aug 20;9(9):30. [doi: 10.1167/tvst.9.9.30] [Medline: 32879786]
6. Chen Z, Zhuang Y, Xu Z, et al. Sensitivity and stability of functional vision tests in detecting subtle changes under multiple simulated conditions. Transl Vis Sci Technol 2021 Jun 1;10(7):7. [doi: 10.1167/tvst.10.7.7] [Medline: 34100925]
7. Anders P, Traber GL, Hall U, et al. Evaluating contrast sensitivity in early and intermediate age-related macular degeneration with the quick contrast sensitivity function. Invest Ophthalmol Vis Sci 2023 Nov 1;64(14):7. [doi: 10.1167/iovs.64.14.7] [Medline: 37934160]
8. Vingopoulos F, Bannerman A, Zhou P, et al. Towards the validation of quantitative contrast sensitivity as a clinical endpoint: correlations with vision-related quality of life in bilateral AMD. Br J Ophthalmol 2024 May 21;108(6):846-851. [doi: 10.1136/bjo-2023-323507] [Medline: 37857454]
9. Kelly SA, Pang Y, Klemencic S. Reliability of the CSV-1000 in adults and children. Optom Vis Sci 2012 Aug;89(8):1172-1181. [doi: 10.1097/OPX.0b013e318264097b] [Medline: 22797510]
10. Arditi A. Improving the design of the letter contrast sensitivity test. Invest Ophthalmol Vis Sci 2005 Jun;46(6):2225-2229. [doi: 10.1167/iovs.04-1198] [Medline: 15914645]
11. Kontsevich LL, Tyler CW. Bayesian adaptive estimation of psychometric slope and threshold. Vision Res 1999 Aug;39(16):2729-2737. [doi: 10.1016/s0042-6989(98)00285-5] [Medline: 10492833]

XSL•FO
RenderX

12. Richman J, Zangalli C, Lu L, Wizov SS, Spaeth E, Spaeth GL. The Spaeth/Richman contrast sensitivity test (SPARCS): design, reproducibility and ability to identify patients with glaucoma. Br J Ophthalmol 2015 Jan;99(1):16-20. [doi: 10.1136/bjophthalmol-2014-305223] [Medline: 25053760]
13. Hou F, Lesmes L, Bex P, Dorr M, Lu ZL. Using 10AFC to further improve the efficiency of the quick CSF method. J Vis 2015;15(9):2. [doi: 10.1167/15.9.2] [Medline: 26161631]
14. Sturm V, Cassel D, Eizenman M. Objective estimation of visual acuity with preferential looking. Invest Ophthalmol Vis Sci 2011 Feb 3;52(2):708-713. [doi: 10.1167/iovs.09-4911] [Medline: 20861485]
15. Jones PR, Kalwarowsky S, Atkinson J, Braddick OJ, Nardini M. Automated measurement of resolution acuity in infants using remote eye-tracking. Invest Ophthalmol Vis Sci 2014 Oct 28;55(12):8102-8110. [doi: 10.1167/iovs.14-15108] [Medline: 25352118]
16. Pérez Roche MT, Yam JC, Liu H, et al. Visual acuity and contrast sensitivity in preterm and full-term children using a novel digital test. Children (Basel) 2022 Dec 31;10(1):87. [doi: 10.3390/children10010087] [Medline: 36670638]
17. Murata N, Toda H, Ubukata H, et al. Development of automated visual acuity measurement using a calibration-free eye-tracking system. Cureus 2024 Jul 12;16(7):e64401. [doi: 10.7759/cureus.64401] [Medline: 39130869]
18. Wen J, Yang B, Li X, Cui J, Wang L. Automated assessment of grating acuity in infants and toddlers using an eye-tracking system. J Vis 2022 Nov 1;22(12):8. [doi: 10.1167/jov.22.12.8] [Medline: 36331256]
19. Denniss J, Scholes C, McGraw PV, Nam SH, Roach NW. Estimation of contrast sensitivity from fixational eye movements. Invest Ophthalmol Vis Sci 2018 Nov 1;59(13):5408-5416. [doi: 10.1167/iovs.18-24674] [Medline: 30452594]
20. Barsingerhorn AD, Boonstra FN, Goossens J. Saccade latencies during a preferential looking task and objective scoring of grating acuity in children with and without visual impairments. Acta Ophthalmol 2019 Sep;97(6):616-625. [doi: 10.1111/aos.14011] [Medline: 30734501]
21. Mooney SW, Alam NM, Prusky GT. Tracking-based interactive assessment of saccades, pursuits, visual field, and contrast sensitivity in children with brain injury. Front Hum Neurosci 2021 Oct 29;15:737409. [doi: 10.3389/fnhum.2021.737409] [Medline: 34776907]
22. Essig P, Sauer Y, Wahl S. Reflexive saccades used for objective and automated measurements of contrast sensitivity in selected areas of visual field. Transl Vis Sci Technol 2022 May 2;11(5):29. [doi: 10.1167/tvst.11.5.29] [Medline: 35639406]
23. Mooney SW, Hill NJ, Tuzun MS, Alam NM, Carmel JB, Prusky GT. Curveball: a tool for rapid measurement of contrast sensitivity based on smooth eye movements. J Vis 2018 Nov 1;18(12):7. [doi: 10.1167/18.12.7] [Medline: 30452585]
24. Mooney SW, Alam NM, Hill NJ, Prusky GT. Gradiate: a radial sweep approach to measuring detailed contrast sensitivity functions from eye movements. J Vis 2020 Dec 2;20(13):17. [doi: 10.1167/jov.20.13.17] [Medline: 33369613]
25. Hyon JY, Yeo HE, Seo JM, Lee IB, Lee JH, Hwang JM. Objective measurement of distance visual acuity determined by computerized optokinetic nystagmus test. Invest Ophthalmol Vis Sci 2010 Feb;51(2):752-757. [doi: 10.1167/iovs.09-4362] [Medline: 19834033]
26. Dakin SC, Turnbull PR. Similar contrast sensitivity functions measured using psychophysics and optokinetic nystagmus. Sci Rep 2016 Oct 4;6:34514. [doi: 10.1038/srep34514] [Medline: 27698486]
27. Manning C, Jones PR, Dekker TM, Pellicano E. Psychophysics with children: investigating the effects of attentional lapses on threshold estimates. Atten Percept Psychophys 2018 Jul;80(5):1311-1324. [doi: 10.3758/s13414-018-1510-2] [Medline: 29582387]
28. Jones PR. Sit still and pay attention: using the Wii Balance-Board to detect lapses in concentration in children during psychophysical testing. Behav Res Methods 2019 Feb;51(1):28-39. [doi: 10.3758/s13428-018-1045-4] [Medline: 29770907]
29. Hosokawa K, Maruya K, Nishida S, Takahashi M, Nakadomari S. Gamified vision test system for daily self-check. Presented at: 2019 IEEE Games, Entertainment, Media Conference (GEM); Jun 18-21, 2019. [doi: 10.1109/GEM.2019.8811563]
30. Elfadaly D, Abdelrazik ST, Thomas PB, Dekker TM, Dahlmann-Noor A, Jones PR. Can psychophysics be fun? Exploring the feasibility of a gamified contrast sensitivity function measure in amblyopic children aged 4-9 years. Front Med (Lausanne) 2020 Aug 26;7:469. [doi: 10.3389/fmed.2020.00469] [Medline: 32984366]
31. Rodriguez-Martinez AC, Tailor-Hamblin VK, Crossland MD, et al. Assessing contrast sensitivity function in CRB1-retinopathies: exploring child-friendly measures of visual function. Transl Vis Sci Technol 2024 Dec 2;13(12):33. [doi: 10.1167/tvst.13.12.33] [Medline: 39699888]
32. Zhuang Y, Gu L, Chen J, et al. The integration of eye tracking responses for the measurement of contrast sensitivity: a proof of concept study. Front Neurosci 2021 Aug 12;15:710578. [doi: 10.3389/fnins.2021.710578] [Medline: 34456676]
33. Gray SI, Bevan C, Campbell S, Cater K. Space Vision: developing a game for vision screening and home-based monitoring. In: Proceedings of the ACM on Human-Computer Interaction 2021, Vol. 5:1-27. [doi: 10.1145/3474699]
34. Lesmes LA, Lu ZL, Baek J, Albright TD. Bayesian adaptive estimation of the contrast sensitivity function: the quick CSF method. J Vis 2010 Mar 30;10(3):17-21. [doi: 10.1167/10.3.17] [Medline: 20377294]
35. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016 Jun;15(2):155-163. [doi: 10.1016/j.jcm.2016.02.012] [Medline: 27330520]
36. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986 Feb 8;1(8476):307-310. [Medline: 2868172]

XSL•FO
RenderX

37. Dorr M, Lesmes LA, Elze T, Wang H, Lu ZL, Bex PJ. Evaluation of the precision of contrast sensitivity function assessment on a tablet device. Sci Rep 2017 Apr 21;7:46706. [doi: 10.1038/srep46706] [Medline: 28429773]
38. Kottner J, Audigé L, Brorson S, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. J Clin Epidemiol 2011 Jan;64(1):96-106. [doi: 10.1016/j.jclinepi.2010.03.002] [Medline: 21130355]
39. Rosenkranz SC, Kaulen B, Zimmermann HG, Bittner AK, Dorr M, Stellmann JP. Validation of computer-adaptive contrast sensitivity as a tool to assess visual impairment in multiple sclerosis patients. Front Neurosci 2021 Feb 23;15:591302. [doi: 10.3389/fnins.2021.591302] [Medline: 33708068]
40. Finn M, Vingopoulos F, Zhao Y, et al. Test-retest repeatability and agreement of the quantitative contrast sensitivity function test: towards the validation of a new clinical endpoint. Graefes Arch Clin Exp Ophthalmol 2024 Mar;262(3):813-822. [doi: 10.1007/s00417-023-06291-y] [Medline: 37955702]
41. Thurman SM, Davey PG, McCray KL, Paronian V, Seitz AR. Predicting individual contrast sensitivity functions from acuity and letter contrast sensitivity measurements. J Vis 2016 Dec 1;16(15):15. [doi: 10.1167/16.15.15] [Medline: 28006065]

Abbreviations

AULCSF: area under the log contrast sensitivity function curve
CoR: coefficient of repeatability
cpd: cycles per degree
CSF: contrast sensitivity function
ETGCSF: eye tracking–based gamified contrast sensitivity function
FRP: fractional rank precision
ICC: intraclass correlation coefficient
logMAR: logarithm of minimum angle of resolution
qCSF: quick contrast sensitivity function
SF: spatial frequency
UCVA: uncorrected visual acuity

Edited by S Brini; submitted 22.07.25; peer-reviewed by AR O'Connor, H Kangari, J Zhou, PJ Bex; revised version received 03.11.25; accepted 03.11.25; published 20.11.25.

Please cite as:
He Y, Zhuang Y, Feng L, Xu X, Deng Y, Li M, Pang Y, Yao Y, Yu W, Xu Z, Zhou Y, Zhong Y, Li Q, Ye Q, Yuan J, Wen Y, Li J
A Novel Eye Tracking–Based Gamified Assessment of Contrast Sensitivity Function in Children: Prospective Development and Reliability Study
JMIR Serious Games 2025;13:e81082
URL: https://games.jmir.org/2025/1/e81082
doi:10.2196/81082

© Yunsi He, Yijing Zhuang, Lei Feng, Xuanyu Xu, Ying Deng, Moruomi Li, Yangfei Pang, Ying Yao, Wentong Yu, Zixuan Xu, Yusong Zhou, Yudan Zhong, Qiuying Li, Qingqing Ye, Junpeng Yuan, Yun Wen, Jinrong Li. Originally published in JMIR Serious Games (https://games.jmir.org), 20.11.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

A Novel Approach Using Serious Game Data to Predict the WISC-V Processing Speed Index in Children With Attention-Deficit/Hyperactivity Disorder: Machine Learning Study

Jun-Su Kim[1*], MS; Yoo Joo Jeong[1,2,3*], BS; Seung-Jae Kim[1], MS; Su Jin Jun[1], BS; Jin-Yeop Park[1], BS; Hyang-Sook Hoe[1,2,3], PhD; Jeong-Heon Song[1], PhD

[1]AI-based Neurodevelopmental Diseases Digital Therapeutics Group, Korea Brain Research Institute (KBRI), 61, Cheomdan-ro, Daegu, Republic of Korea

[2]Neurodegenerative Disease Group, Korea Brain Research Institute (KBRI), Daegu, Republic of Korea

[3]Department of Brain and Cognitive Sciences, Gyeongbuk Institute of Science & Technology, Daegu, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Jeong-Heon Song, PhD
AI-based Neurodevelopmental Diseases Digital Therapeutics Group, Korea Brain Research Institute (KBRI), 61, Cheomdan-ro, Daegu, Republic of Korea

Related Article:

This is a corrected version. See correction statement: https://games.jmir.org/2025/1/e88500

Abstract

Background: The processing speed index (PSI) of the Korean Wechsler Intelligence Scale for Children-Fifth Edition (K-WISC-V) is highly correlated with symptoms of attention-deficit/hyperactivity disorder (ADHD) and is an important indicator of cognitive function. However, restrictions on the frequency of testing prevent short-term PSI assessments. An accessible, objective technique for predicting PSI scores would enable better short-term monitoring and intervention for children with ADHD.

Objective: To enable objective and accessible monitoring of cognitive function beyond traditional clinical assessments, this study aimed to develop a machine learning model that predicts the PSI scores of children with ADHD using behavioral data from serious games.

Methods: Sixty-eight children (6-13 y of age) with ADHD were recruited, and after excluding incomplete data, 59 participants were included in the final analysis. The participants completed an initial PSI assessment using the K-WISC-V followed by 25 minutes of engagement with serious game content. Data from the game sessions were used to train machine learning models, and the models' performance in predicting PSI scores was evaluated using the root mean squared error (RMSE), mean absolute error (MAE), and mean absolute percent error (MAPE), with K-fold cross-validation (k=4) applied to ensure robustness.

Results: Among the individual machine learning models, support vector regression (SVR) had the best performance, with the lowest RMSE of 11.288, MAE of 7.874, and MAPE of 7.375%. The best overall performance was achieved by the ensemble integrating AdaBoost, Elastic Net, and SVR, which recorded the lowest RMSE of 10.072, MAE of 6.798, and MAPE of 6.611%. The predictive accuracy of this ensemble model was highest for PSI scores near the mean value of 100, demonstrating its reliability for clinical applications.

Conclusions: The developed PSI prediction model has the potential to serve as an objective and accessible tool for monitoring cognitive function in children with ADHD. As a complement to traditional assessments, this approach allows continuous tracking of symptom changes and can support more personalized treatment planning in both clinical and everyday settings, which may improve accessibility and adherence. However, the findings need to be validated in larger, more diverse populations, and the long-term feasibility of using serious games in clinical and educational settings must be further examined.

Trial Registration: The Korea Clinical Research Information Service KCT0009326; https://cris.nih.go.kr/cris/search/detailSearch.do?seq=26799

(JMIR Serious Games 2025;13:e73408) doi:10.2196/73408

XSL•FO
RenderX

KEYWORDS

attention deficit hyperactivity disorder; ADHD; prediction; machine learning; symptom tracking; processing speed; Serious games; digital therapeutics

Introduction

Attention-deficit/hyperactivity disorder (ADHD) is a neurodevelopmental disorder that progresses throughout childhood and continues into adulthood [1]. Accurate assessment and management are essential for determining the best time to intervene and improve treatment outcomes, but the diagnosis of ADHD and other neurodevelopmental disorders often relies on subjective clinical evaluations, which can result in inconsistent or inaccurate diagnoses [2-4]. Additionally, symptom evaluation outside clinical environments, such as at home or school, typically depends on observation rather than objective measurements [5]. Processing speed (PS), that is, the ability to respond to a given stimulus within a limited time frame, may offer valuable insights for treatment and monitoring [6]. Individuals with ADHD exhibit slower PS, which has been shown to hinder an individual's ability to process and respond to information efficiently, negatively affecting academic performance, adaptability, anxiety levels, and social competence [7,8]. Lower scores on PS measures, such as the processing speed index (PSI), are correlated with greater severity of attention deficits [9].

The PSI concept was first introduced in the Third Edition of the Wechsler Intelligence Scale for Children (WISC-III), and a formal composite score was established in the Fourth Edition (WISC-IV) [10,11]. The WISC is one of the most widely used and highly validated tools for assessing cognitive function in children, making its index scores highly reliable [12-14]. The latest version, WISC-V, includes the PSI as one of its indices and is frequently used with children in both clinical and educational settings [12]. Due to its strong psychometric properties, the PSI is considered a reliable indicator of PS. Beyond its utility in diagnosis, the PSI is also associated with daily functioning and provides a useful way to quantify ADHD symptoms, highlighting its importance in managing ADHD. However, PSI assessment has two key limitations. First, for effective PSI measurement, trained professionals, such as clinical psychologists or physicians, must administer and interpret specialized tests. Second, repeated testing can lead to practice effects, which may reduce data validity [15-17]. Consequently, the recommended usage interval for WISC is 1-to 2.6 years, making it inappropriate for continuous monitoring [18-20].

To overcome the methodological limitations of measuring the PSI, this study aims to develop a machine learning model that predicts PSI scores in children with ADHD. The model uses behavioral characteristic data collected from Neuro-World, an artificial intelligence-based serious game classified as Software as a Medical Device (SaMD). Serious games and other digital therapeutics (DTx) have emerged as promising tools for treating neurodevelopmental disorders like ADHD [21-24]. DTx allows continuous, location-independent monitoring, improving adherence and enabling real-time tracking [25,26]. These advantages of DTx may also be useful for cognitive assessment. For example, a recent study demonstrated the feasibility of using mobile platforms for assessing PS and working memory via momentary assessment methods [27]. Building on this, large-scale population studies have used mobile cognitive games to evaluate PS across different age groups and cognitive domains [28]. Although digital tools have shown promise for assessing PS through mobile technologies, most have not been validated against established neuropsychological indices such as the PSI. Moreover, while some research has used machine learning models to predict mild cognitive impairment (MCI) based on Mini-Mental State Examination (MMSE) scores, no study has specifically focused on predicting the PSI from behavioral data [29,30]. In addition, preliminary studies have primarily focused on detecting age-related differences or demonstrating feasibility rather than predicting clinically meaningful cognitive metrics.

Based on the well-established association between ADHD severity and a lower PSI [8], as well as prior evidence of improvements in WISC-V scores following Neuro-World training [31], this study aimed to monitor and track cognitive function related to PS. A systematic evaluation of multiple machine learning algorithms identified the model that exhibited the highest level of performance while maintaining the accuracy and reliability required for PSI prediction from game-derived behavioral data.

Methods

Participants

This study was conducted with pediatric participants aged 6-13 years with a diagnosis of ADHD. A total of 34 children with ADHD were recruited between December 2021 and December 2022 (IRB no. 2021-10-080), and an additional 34 children diagnosed with ADHD were recruited between December 2022 and March 2023 (CRIS Registration no. KCT0009326). However, 4 participants were excluded as they refused to participate. The total sample size of the study was 64 participants: 51 (79.7%) male and 13 (20.3%) female participants. The average age of the participants was 8.5 (SD 1.7) years. The participants were all South Korean and were recruited from a single clinical institute. In this study, demographic factors such as socioeconomic status and the education level of the legal guardian were not collected.

Ethical Considerations

This study was approved by two different institutional review boards (IRBs) due to the different recruitment phases. The initial phase of the study was reviewed and approved by the IRB of Keimyung University Dongsan Hospital (IRB no. 2021-10-080), which is certified by the Korea Laboratory Accreditation Scheme (KOLAS), a nationally recognized accreditation body in South Korea. The subsequent phase of the study was registered with the Korea Clinical Research Information Service (CRIS Registration no. KCT0009326), which is accredited by the World Health Organization (WHO). Participants were

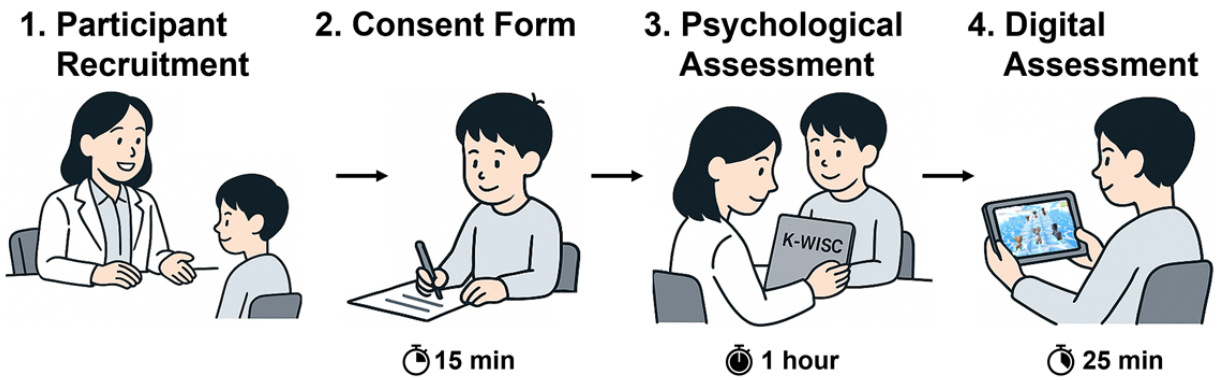
XSL•FO
RenderX

recruited from the Department of Psychiatry at Keimyung University Dongsan Hospital. Before study enrollment, written informed consent was obtained from the legal guardians of all participants, and the study procedures were explained to the participants to secure their assent. The consent form specified that the collected data could be used not only for the primary study but also for future research and secondary analyses without requiring additional consent. All collected data were deidentified and anonymized to ensure confidentiality. As compensation, participants were provided with coverage for transportation costs, and their legal guardians received professional psychological assessment results and evaluation reports. No identifiable images of participants are included in this manuscript or supplementary materials, and no personal information was disclosed.

Experimental Methodology

To minimize disruptions to routine that might affect performance on attention-based activities, we requested that the participants follow their usual daily routines and take their prescribed medications as scheduled. After the participants completed consent forms, they underwent a psychological assessment using the K-WISC-V, which took 1 hour. The participants were provided with a 10-minute break after the psychological assessment and performed gameplay with the developed serious game content for 25 minutes. The overall clinical trial process is presented in Figure 1.

Figure 1. Clinical trial process. A total of 64 children diagnosed with ADHD participated in this study. After completing consent forms, the participants underwent a 1-hour psychological assessment using K-WISC-V. After the K-WISC-V assessment, the children performed a digital assessment using Neuro-World software for 25 minutes. Artificial intelligence–generated image (Generator: ChatGPT, OpenAI; July 14, 2025; Requestor: Jun-Su Kim). ADHD: attention-deficit/hyperactivity disorder; K-WISC-V: Korean Wechsler Intelligence Scale for Children-Fifth Edition.



To ensure an objective setting for gameplay, all participants performed the serious game content independently in a quiet environment free from external distractions and noise. To ensure participant comfort, a pause button was provided, which allowed participants to stop if they experienced dizziness or required a break.

K-WISC-V: Processing Speed Index (PSI)

WISC is widely recognized as a comprehensive instrument for assessing cognitive functioning in children [31]. The WISC-V has been translated and culturally adapted for use in Korea, and this version is known as the Korean Wechsler Intelligence Scale for Children-Fifth Edition (K-WISC-V). The K-WISC-V is targeted to children and adolescents (aged 6 to 16 years and 11 mo) and includes core indices for the verbal comprehension index (VCI), perceptual reasoning index (PRI), working memory index (WMI), and PSI [32,33]. Each assessment is conducted individually, with scores calculated based on established criteria and then converted to age- and gender-specific scaled scores. The index scores are standardized with a mean of 100 and an SD of 15, enabling comparisons of an individual's abilities with those in the same age group [34].

PSI, a critical component of the K-WISC-V, assesses visual PS and psychomotor efficiency by measuring reaction time, spatial navigation, discrimination, and accuracy. Attention and hand-eye coordination are the major contributing factors to the PSI [35]. The PSI captures learning efficiency, attentional control, and executive functioning, which are related to various cognitive dimensions in children. Higher PSI scores indicate faster and more accurate visual decision-making, reflecting greater efficiency in time management for learning tasks and assessments [36]. Symbol search and coding, the two core subtests of PSI, are described in Table 1.

Table . Detailed description of core PSI subtests.

Subtest	Description
Symbol Search	A target symbol is given to the participants, who are asked to identify and mark the matching symbols in a given set within a limited time. This task assesses visual search speed and coordination-discrimination ability.
Coding	A set of numbers is matched with symbols, and participants are required to write down the symbols corresponding to a given set of numbers within a limited time. This task evaluates visuomotor coordination, associative learning ability, and visual short-term memory.

XSL•FO
RenderX

Serious Game Contents for PS Measurement

PS refers to the ability to quickly complete simple tasks while maintaining focus. Cognitive function assessments typically use a brief neuropsychological task that requires approximately 3-5 minutes [37], but neuropsychological tasks used to evaluate sustained attention generally require continuous engagement for 15 to 20 minutes [37]. Based on these observations, we designed serious game sessions with a total duration of 25 minutes. Several recent studies have found that interventions using the software Neuro-World, developed by Woorisoft Co, Ltd (Republic of Korea), improve decision-making ability, overall cognitive PS, working memory, and ADHD-related symptoms, including attention deficit, hyperactivity, and internalization and externalization disorder, in children with ADHD [38-40]. Building on these findings, Neuro-World was used in this study to monitor and track cognitive function related to PS in pediatric participants diagnosed with ADHD.

As shown in Figure 2, Neuro-World comprises 5 independent serious game content modules. Each module takes approximately 5 minutes to complete; thus, the whole digital assessment procedure involves a total of 25 minutes of engagement. The structure of the assessment was designed to assess both cognitive performance and sustained attention within a single session. The sequence of content modules was randomized, and the next content module began automatically after completion of the previous content module. To address potential device or platform bias, all participants completed the game-based assessments using standardized Android tablets (≥9.7-inch screens) with identical touch screen interfaces. No external input devices were used. Table 2 summarizes the technical specifications of the devices and software.

Figure 2. Five Neuro-World serious game content modules used to evaluate processing speed. Content 1: identifying and moving animals; content 2: rearranging sequences; content 3: avoiding obstacles; content 4: dual obstacle avoidance; and content 5: memory matching.



Table . Device and software specifications used in the Neuro-World-based digital assessment.

Component	Specification
Item name	Cognitive Therapy Software
Model name	Neuro-World DTx-ADHD
Operating system	Android
Device type	Tablet (9.7 inches or larger)
Input method	Touch screen

Behavioral Feature Extraction From the Serious Game

The following methodologies were used to extract behavioral characteristics from the 5 distinct content modules. First, the complexity of each content module was dynamically adjusted based on the participant's performance:

- Play speed: As the difficulty level increased, the speed of content progression accelerated, intensifying the sense of urgency during task performance.
- Number of items to remember: The cognitive load was gradually increased by requiring the participants to memorize a larger number of items.

XSL•FO
RenderX

Previous studies have shown that the variability of response time is strongly correlated with ADHD [41-43]. Based on these observations, the game content was designed to gradually reduce the effective time available for problem-solving as the difficulty increased, requiring participants to process information more quickly and efficiently.

Second, a series of data were collected during the cognitive assessment using the Neuro-World content modules: 1, total number of trials conducted in one content module; 2, accuracy rate; 3, total score; 4, total number of screen touches; 5, unnecessary actions during content module performance, calculated as the difference between the total number of screen touches and the number of touches required for correct answers; 6, highest level of performance; 7, average level of performance; 8, SD of performance level; 9 and 10, average time to touch the screen after a task was presented (ie, response time) and its SD; 11 and 12, average time to the final touch required to complete each game (ie, task completion consistency) and its SD; and 13 and 14, average decision-making time, calculated as the difference between the final and first touch times and its SD.

Exploratory studies have suggested that individuals with ADHD exhibit deficits in sustained attention [44-47]. Based on these findings, we divided the total content performance into 2 phases and aimed to quantitatively assess whether the participants' performance improved over time or deteriorated due to reduced concentration, thereby capturing changes in their cognitive task-solving processes. The 14 variables were recorded separately in the 2 phases: 1:14 (initial phase) and 15:28 (later phase). The 28 features were extracted from each of the 5 content modules, resulting in a total of 140 variables (1:140) [48].

Data Preprocessing

Machine learning models, including support vector regression (SVR), random forest, and AdaBoost, generally require a sample size that is at least 10- 20 times larger than the number of input features to ensure stable learning performance [49]. In this study, the initial features consisted of 140 variables extracted from serious game performance data. To minimize the risk of overfitting and enhance the stability of the prediction model, dimensionality reduction was required. Accordingly, to extract the appropriate variables from the data to enhance the model's predictive performance, the correlation-based feature selection (CFS) method was applied. CFS is a filter-based method that identifies an optimal set of variables by maximizing the correlations of the variables with the PSI while minimizing correlations between variables [50]. The effectiveness of the selected variable set in CFS is measured using <Equation 1>. In this equation, *Merits* represent the cumulative correlation between the variables and the external variable; *k* denotes the number of variables; rzi- represents the average correlation between variables and the external variable; and rii- denotes the average intercorrelation between variables [51-53]. The value of *Merits* corresponds to the usefulness of the variable set for prediction; a higher value indicates a stronger correlation with the target class and reduced redundancy among variables. A higher value of *Merits* therefore signifies not only that the variable set is more closely aligned with the target class but also that it contributes to improved model performance by reducing irrelevant or redundant information.

(1)Merits=krzi-k+kk-1rii-

Using the CFS method, 6 variables were selected from 140 content variables. In addition, the variables gender and age were included, resulting in a total of 8 variables used as inputs for the machine learning models in the analysis. The selected variables are listed in Table 3.

Table . List of input variables for the machine learning model.

Index	Variable (n)	Descriptions
1	Con1_answer_rate_back (16)	Accuracy rate in the later phase of content 1
2	Con2_touch_count_diff_back (47)	Number of unnecessary touches in the later phase of content 2
3	Con3_user_touch_count_back (74)	Total number of touches in the later phase of content 3
4	Con5_total_score_front (115)	Total score in the initial phase of content 5
5	Con3_first_touch_time_mean_front (65)	Mean response time to the first touch in the initial phase of content 3
6	Con1_decision_time_mean_front (13)	Mean decision-making time in the initial phase of content 1
7	Gender	Participant's gender
8	Age	Participant's age

Machine Learning Models Used for Predicting the PSI

To predict the PSI scores of children diagnosed with ADHD, we collected performance data from serious games, trained various machine learning models using the selected variables, and compared the predictive performance of the models. We utilized 4 different machine learning models: random forest, AdaBoost, SVR, and Elastic Net. The characteristics of each model are described below.

Random forest is an algorithm that enhances prediction performance by ensemble learning with multiple decision trees

XSL•FO
RenderX

[54]. The random forest algorithm constructs regression trees, processes an input vector composed of feature values corresponding to specific training instances, and averages the predicted outputs from each tree [55]. After creating several training data subsets, a bagging method is applied to reduce correlations among different decision trees to reduce variance [56]. This bagging characteristic enhances the predictive accuracy and stability of the random forest algorithm [54].

AdaBoost (Adaptive Boosting) is a boosting algorithm that iteratively trains a series of weak learners and combines them to form a robust ensemble model [57]. Unlike other boosting methods, AdaBoost applies the relative instead of absolute error, giving more weight to examples with lower values [58]. Additionally, AdaBoost continues to run even if the error rate exceeds 0.5 and supports user-defined iterations, allowing performance improvement over initial iterations [58].

SVR solves regression problems by extending binary classification through convex optimization formulations [59]. SVR generalizes support vector machines to adapt to regression problems and introduces an ε-insensitive region (tube) to reformulate the optimization problem [60]. Unlike other multivariate regression models, support vector machines can learn nonlinear functions with the kernel trick, which maps independent variables in a higher dimensional feature space [61]. The ε-insensitive region approximates continuous functions well and balances model complexity with prediction errors [60].

Elastic Net, proposed by Zou and Hastie in 2005, was developed to estimate the regression coefficient ß in situations where p>n [62]. Elastic Net combines the concepts of Lasso and Ridge Regression to estimate ß [63]. The l2-norm penalty tends to reduce coefficients toward zero while retaining all predictors in the Elastic Net model, potentially leading to multicollinearity, whereas the l1-norm penalty automatically selects key information and continuously reduces redundant information [64]. By combining both penalties, Elastic Net effectively mitigates multicollinearity and enables the group selection of correlated features [65,66].

Machine Learning Model Training and Hyperparameter Optimization for PSI Prediction

To develop an optimal PSI prediction model, we trained the machine learning model and specifically optimized the hyperparameters of each model. No power analysis was conducted because of the exploratory nature of the study, but the use of dimensionality reduction and cross-validation mitigated overfitting and supported generalizability. To prevent issues such as limited generalizability due to the small sample size, we implemented a robust machine learning framework incorporating dimensionality reduction, k-fold cross-validation, hyperparameter tuning via grid search, and random shuffling to mitigate overfitting. Of the initial dataset comprising 64 participants, 5 were excluded because of missing data. Thus, 59 datasets were available for analysis, divided into a training set of 47 (80%) and an independent test set of 12 (20%).

Given the relatively small sample size of this machine learning algorithm study, we also used k-fold cross-validation to ensure robust model validation. The number of folds (k) can be flexibly determined based on the dataset size and the model's structural complexity and inherent characteristics [67-69]. Following this rationale, we experimented with different k values (k=3, 4, 5, and 6) and chose k=4 to achieve a balanced distribution among the training, validation, and test sets. In each fold, approximately 35-36 samples were used for training and 11-12 samples were used for validation. This configuration ensured that the size of the validation set closely matched the final test set, allowing for more effective parameter tuning under conditions similar to final testing.

Generalizability was further enhanced by performing randomized shuffling in each fold (shuffle=True), ensuring that the final test set remained completely unseen throughout training and tuning. A grid search was performed within each fold to optimize hyperparameters, and the optimal settings for each model are presented in Table 4. For example, the random forest model achieved the best performance with "max_features=log2", "n_estimators=50", "max_depth=10", and "min_samples_split=2".

XSL•FO
RenderX

Table . Hyperparameter grids and selected optimal values for each machine learning model.

Model and hyperparameter		Grid search range	Optimal value
Random forest			
	max_features	['sqrt', 'log2']	log2
	n_estimators	[50, 100, 200, 300]	50
	max_depth	[3,5,7,10]	10
	min_samples_split	[2-5]	2
AdaBoost			
	n_estimators	[50, 100, 200, 300]	200
	max_depth	[3-10]	5
	learning_rate	[0.01, 0.05, 0.1, 0.5, 1.0]	1.0
Elastic Net			
	alpha	[0.1, 0.5, 1.0, 2.0, 5.0]	0.1
	l1_ratio	[0.1, 0.3, 0.5, 0.7, 0.9]	0.9
Support vector regression			
	C	[0.1, 1, 10, 100]	10.0
	epsilon	[0.01, 0.1, 1]	1.00
	kernel	['linear', 'rbf', 'poly']	linear

Results

The SVR Machine Learning Model Effectively Predicts PSI Scores From the Serious Game Performance Outcomes of Children Diagnosed With ADHD

The predictive performance of the 4 machine learning models was compared to determine which was most effective in predicting the PSI scores of children diagnosed with ADHD from their Neuro-World content performance outcomes (Table 5). SVR demonstrated the best overall performance on the training set, achieving a root mean squared error (RMSE) of 13.107, mean absolute error (MAE) of 9.933, and mean absolute percent error (MAPE) of 9.792% (Table 5). The Elastic Net model exhibited the second highest performance on the training set, with an RMSE of 13.458, MAE of 10.580, and MAPE of 10.352% (Table 5). Random forest and AdaBoost had higher errors on the training set, with RMSE values of 14.307 and 14.269, respectively.

Table . Comparison of machine learning model performance for predicting processing speed index (PSI) scores of children diagnosed with attention-deficit/hyperactivity disorder (ADHD).

	Training set			Test set		
Model	RMSE[a]	MAE[b]	MAPE[c] (%)	RMSE	MAE	MAPE (%)
Random forest	14.307	11.813	11.828	15.517	13.165	12.314
AdaBoost	14.269	11.676	11.738	14.511	12.092	11.324
Elastic Net	13.458	10.58	10.352	12.488	8.328	7.993
SVR[d]	13.107	9.933	9.792	11.288	7.874	7.375

[a]RMSE: root mean squared error.

[b]MAE: mean absolute error.

[c]MAPE: mean absolute percentage error.

[d]SVR : support vector regression.

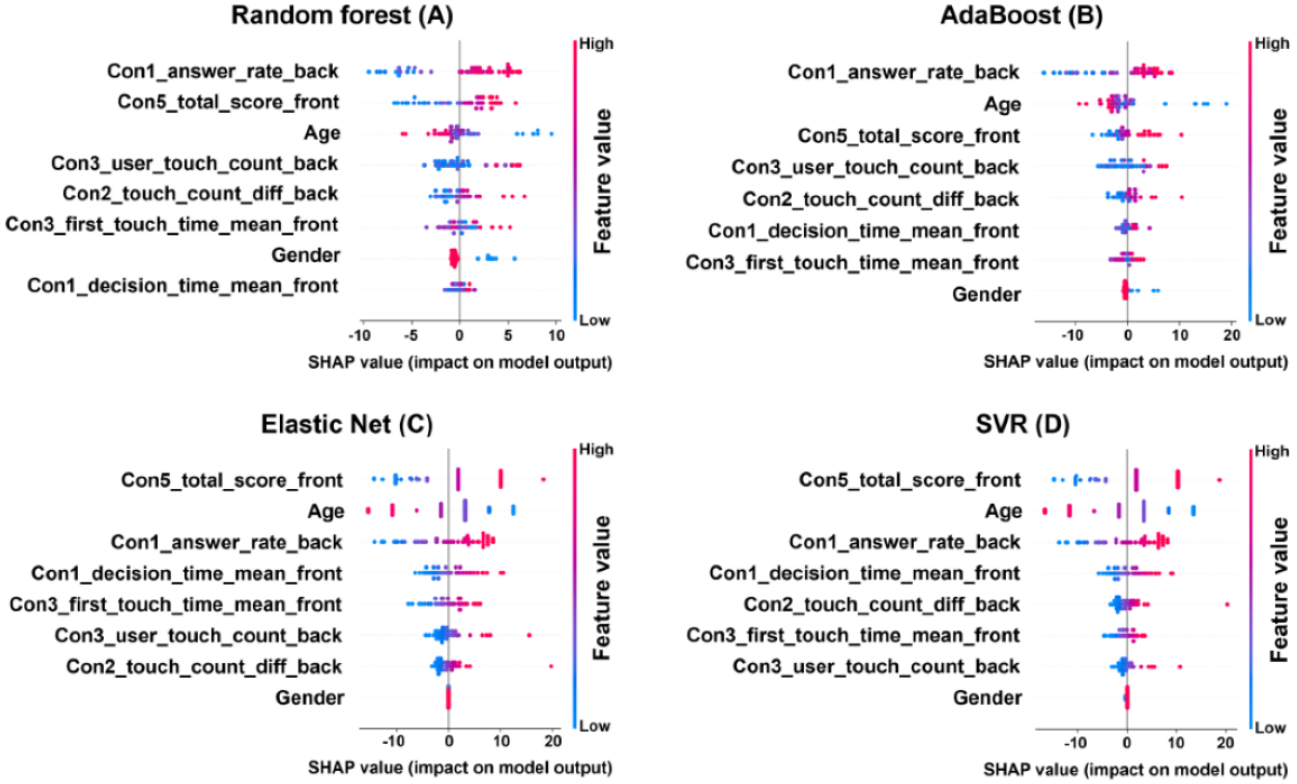
We then investigated the performance of each model on the test set and found that SVR continued to outperform the other models, with an RMSE of 11.288, MAE of 7.874, and MAPE of 7.375% (Table 5). As with the training set, Elastic Net was second, with an RMSE of 12.488, MAE of 8.328, and MAPE of 7.993%, demonstrating the ability of this model to provide accurate predictions (Table 5). Random forest and AdaBoost again displayed weaker performance. Random forest had an RMSE of 15.517, MAE of 13.165, and MAPE of 12.314%, while AdaBoost had an RMSE of 14.511, MAE of 12.092, and MAPE of 11.324% (Table 4). Overall, SVR emerged as the most effective model for predicting PSI across both the training and test sets.

To improve the interpretability of the machine learning model, SHAP (SHapley Additive exPlanations) analysis was conducted

XSL•FO
RenderX

to assess the contribution of each feature to PSI prediction. Figure 3 presents a summary plot of SHAP analysis results for all machine learning models based on the top 8 input variables for each model (detailed descriptions of these 8 input variables are provided in Table 3). Across all models, "Con5_total_score_front", "Con1_answer_rate_back", "age", and "Con1_decision_time_mean_front" consistently emerged as the main prediction variables. While the relative importance of individual features varied across models, the behavioral performance indices derived from the serious games, including response rate, touch time, and decision time, consistently demonstrated a high degree of predictive efficacy. Gender had a negligible impact on all models. The results of the study support the feasibility of using behavioral performance metrics derived from digital games to predict the PSI scores of children diagnosed with ADHD.

Figure 3. SHAP plot for each machine learning model: (A) Random forest, (B) AdaBoost, (C) Elastic Net, and (D) support vector regression (SVR). Each model represents the top 8 input variables contributing to processing speed index (PSI) prediction. Each dot indicates the SHAP value for an individual participant, and the color indicates the initial value (red: high, blue: low). Features are aligned in order of absolute average SHAP value, and higher values indicate a larger effect of the factor on model output. SHAP: SHapley Additive exPlanations.



The Ensemble of AdaBoost and Elastic Net Model Exhibits High Training Performance for Predicting PSI Scores of Children With ADHD

The ensemble technique combines the prediction results of multiple individual machine learning models to achieve superior accuracy and stability performance compared with a single machine learning model [70]. Therefore, to enhance predictive performance by leveraging the strengths of various machine learning models, we applied an ensemble technique using the performance dataset of children with ADHD playing serious game content modules. Specifically, ensemble combinations of random forest, AdaBoost, Elastic Net, and SVR were applied to the training set. The top 5 model combinations with the lowest RMSE scores are presented in Table 6. The combination of AdaBoost and Elastic Net achieved the lowest RMSE of 12.738, indicating the strongest PSI score predictive performance. This combination also demonstrated competitive accuracy, with an MAE of 10.116 and MAPE of 10.086% (Table 6). By contrast, the combination of random forest and Elastic Net had the lowest MAE and MAPE scores of 9.878 and 9.816%, respectively, along with an RMSE of 12.781 (Table 6). These findings showed that combinations of tree-based models (AdaBoost, random forest) with a regression-based model (Elastic Net) outperformed the individual models by leveraging complementary strengths, improving overall predictive performance.

XSL•FO
RenderX

Table . Top 5 ensemble model combinations ranked by RMSE performance on the training set for predicting PSI[a] scores of children with ADHD[b].

Combination	Training set		
	RMSE[c]	MAE[d]	MAPE[e] (%)
[AdaBoost, Elastic Net]	12.738	10.116	10.086
[AdaBoost, Elastic Net, SVR[f]]	12.770	10.013	9.938
[Random forest, Elastic Net]	12.781	9.878	9.816
[Random forest, AdaBoost, Elastic Net, SVR]	12.802	10.135	10.071
[Random forest, Elastic Net, SVR]	12.821	9.930	9.829

[a]PSI:processing speed index.

[b]ADHD: attention-deficit/hyperactivity disorder.

[c]RMSE: root mean squared error.

[d]MAE: mean absolute error.

[e]MAPE: mean absolute percentage error.

[f]SVR: support vector regression.

The Ensemble of AdaBoost, Elastic Net, and SVR Exhibits High Testing Performance for Predicting PSI Scores of Children With ADHD

The performance of the ensemble machine learning combinations was then tested using the remaining 12 datasets of Neuro-World content performance of children with ADHD. The combination of AdaBoost, Elastic Net, and SVR resulted in the lowest RMSE of 10.072, with MAE and MAPE values of 6.798 and 6.611%, respectively, indicating high performance on the test set (Table 7). The combination of AdaBoost and Elastic Net, which showed the highest training performance for predicting PSI scores, exhibited the second highest RMSE of 10.089, with MAE and MAPE values of 7.126 and 6.946%, respectively (Table 7). By contrast, the combination of random forest and Elastic Net, which performed strongly on the training set and had the lowest MAE of 9.878, exhibited higher error rates on the test set, with an MAE of 7.788. Together, these results suggest that the PSI prediction accuracy of the ensemble model of AdaBoost and Elastic Net, which provided consistent performance on both the training and test sets, was enhanced by the inclusion of SVR when applied to the test set.

Table . Test set results of the top 5 ensemble model combinations from the training set for predicting PSI[a] scores of children with ADHD[b].

Combination	Test set		
	RMSE[c]	MAE[d]	MAPE[e] (%)
[AdaBoost, Elastic Net]	10.089	7.126	6.946
[AdaBoost, Elastic Net, SVR[f]]	10.072	6.798	6.611
[Random forest, Elastic Net]	10.432	7.788	7.578
[Random forest, AdaBoost, Elastic Net, SVR]	10.384	7.411	7.272
[Random forest, Elastic Net, SVR]	10.203	7.202	7.035

[a]PSI; processing speed index.

[b]ADHD: attention-deficit/hyperactivity disorder.

[c]RMSE: root mean squared error.

[d]MAE: mean absolute error.

[e]MAPE: mean absolute percentage error.

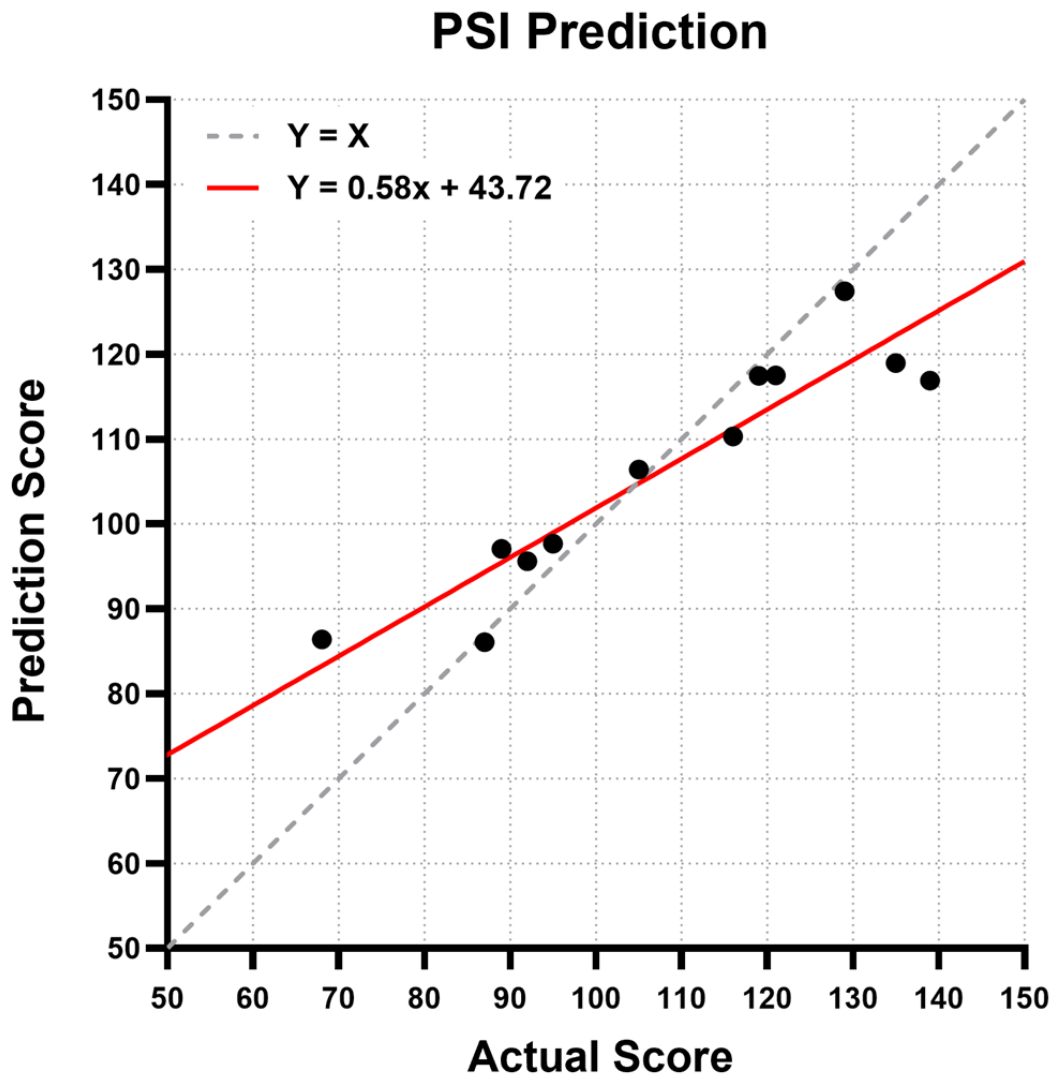
[f]SVR: support vector regression.

Comparison of the PSI Score Predictions of the Ensemble Model of AdaBoost, Elastic Net, and SVR With Actual PSI Scores

Since we found that the ensemble model of AdaBoost, Elastic Net, and SVR exhibited high test performance for predicting the PSI scores of children with ADHD, we visualized the prediction scores and actual PSI values on a graph (Figure 4). The x-axis presents the actual PSI scores, while the y-axis presents the scores predicted using the ensemble model of AdaBoost, Elastic Net, and SVR. The data points on the dashed line (y=x) indicate accurate prediction of the actual PSI values. The regression line, expressed as y=0.58x+43.72, reflects the relationship between the predicted and actual PSI scores.

XSL•FO
RenderX

Figure 4. Comparison of actual processing speed index (PSI) values with those predicted by ensemble application. The dashed line (y=x) represents perfect correspondence of the predicted and actual PSI values, and the red line (y=0.58x+43.72) is the regression line of the relationship between the predicted and actual PSI scores.



PSI scores typically follow a normal distribution with a mean value of 100, and the intersection of the regression line with the y=x line near a PSI score of 100 suggests higher prediction accuracy for PSI values close to the mean. This alignment highlights the model's strong performance in accurately predicting PSI scores within the most densely distributed range. Furthermore, most data points closely follow the y=x line, underscoring the model's reliability and robust predictive capability.

Discussion

Principal Findings

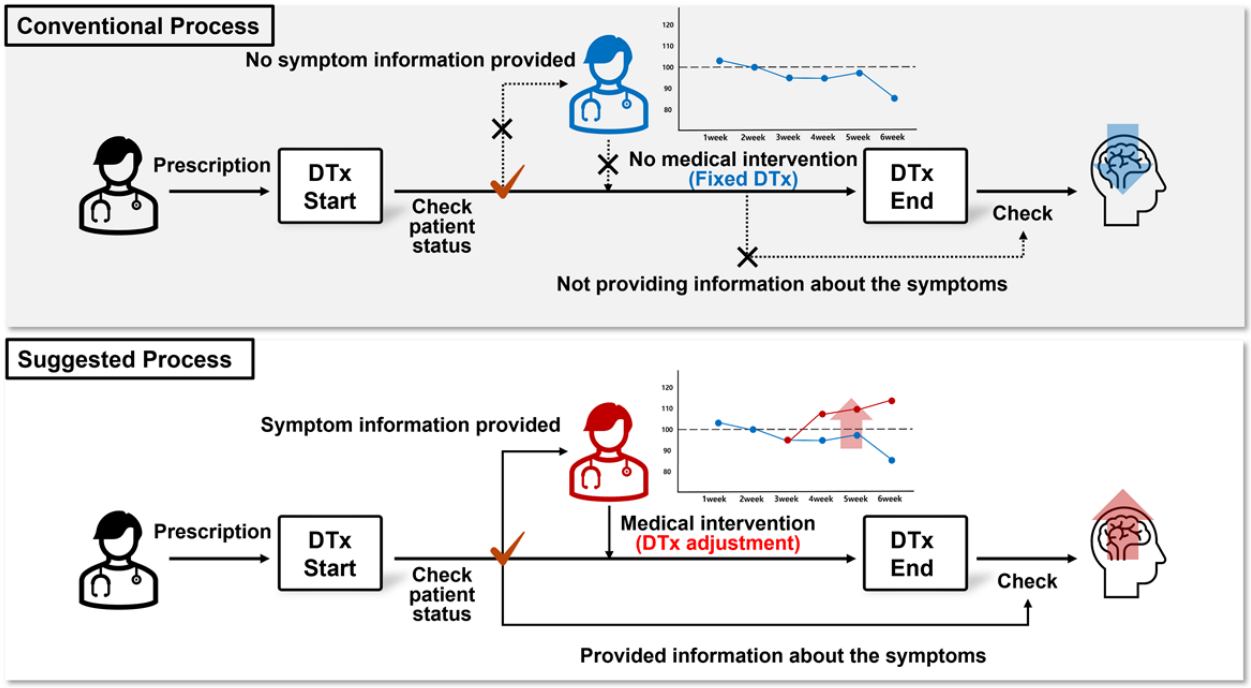
Objective methods for tracking ADHD symptoms remain limited in both clinical and nonclinical settings. Due to the lack of standardized measurement tools, symptom monitoring often relies on observation, especially in home and school environments. Although the results of WISC-V assessments are closely linked to daily functioning and thus can be used to quantify symptoms and inform management strategies, such clinical assessments can only be conducted by qualified professionals and cannot be repeated for approximately 2 years. Because traditional clinical assessment tools can only be readministered after a considerable period following the initial psychological evaluation, it is difficult to continuously and objectively monitor symptom improvement. While serious games can be effective means of monitoring progress between hospital visits, most are primarily designed for training and rely on subjective evaluations or intrinsic performance metrics such as task duration and accuracy rates, which have a limited correlation with actual symptom improvement.

To overcome these limitations, this study aimed to integrate serious game performance data with clinically relevant assessment outcomes for ADHD. Specifically, we developed a machine learning-based algorithm to predict PSI scores from behavioral data collected from 59 children aged 6 - 13 years

XSL•FO
RenderX

diagnosed with ADHD and evaluated the algorithm's performance. We applied the CFS method to extract 6 variables from the serious game performance data, which were input along with gender and age into machine learning models for analysis. Given the relatively small sample size, we applied cross-validation with varying k values (3, 4, 5, and 6) to enhance the robustness of model evaluation. Among these, k=4 provided the most stable performance without evidence of overfitting. Among the individual machine learning models, SVR was the most effective in predicting the PSI scores of children with ADHD. The combination of AdaBoost and Elastic Net model had the highest training performance for predicting PSI scores, while the ensemble of AdaBoost, Elastic Net, and SVR model had the highest test performance. Ultimately, we developed a model capable of predicting PSI scores using game performance data, enabling estimation of the PSI comparable to that obtained from expert-administered WISC-V assessments (Figure 5).

Figure 5. Limitations of traditional digital therapeutics (DTx) monitoring and the need for effective tracking.



How can our PSI prediction model be applied in clinical practice or ADHD interventions? The predicted PSI scores can be used to assess symptom improvement using clinically meaningful metrics without time or location constraints. Thus, the model may be feasibly integrated into ADHD care to support ongoing monitoring of cognitive function by complementing or partially replacing standardized assessments such as the full WISC-V, allowing more frequent and accessible tracking of cognitive changes. For instance, regular evaluation of PS using the model may give clinicians continual insights into the treatment responses of pediatric patients with ADHD, enabling timely and precise treatment adjustments. Prior studies have demonstrated the clinical value of machine learning-based predictive models. For instance, a cardiovascular disease risk prediction model developed using machine learning outperformed traditional models in terms of accuracy [71]. Another risk prediction model for cardiovascular disease also demonstrated improved predictive accuracy, particularly in high-risk subgroups such as individuals with diabetes [72]. Machine learning has been shown to enhance both discrimination and calibration in cardiovascular risk prediction, underscoring its potential to improve personalized care in clinical settings [73]. Collectively, these findings support the applicability of machine learning-based PSI prediction for not only diagnosing but also managing and monitoring ADHD.

Our comparison of individual machine learning models revealed that SVR had the highest accuracy for predicting PSI scores, followed by Elastic Net, while random forest and AdaBoost had higher errors on both the training and test sets. Although SVR achieved the highest accuracy in both the training and test evaluations, the 3 other algorithms also exhibited strengths. It is difficult for a single model to achieve sufficient predictive stability and generalizability, and previous studies have shown that leveraging multiple models can substantially improve predictive performance [74]. For example, when multiple models were combined to analyze digital therapeutic log data, accuracy comparisons yielded determination coefficients (R^2) ranging from 0.82 to 0.86 [74]. These results support the idea that algorithmic diversity is a key factor in the enhanced efficacy of ensemble methods. When we adopted an ensemble approach in this study, we found that the combination of AdaBoost and Elastic Net outperformed SVR, and the best overall performance was achieved by the ensemble integrating AdaBoost, Elastic Net, and SVR, which was superior to any individual model. These improvements suggest that integrating diverse algorithmic techniques—such as adaptive boosting, regularized linear regression, and complementary model tuning—can significantly improve the reliability of PSI score predictions, a critical factor in ADHD symptom tracking.

XSL•FO
RenderX

The regression line of the PSI prediction accuracy of our model intersected y=x at a PSI score of approximately 100, indicating stable predictive performance in the clinically significant range (approximately 85 - 115). This is particularly important for effective monitoring and timely interventions, underscoring the practical applicability of our ensemble model. Similarly, HT Jung et al found that the regression line for MMSE score prediction for MCI crossed y=x near the critical cutoff of 27, indicating high accuracy in distinguishing normal cognition from MCI as well as stability in clinically meaningful ranges [48].

Our findings suggest that the PSI prediction model using serious games can contribute to the quantitative and objective tracking of ADHD symptoms. The ensemble approach effectively compensates for the limitations of individual models, improving predictive stability and accuracy. This method holds promise for clinical use to support continuous patient monitoring and optimization of treatment plans. Furthermore, by integrating this tool into everyday environments such as homes or schools, clinicians may gain valuable insights into symptom tracking and treatment progress beyond those attainable by traditional approaches.

Limitations

Notwithstanding its contributions, this study has several limitations.

First, this study included 64 participants (59 after excluding incomplete data), a relatively small sample size for developing a generalizable PSI prediction model. Riley et al [75] suggest sample sizes ranging from 100 to several hundred participants to ensure adequate generalizability in clinical machine learning. Although cross-validation was applied to minimize bias, the limited sample size may still restrict reproducibility. In future work, we will validate the generalizability of our prediction model by using a more substantial dataset.

Second, this study focused on pediatric participants aged 6 to 13 years with ADHD, all recruited from South Korea. As the sample was limited geographically and demographically, it is possible that testing of our prediction model on datasets collected from participants from other geographical locations, socioeconomic levels, or backgrounds may yield different outcomes. Although standardized diagnostic procedures were implemented to ensure internal validity, future studies incorporating more diverse populations are needed to improve external validity.

Third, this study used serious game performance data as an alternative to traditional assessment tools. However, repeated use of the serious game may introduce practice effects similar to those observed in the WISC-V assessments. To mitigate practice effects, we used randomization techniques to modulate game complexity in each session. The game uses adaptive mechanisms to adjust difficulty based on individual performance, providing a balanced experience. Nevertheless, there is a paucity of research on the long-term use of game-based assessments and their feasibility. It is imperative to assess the extent of the learning effect and recalibrate the model in the future.

Future Work

To further enhance our PSI prediction model and facilitate its integration into clinical practice, future studies will use larger and more diverse datasets encompassing varied geographic and demographic populations. Expanding the dataset will also strengthen external validity and improve applicability across broader clinical contexts. In addition, longitudinal studies assessing the long-term reliability of the PSI prediction model by tracking PSI changes over time could further validate its clinical utility in monitoring ADHD-related cognitive function and informing personalized interventions. Finally, incorporating physiological signals, including multimodal data from electroencephalogram, electrocardiogram, and functional near-infrared spectroscopy, may enhance the robustness of the PSI prediction model and advance its potential for medical application in ADHD monitoring and intervention.

Conclusion

This study introduces a novel approach for predicting K-WISC-V PSI scores using data from serious games designed specifically for children with ADHD. The experimental results showed that among the individual models, SVR had the highest performance, followed by Elastic Net, AdaBoost, and random forest. Furthermore, the ensemble method combining AdaBoost and Elastic Net provided stable and reliable predictive accuracy, highlighting the potential of model integration to enhance overall performance. Our findings suggest that this approach can complement existing subjective methods of treatment monitoring to facilitate continuous symptom tracking and longitudinal monitoring and contribute to the optimization of treatment planning. Integrating PSI prediction with serious game content may enhance patient accessibility and provide a quantitative and reliable tool for clinicians to objectively track and assess a patient's condition in their everyday environment. Future research should aim to validate these findings using larger datasets including patients of diverse ages and ADHD symptom severities. Incorporating data from participants with varied demographic and clinical characteristics will enhance the model's generalizability and applicability in real-world clinical environments.

Acknowledgments

JHS and HSH contributed equally as co-corresponding authors on this work, and the latter can be reached at sookhoe72@kbri.re.kr.

This work was supported by the KBRI basic research program through KBRI funded by the Ministry of Science, ICT & Future Planning (grant 25-BR-02- - 04 and 25-BR-05- - 01, 25-BR-06- - 01, H.S.H.) and a grant of the Korea Dementia Research Project through the Korea Dementia Research Center (KDRC) funded by the Ministry of Health & Welfare and Ministry of Science and ICT, Republic of Korea (grant RS-2024- - 00343370, H.S.H.). This work was also supported by the Korea Ministry

XSL•FO
RenderX

of Science and ICT's Special Account for Regional Balanced Development for Commercialization supervised by the NIPA (National IT Industry Promotion Agency) to support digital medical devices for artificial intelligence–based Neurodevelopmentalneurodevelopmental disorders (H0301-24-1001, H.S.H.).

Data Availability

The datasets generated and analyzed in this study are not publicly available as they are restricted to the scope originally approved by the institutional review board (IRB). In addition, even for deidentified datasets with all personal information removed, additional IRB review and approval would be required to assess the feasibility of data sharing; therefore, the data cannot be shared at this stage.

Authors' Contributions

JSK: conceptualization, formal analysis, investigation, visualization, writing (original draft, review and editing); YJJ: formal analysis, writing (original draft, review and editing); SJK: formal analysis, visualization, writing (original draft); SJJ: formal analysis, investigation; JYP: formal analysis, investigation; HSH: supervision, writing (review and editing); JHS: conceptualization, methodology, supervision, visualization, writing (review and editing)

Conflicts of Interest

None declared.

References

1. Mahone EM, Denckla MB. Attention-deficit/hyperactivity disorder: a historical neuropsychological perspective. J Int Neuropsychol Soc 2017 Oct;23(9-10):916-929. [doi: 10.1017/S1355617717000807] [Medline: 29198277]
2. A 14-month randomized clinical trial of treatment strategies for attention-deficit/hyperactivity disorder. The MTA Cooperative Group. Multimodal Treatment Study of Children with ADHD. Arch Gen Psychiatry 1999 Dec;56(12):1073-1086. [doi: 10.1001/archpsyc.56.12.1073] [Medline: 10591283]
3. Felt BT, Biermann B, Christner JG, Kochhar P, Harrison RV. Diagnosis and management of ADHD in children. Am Fam Physician 2014 Oct 1;90(7):456-464. [Medline: 25369623]
4. Chang TM, Wu RC, Yang RC, et al. Objective diagnosis of ADHD through movement analysis by using a smart chair with piezoelectric material. Pediatr Neonatol 2023 Jan;64(1):46-52. [doi: 10.1016/j.pedneo.2022.06.007] [Medline: 36089537]
5. Hall CL, Bellato A, Kirk JD, Hollis C. The clinical utility of QbTest in supporting the assessment and monitoring of attention-deficit/hyperactivity disorder (ADHD): what do paediatricians need to know? Paediatr Child Health (Oxford) 2023 Sep;33(9):259-264. [doi: 10.1016/j.paed.2023.06.006]
6. Cook NE, Braaten EB, Surman CBH. Clinical and functional correlates of processing speed in pediatric attention-deficit/hyperactivity disorder: a systematic review and meta-analysis. Child Neuropsychol 2018 Jul;24(5):598-616. [doi: 10.1080/09297049.2017.1307952] [Medline: 28345402]
7. Rommelse N, Luman M, Kievit R. Slow processing speed: a cross-disorder phenomenon with significant clinical value, and in need of further methodological scrutiny. Eur Child Adolesc Psychiatry 2020 Oct;29(10):1325-1327. [doi: 10.1007/s00787-020-01639-9] [Medline: 32915272]
8. Shanahan MA, Pennington BF, Yerys BE, et al. Processing speed deficits in attention deficit/hyperactivity disorder and reading disability. J Abnorm Child Psychol 2006 Oct;34(5):585-602. [doi: 10.1007/s10802-006-9037-8] [Medline: 16850284]
9. Anker E, Ogrim G, Heir T. Verbal working memory and processing speed: correlations with the severity of attention deficit and emotional dysregulation in adult ADHD. J Neuropsychol 2022 Mar;16(1):211-235. [doi: 10.1111/jnp.12260] [Medline: 34218514]
10. Kaufman AS. Intelligent Testing with the WISC-III: John Wiley & Sons; 1994.
11. Calhoun SL, Mayes SD. Processing speed in children with clinical disorders. Psychol Schs 2005 Apr;42(4):333-343. [doi: 10.1002/pits.20067]
12. Dale BA, Finch WH, Shellabarger KAR, Davis A. Wechsler Intelligence Scale for Children, Fifth Edition profiles of children with autism spectrum disorder using a classification and regression trees analysis. J Psychoeduc Assess 2021 Oct;39(7):783-799. [doi: 10.1177/07342829211025924]
13. McGill RJ, Ward TJ, Canivez GL. Use of translated and adapted versions of the WISC-V: Caveat emptor. Sch Psychol Int 2020 Jun;41(3):276-294. [doi: 10.1177/0143034320903790]
14. Wilson CJ, Bowden SC, Byrne LK, Vannier LC, Hernandez A, Weiss LG. Cross-national generalizability of WISC-V and CHC broad ability constructs across France, Spain, and the US. J Intell 2023 Aug 7;11(8):159. [doi: 10.3390/jintelligence11080159] [Medline: 37623542]
15. Hagemeister C. How useful is the power law of practice for recognizing practice in concentration tests? Eur J Psychol Assess 2007 Jan;23(3):157-165. [doi: 10.1027/1015-5759.23.3.157]

XSL•FO
RenderX

16. Lievens F, Reeve CL, Heggestad ED. An examination of psychometric bias due to retesting on cognitive ability tests in selection settings. J Appl Psychol 2007 Nov;92(6):1672-1682. [doi: 10.1037/0021-9010.92.6.1672] [Medline: 18020804]
17. Roediger HL III, Butler AC. The critical role of retrieval practice in long-term retention. Trends Cogn Sci (Regul Ed) 2011 Jan;15(1):20-27. [doi: 10.1016/j.tics.2010.09.003]
18. Canivez GL, Watkins MW. Long-term stability of the Wechsler Intelligence Scale for Children--Third Edition. Psychol Assess 1998;10(3):285-291. [doi: 10.1037//1040-3590.10.3.285]
19. Ryan JJ, Glass LA, Bartels JM. Stability of the WISC-IV in a sample of elementary and middle school children. Appl Neuropsychol 2010 Jan;17(1):68-72. [doi: 10.1080/09084280903297933] [Medline: 20146124]
20. Watkins MW, Canivez GL, Dombrowski SC, et al. Long-term stability of Wechsler Intelligence Scale for Children-fifth edition scores in a clinical sample. Appl Neuropsychol Child 2022;11(3):422-428. [doi: 10.1080/21622965.2021.1875827] [Medline: 33556254]
21. Ahuja J, Jaka S, Saeed S, Gunturu S. Download to heal: navigating the pixelated path of digital therapeutics in psychiatric care. Braz J Psychiatry 2024 Sep 27. [doi: 10.47626/1516-4446-2024-3802] [Medline: 39329484]
22. Suravarapu ST, Yetukuri K. Software programmes employed as medical devices. World J Curr Med Pharm Res 2023:67-74. [doi: 10.37022/wjcmpr.v5i3.268]
23. Craven MP, Groom MJ. Computer games for user engagement in attention deficit hyperactivity disorder (ADHD) monitoring and therapy. Presented at: 2015 International Conference on Interactive Technologies and Games (iTAG); Oct 22-23, 2015. [doi: 10.1109/iTAG.2015.9]
24. Shah A, Kraemer KR, Won CR, Black S, Hasenbein W. Developing digital intervention games for mental disorders: a review. Games Health J 2018 Aug;7(4):213-224. [doi: 10.1089/g4h.2017.0150] [Medline: 30106642]
25. Kollins SH, Childress A, Heusser AC, Lutz J. Effectiveness of a digital therapeutic as adjunct to treatment with medication in pediatric ADHD. NPJ Digit Med 2021 Mar 26;4(1):58. [doi: 10.1038/s41746-021-00429-0] [Medline: 33772095]
26. Sverdlov O, van Dam J, Hannesdottir K, Thornton-Wells T. Digital therapeutics: an integral component of digital innovation in drug development. Clin Pharmacol Ther 2018 Jul;104(1):72-80. [doi: 10.1002/cpt.1036] [Medline: 29377057]
27. Daniëls NEM, Bartels SL, Verhagen SJW, Van Knippenberg RJM, De Vugt ME, Delespaul P. Digital assessment of working memory and processing speed in everyday life: feasibility, validation, and lessons-learned. Internet Interv 2020 Mar;19(100300):100300. [doi: 10.1016/j.invent.2019.100300] [Medline: 31970080]
28. Bonnechère B. Evaluation of processing speed of different cognitive functions across the life span using cognitive mobile games. Games Health J 2022 Apr;11(2):132-140. [doi: 10.1089/g4h.2021.0144] [Medline: 35180366]
29. Zhu F, Panwar B, Dodge HH, et al. COMPASS: A computational model to predict changes in MMSE scores 24-months after initial assessment of Alzheimer's disease. Sci Rep 2016 Oct 5;6(1). [doi: 10.1038/srep34567]
30. Farzana S, Parde N. Exploring MMSE score prediction using verbal and non-verbal cues. Presented at: Interspeech 2020; Oct 25-29, 2020. [doi: 10.21437/Interspeech.2020-3085]
31. Twomey C, O'Connell H, Lillis M, Tarpey SL, O'Reilly G. Utility of an abbreviated version of the stanford-binet intelligence scales (5th ed.) in estimating "full scale" IQ for young children with autism spectrum disorder. Autism Res 2018 Mar;11(3):503-508. [doi: 10.1002/aur.1911] [Medline: 29282895]
32. Kadiyala PK. Mnemonics for diagnostic criteria of DSM V mental disorders: a scoping review. Gen Psychiatr 2020;33(3):e100109. [doi: 10.1136/gpsych-2019-100109] [Medline: 32478284]
33. Audras-Torrent L, Miniarikova E, Couty F, et al. WISC-V profiles and their correlates in children with autism spectrum disorder without intellectual developmental disorder: report from the ELENA cohort. Autism Res 2021 May;14(5):997-1006. [doi: 10.1002/aur.2444] [Medline: 33247546]
34. Salonen J, Slama S, Haavisto A, Rosenqvist J. Comparison of WPPSI-IV and WISC-V cognitive profiles in 6-7-year-old Finland-Swedish children - findings from the FinSwed study. Child Neuropsychol 2023 Jul;29(5):687-709. [doi: 10.1080/09297049.2022.2112163] [Medline: 35996965]
35. Sattler JM, Dumont R. Assessment of Children: WISC-IV and WPPSI-III Supplement: Jerome M. Sattler; 2004.
36. Salthouse TA. The processing-speed theory of adult age differences in cognition. Psychol Rev 1996 Jul;103(3):403-428. [doi: 10.1037/0033-295x.103.3.403] [Medline: 8759042]
37. Tucha L, Fuermaier ABM, Koerts J, et al. Sustained attention in adult ADHD: time-on-task effects of various measures of attention. J Neural Transm (Vienna) 2017 Feb;124(Suppl 1):39-53. [doi: 10.1007/s00702-015-1426-0] [Medline: 26206605]
38. Kim SC. K-WISC-V processing speed index analysis to verify the effectiveness of ADHD symptom improvement using pediatric digital content. Appl Sci (Basel) 2024 May;14(9):3792. [doi: 10.3390/app14093792]
39. Kim SC, Lee H, Lee HS, Kim G, Song JH. Adjuvant therapy for attention in children with ADHD using game-type digital therapy. Int J Environ Res Public Health 2022 Nov 14;19(22):14982. [doi: 10.3390/ijerph192214982] [Medline: 36429699]
40. Kim SC, Lee HS. Effect of game-based cognitive training programs on cognitive learning of children with intellectual disabilities. Appl Sci (Basel) 2021 Sep;11(18):8582. [doi: 10.3390/app11188582] [Medline: 36003951]
41. Adams ZW, Roberts WM, Milich R, Fillmore MT. Does response variability predict distractibility among adults with attention-deficit/hyperactivity disorder? Psychol Assess 2011 Jun;23(2):427-436. [doi: 10.1037/a0022112] [Medline: 21443365]

XSL•FO
RenderX

42. Cross-Villasana F, Finke K, Hennig-Fast K, et al. The speed of visual attention and motor-response decisions in adult attention-deficit/hyperactivity disorder. Biol Psychiatry 2015 Jul 15;78(2):107-115. [doi: 10.1016/j.biopsych.2015.01.016] [Medline: 25773661]
43. Kofler MJ, Rapport MD, Sarver DE, et al. Reaction time variability in ADHD: a meta-analytic review of 319 studies. Clin Psychol Rev 2013 Aug;33(6):795-811. [doi: 10.1016/j.cpr.2013.06.001] [Medline: 23872284]
44. Epstein JN, Conners CK, Sitarenios G, Erhardt D. Continuous performance test results of adults with attention deficit hyperactivity disorder. Clin Neuropsychol 1998 May;12(2):155-168. [doi: 10.1076/clin.12.2.155.2000]
45. Epstein JN, Erkanli A, Conners CK, Klaric J, Costello JE, Angold A. Relations between continuous performance test performance measures and ADHD behaviors. J Abnorm Child Psychol 2003 Oct;31(5):543-554. [doi: 10.1023/a:1025405216339] [Medline: 14561061]
46. López J, López V, Rojas D, et al. Effect of psychostimulants on distinct attentional parameters in attentional deficit/hyperactivity disorder. Biol Res 2004;37(3):461-468. [doi: 10.4067/s0716-97602004000300010] [Medline: 15515970]
47. Stins JF, Tollenaar MS, Slaats-Willemse DIE, et al. Sustained attention and executive functioning performance in attention-deficit/hyperactivity disorder. Child Neuropsychol 2005 Jun;11(3):285-294. [doi: 10.1080/09297040490916938] [Medline: 16036452]
48. Jung HT, Lee H, Kim K, et al. Predicting cognitive impairment level after a serious game-based therapy in chronic stroke survivors. In: Jung HT, Lee H, Kim K, editors. Presented at: 2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI); May 19-22, 2019. [doi: 10.1109/BHI.2019.8834484]
49. Rodriguez-Galiano V, Sanchez-Castillo M, Chica-Olmo M, Chica-Rivas M. Machine learning predictive models for mineral prospectivity: an evaluation of neural networks, random forest, regression trees and support vector machines. Ore Geol Rev 2015 Dec;71:804-818. [doi: 10.1016/j.oregeorev.2015.01.001]
50. Hall MA. Correlation-Based Feature Selection for Machine Learning: The University of Waikato; 1999.
51. Ghiselli EE. Theory of Psychological Measurement: McGraw-Hill; 1964.
52. Hogarth RM, Jungermann H, editors. Methods for aggregating opinions. In: Decision Making and Change in Human Affairs Theory and Decision Library: Springer, Dordrecht, Vol. 16. [doi: 10.1007/978-94-010-1276-8_16]
53. Zajonc RB. A note on group judgements and group size. Human Relations 1962 May;15(2):177-180. [doi: 10.1177/001872676201500206]
54. Breiman L. Random forests. Mach Learn 2001 Oct;45(1):5-32. [doi: 10.1023/A:1010933404324]
55. Hastie T. The Elements of Statistical Learning: Data Mining, Inference, and Prediction: Springer; 2009. [doi: 10.1007/978-0-387-84858-7]
56. Zeini HA, Al-Jeznawi D, Imran H, Bernardo LFA, Al-Khafaji Z, Ostrowski KA. Random forest algorithm for the strength prediction of geopolymer stabilized clayey soil. Sustainability 2023 Jan;15(2):1408. [doi: 10.3390/su15021408]
57. Freund Y, Schapire RE, editors. Experiments with a new boosting algorithm. In: ICML'96: Proceedings of the Thirteenth International Conference on International Conference on Machine Learning: Morgan Kaufmann Publishers Inc; 1996:148-156.
58. Solomatine DP. AdaBoost.RT: a boosting algorithm for regression problems. Presented at: 2004 IEEE International Joint Conference on Neural Networks (IEEE Cat No 04CH37541); Jul 25-29, 2004. [doi: 10.1109/IJCNN.2004.1380102]
59. Vapnik V. Statistical Learning Theory: John Wiley Sons; 1998, Vol. 2:831-842.
60. Awad M, Khanna R. Support vector regression. In: Efficient Learning Machines: Theories, Concepts, Applications for Engineers System Designers: Springer; 2015:67-80. [doi: 10.1007/978-1-4302-5990-9_4]
61. Poirier C, Lavenu A, Bertaud V, et al. Real time influenza monitoring using hospital big data in combination with machine learning methods: comparison study. JMIR Public Health Surveill 2018 Dec 21;4(4):e11361. [doi: 10.2196/11361] [Medline: 30578212]
62. Zou H, Hastie T. Regularization and variable selection via the Elastic Net. J R Stat Soc Series B Stat Methodol 2005 Apr 1;67(2):301-320. [doi: 10.1111/j.1467-9868.2005.00503.x]
63. Gana R. Ridge regression and the Elastic Net: how do they do as finders of true regressors and their coefficients? Mathematics 2022;10(17):3057. [doi: 10.3390/math10173057]
64. Tibshirani R. Regression shrinkage and selection via the lasso. J R Stat Soc Series B Stat Methodol 1996 Jan 1;58(1):267-288. [doi: 10.1111/j.2517-6161.1996.tb02080.x]
65. Zhang Z, Lai Z, Xu Y, Shao L, Wu J, Xie GS. discriminative elastic-net regularized linear regression. IEEE Trans Image Process 2017 Mar;26(3):1466-1481. [doi: 10.1109/TIP.2017.2651396] [Medline: 28092552]
66. Gopukumar D, Ghoshal A, Zhao H. Predicting readmission charges billed by hospitals: machine learning approach. JMIR Med Inform 2022 Aug 30;10(8):e37578. [doi: 10.2196/37578] [Medline: 35896038]
67. Mladenova T, Valova I. Comparative analysis between the traditional k-nearest neighbor and modifications with weight-calculation. Presented at: 2022 International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT); Oct 20-22, 2022. [doi: 10.1109/ISMSIT56059.2022.9932693]
68. Li JS. Asymptotics of k-fold cross validation. J Artif Intell Res 2023;78:491-526. [doi: 10.1613/jair.1.13974]
69. Moon J, Ke FF, Sokolikj Z. Automatic assessment of cognitive and emotional states in virtual reality‐based flexibility training for four adolescents with autism. Brit J Educational Tech 2020 Sep;51(5):1766-1784. [doi: 10.1111/bjet.13005]
70. Rokach L. Ensemble-based classifiers. Artif Intell Rev 2010 Feb;33(1-2):1-39. [doi: 10.1007/s10462-009-9124-7]

XSL•FO
RenderX

71. Weng SF, Reps J, Kai J, Garibaldi JM, Qureshi N. Can machine-learning improve cardiovascular risk prediction using routine clinical data? PLoS ONE 2017;12(4):e0174944. [doi: 10.1371/journal.pone.0174944] [Medline: 28376093]
72. Alaa AM, Bolton T, Di Angelantonio E, Rudd JHF, van der Schaar M. Cardiovascular disease risk prediction using automated machine learning: a prospective study of 423,604 UK Biobank participants. PLoS One 2019;14(5):e0213653. [doi: 10.1371/journal.pone.0213653] [Medline: 31091238]
73. Cho SY, Kim SH, Kang SH, et al. Pre-existing and machine learning-based models for cardiovascular risk prediction. Sci Rep 2021 Apr 26;11(1):8886. [doi: 10.1038/s41598-021-88257-w] [Medline: 33903629]
74. Kim Y, Yang J, Rim D, Oh U. Prediction of cognitive ability utilizing a machine learning approach based on digital therapeutics log data. Int J Adv Smart Converg 2023;12(2):17-24. [doi: 10.7236/IJASC.2023.12.2.17]
75. Riley RD, Debray TPA, Collins GS, et al. Minimum sample size for external validation of a clinical prediction model with a binary outcome. Stat Med 2021 Aug 30;40(19):4230-4251. [doi: 10.1002/sim.9025] [Medline: 34031906]

Abbreviations

ADHD: attention-deficit/hyperactivity disorder
CFS: correlation feature selection
IRB: institutional review board
K-WISC-V: Korean Wechsler Intelligence Scale for Children-Fifth Edition
MAE: mean absolute error
MAPE: mean absolute percentage error
PSI: processing speed index
RMSE: root mean squared error
SVR: support vector regression
WISC: Wechsler Intelligence Scale for Children

Edited by A Coristine; submitted 04.03.25; peer-reviewed by I Navarro-Soria, R Singh, Z Yu; revised version received 29.08.25; accepted 30.08.25; published 14.10.25.

Please cite as:
Kim JS, Jeong YJ, Kim SJ, Jun SJ, Park JY, Hoe HS, Song JH
A Novel Approach Using Serious Game Data to Predict the WISC-V Processing Speed Index in Children With Attention-Deficit/Hyperactivity Disorder: Machine Learning Study
JMIR Serious Games 2025;13:e73408
URL: https://games.jmir.org/2025/1/e73408
doi:10.2196/73408

© Jun-Su Kim, Yoo Joo Jeong, Seung-Jae Kim, Su Jin Jun, Jin-Yeop Park, Hyang-Sook Hoe, Jeong-Heon Song. Originally published in JMIR Serious Games (https://games.jmir.org), 14.10.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Discriminative Power of the Serious Game Attention Slackline in Children and Adolescents With and Without Attention-Deficit/Hyperactivity Disorder: Validation Study

Nicolás Ruiz-Robledillo[1], PhD; Ignacio Lucas[1], PhD; Rosario Ferrer-Cascales[1], PhD; Natalia Albaladejo-Blázquez[1], PhD; Javier Sanchis[2], PhD; Juan Trujillo[2], PhD

[1]Department of Health Psychology, Faculty of Health Science, University of Alicante, Alicante Institute for Health and Biomedical Research (ISABIAL), Carretera San Vicente Del Raspeig S/N, San Vicente Del Raspeig, Alicante, Spain

[2]Lucentia Research Group, University of Alicante, Alicante Institute for Health and Biomedical Research (ISABIAL), San Vicente Del Raspeig, Alicante, Spain

Corresponding Author:
Ignacio Lucas, PhD
Department of Health Psychology, Faculty of Health Science, University of Alicante, Alicante Institute for Health and Biomedical Research (ISABIAL), Carretera San Vicente Del Raspeig S/N, San Vicente Del Raspeig, Alicante, Spain

Abstract

Background: Attention-deficit/hyperactivity disorder (ADHD) is a prevalent neurodevelopmental condition characterized by inattention, hyperactivity, and impulsivity, significantly impacting the psychological, social, and academic well-being of affected children and adolescents. Traditional ADHD diagnostic methods often rely on subjective reports, which can be biased. Recent advancements in serious games offer the potential for objective assessment tools.

Objective: This study aimed to evaluate the discriminative power and concurrent validity of the serious game Attention Slackline in distinguishing children and adolescents with ADHD from those without the condition and in correlating game performance with standardized ADHD assessment scales.

Methods: A sample of 32 children and adolescents diagnosed with ADHD and 39 healthy controls participated in the study. Participants were divided into 2 age groups: children (aged 6 - 11 years) and adolescents (aged 12 - 17 years). The serious game Attention Slackline was administered alongside established ADHD assessment scales, including the Child and Adolescent Assessment System and the ADHD Rating Scale IV. Group differences were analyzed using multivariate analysis of covariance, and effect sizes were reported using Cohen *d*. Correlations between game performance and ADHD symptoms were calculated using Pearson *r*.

Results: Children with ADHD demonstrated significantly worse performance in Attention Slackline than the controls (t_{65}=−2.26; *P*=.03; |*d*|=0.901), whereas no significant differences were observed in adolescents (t_{65}=0.75; *P*=.73; |*d*|=0.191). Task performance was negatively correlated with family-reported hyperactivity/impulsivity symptoms in children across both tests (*r*=−0.43 and *r*=−0.51), but no significant correlations were observed in adolescents.

Conclusions: The findings support the validity of Attention Slackline for assessing hyperactivity/impulsivity symptoms in children with ADHD. However, its efficacy decreases in adolescents, potentially due to developmental factors, such as compensatory strategies and ceiling effects in task performance. The gamified nature of the tool enhances engagement, which is crucial for young populations, while maintaining its diagnostic utility in measuring impulsivity. The age-dependent validity aligns with previous research indicating that continuous performance test paradigms are less effective in older populations due to developmental maturation. Attention Slackline shows potential as a complementary tool for ADHD diagnosis in children, offering an engaging and objective assessment of hyperactivity/impulsivity. Future research should aim to establish clinical cutoff points and refine the task's complexity to align with individual characteristics.

(JMIR Serious Games 2025;13:e65170) doi:10.2196/65170

KEYWORDS

attention-deficit/hyperactivity disorder; ADHD; impulsivity; children; adolescents; neuropsychology; assessment; Attention Slackline; serious game; child; young adults; neurodevelopmental

XSL•FO
RenderX

Introduction

Attention-deficit/hyperactivity disorder (ADHD) is defined as a neurodevelopmental disorder that must be present before the age of 12 years. It is characterized by a persistent pattern of inattention and hyperactivity, present individually or in combination, that may be maladaptive and interfere with the patient's daily functioning [1,2]. According to data reported by the American Psychiatric Association, ADHD may be present in between 4% and 13.3% of children and adolescents worldwide, depending on the assessment protocol and the number of informants involved in the diagnostic process [2-4]. Globally, the estimated prevalence of ADHD in children is 7.6% [5]. Specifically, in Spain, a meta-analysis estimates its prevalence to be 6.8% [6]. ADHD is more frequently diagnosed in boys than in girls, reflecting its higher prevalence rate among males [7]. However, the impacts of ADHD in several dimensions regarding daily living, such as social functioning or time perception, are higher in females than in males [8].

ADHD is a prevalent disorder in the pediatric and adolescent population, characterized by significant difficulties associated with inattention, hyperactivity and impulsivity, emotional problems, and cognitive deficits [2,9,10]. This disorder represents one of the most common causes of conduct problems among children [9,11]. In this sense, children with a diagnosis of ADHD may experience a significant deterioration in psychological, social, and academic well-being [12,13]. Given its high prevalence, associated comorbidities, and chronic nature, ADHD represents a major public health concern. [14,15].

In relation to the diagnosis of ADHD, it is often necessary to rely on clinical observations, subjective reports, and objective measures of symptomatology. Clinical observations by specialists are useful but limited to the clinical environment and, thus, may not reflect the whole picture of symptoms. Assessment of ADHD using self-report ratings reveals age-related differences in perception accuracy. Children with ADHD often exhibit a positive illusory bias, overestimating their competence across various domains [16]. Conversely, adolescents with ADHD demonstrate greater awareness of their limitations, reflecting a developmental shift in self-perception [17]. Due to the difficulty of using self-reports to diagnose children with ADHD, assessment tools also consider the family or teacher's perception of symptoms, particularly at younger ages. Therefore, it is essential to have assessment tools that allow an adequate differential diagnosis due to the subjectivity of the symptoms as reported by families and teachers, in addition to the possible overlap of diagnostic criteria of other attentional and behavioral disorders [1,14,18,19]. In this sense, neuropsychological assessment tests represent an advancement as they are specifically designed to report valid and reliable data that allow an appropriate diagnosis to be made for each patient [20,21]. Numerous studies have reported the development of standardized neuropsychological batteries that have been used in daily clinical practice in recent years and have been a turning point for the assessment and clinical diagnosis of ADHD. These neuropsychological batteries also allow for greater precision in the description of the cognitive deficits that can appear in children with ADHD and thus serve as a support for decision-making in clinical treatment [21,22]. One of the most frequently used objective paradigms for the assessment of ADHD is the continuous performance test (CPT). The CPT is a paradigm where the participant is required to respond to certain stimuli and inhibit responses to other specific stimuli. Performance in CPT has been shown to be worse in children with ADHD than their nonclinical counterparts [23]. However, this difference between them is reduced with maturation [24]. In fact, CPT has shown better discriminative power for children [25] than for adults [26].

Likewise, new technologies have been used to identify the neuropsychological profile of children with ADHD, as well as for their treatment and rehabilitation, which has shown great advantages over more traditional methodologies, reducing costs and time for diagnosis [27]. These types of tools also allow recreating different stimuli from the real world based on an ecological approach, in addition to flexibly customizing different visual and auditory stimuli in the virtual environment, such as controlling the time, order, and type of stimulus [28]. In this sense, it has been demonstrated how the use of technological measures in neuropsychological studies is valid and reliable when detecting problems in cognitive functioning [29]. A specific example of these technologies is serious games, which are defined as a type of game designed and developed with a primary purpose that goes beyond mere entertainment [30,31]. In the field of health, serious games are designed to achieve multiple objectives, including evaluation [32], diagnosis [33], and treatment of numerous pathologies [34-36]. In the context of diagnosis and evaluation, the majority of analyzed serious games have exhibited high levels of accuracy, comparable to the established classical methods [32,37,38], with even higher accuracy in some cases [33]. Specifically, in the case of cognitive assessment, serious games are a promising tool, especially for the assessment and diagnosis of neurodevelopmental disorders [39]. In the particular case of ADHD, different systematic reviews have shown that serious games are effective tools for the valid and reliable evaluation of ADHD core symptoms and associated neuropsychological dysfunctions [40,41]. Furthermore, in June 2020, the US Food and Drug Administration approved the first serious game designed for the treatment of children with ADHD [42], as part of the therapeutic process. Furthermore, more recently, other serious games have been developed as auxiliary tools for the assessment and treatment of ADHD [43-45]. However, to our knowledge, there is a gap in previous research, as the effect of relevant sociodemographic variables implicated in the diagnosis of ADHD, such as age, on serious game performance has been studied very little. Furthermore, as indicated above, analysis of the correlation between children's and their families' own assessment of ADHD and the serious game performance is essential. Therefore, as indicated in a recent review, research in this area is at an early stage and further studies are needed [39].

In this regard, Attention Slackline is a recently developed serious game designed to assess several ADHD symptoms. It has shown promising results in the evaluation of attentional processes in children and adolescents [46]. Although the previous pilot testing study regarding the validity of Attention Slackline

XSL•FO
RenderX

reported promising results, further research is needed to validate this serious game considering wider samples, other criterion variables, and different evaluation methodologies, including self- and informant-reported tests. In this respect, one of the main issues to be considered is criterion validity.

This type of validity is commonly subdivided into 2 distinct categories: concurrent and predictive validity [47]. Concurrent validity is based on the correlation between the test to be validated and another criterion when both are measured simultaneously. Conversely, predictive validity appraises the capacity of a test to predict performance, behavior, or outcomes pertinent to the construct under assessment. In this regard, predictive validity can be regarded as an indicator of the discriminative power of a tool, allowing us to ascertain the tool's capacity to discriminate between groups based on test results, for instance, between individuals with and without a diagnosis of ADHD. In this regard, few studies have tested the use of serious games in the evaluation of ADHD symptomatology, both in clinical and nonclinical populations, analyzing the concurrent validity of the data derived from the serious game with other subjective measures. Furthermore, little is known regarding the discriminative power of serious games in the identification of individuals with and without ADHD, especially considering the influence of significant sociodemographic factors, such as the age of the participants [48].

Taking all of this into account, the main aim of this study was to assess the discriminative power of the serious game Attention Slackline (developed by our research group) in samples of children and adolescents with and without ADHD. Furthermore, this study aimed to assess the concurrent validity of the task with self-reports and familial reports of attention and impulsivity or hyperactivity symptoms. Based on the review evidence on ADHD and CPT paradigms, we hypothesized that the task performance in the serious game Attention Slackline would be worse in those individuals with an ADHD diagnosis [40,41], especially in the younger group [16,17]. Moreover, we hypothesized that there would be a correlation between reported ADHD symptomatology and task performance [40,41].

Methods

Participants

The study included 32 patients diagnosed with ADHD and 39 healthy controls, all aged between 6 and 17 years. Participants younger than 12 years were classified into the children group (mean age 8.97, SD 1.61 years), whereas those aged 12 years or older were classified into the adolescent group (mean age 14.05, SD 1.66 years). Consequently, the children group comprised 10 patients and 20 controls, while the adolescents group consisted of 22 patients and 19 controls. The patients with ADHD were referred from a unit specialized in neurodevelopmental disorders from the Child and Adolescent Mental Health Service of the General University Hospital Dr. Balmis of Alicante (Spain). All patients had been diagnosed with ADHD by experienced psychologists in accordance with the criteria set forth in the *DSM-5* (*Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition*). This diagnosis was reached through clinical interviews with the patients, their families, and teachers, as well as through the administration of structured tests. Healthy controls were recruited via personal referrals and recommendations from participants and study team members. Individuals with any other mental condition that can impair cognitive performance, such as intellectual disability or autism spectrum disorder, were excluded from the study.

Instruments

Attention Slackline

Attention Slackline is based on a CPT paradigm [46], commonly used for assessing ADHD due to its ability to measure core symptoms such as impulsivity and inattention. The game mechanics were specifically designed to target these symptoms by requiring sustained attention and controlled responses, aligning with established neuropsychological assessment methods to ensure validity. Set in a snowy valley, the game's narrative involves rescuing a friend isolated on a distant mountain. The player's character, whose gender is not specified, must walk across a slackline to the opposite mountain as quickly as possible. The game is crafted to be child-friendly, with no references to hazardous situations, ensuring the character never falls from the slackline. In the Attention Slackline game, the player sees a target flag with a specific pattern in the top-left corner of the screen. Multiple flags appear on the distant mountain where the friend is located. The player's task is to press a button when the pattern of a flag on the mountain matches the target flag. If the button is pressed correctly, the character progresses along the slackline; if not, the character slows down (Figure 1). Each game level lasts for 5 minutes and concludes with a motivational message, which is always positive regardless of the game's outcome.

XSL•FO
RenderX

Figure 1. Screenshot of the Attention Slackline task.



The version of the game used in this study comprised 3 blocks of increasing complexity, as it can be observed in Figure 2. First, Block 1 introduces the basic task of pattern recognition. A target flag with a unique pattern is displayed in the top-left corner, as previously explained. Second, Block 2 increases complexity by requiring the player to track 2 target flags simultaneously. The player must respond whenever either flag appears on the mountain, testing their ability to divide attention and respond to multiple stimuli. Finally, Block 3, the final block, involves a sequential pattern recognition task, demanding higher cognitive effort. The player must identify a sequence of 2 target flags in the correct order among those on the mountain. Correct responses are registered when the player presses the button after the second flag in the sequence. Omissions are considered when the participant fails to respond to a stimulus that should have prompted a response. Commissions, on the other hand, are considered when a participant responds to an incorrect stimulus.

Figure 2. The 3 blocks of Attention Slackline used in this study.



XSL•FO
RenderX

Child and Adolescent Assessment System

The attention problems and hyperactivity/impulsivity subscales of the Child and Adolescent Assessment System (SENA) [49] were applied in this study. The SENA is a well-validated instrument in Spanish adjusted by age for the assessment of emotional and behavioral problems in children and adolescents. Both the self-assessed version and the parent rating version were used.

ADHD Rating Scale IV

The parent's report of the Spanish version of the ADHD Rating Scale IV (ADHD-RS) [50,51] was used for the evaluation of ADHD symptomatology. This is a well-validated instrument for the assessment of ADHD that includes inattention and hyperactivity/impulsivity subscales, as well as a combined score summing both.

Procedure

First, the sample was selected following the inclusion and exclusion criteria described above. Professionals with expertise in ADHD from a local hospital, together with the project research team, selected the sample according to these criteria. Second, the families were contacted to explain the objectives of the project and what their participation in it would consist of. Once the families decided to participate, the informed consent form was signed by the father, mother, or the guardian of the minor. Once the informed consent was signed, participants were scheduled for 2 neuropsychological assessment sessions lasting approximately 1.5 hours. These assessment sessions took place in the University of Alicante laboratory responsible for developing this project. This laboratory consists of 4 evaluation rooms with the requirements to avoid distractions and with the appropriate sound, light, and temperature conditions to be able to carry out an evaluation of these characteristics. The neuropsychological assessment was carried out by experts in neuropsychological assessment and ADHD diagnosis. Participants were asked to abstain from taking medication before the assessment.

Ethical Considerations

This study was approved by the ethical committee of the Alicante Institute for Health and Biomedical Research (ISABIAL; PI2022-139) and the University of Alicante (UA-2023-05-17_1). For children more than 12 years of age and less than 18 years of age, in addition to the consent of the parents or legal guardians, informed consent was also requested.

Statistical Analysis

Statistical analyses were performed using Jamovi (version 2.3.28) for Windows. To compare age between groups, the Mann-Whitney *U* test was used, while sex differences were assessed using the chi square test. Multivariate analysis of covariance (ANCOVA) was used to evaluate differences in task performance between the clinical (ADHD) and control groups in both children and adolescents, controlling for age and sex. Task performance was calculated by using the formula , where errors were defined as the sum of both omissions and commissions. Normality and homogeneity of variances were checked using the Shapiro-Wilk and Levene tests, respectively. Effect sizes for the ANCOVA were calculated using partial eta squared (η^2), with thresholds for low, moderate, and high effect sizes set at 0.06, 0.1, and 0.25 [52], respectively. Post hoc comparisons were performed using 2-tailed *t* tests to compare ADHD and control groups for both children and adolescents, with Cohen *d* used to interpret effect sizes: small (0.2), medium (0.5), and large (0.8) [53]. Pearson correlation coefficient analysis was conducted to examine the relationship between task performance and both family- and self-reported measures of attentional problems and hyperactivity. The level of confidence for all tests was set at .05.

Results

Age and Sex by Groups

The mean age of children in the ADHD group was comparable to that of the control group. However, there was a marginally significant difference in the number of boys and girls: 90% (9/10) of children in the ADHD group were boys, compared with 55% (11/20) in the control group (*P*=.06). The mean age of adolescents in the ADHD group was similar to that of adolescents in the control group. However, there was a difference in the number of boys and girls: 68% (15/22) of adolescents in the ADHD group were boys, compared with 26% (5/19) in the control group. (*P*<.001; Table 1). Table 2 shows the raw results of the task, divided by age and group.

XSL•FO
RenderX

Table . Sample age and sex, divided by group.

	ADHD[a] group	Control group	*P* value
Children (n=30)			
Age (years), mean (SD)	9 (1.56)	8.95 (1.67)	.97[b]
Sex, n (%)[c]			.06[d]
Male	9 (90)	11 (55)	
Female	1 (10)	9 (45)	
Adolescents (n=41)			
Age (years), mean (SD)	14.1 (1.64)	13.9 (1.72)	.72[b]
Sex, n (%)[c]			<.001[d,e]
Male	15 (68)	5 (26)	
Female	7 (32)	14 (74)	

[a]ADHD: attention-deficit/hyperactivity disorder.

[b]Mann-Whitney *U*.

[c]n (%) within group.

[d]χ^2.

[e]*P*<.05.

XSL•FO
RenderX

Table . Task results, divided by age and group.

	Mean (SD)
Hits	
Children – ADHD[a]	51 (5.27)
Children – Control	52.9 (5.22)
Adolescents – ADHD	54.41 (3.05)
Adolescents – Control	53.47 (3.79)
Omissions	
Children – ADHD	6 (5.27)
Children – Control	4.1 (5.22)
Adolescents – ADHD	2.59 (3.05)
Adolescents – Control	3.53 (3.79)
Commissions	
Children – ADHD	19.8 (16.65)
Children – Control	9.25 (5.84)
Adolescents – ADHD	6.64 (5.31)
Adolescents – Control	6.32 (4.36)
RT[b] hits	
Children – ADHD	1.04 (0.09)
Children – Control	1.02 (0.1)
Adolescents – ADHD	0.96 (0.09)
Adolescents – Control	0.96 (0.13)
RT commissions	
Children – ADHD	2.04 (0.62)
Children – Control	2.01 (0.75)
Adolescents – ADHD	2.12 (0.94)
Adolescents – Control	2.53 (2.51)
Task performance	
Children – ADHD	38.21 (32.62)
Children – Control	61.32 (24.46)
Adolescents – ADHD	72.38 (19.28)
Adolescents – Control	70.17 (20.7)

[a]ADHD: attention-deficit/hyperactivity disorder.

[b]RT: reaction time.

Differences in Task Performance Between Clinical and Control Groups

The ANCOVA analysis comparing task performance across 2 age groups (children and adolescents) with and without an ADHD diagnosis, controlling for age and sex, revealed no significant main effect between the ADHD and control groups ($F_{1,65}$=1.47; P=.23; η^2_p=0.022). However, there was a significant group-by-age group interaction ($F_{1,65}$=5.33; P=.02; η^2_p=0.08), indicating that the difference in task performance between the ADHD and control groups varied by age. Age itself was a significant covariate in predicting task performance, ($F_{1,65}$=7.13, P=.01, η^2_p=0.10) while sex was not ($F_{1,65}$=2.84, P=.10, η^2_p=0.042). The assumptions of homogeneity of variances and normality were met, confirming the validity of the analysis. Post hoc analyses showed that children with ADHD performed significantly worse than controls (t_{65}=−2.26, P=.03, Cohen d=0.9), whereas no significant differences were found between adolescents with ADHD and their control counterparts (t_{65}=0.75, P=.73, Cohen d=0.19; Table 3, Figure 3).

XSL•FO
RenderX

Table . Post hoc analysis of task performance differences between attention-deficit/hyperactivity disorder and control groups by age groups, using age and sex as covariates.

Age group	*t* test *(df)*	*P* value	|*d*|[a]
Children			
ADHD[b] vs Control	−2.26 (65)	.03[c]	0.90[d]
Adolescents			
ADHD vs Control	0.75 (65)	.73	0.19

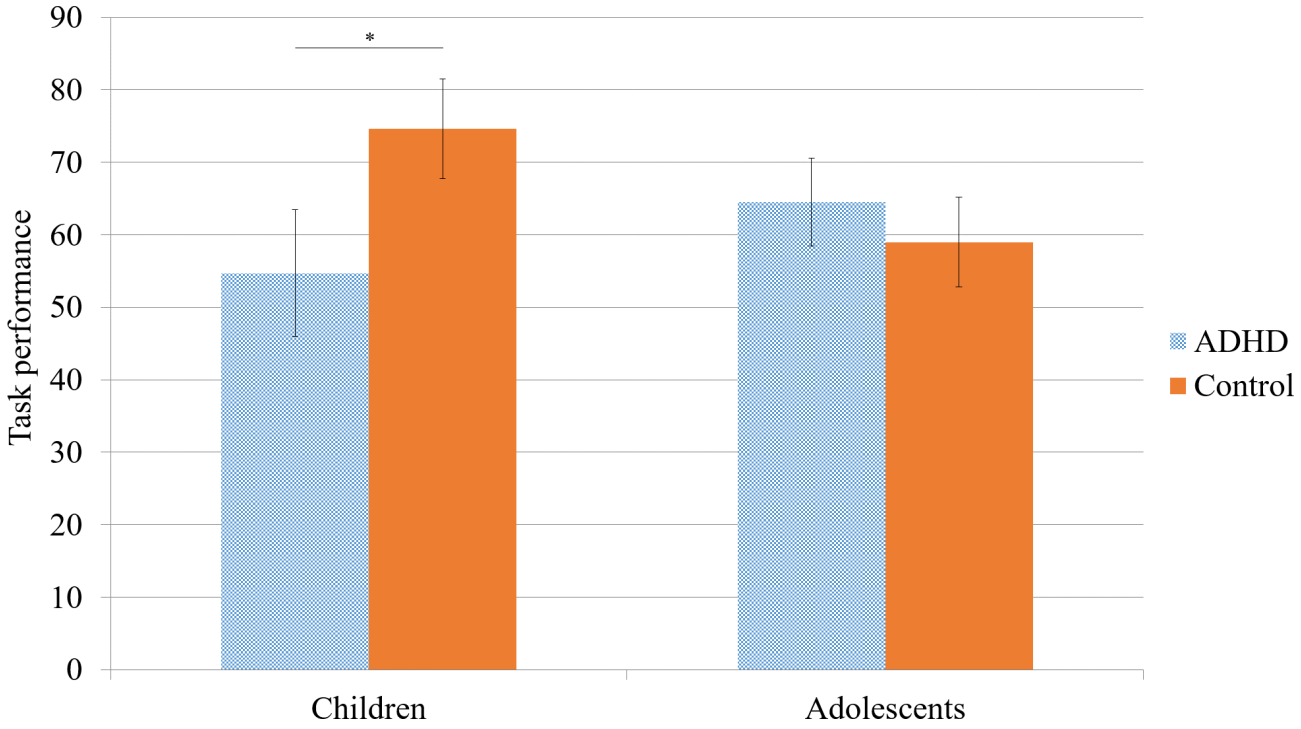
[a]Cohen *d* (large effect size: |*d*|>0.80).

[b]ADHD: attention-deficit/hyperactivity disorder.

[c]*P*<.05.

[d] |*d*|>0.8.

Figure 3. Estimated marginal means of task performance for children and adolescents with and without attention-deficit/hyperactivity disorder (ADHD). Error bars represent SE. *P<.05.



Correlation Analysis

Table 4 presents the results of the Pearson correlation analysis of the relationship between task performance and self- and family-reported measures of attention and impulsivity problems for both children and adolescents. For children, there was a significant negative correlation between task performance and family-reported hyperactivity/impulsivity problems of the SENA instrument (*r*=−0.42, *P*=.02) and family-reported hyperactivity/impulsivity from the ADHD-RS (*r*=−0.51, *P*=.01). Furthermore, task performance showed a marginally significant negative correlation with family-reported combined ADHD symptoms (*r*=−0.37, *P*=.06) of the SENA. No significant correlations were observed between task performance and self-reported attention problems, self-reported hyperactivity/impulsivity, or family-reported attention problems. In the case of adolescents, none of the correlations between task performance and the measures of attention and impulsivity problems reached significance.

XSL•FO
RenderX

Table . Correlation analysis between task performance and self or family reports.

	SENA[a] attention problems (self)	SENA hyperactivity-impulsivity (self)	SENA attention problems (family)	SENA hyperactivity-impulsivity (family)	ADHD-RS[b] inattention (family)	ADHD-RS hyperactivity-impulsivity (family)	ADHD-RS combined (family)
Children							
r	−0.12	0.02	−0.26	−0.43[c]	−0.19	−0.51[c]	−0.37
P value	.57	.93	.18	.02[c]	.33	.01[c]	.06
Adolescents							
r	−0.16	−0.29	0.04	−0.03	0.09	0.16	0.14
P value	.33	.08	.83	.87	.60	.35	.41

[a]SENA: Child and Adolescents Assessment System.

[b]ADHD-RS: ADHD Rating Scale IV.

[c]$P<.05$.

Discussion

Principal Findings

The objective of this study was to check the validity of a newly developed serious game for the assessment of children and adolescents with ADHD [46]. The first objective was to examine the discriminative power of the serious game in groups of children and adolescents with and without ADHD. Performance on the task was significantly worse in those children with ADHD compared with their nonclinical counterparts. Nevertheless, this difference was not observed in adolescents. These results support the use of Attention Slackline for assessing ADHD in children but not in adolescents. The second objective of the study was to check the concurrent validity of the serious game with self- and family-reported attention and hyperactivity-impulsivity symptomatology. In the children's group, the task performance was related to the measures of hyperactivity-impulsivity reported by relatives but not to attention scales or any self-reported values. Also, these results indicate that, for adolescents, task performance was not significantly associated with either self- or family-reported measures of attention and impulsivity. These findings provide support for the use of Attention Slackline as a reliable method to assess hyperactivity/impulsivity in children.

The current gold standard for ADHD assessment incorporates semistructured interviews, clinical observations, and neuropsychological tests, often complemented by teacher-parent rating scales [54]. Although CPTs are recognized as valid for diagnosing ADHD in children [55], their utility appears more limited in older populations [26]. For example, Berger et al [25] reported that children with ADHD scored significantly lower than controls on attention, timing, hyperactivity, and impulsivity indices. However, CPT performance tends to improve with age [56]. Therefore, the adolescent sample included in this study may have already developed sufficient skills to achieve a good performance in this paradigm.

Gamified interventions have shown promise for enhancing engagement and improving executive functions in children with ADHD [57]. Video game–based approaches have demonstrated utility in both assessment and treatment [40], and serious games can target the neurocognitive processes associated with neurodevelopmental disorders [39]. They also offer opportunities for customization to individual patient needs [30] and can be used to assist with both diagnosis and intervention [45]. In line with these findings, our data suggest that Attention Slackline may complement conventional methods, particularly by capturing impulsivity and hyperactivity. However, older individuals might perform near-typical levels on attention-focused tasks due to the development of compensatory strategies, consistent with concerns regarding the diagnostic power of CPTs in adults [26]. These observations highlight the importance of interpreting gamified CPT performance in the context of multiple measures and suggest that further refinements may be necessary to address subtle attentional deficits.

One of the limitations of the CPT paradigms is the ceiling effect [58,59], with a significant proportion of individuals achieving the highest score, which may be particularly prevalent in the older age groups. The paradigm used in this study was the same across all participants. As expected, task performance improved with age [56,60], with both groups of adolescents exhibiting high performance. This suggests the task may be too simple for adolescents, limiting its ability to differentiate between groups. It is possible that adolescents with ADHD may have developed compensatory strategies, which may explain the lack of correlation between task performance and ADHD symptomatology. To address this limitation, the task may be adapted by introducing context-specific distractors, such as visual or auditory stimuli, to better challenge cognitive control mechanisms in adolescents [60]. Another adaptation could involve increasing the complexity of task parameters, such as response speed requirements or number of stimuli. Increasing task complexity or tailoring its demands to developmental levels may enhance its sensitivity and applicability for this age group.

Previous CPT assessments have yielded inconclusive results with regard to parent and rating scales of ADHD symptoms [55]. The findings of the Attention Slackline paradigm in this study indicate that this paradigm is associated with the impulsivity/hyperactivity features observed in children as reported by their families. These results are congruent with the positive illusory bias, a cognitive distortion in which individuals,

XSL•FO
RenderX

particularly children with ADHD, tend to overestimate their own competencies and abilities compared with objective assessments [16]. Attention Slackline appears to be a more effective method for assessing hyperactivity and impulsivity symptoms in children with ADHD than attention deficits. This could be attributed to the relatively brief duration of the task and the gamified environment. The use of a gamified environment may have augmented their engagement with the task, thereby reducing attention problems [40,45,57]. However, the use of gamified tasks would permit the observation of their inclination to respond impulsively, without fatigue, due to the duration influencing their impulsive tendency. This is why a serious game such as Attention Slackline can serve as a complement to the diagnosis of ADHD [55], serving to assess attention and impulsivity problems.

Serious games, such as Attention Slackline, are emerging as a promising addition to the clinical practice toolkit for ADHD by offering advantages over traditional assessment methods. In a narrative review, Aloisio et al [43] noted that these tools can reduce evaluation costs and time while preserving accuracy and promoting patient engagement. A randomized controlled trial by Kollins et al [42] provided further evidence by demonstrating that a digital intervention significantly reduced ADHD symptom severity in children, suggesting strong potential for real-world clinical impact. Schena et al [44] supported these findings through preliminary results showing that a virtual reality-based serious game improved core ADHD symptoms in an experimental setting, highlighting the adaptability of this approach. Zheng et al [45] additionally emphasized that serious games can facilitate both assessment and treatment, especially when integrated with large-scale data analysis. Collectively, these studies underscore the capacity of serious games to improve ADHD evaluation and intervention by offering scalable, flexible, and engaging solutions.

Limitations

The results of this study should be interpreted in light of its limitations. First, the small size of the female population included in the sample, especially in the children's group. Second, although the participants were instructed to abstain from taking medication before the assessment if they were under medical treatment for ADHD, possible sustained long-term effects of medication could not be controlled.

Future Works

Future studies should aim to expand the sample size of children to determine clinically relevant cutoff points for Attention Slackline, enhancing its utility in clinical contexts [61,62]. In addition, research should explore whether a more challenging CPT could effectively distinguish between adolescents with and without ADHD. This could involve increasing the speed or duration of the task, or introducing stimuli as distractors [25]. The use of computer-gamified paradigms for ADHD assessment also opens up opportunities for incorporating machine-learning technology to evaluate performance in serious games, which may significantly improve the accuracy of ADHD diagnosis. Furthermore, future research should aim to differentiate between subtypes of ADHD [40], potentially allowing for more tailored and precise interventions based on the specific subtype diagnosed.

Conclusions

This study provides preliminary evidence to support the use of Attention Slackline as a serious game for assessing impulsivity in children with and without ADHD diagnosis. The results suggest that the serious game is an effective tool for distinguishing children with ADHD from those without the disorder. The use of a gamified task has the potential to engage children in the task and demonstrate their impulsive tendencies in responding. Early identification of ADHD symptoms is crucial to facilitate timely and appropriate interventions. The use of a serious game such as Attention Slackline as an additional assessment tool may provide a useful objective measure of impulsivity in children.

Acknowledgments

This work has been funded by the BALLADEER (PROMETEO/2021/088) project, a big data analytical platform for the diagnosis and treatment of attention-deficit/hyperactivity disorder featuring extended reality, funded by Conselleria de Innovación, Universidades Ciencia y Sociedad Digital (Generalitat Valenciana). This work has also been funded by Grant RED2022-134656-T funded by MCIN/AEI/10.13039/501100011033. This work has also been funded by Program for the Promotion of R + D + I (UAIND20-03B), funded by Vicerrectorado de Investigación y Transferencia de Conocimiento de la Universidad de Alicante. This work has also been funded by the AETHER-UA project (PID2020-112540RB-C43), a smart data holistic approach for context-aware data analytics: smarter machine learning for business modelling and analytics, funded by Spanish Ministry of Science and Innovation.

Conflicts of Interest

None declared.

References

1. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, Text Revision (DSM-5-TR): American Psychiatric Association Publishing; 2022. [doi: 10.1176/appi.books.9780890425787]

XSL•FO
RenderX

2. Wolraich ML, Hagan JF Jr, Allan C, et al. Clinical practice guideline for the diagnosis, evaluation, and treatment of attention-deficit/hyperactivity disorder in children and adolescents. Pediatrics 2019 Oct;144(4):e20192528. [doi: 10.1542/peds.2019-2528] [Medline: 31570648]
3. Willcutt EG. The prevalence of DSM-IV attention-deficit/hyperactivity disorder: a meta-analytic review. Neurotherapeutics 2012 Jul;9(3):490-499. [doi: 10.1007/s13311-012-0135-8] [Medline: 22976615]
4. Zulueta A, Díaz-Orueta U, Crespo-Eguilaz N, et al. Virtual reality-based assessment and rating scales in ADHD diagnosis. Psicología Educativa 2018 Dec 20;25(1):13-22. [doi: 10.5093/psed2018a18]
5. Salari N, Ghasemi H, Abdoli N, et al. The global prevalence of ADHD in children and adolescents: a systematic review and meta-analysis. Ital J Pediatr 2023 Apr 20;49(1):48. [doi: 10.1186/s13052-023-01456-1] [Medline: 37081447]
6. Catalá-López F, Peiró S, Ridao M, et al. Prevalence of attention deficit hyperactivity disorder among children and adolescents in Spain: a systematic review and meta-analysis of epidemiological studies. BMC Psychiatry 2012 Oct;12:168. [doi: 10.1186/1471-244X-12-168] [Medline: 23057832]
7. Sayal K, Prasad V, Daley D, et al. ADHD in children and young people: prevalence, care pathways, and service provision. Lancet Psychiatry 2018 Feb;5(2):175-186. [doi: 10.1016/S2215-0366(17)30167-0] [Medline: 29033005]
8. Faheem M, Akram W, Akram H, et al. Gender-based differences in prevalence and effects of ADHD in adults: a systematic review. Asian J Psychiatr 2022 Sep;75:103205. [doi: 10.1016/j.ajp.2022.103205] [Medline: 35878424]
9. Barkley RA. Attention-Deficit Hyperactivity Disorder: A Handbook for Diagnosis and Treatment: The Guilford Press; 2018.
10. Thome J, Ehlis AC, Fallgatter AJ, et al. Biomarkers for attention-deficit/hyperactivity disorder (ADHD). A consensus report of the WFSBP task force on biological markers and the World Federation of ADHD. World J Biol Psychiatry 2012 Jul;13(5):379-400. [doi: 10.3109/15622975.2012.690535] [Medline: 22834452]
11. Pollak Y, Dekkers TJ, Shoham R, et al. Risk-taking behavior in attention deficit/hyperactivity disorder (ADHD): a review of potential underlying mechanisms and of interventions. Curr Psychiatry Rep 2019 Mar 22;21(5):33. [doi: 10.1007/s11920-019-1019-y] [Medline: 30903380]
12. Lopez-Villalobos JA, Garrido-Redondo M, Sacristan-Martin AM, et al. Children's and adolescents' perception of their quality of life in cases of attention deficit hyperactivity disorder with and without pharmacological treatment and in controls. Rev Neurol 2018 Sep 16;67(6):195-202. [doi: 10.33588/rn.6706.2017517] [Medline: 30183059]
13. Peasgood T, Bhardwaj A, Biggs K, et al. The impact of ADHD on the health and well-being of ADHD children and their siblings. Eur Child Adolesc Psychiatry 2016 Nov;25(11):1217-1231. [doi: 10.1007/s00787-016-0841-6] [Medline: 27037707]
14. Bélanger SA, Andrews D, Gray C, et al. ADHD in children and youth: part 1-etiology, diagnosis, and comorbidity. Paediatr Child Health 2018 Nov;23(7):447-453. [doi: 10.1093/pch/pxy109] [Medline: 30681669]
15. Elwin M, Elvin T, Larsson JO. Symptoms and level of functioning related to comorbidity in children and adolescents with ADHD: a cross-sectional registry study. Child Adolesc Psychiatry Ment Health 2020;14:30. [doi: 10.1186/s13034-020-00336-4] [Medline: 32760444]
16. Owens JS, Goldfine ME, Evangelista NM, et al. A critical review of self-perceptions and the positive illusory bias in children with ADHD. Clin Child Fam Psychol Rev 2007 Dec;10(4):335-351. [doi: 10.1007/s10567-007-0027-3] [Medline: 17902055]
17. Colomer C, Wiener J, Varma A. Do adolescents with ADHD have a self-perception bias for their ADHD symptoms and impairment? Can J Sch Psychol 2020 Dec;35(4):238-251. [doi: 10.1177/0829573520936457]
18. Marshall P, Hoelzle J, Nikolas M. Diagnosing attention-deficit/hyperactivity disorder (ADHD) in young adults: a qualitative review of the utility of assessment measures and recommendations for improving the diagnostic process. Clin Neuropsychol 2021 Jan;35(1):165-198. [doi: 10.1080/13854046.2019.1696409] [Medline: 31791193]
19. Peterson BS, Trampush J, Brown M, et al. Tools for the diagnosis of ADHD in children and adolescents: a systematic review. Pediatrics 2024 Apr 1;153(4):e2024065854. [doi: 10.1542/peds.2024-065854] [Medline: 38523599]
20. Koziol LF, Budding D. Requiem for a diagnosis: attention-deficit hyperactivity disorder. Appl Neuropsychol Child 2012;1(1):2-5. [doi: 10.1080/21622965.2012.665774] [Medline: 23428272]
21. Lange KW, Hauser J, Lange KM, et al. Utility of cognitive neuropsychological assessment in attention-deficit/hyperactivity disorder. Atten Defic Hyperact Disord 2014 Dec;6(4):241-248. [doi: 10.1007/s12402-014-0132-3] [Medline: 24639037]
22. Wright AJ. Psychological and neuropsychological underpinnings of attention-deficit/hyperactivity disorder assessment. Clin Child Psychol Psychiatry 2021 Jul;26(3):783-794. [doi: 10.1177/1359104521996765] [Medline: 33624519]
23. Losier BJ, McGrath PJ, Klein RM. Error patterns on the continuous performance test in non-medicated and medicated samples of children with and without ADHD: a meta-analytic review. J Child Psychol Psychiatry 1996 Nov;37(8):971-987. [doi: 10.1111/j.1469-7610.1996.tb01494.x] [Medline: 9119944]
24. Berger I, Slobodin O, Aboud M, et al. Maturational delay in ADHD: evidence from CPT. Front Hum Neurosci 2013;7:691. [doi: 10.3389/fnhum.2013.00691] [Medline: 24298243]
25. Berger I, Slobodin O, Cassuto H. Usefulness and validity of continuous performance tests in the diagnosis of attention-deficit hyperactivity disorder children. Arch Clin Neuropsychol 2017 Feb;32(1):81-93. [doi: 10.1093/arclin/acw101] [Medline: 28122767]

XSL•FO
RenderX

26. Varela JL, Magnante AT, Miskey HM, et al. A systematic review of the utility of continuous performance tests among adults with ADHD. Clin Neuropsychol 2024 Oct;38(7):1524-1585. [doi: 10.1080/13854046.2024.2315740] [Medline: 38424025]
27. Valentine AZ, Brown BJ, Groom MJ, et al. A systematic review evaluating the implementation of technologies to assess, monitor and treat neurodevelopmental disorders: A map of the current evidence. Clin Psychol Rev 2020 Aug;80:101870. [doi: 10.1016/j.cpr.2020.101870] [Medline: 32712216]
28. Tan Y, Zhu D, Gao H, et al. Virtual classroom: an ADHD assessment and diagnosis system based on virtual reality. Presented at: 2019 IEEE International Conference on Industrial Cyber Physical Systems (ICPS); May 6-9, 2019; Taipei, Taiwan p. 203-208. [doi: 10.1109/ICPHYS.2019.8780300]
29. Negu A, Matu SA, Sava FA, et al. Virtual reality measures in neuropsychological assessment: a meta-analytic review. Clin Neuropsychol 2016 Feb;30(2):165-184. [doi: 10.1080/13854046.2016.1144793] [Medline: 26923937]
30. Damaševičius R, Maskeliūnas R, Blažauskas T. Serious games and gamification in healthcare: a meta-review. Information 2023;14(2):105. [doi: 10.3390/info14020105]
31. Laamarti F, Eid M, El Saddik A. An overview of serious games. Int J Comput Games Technol 2014;2014:1-15. [doi: 10.1155/2014/358152]
32. Valladares-Rodríguez S, Pérez-Rodríguez R, Anido-Rifón L, et al. Trends on the application of serious games to neuropsychological evaluation: A scoping review. J Biomed Inform 2016 Dec;64:296-319. [doi: 10.1016/j.jbi.2016.10.019] [Medline: 27815228]
33. Zygouris S, Iliadou P, Lazarou E, et al. Detection of mild cognitive impairment in an at-risk group of older adults: can a novel self-administered serious game-based screening test improve diagnostic accuracy? J Alzheimers Dis 2020;78(1):405-412. [doi: 10.3233/JAD-200880] [Medline: 32986676]
34. Hong JS, Kim SM, Aboujaoude E, et al. Investigation of a mobile “serious game” in the treatment of obsessive-compulsive disorder: a pilot study. Games Health J 2018 Oct;7(5):317-326. [doi: 10.1089/g4h.2017.0158] [Medline: 30129775]
35. Mack I, Reiband N, Etges C, et al. The kids obesity prevention program: cluster randomized controlled trial to evaluate a serious game for the prevention and treatment of childhood obesity. J Med Internet Res 2020 Apr 24;22(4):e15725. [doi: 10.2196/15725] [Medline: 32329742]
36. Mena-Moreno T, Munguía L, Granero R, et al. Cognitive behavioral therapy plus a serious game as a complementary tool for a patient with Parkinson disease and impulse control disorder: case report. JMIR Serious Games 2022 Sep 9;10(3):e33858. [doi: 10.2196/33858] [Medline: 36083621]
37. Isernia S, Cabinio M, Di Tella S, et al. Diagnostic validity of the smart aging serious game: an innovative tool for digital phenotyping of mild neurocognitive disorder. J Alzheimers Dis 2021;83(4):1789-1801. [doi: 10.3233/JAD-210347] [Medline: 34459394]
38. Tong T, Chignell M, Tierney MC, Lee J. A serious game for clinical assessment of cognitive status: validation study. JMIR Serious Games 2016 May 27;4(1):e7. [doi: 10.2196/games.5006] [Medline: 27234145]
39. Vacca RA, Augello A, Gallo L, et al. Serious games in the new era of digital-health interventions: a narrative review of their therapeutic applications to manage neurobehavior in neurodevelopmental disorders. Neurosci Biobehav Rev 2023 Jun;149:105156. [doi: 10.1016/j.neubiorev.2023.105156] [Medline: 37019246]
40. Peñuelas-Calvo I, Jiang-Lin LK, Girela-Serrano B, et al. Video games for the assessment and treatment of attention-deficit/hyperactivity disorder: a systematic review. Eur Child Adolesc Psychiatry 2022 Jan;31(1):5-20. [doi: 10.1007/s00787-020-01557-w] [Medline: 32424511]
41. Rodrigo-Yanguas M, González-Tardón C, Bella-Fernández M, et al. Serious video games: angels or demons in patients with attention-deficit hyperactivity disorder? A quasi-systematic review. Front Psychiatry 2022;13:798480. [doi: 10.3389/fpsyt.2022.798480] [Medline: 35573357]
42. Kollins SH, DeLoss DJ, Cañadas E, et al. A novel digital intervention for actively reducing severity of paediatric ADHD (STARS-ADHD): a randomised controlled trial. Lancet Digit Health 2020 Apr;2(4):e168-e178. [doi: 10.1016/S2589-7500(20)30017-0] [Medline: 33334505]
43. Aloisio A, Marzano G, Bonanni R, Ricci M. Serious games for ADHD: a narrative review. Presented at: Proceedings of the International Workshop on Digital Innovations for Learning and Neurodevelopmental Disorders; May 24-25, 2024; Rome, Italy.
44. Schena A, Garotti R, D’Alise D, et al. IAmHero: preliminary findings of an experimental study to evaluate the statistical significance of an intervention for ADHD conducted through the use of serious games in virtual reality. Int J Environ Res Public Health 2023 Feb 15;20(4):3414. [doi: 10.3390/ijerph20043414] [Medline: 36834109]
45. Zheng Y, Li R, Li S, et al. A review on serious games for ADHD. arXiv. Preprint posted online on Mar 31, 2021. [doi: 10.48550/arXiv.2105.02970]
46. Teruel MA, Sanchis J, Ruiz-Robledillo N, et al. Measuring attention of ADHD patients by means of a computer game featuring biometrical data gathering. Heliyon 2024 Mar 15;10(5):e26555. [doi: 10.1016/j.heliyon.2024.e26555] [Medline: 38434359]
47. Piedmont RL. Criterion validity. In: Encyclopedia of Quality of Life and Well-Being Research: Springer International Publishing; 2023:1495-1495. [doi: 10.1007/978-3-031-17299-1_618]

XSL•FO
RenderX

48. Frisira E, Holland J, Sayal K. Systematic review and meta-analysis: relative age in attention-deficit/ hyperactivity disorder and autism spectrum disorder. Eur Child Adolesc Psychiatry 2025 Feb;34(2):381-401. [doi: 10.1007/s00787-024-02459-x] [Medline: 38767699]
49. Sánchez-Sánchez F, Fernández-Pinto I, Santamaría P, Carrasco MA, Del Barrio V. SENA, Sistema de Evaluación de Niños y Adolescentes: proceso de desarrollo y evidencias de fiabilidad y validez [Article in Spanish]. Revista de Psicología Clínica con Niños y Adolescentes 2016;3(2) [FREE Full text]
50. DuPaul GJ, Power TJ, Anastopoulos AD, Reid R. ADHD Rating Scale—IV: Checklists, Norms, and Clinical Interpretation: The Guilford Press; 1998.
51. Servera M, Cardo E. ADHD Rating Scale-IV in a sample of Spanish schoolchildren: normative data and internal consistency for teachers and parents. Rev Neurol 2007;45(7):393-399. [Medline: 17918104]
52. Levine TR, Hullett CR. Eta squared, partial eta squared, and misreporting of effect size in communication research. Human Comm Res 2002 Oct;28(4):612-625. [doi: 10.1111/j.1468-2958.2002.tb00828.x]
53. Kelley K, Preacher KJ. On effect size. Psychol Methods 2012 Jun;17(2):137-152. [doi: 10.1037/a0028086] [Medline: 22545595]
54. Staff AI, Oosterlaan J, van der Oord S, et al. The validity of teacher rating scales for the assessment of ADHD symptoms in the classroom: a systematic review and meta-analysis. J Atten Disord 2021 Sep;25(11):1578-1593. [doi: 10.1177/1087054720916839] [Medline: 32390490]
55. Hall CL, Valentine AZ, Groom MJ, et al. The clinical utility of the continuous performance test and objective measures of activity for diagnosing and monitoring ADHD in children: a systematic review. Eur Child Adolesc Psychiatry 2016 Jul;25(7):677-699. [doi: 10.1007/s00787-015-0798-x] [Medline: 26620873]
56. Brocki KC, Tillman CM, Bohlin G. CPT performance, motor activity, and continuous relations to ADHD symptom domains: a developmental study. Eur J Dev Psychol 2010 Mar;7(2):178-197. [doi: 10.1080/17405620801937764]
57. Alabdulakareem E, Jamjoom M. Computer-assisted learning for improving ADHD individuals' executive functions through gamified interventions: a review. Entertain Comput 2020 Mar;33:100341. [doi: 10.1016/j.entcom.2020.100341]
58. Kahn PV, Walker TM, Williams TS, et al. Standardizing the use of the continuous performance test in schizophrenia research: a validation study. Schizophr Res 2012 Dec;142(1-3):153-158. [doi: 10.1016/j.schres.2012.09.009] [Medline: 23040735]
59. Hsieh PC, Chu CL, Yang YK, et al. Norms of performance of sustained attention among a community sample: Continuous performance test study. Psychiatry Clin Neurosci 2005 Apr;59(2):170-176. [doi: 10.1111/j.1440-1819.2005.01353.x] [Medline: 15823163]
60. Slobodin O, Cassuto H, Berger I. Age-related changes in distractibility: developmental trajectory of sustained attention in ADHD. J Atten Disord 2018 Dec;22(14):1333-1343. [doi: 10.1177/1087054715575066] [Medline: 25791438]
61. Brevik EJ, Lundervold AJ, Haavik J, et al. Validity and accuracy of the adult attention-deficit/hyperactivity disorder (ADHD) Self-Report Scale (ASRS) and the Wender Utah Rating Scale (WURS) symptom checklists in discriminating between adults with and without ADHD. Brain Behav 2020 Jun;10(6):e01605. [doi: 10.1002/brb3.1605] [Medline: 32285644]
62. Faraone SV, Newcorn JH, Antshel KM, et al. The groundskeeper gaming platform as a diagnostic tool for attention-deficit/hyperactivity disorder: sensitivity, specificity, and relation to other measures. J Child Adolesc Psychopharmacol 2016 Oct;26(8):672-685. [doi: 10.1089/cap.2015.0174] [Medline: 27105181]

Abbreviations

ADHD: attention-deficit/hyperactivity disorder
ADHD-RS: ADHD Rating Scale IV
ANCOVA: analysis of covariance
CPT: continuous performance test
***DSM-5*:** *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition*
SENA: Child and Adolescent Assessment System (Sistema de Evaluación de Niños y Adolescentes)

Edited by A Bustillo; submitted 07.08.24; peer-reviewed by F Yang, SC Kim; revised version received 10.02.25; accepted 10.02.25; published 13.05.25.

Please cite as:
Ruiz-Robledillo N, Lucas I, Ferrer-Cascales R, Albaladejo-Blázquez N, Sanchis J, Trujillo J
Discriminative Power of the Serious Game Attention Slackline in Children and Adolescents With and Without Attention-Deficit/Hyperactivity Disorder: Validation Study
JMIR Serious Games 2025;13:e65170
URL: https://games.jmir.org/2025/1/e65170
doi:10.2196/65170

XSL•FO
RenderX

© Nicolás Ruiz-Robledillo, Ignacio Lucas, Rosario Ferrer-Cascales, Natalia Albaladejo-Blázquez, Javier Sanchis, Juan Trujillo. Originally published in JMIR Serious Games (https://games.jmir.org), 13.5.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Machine Learning Analysis of Engagement Behaviors in Older Adults With Dementia Playing Mobile Games: Exploratory Study

Melika Torabgar[1], BE, MSc; Mathieu Figeys[1], PhD; Shaniff Esmail[1], OTR, PhD; Eleni Stroulia[2], PhD; Adriana M Ríos Rincón[1], OTR, PhD

[1]Department of Occupational Therapy, University of Alberta, Edmonton, AB, Canada

[2]Department of Computing Science, University of Alberta, Edmonton, AB, Canada

Corresponding Author:
Adriana M Ríos Rincón, OTR, PhD
Department of Occupational Therapy, University of Alberta, Edmonton, AB, Canada

Abstract

Background: The prevalence of dementia is expected to rise with an aging population, necessitating accessible early detection methods. Serious games have emerged as potential cognitive screening tools. They provide not only an engaging platform for assessing cognitive function but also serve as valuable indicators of cognitive health through engagement levels observed during play.

Objective: This study aims to examine the differences in engagement-related behaviors between older adults with and without dementia during serious gaming sessions. Further, it seeks to identify the key contributors that enhance the effectiveness of machine learning for dementia classification based on engagement-related behaviors.

Methods: This was an exploratory proof-of-concept study. Over 8 weeks, 20 older adults, 6 of whom were living with dementia, were enrolled in a single-case design study. Participants played 1 of 4 "Vibrant Minds" serious games (Bejeweled, Whack-A-Mole, Mah-jong, and Word-Search) over 8 weeks (16 30-min sessions). Throughout the study, sessions were recorded to analyze engagement-related behaviors. This paper reports on the analysis of the engagement-related behaviors of 15 participants. The videos of these 15 participants (10 cognitively intact, 5 with dementia) were analyzed by 2 independent raters, individually annotating engagement-related behaviors at 15-second intervals using a coding system. This analysis resulted in 1774 data points categorized into 47 behavior codes, augmented by 54 additional features including personal characteristics, technical issues, and environmental factors. Each engagement-related behavior was compared between older adults living with dementia and older adults without dementia using the χ^2 test with a 2×2 contingency table with a significance level of .05. Codes underwent one-hot encoding and were processed using random forest classifiers to distinguish between participant groups.

Results: Significant differences in 64% of engagement-related behaviors were found between groups, notably in torso movements, voice modulation, facial expressions, and concentration. Including engagement-related behaviors, environmental disturbances, technical issues, and personal characteristics resulted in the best model for classifying cases of dementia correctly, achieving an F_1-score of 0.91 (95% CI 0.851 - 0.963) and an area under the receiver operating curve of 0.99 (95% CI 0.984 - 1.000).

Conclusions: Key features distinguishing between older adults with and without dementia during serious gameplay included torso, voice, facial, and concentration behaviors, as well as age. The best performing machine learning model identified included features of engagement-related behavios, environmental disturbances, technical challenges, and personal attributes. Engagement-related behaviors observed during serious gaming offer crucial markers for identifying dementia. Machine learning models that incorporate these unique behavioral markers present a promising, noninvasive approach for early dementia screening in a variety of settings.

(JMIR Serious Games 2025;13:e54797) doi:10.2196/54797

KEYWORDS

dementia; gaming; engagement; cognition; machine learning; games; cognitive; screening; classification; Alzheimer disease; gerontology; geriatric; older adult; elderly; aging

Introduction

The world's population is aging at an unprecedented rate, which will impact the prevalence rate of older adults living with dementia. The World Health Organization projects that there will be over 132 million people living with dementia by 2050 [1]. Dementia not only affects the individual's cognitive functioning but also creates several challenges for families, caregivers, and health care systems. Moreover, the economic

XSL•FO
RenderX

burden of dementia is burdensome, with global estimated economic costs of nearly US$1 trillion in 2016 [2].

Dementia encompasses a group of disorders characterized by a gradual decline in memory, reasoning, communication, and daily functioning, typically due to brain damage. Alzheimer disease (AD) is the most common form, followed by vascular dementia and dementia with Lewy bodies [3]. Cognitive impairments, such as dementia, are typically assessed and monitored by health care professionals during routine clinical examinations or when an older adult or their family members identify symptoms of cognitive decline and seek assistance [3]. Limitations of current screening have spurred research into the development of alternative, more naturalistic methods to monitor for cognitive decline. Early identification and the prevention of cognitive decline are critical in reducing the burden of dementia on individuals, families, health care systems, and larger economies and societies.

The use of machine learning (ML) in assessing cognitive decline is an emerging field of study. A recent scoping review by Tolks et al [4] identified a limited evidence base (16 articles) on the use of artificial intelligence in health-related serious games, predominantly using ML methods. Among these, only 2 studies focused on mild cognitive impairment, with none specifically targeting dementia. Beyond serious gaming, Bayat et al [5] developed an ML model with a 91% accuracy in identifying older individuals with and without preclinical AD (the phase that occurs prior to the onset of symptoms) by analyzing driving behaviors recorded through in-vehicle global positioning system data loggers. Similarly, Padhee et al [6] built an ML model with an 81% accuracy rate to distinguish individuals with AD from those without AD based on verbal utterances while describing a picture.

Computer serious games performance has been used to distinguish dementia and mild cognitive impairment in older adults using ML techniques [7-9]. While using game performance and ML techniques to distinguish dementia in older adults can have some potential benefits, there are also disadvantages to consider. First, there is variability in the computer games designed to assess cognitive abilities in aspects such as the cognitive domains targeted, topics or themes of the game, levels of difficulty, and protocols. This variability prevents the generalization of the results, limiting the performance of an ML algorithm to a specific game and under specific conditions. Second, older adults with dementia may have varying levels of interest, motivation, and familiarity with the computer game being used for assessment. Thus, assessments that rely exclusively on game performance can lead to unreliable or biased results. In our study, we explore engagement behaviors instead of game performance to distinguish dementia in older adults.

Engagement is a complex construct that has behavioral, cognitive, and affective dimensions [10]. The behavioral dimension includes effort and task persistence, whereas the cognitive dimension includes attention and concentration. The affective dimension includes valence, arousal, as well as discrete emotions such as curiosity and interest [11]. The multifaceted nature of engagement suggests that behaviors related to engagement could potentially provide insights into the presence and progression of dementia.

This study aims to contribute to the ongoing efforts to develop new and objective methods of assessing cognitive decline in older adults. In doing so, this study tested two hypotheses: (1) there are significant differences in engagement-related behaviours of older adults living with dementia and those without dementia while playing mobile games and (2) a dataset of engagement-related behaviours, along with other characteristics such as personal characteristics, environmental disturbances, and technical issues, generates distinct patterns for the presence of dementia in older adults while playing mobile games.

Methods

Study Design

In total, 20 older adults, including 6 individuals living with dementia, participated in a single-case design study [12]. This paper focuses on analyzing engagement-related behaviors observed in 15 participants who were video-recorded during their game sessions.

Participants

Participants were eligible if they were 60 years or older, cognitively intact or self-reported as having dementia, and able to attend the scheduled sessions. All participants were recruited from the community, and the study was conducted in 2 settings: a community organization providing support to individuals with dementia and their families and a supportive living facility, both located in Edmonton, Alberta, Canada [12]. Each participant was randomly assigned to 1 of 4 “Vibrant Minds” [13] mobile games developed for cognitive assessment and intervention. The method used for random allocation involved a lottery-style draw, a well-recognized approach to ensure randomization. Each participant randomly selected a card from a set of 4 cards, each featuring the logo of one of the “Vibrant Minds” mobile games (Bejeweled, Whack-A-Mole, Mah-jong, or Word-Search). To maintain blinding during the selection, the cards were placed face down with identical backs, ensuring that participants were unaware of the game assignment prior to selection. This approach was administered and supervised by a member of the research team to ensure fairness and adherence to the protocol. Once assigned, participants continued playing the same game throughout the 8-week study, which included 16 sessions lasting 30 minutes each. These games were designed to provide cognitive stimulation for older adults and were played on a 10-inch tablet in each session. Sessions took place in a designated room where participants were seated at a table. They were held twice weekly, typically on the same days and at the same time whenever feasible. While participants completed the activities individually, they shared the same room. Each session was supervised by 2-5 research assistants who prepared the space, materials, and equipment; provided instructions; and assisted participants as needed. Sessions lasted 30 minutes, with participants receiving consistent verbal instructions: “We prepared these activities for you. We want you to do these for 30 minutes.” Research assistants collected demographic data (eg, age, gender, education, technology literacy, and experience

XSL•FO
RenderX

with serious games) and administered cognitive measures (eg, Montreal Cognitive Assessment [14]) at each phase of the single-case design. Cognitive measures assessed the intervention's potential impact on cognitive functions, with results reported elsewhere [12].

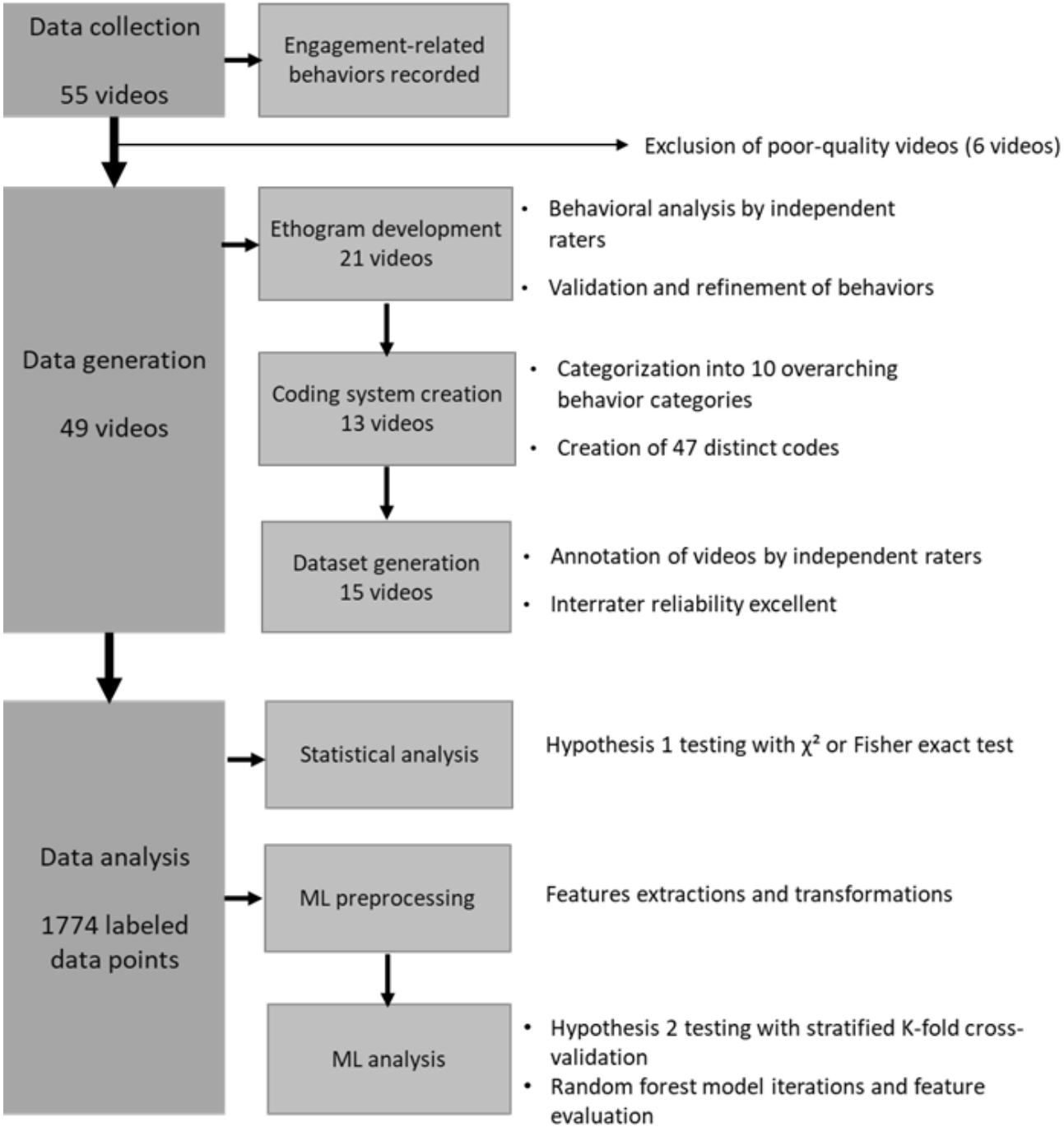
Data Source

Overview

Figure 1 illustrates the study phases. Throughout the study, engagement-related behaviors were recorded for all participants during at least 1 session, resulting in 55 videos. Participants for each session were randomly selected using a lottery-style draw, and each recording lasted 30 minutes. The video footage captured each participant's face, trunk, and upper limbs [15]. In total, 6 videos were excluded due to poor quality, leaving 49 usable recordings.

Our approach consisted of 3 phases: (1) the development of an ethogram (a detailed description and analysis of behaviors), (2) the creation of a coding system (a structured framework for categorizing and labeling behaviors to enable systematic analysis), and (3) the generation of the dataset of engagement-related behaviors. This methodology has been applied in prior research involving individuals living with dementia [16]. A total of 34 videos (70% of the 49 usable recordings) were used during the first 2 phases to ensure a comprehensive and rigorous foundation for subsequent analyses [15].

Figure 1. Study flow chart. ML: machine learning.



XSL•FO
RenderX

Ethogram Development

In total, 2 independent raters analyzed video sequences to identify and describe behaviors indicative of engagement, disengagement, or neutrality. A second set of raters reviewed the identified behaviors to validate those with clear relevance to engagement. Ambiguous behaviors were excluded through team discussions to ensure clarity and consistency.

Coding System Creation

A separate researcher reviewed the validated behaviors and organized them into 10 overarching categories: gaze, eyes, head, torso, limbs, face, voice, gameplay, concentration, and breath. Within each category, behaviors were grouped by similarities, resulting in 47 distinct codes that formed a structured framework for behavior analysis.

Dataset Generation

The remaining 15 videos (30%) were analyzed to create a dataset of engagement-related behaviors. These videos represented a subset of 15 participants, comprising 10 cognitively intact older adults and 5 individuals with mild-to-moderate dementia. In total, 2 independent raters annotated the videos, identifying engagement-related behaviors at 15-second intervals using the coding system. This process resulted in a dataset with 1774 labeled data points. Interrater reliability for the coding system was excellent (κ=0.87) [15].

All raters underwent training in engagement theory and behavior analysis, led by a senior member of the research team. Importantly, the raters were blinded to the demographic information of participants in the videos, including their dementia status.

Ethical Considerations

Potential participants and their caregivers were given recruitment flyers with the research team's contact details. Cognitively healthy older adults, individuals with mild dementia, or their substitute decision-makers contacted the research team to express interest. A team member then scheduled in-person or telephone meetings to explain the study and confirm. Written consent was obtained from participants or their substitute decision-makers. The University of Alberta Research Ethics Board reviewed and approved this study (Pro00069138). Three times during the study, participants were given a coffee shop gift card, ranging in value from CAD $10 to $25 (US $6.99 to $17.47). All personal identifying information was removed (deidentified) before analysis.

ML Preprocessing

The input data encompassed 54 features, which incorporated the 47 engagement-related behaviors along with personal characteristics (ie, age, gender, highest level of education, technology literacy, and prior experience with serious games), environmental disturbances (ie, any disruption during the game session), and technical issues. Age was transformed into a categorical variable using the visual binning method in SPSS (version 23; IBM). A one-hot encoding method was used to convert the categorical codes into binary values.

Statistical Analysis

To test hypothesis 1, each engagement-related behavior was compared between older adults living with dementia and older adults without dementia using the χ^2 test with a 2×2 contingency table with a significance level of .05. For expected frequencies of less than 5 occurrences, Fisher exact test was employed [17]. The analyses were performed using SPSS Statistics (version 23.0; IBM Corp.).

ML Analysis

To test hypothesis 2, the preprocessed dataset underwent stratified K-fold cross-validation [18], due to the ratio of older adults living with dementia versus older adults without dementia being 0.33 [19]. The dataset was divided into 4 different folds, with the model trained on three folds and tested on the remaining fold in each iteration. The training dataset constituted 75% of the data, and the test dataset comprised 25%. Results were obtained by averaging values from test samples across 4 iterations.

A random forest (RF) classifier, a supervised ML model, was used to distinguish between relevant engagement-related behaviors of older adults with and without dementia. In total, 4 RF models were trained, each with different sets of input variables. The models were trained on the parameters listed below, and the best-performing model was chosen: (1) engagement-related behavior features; (2) engagement-related behaviors features and personal features; (3) engagement-related behaviors features, environmental disturbance features, and technical issues features; (4) engagement-related behavior features, environmental disturbance features, technical issues features, and personal features.

Precision, recall, and accuracy (F_1) scores with 95% confidence intervals were calculated and compared across the 4 models [20]. To evaluate the performance of each model, a receiver operating curve was generated to visualize the true positive rate against the false positive rate at various thresholds, and the area under the receiver operating curve (AUC) was computed for each model. A high AUC (ranging from 0 to 1) indicates a model with a good fit in distinguishing between conditions [21,22].

The feature importance method was used to identify older adults with dementia from those without dementia [23]. This method assigns a score to each feature, with higher scores indicating greater importance for the machine's prediction. The RF Gini index was used to calculate the decrease in node impurity, which measures how the features of a dataset should be divided into nodes in constructing a decision tree [24,25]. The relevant features with a score of at least 0.05 in a decrease in mean impurity were selected, as these relevant features are important for identifying older adults with or without dementia.

Results

Demographics

Older adults without dementia were, on average, 3.1 years older than participants with dementia. The ratio of women to men was equal in both groups. A high school diploma was the most common highest level of education for older adults in both

XSL•FO
RenderX

groups (5/15, 33%). Computers were the most used information and communication technologies by older adults without dementia (6/10, 60%). Additionally, both groups had similar previous experiences with serious gaming (8/15, 60%), as shown in Table 1.

Table . Participant demographics. Participants living with dementia were on average 76.6 (SD 6.8) years old. Participants without dementia were on average 79.7 (SD 8.8) years old. The mean Montreal Cognitive Assessment (MoCA) [12] was 21.75 (SD 7.48); mean MoCA for older adults living with dementia was 13.60 (SD 7.82); mean MoCA older adults without dementia was 25.45 (SD 3.36).

Variable		Dementia (n=5), n (%)	Without dementia (n=10), n (%)
Sex			
	Female	3 (60)	6 (60)
	Male	2 (40)	4 (40)
Highest educational level			
	Elementary school	0 (0)	1 (10)
	High school diploma	2 (40)	3 (30)
	Trade or vocational training	1 (20)	0 (0)
	College diploma	1 (20)	2 (20)
	Bachelor's degree	1 (20)	3 (30)
	Master's degree	0 (0)	1 (10)
Technology literacy (ICT[a] used)			
	Computer	2 (40)	6 (60)
	Computer and tablet	2 (40)	0 (0)
	Computer and smartphone	0 (0)	1 (10)
	Computer, tablet, and smartphone	1 (20)	3 (30)
Previous experience with serious games		3 (60)	6 (60)

[a]ICT: information and communication technology.

Descriptive Analysis of Engagement-Related Behaviors in Older Adults

The descriptive analysis revealed that the most common behaviors, occurring over 70% of the time in both groups, were related to gaze (ie, gaze toward game) (1767/1774, 99.6%), eyes (ie, scanning behavior) (1760/1774, 99.2%), head (ie, scanning behavior) (1765/1774, 99.5%), face (ie, neutral expression) (1603/1774, 90.4%), gameplay (ie, physical interaction with screen) (1755/1774, 99%), and concentration (ie, not distracted by external stimuli) (1611/1774, 90.8%).

Notably, older adults without dementia exhibited lean-forward behaviors and eyebrow movement more frequently compared to those with dementia. In contrast, older adults with dementia demonstrated higher frequencies of upright posture, voiced utterance, and playing while doing something else compared to those without dementia.

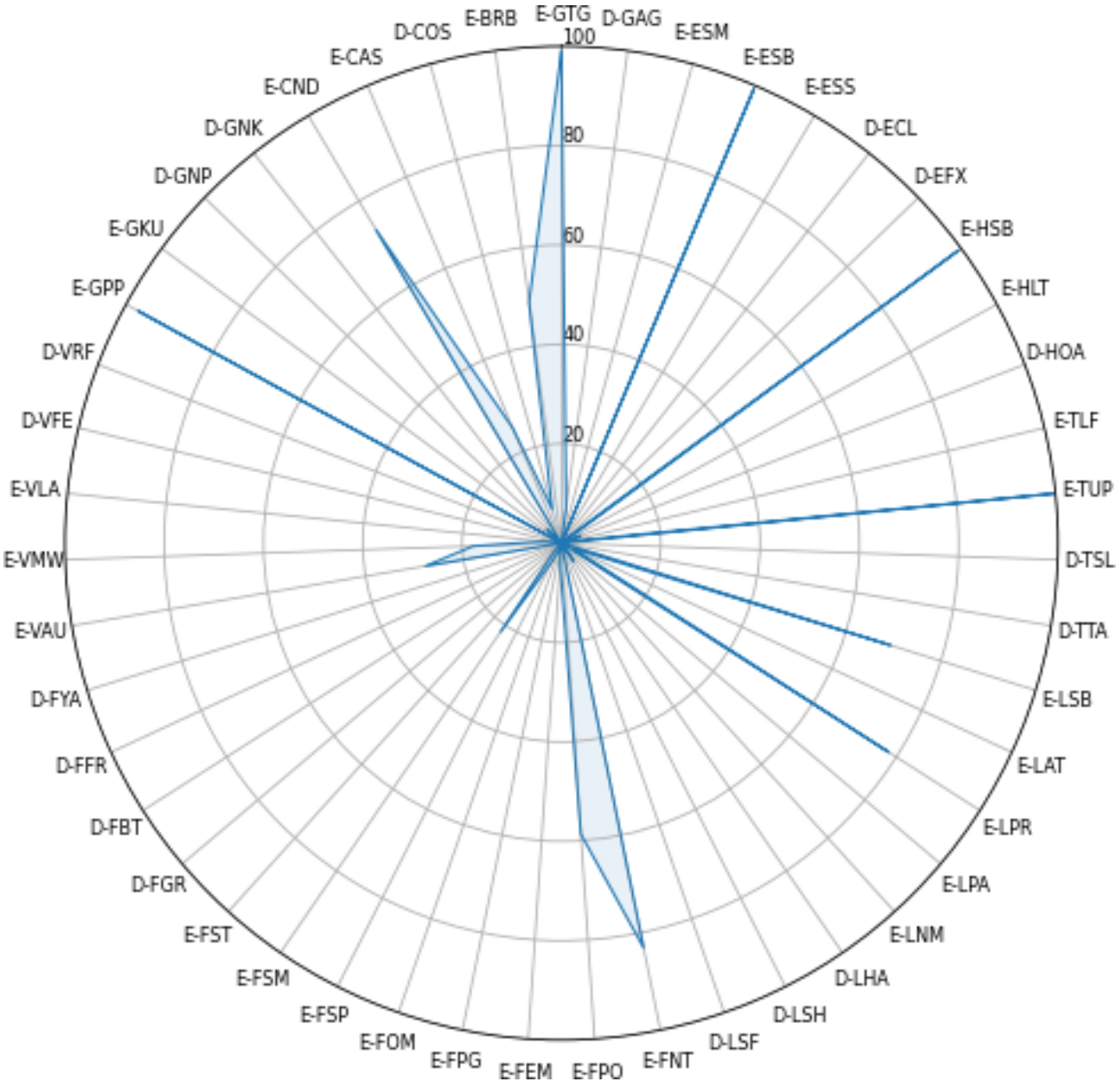
Certain behaviors had zero frequencies in older adults with dementia. Examples are engagement behaviors such as saccadic eye movements and keeping up with the game as well as disengagement behaviors such as slouched posture or yawning.

Differences in Behavior Frequencies Across Dementia Status

A comparison of the frequencies of the 47 engagement-related behaviors, along with their respective statistics and *P* values, is presented in Table S1 in Multimedia Appendix 1. A total of 30 out of 47 (64%) engagement-related behaviors showed statistically significant differences in frequencies between older adults living with dementia and those without dementia. Among the different behavioral categories, the face category exhibited the highest number of behaviors, with 8 out of 12 (66.6%), showing significant differences between the 2 groups. Figures 2 and 3 display graphical representations detailing the variance in behavioral frequencies among older adults in each group, depicted as distinctive behavioral fingerprints.

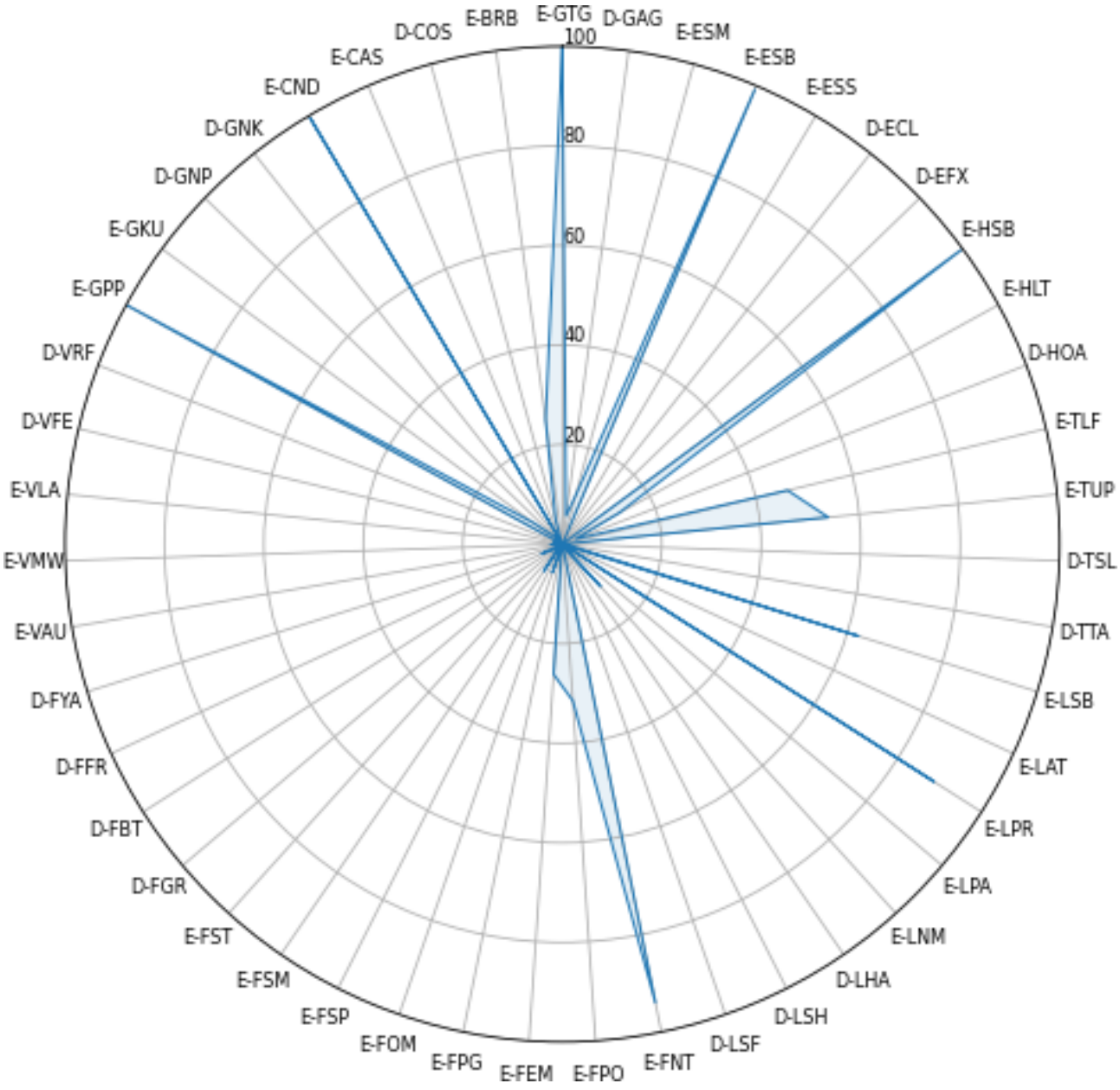
XSL•FO
RenderX

Figure 2. Radar plot of the behavioral frequencies of older adults with dementia. D-COS: stopping play to attend to another stimulus; D-ECL: closed eyes; D-EFX: eyes fixed on point; D-FBT: angry expression; D-FFR: frowning; D-FGR: pained expression; D-FYA: yawning; D-GAG: gaze away from the game; D-GNK: not keeping up with the game; D-GNP: not physically interacting with the screen; D-HOA: head oriented away from the display; D-LHA: play hand away from device; D-LSH: play hand movements unrelated to the game; D-LSF: smacking face; D-TSL: slouched posture; D-TTA: turned away from the game; D-VFE: frustrated exclamation; D-VRF: refusal; E-BRB: rhythmic breathing; E-CAS: playing while doing something else; E-CND: not distracted by external stimuli; E-ESB: scanning behaviors; E-ESM: saccadic eye movements; E-ESS: squinting at screen; E-FEM: eyebrow movement; E-FNT: neutral expression; E-FOM: open mouth; E-FPG: playful grimace; E-FPO: lip behavior; E-FSM: smiling; E-FSP: surprised expression; E-FST: tongue behavior; E-GKU: keeping up with the game; E-GPP: physical interaction with the screen; E-GTG: gaze toward game; E-HLT: head learning toward game; E-HSB: scanning behavior; E-LAT: adjust the position of the tablet; E-LNM: non-play hand at mouth; E-LPA: play arm position adjustment; E-LPR: play hand ready; E-LSB: limbs scanning behavior; E-TLF: leaning forward; E-TUP: upright posture; E-VAU: voiced utterance; E-VLA: laughing; E-VMW: voiceless utterance.



XSL•FO
RenderX

Figure 3. Radar plot of the behavioral frequencies of older adults without dementia. D-COS: stopping play to attend to another stimulus; D-ECL: closed eyes; D-EFX: eyes fixed on point; D-FBT: angry expression; D-FFR: frowning; D-FGR: pained expression; D-FYA: yawning; D-GAG: gaze away from the game; D-GNK: not keeping up with the game; D-GNP: not physically interacting with the screen; D-HOA: head oriented away from the display; D-LHA: play hand away from device; D-LSH: play hand movements unrelated to the game; D-LSF: smacking face; D-TSL: slouched posture; D-TTA: turned away from the game; D-VFE: frustrated exclamation; D-VRF: refusal; E-BRB: rhythmic breathing; E-CAS: playing while doing something else; E-CND: not distracted by external stimuli; E-ESB: scanning behaviors; E-ESM: saccadic eye movements; E-ESS: squinting at screen; E-FEM: eyebrow movement; E-FNT: neutral expression; E-FOM: open mouth; E-FPG: playful grimace; E-FPO: lip behavior; E-FSM: smiling; E-FSP: surprised expression; E-FST: tongue behavior; E-GKU: keeping up with the game; E-GPP: physical interaction with the screen; E-GTG: gaze toward game; E-HLT: head learning toward game; E-HSB: scanning behavior; E-LAT: adjust the position of the tablet; E-LNM: non-play hand at mouth; E-LPA: play arm position adjustment; E-LPR: play hand ready; E-LSB: limbs scanning behavior; E-TLF: leaning forward; E-TUP: upright posture; E-VAU: voiced utterance; E-VLA: laughing; E-VMW: voiceless utterance.



Random Forest Models

In total, 4 different RF models were trained and evaluated using F_1-scores to identify the best model. The higher F_1-score indicated a better balance between precision and recall, and the higher AUC indicated superior performance in classifying between the 2 groups. Table 2 highlights the trained model performances by variable. Model 4 achieved the highest F_1-score (91%) and AUC (99%) among all the models, making it the best suitable model (see Figure 4).

XSL•FO
RenderX

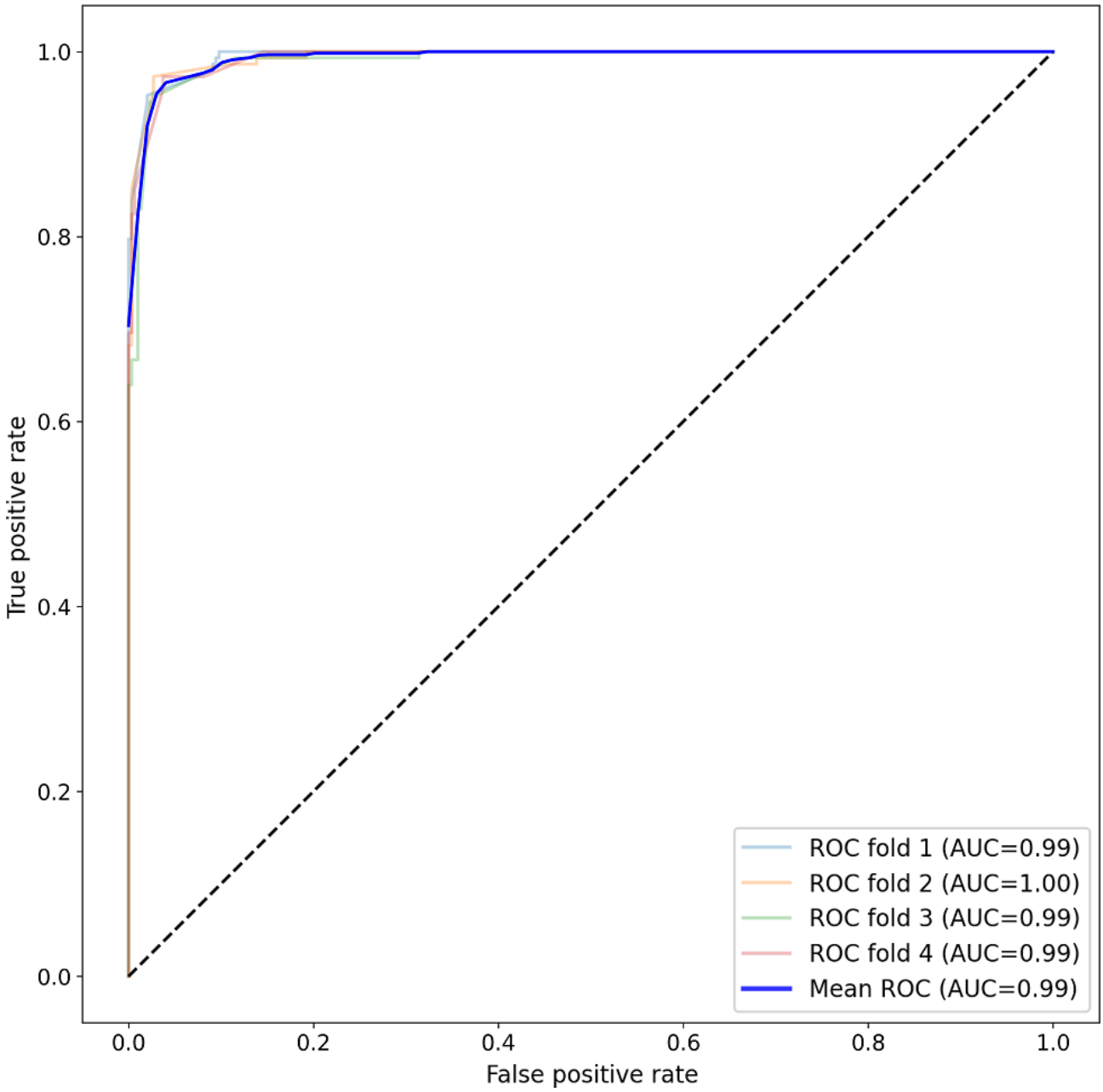
Table . Model performance by input variables.

Model	Input variables	Precision (95% CI)	Recall (95% CI)	F_1-score (95% CI)	AUC[a] (95% CI)
1	Engagement-related behaviors	0.69 (0.654 - 0.715)	0.93 (0.843 - 1.000)	0.78 (0.770 - 0.813)	0.96 (0.939 - 0.970)
2	Engagement-related behaviors and personal features	0.81 (0.780 - 0.839)	0.98 (0.958 - 1.000)	0.88 (0.872 - 0.902)	0.99 (0.984 - 1.000)
3	Engagement-related behaviors, environmental disturbance, and technical issues	0.68 (0.634 - 0.730)	0.94 (0.875 - 1.000)	0.79 (0.774 - 0.815)	0.96 (0.949 - 0.965)
4	Engagement-related behaviors, environmental disturbance, technical issues, and personal features	0.85 (0.724 - 0.990)	0.97 (0.930 - 1.000)	0.91 (0.851 - 0.963)	0.99 (0.984 - 1.000)

[a]AUC: area under the receiver operating curve.

XSL•FO
RenderX

Figure 4. The AUC for engagement-related behaviors, personal features, environmental disturbance, and technical issues. ROC: receiver operating curve; AUC: area under the receiver operating curve.

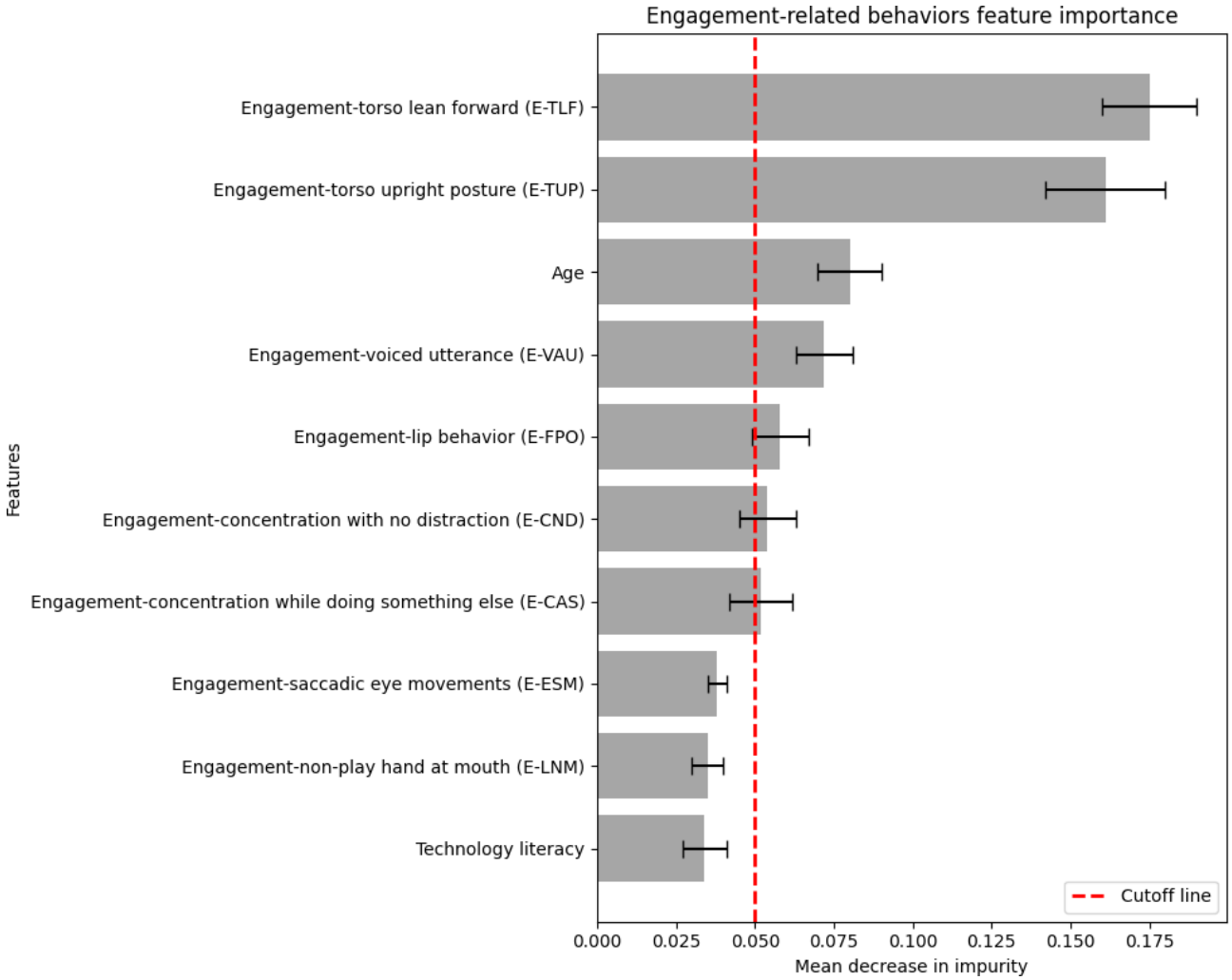


Importance of Features in Identifying Older Adults With Dementia

Ranked importance of features in identifying older adults with dementia was quantified by impurity (disorder within the data). The measure of the mean decrease in impurity, visualized in Figure 5, demonstrated that behaviors with higher impurity were considered more relevant, while behaviors with decreased impurity were deemed less important or irrelevant to older adults with dementia. The engagement-related behaviors of torso lean forward and torso upright posture were found to be the most relevant in distinguishing between older adults with and without dementia. Age was also determined as a relevant factor. The feature importance analysis highlights key behavioral indicators that differentiate older adults with and without dementia. Six primary behavioral categories—Face, Concentration, Torso, Voice, Eyes, and Limbs—were identified, each containing specific behaviors relevant to distinguishing between the two groups. Certain features, such as lip behavior (Face), playing while doing something else (Concentration), upright posture and E-TLF (Torso), and voiced utterance (Voice), were more commonly associated with older adults living with dementia. In contrast, features like not distracted by external stimuli (Concentration), and leaning forward (Torso), saccadic eye movements (Eyes), and non-play hand at mouth (Limbs) were primarily observed in older adults without dementia..

XSL•FO
RenderX

Figure 5. A bar graph representing the feature importance by mean decrease in data impurity and their confidence intervals. Behaviors with higher impurity were considered more relevant.



Discussion

Overview

This study explored the use of engagement-related behaviors demonstrated by older adults while playing serious computer games to provide valuable information about the presence of dementia. Our results support hypothesis 1, as 64% of engagement-related behaviors were significantly different between older adults with and without dementia. Using ML, key features distinguishing between older adults with and without dementia during serious gameplay include behaviors in categories of torso, voice, facial expression, concentration, and age, supporting hypothesis 2.

Serious Gaming, Engagement-Related Patterns, and Dementia Differentiation

In our study, we observed engagement-related behaviors among older adults in behaviors such as directing attention toward the game, head scanning behavior or scanning the screen. These findings suggest that certain elements of engagement in gameplay are universal, regardless of an individual's cognitive status. Similarly, Perugia et al [16] identified head movements as the most prominent engagement-related behavior in older adults, followed by torso movement, arm or hand movement, and gaze directed toward the activity when older adults with dementia interacted with a robot and played a board game.

The behaviors that were significantly different in frequency between the 2 groups, were also found to be important features to distinguish older adults with dementia from those without dementia. These behaviors are leaning forward torso behavior, upright torso behavior, voice utterances, lip behavior, not being distracted by external stimuli, and playing while doing something else. The present study observed that older adults with dementia demonstrated a higher percentage of upright torso behavior compared to older adults without dementia. Conversely, older adults without dementia exhibited a higher percentage of leaning forward torso behavior compared to those with dementia. Perugia et al [16] found that leaning forward was a common behavior in older adults with dementia. This discrepancy between Perugia et al's [16] results and ours could be attributed to the use of different tasks. While Perugia et al [16] used board games that typically necessitate leaning forward, our study employed tablet-based games, allowing for a range of postures facilitated by tablet stands and varied grasps and angles of holding a tablet. It is known that the environment (eg, use of tablet stands) can influence engagement-related behaviors in older adults with dementia [16,26]. In our study, 80% of our participants in each group (ie, 4 out of 5 older adults living with

XSL•FO
RenderX

dementia and 8 out of 10 older adults without dementia) decided to use tablet stands to angle the tablet screen to improve the view of the game. Therefore, in the present study, the environmental condition seems not to have impacted the torso behaviors of older adults, as they equally preferred to use the tablet stands to angle the tablet in the 2 groups. Another explanation could be that people living with dementia may experience increase anxiety when leaving home [27]. This could explain that our participants living with dementia felt over-attentive or less comfortable with the environment than older adults without dementia, resulting in sitting upright as a predominant behavior in participants with dementia.

Approximately one-third of the time intervals involved older adults with dementia exhibiting voiced utterances, whereas this finding was observed in less than 1% of the time intervals for older adults without dementia. Similarly, the results showed that older adults living with dementia had nearly twice the frequencies in lip behavior than older adults without dementia, suggesting that lip behaviors might be more common in older adults with dementia. These utterances, such as “oh,” “oops,” “um,” etc, and lip behavior (eg, mouthing words), may stem from changes in speech-language processes associated with dementia, impacting verbal communication [28,29].

The study revealed that engaging in concurrent activities while playing was more prevalent in older adults living with dementia compared to those without dementia. This finding suggests that individuals with dementia may be prone to distraction during gameplay, which can be attributed to cognitive decline, reduced ability to sustain focus, and impaired capacity to disregard distractors [26,30]. These findings emphasize the significance of considering the impact of external stimuli on engagement-related behaviors in individuals with dementia and underscore the necessity for tailored interventions that address attentional challenges.

The study revealed the absence of certain engagement-related behaviors in older adults with dementia compared to cognitively intact older adults, such as saccadic eye movements, play arm position adjustment, and keeping up with the game. Saccadic eye movements, rapid changes in gaze directions, appear to decline and become slower with cognitive impairment and dementia [31,32]. The lack of arm adjustment may reflect both the cognitive and perceptual-motor changes commonly associated with dementia, which can affect visuospatial skills and the ability to adjust body movements [33,34]. The inability to keep up with the game might signal a reduction in processing speed, a common cognitive deficit in dementia [35] in line with results from a separate serious game intervention for older adults with dementia by Tziraki et al [36] , as well as changes in vigilance, sustained attention, and fatigue [37].

Classification Model of Dementia Status Using Serious-Gaming Engagement in Older Adults

This research further examined the potential of using ML to differentiate older adults with and without dementia based on their engagement-related behaviors while playing serious computer games. Model 4 emerged as the most effective to distinguish the presence of dementia based on engagement-related behaviors alongside personal, technical, and environmental factors. The identified engagement-related behaviors may aid in recognizing early signs of dementia, complementing standardized cognitive screening tools and information provided by older adults and their families.

Our study achieved similar accuracies to other studies that used ML to distinguish older adults with and without dementia [5,6]. The comparable accuracies achieved in existing ML studies suggest that playing mobile games can serve as a safe and cost-effective means of identifying older adults with and without dementia based on their engagement-related behaviors. This approach is less complex, risky, and costly compared to other approaches, such as using driving-related assessments, making it more accessible for older adults and their caregivers, particularly in rural or low-income settings.

Limitations

This proof-of-concept study has several limitations that warrant consideration and guide future research. A primary limitation is the small sample size, which affects the accuracy and generalizability of the ML model’s predictions for older adults living with dementia. ML models typically require large and diverse datasets to achieve robust performance [38]. Future studies should aim to recruit larger and more diverse participant samples to enhance generalizability. Additionally, the potential for sample bias cannot be overlooked. Engagement behaviors are highly context-dependent and influenced by environmental factors, yet this study was conducted in only 2 settings. To address this, a multicenter approach, involving participants from various demographics, cultural backgrounds, and care environments, would provide a more comprehensive understanding of engagement-related behaviors and improve the applicability of the findings.

Despite the rigorous development of the coding system used to generate the dataset of engagement-related behaviors, some limitations persist. Strategies to minimize bias included a systematic 3-stage approach, rater training in engagement theory, and the use of independent raters. However, biases inherent to human judgment and the specific contexts of this study may still affect the coding system’s reliability and generalizability. To strengthen its validity, future research should test the coding system with different populations, raters, environments, and games. External validation across multiple research teams would further enhance its reliability and reproducibility.

Furthermore, the study highlights the need to explore engagement behaviors in diverse settings, such as home environments, community centers, and clinical care facilities. Investigating how context influences engagement will help refine ML models to account for environmental variability, making them more robust and broadly applicable.

Finally, while this study supports the potential of digital assessments, further work is needed to validate and standardize these tools. Ensuring their reliability, validity, and clinical utility for detecting cognitive decline in older adults remains a crucial step for their adoption in practice [3,39].

XSL•FO
RenderX

Conclusions

Key features distinguishing between older adults with and without dementia during serious gameplay included torso, voice, facial, and concentration behaviors, as well as age. The best-performing ML model identified included features of engagement-related behaviors, environmental disturbances, technical challenges, and personal attributes. These findings have important implications for dementia research, suggesting opportunities for the ongoing development of engagement-targeted assessments and therapeutic interventions through the integration of serious gaming and ML.

Acknowledgments

This research was funded by Aging Gracefully across Environments using Technology to Support Wellness Networks of Centers of Excellence (AGE-WELL), Canada's technology and aging network, under grant AW-PP2019-PP4. The opinions expressed in this publication are those of the authors and do not necessarily reflect those of AGE-WELL.The authors attest that there was no use of generative artificial intelligence technology in generation of text, figures, or other information content in this manuscript.

Data Availability

The datasets generated during or analyzed during this study are not publicly available due to limitations imposed by the study's ethical approval but are available from the corresponding author on reasonable request.

Authors' Contributions

MT was involved in the investigation, formal analysis, methodology, and writing – original draft; MF was involved in the writing – review & editing; SE was involved in the supervision and writing – review & editing; ES was involved in the funding acquisition, supervision, and writing – review & editing; AMRR was involved in the conceptualization, funding acquisition, supervision, and writing – review & editing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary Table 1: χ^2 and Fisher exact tests.
[DOCX File, 70 KB - games_v13i1e54797_app1.docx]

References

1. World Health Organisation. Global action plan on the public health response to dementia 2017 - 2025. : Geneva: World Health Organization; 2017.
2. Xu J, Zhang Y, Qiu C, Cheng F. Global and regional economic costs of dementia: a systematic review. The Lancet 2017 Dec;390:S47. [doi: 10.1016/S0140-6736(17)33185-9]
3. Ford E, Milne R, Curlewis K. Ethical issues when using digital biomarkers and artificial intelligence for the early detection of dementia. Wiley Interdiscip Rev Data Min Knowl Discov 2023;13(3):e1492. [doi: 10.1002/widm.1492] [Medline: 38439952]
4. Tolks D, Schmidt JJ, Kuhn S. The Role of AI in Serious Games and Gamification for Health: Scoping Review. JMIR Serious Games 2024 Jan 15;12:e48258. [doi: 10.2196/48258] [Medline: 38224472]
5. Bayat S, Babulal GM, Schindler SE, et al. GPS driving: a digital biomarker for preclinical Alzheimer disease. Alzheimers Res Ther 2021 Jun 14;13(1):115. [doi: 10.1186/s13195-021-00852-1] [Medline: 34127064]
6. Padhee S, Illendula A, Sadler M, et al. Predicting early indicators of cognitive decline from verbal utterances. arXiv. Preprint posted online on Nov 19, 2020. [doi: 10.48550/arXiv.2012.02029]
7. Mezrar S, Bendella F. Machine learning and Serious Game for the Early Diagnosis of Alzheimer's Disease. Simul Gaming 2022 Aug;53(4):369-387. [doi: 10.1177/10468781221106850]
8. Boletsis C, McCallum S. Smartkuber: A Serious Game for Cognitive Health Screening of Elderly Players. Games Health J 2016 Aug;5(4):241-251. [doi: 10.1089/g4h.2015.0107] [Medline: 27192473]
9. Karapapas C, Goumopoulos C. Mild Cognitive Impairment Detection Using Machine Learning Models Trained on Data Collected from Serious Games. Appl Sci (Basel) 2021;11(17):8184. [doi: 10.3390/app11178184]
10. Hookham G, Nesbitt K. A systematic review of the definition and measurement of engagement in serious games. Presented at: ACSW 2019; Jan 29-31, 2019; Sydney, Australia. [doi: 10.1145/3290688.3290747]
11. Mills C, D'mello S, Lehman B, Bosch N, Strain A, Graesser A. What makes learning fun? Exploring the influence of choice and difficulty on mind wandering and engagement during learning. Presented at: 16th International Conference on Artificial Intelligence in Education, AIED 2013; Jul 9-13, 2013; Memphis, TN, United States. [doi: 10.1007/978-3-642-39112-5_8]

XSL•FO
RenderX

12. Ríos Rincón AM, Daum C, Miguel Cruz A, Liu L, Stroulia E. Feasibility and Acceptability of a Serious Mobile-Game Intervention for Older Adults. Phys Occup Ther Geriatr 2022 Jul 3;40(3):295-318. [doi: 10.1080/02703181.2022.2030849]
13. University of Alberta. VibrantMinds. 2024. URL: https://vibrant-minds.org/vibrantminds2/start [accessed 2025-01-06]
14. Nasreddine ZS, Phillips NA, Bédirian V, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. J Am Geriatr Soc 2005 Apr;53(4):695-699. [doi: 10.1111/j.1532-5415.2005.53221.x] [Medline: 15817019]
15. Rios Rincon AM, Liu L, Daum C, Comeau A, Rincon Martinez D, Catalina L. Engagement is older adults during gameplay: an ethogram. Presented at: CAOT Conference; May 6-9, 2020.
16. Perugia G, van Berkel R, Díaz-Boladeras M, Català-Mallofré A, Rauterberg M, Barakova E. Understanding Engagement in Dementia Through Behavior. The Ethographic and Laban-Inspired Coding System of Engagement (ELICSE) and the Evidence-Based Model of Engagement-Related Behavior (EMODEB). Front Psychol 2018;9:690. [doi: 10.3389/fpsyg.2018.00690] [Medline: 29881360]
17. Portney L, Watkins M. Foundations of Clinical Research: Applications to Practice, 3rd edition: Pearson/Prentice Hall; 2008.
18. Mate Y, Potdar A, Priya R. Ensemble methods with bidirectional feature elimination for prediction and analysis of employee attrition rate during COVID-19 pandemic. Presented at: ICRTAC-AIT 2020; Dec 17-18, 2020; Chennai, India. [doi: 10.1007/978-981-16-6448-9_11]
19. James D, Vimina ER. Machine learning-based early diabetes prediction. In: Intelligent Sustainable Systems: Springer; 2022:661-678. [doi: 10.1007/978-981-16-2422-3_52]
20. Priya A, Garg S, Tigga NP. Predicting Anxiety, Depression and Stress in Modern Life using Machine Learning Algorithms. Procedia Comput Sci 2020;167:1258-1267. [doi: 10.1016/j.procs.2020.03.442]
21. Arief Subchan M, Andayani NN. Classification of maize genotype using logistic regression. IOP Conf Ser: Earth Environ Sci 2021 Nov 1;911(1):012017. [doi: 10.1088/1755-1315/911/1/012017]
22. Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. J Thorac Oncol 2010 Sep;5(9):1315-1316. [doi: 10.1097/JTO.0b013e3181ec173d] [Medline: 20736804]
23. Artasanchez A, Joshi P. Artificial Intelligence With Python, 2nd edition: Packt Publishing; 2020.
24. Ronaghan S. The mathematics of decision trees, random forest and feature importance in Scikit-learn and Spark. Medium. 2018. URL: https://medium.com/towards-data-science/the-mathematics-of-decision-trees-random-forest-and-feature-importance-in-scikit-learn-and-spark-f2861df67e3 [accessed 2025-02-06]
25. Kaur M. An approach for sentiment analysis using gini index with random forest classification. . 2020 p. 541-554. [doi: 10.1007/978-3-030-37218-7_62]
26. Hamdy RC, Kinser A, Depelteau A, Kendall-Wilson T, Lewis JV, Whalen K. Patients with Dementia Are Easily Distracted. Gerontol Geriatr Med 2017;3. [doi: 10.1177/2333721417735938] [Medline: 29242811]
27. Brittain K, Corner L, Robinson L, Bond J. Ageing in place and technologies of place: the lived experience of people with dementia in changing social, physical and technological environments. Sociol Health Illn 2010 Feb 1;32(2):272-287. [doi: 10.1111/j.1467-9566.2009.01203.x] [Medline: 20003041]
28. Hubbard G, Downs MG, Tester S. Including older people with dementia in research: challenges and strategies. Aging Ment Health 2003 Sep;7(5):351-362. [doi: 10.1080/1360786031000150685] [Medline: 12959804]
29. Martínez-Sánchez F, Meilán JJG, Carro J, Ivanova O. A Prototype for the Voice Analysis Diagnosis of Alzheimer's Disease. J Alzheimers Dis 2018;64(2):473-481. [doi: 10.3233/JAD-180037] [Medline: 29914025]
30. Wargnier P, Malaise A, Jacquemot J, et al. Towards attention monitoring of older adults with cognitive impairment during interaction with an embodied conversational agent. Presented at: 2015 3rd IEEE VR International Workshop on Virtual and Augmented Assistive Technology (VAAT); Mar 23, 2015; Arles, France. [doi: 10.1109/VAAT.2015.7155406]
31. Molitor RJ, Ko PC, Ally BA. Eye movements in Alzheimer's disease. J Alzheimers Dis 2015;44(1):1-12. [doi: 10.3233/JAD-141173] [Medline: 25182738]
32. Zeeman M, Figeys M, Brimmo T, Burnstad C, Hao J, Kim ES. Task-Evoked Pupillary Response as A Potential Biomarker of Dementia and Mild Cognitive Impairment: A Scoping Review. Am J Alzheimers Dis Other Demen 2023;38:15333175231160010. [doi: 10.1177/15333175231160010] [Medline: 36896819]
33. Cipriani G, Lucetti C, Danti S, Nuti A. Apathy and dementia. Nosology, assessment and management. J Nerv Ment Dis 2014 Oct;202(10):718-724. [doi: 10.1097/NMD.0000000000000190] [Medline: 25265266]
34. Duong S, Patel T, Chang F. Dementia: What pharmacists need to know. Can Pharm J (Ott) 2017;150(2):118-129. [doi: 10.1177/1715163517690745] [Medline: 28405256]
35. Salthouse TA. Aging and measures of processing speed. Biol Psychol 2000 Oct;54(1-3):35-54. [doi: 10.1016/s0301-0511(00)00052-1] [Medline: 11035219]
36. Tziraki C, Berenbaum R, Gross D, Abikhzer J, Ben-David BM. Designing Serious Computer Games for People With Moderate and Advanced Dementia: Interdisciplinary Theory-Driven Pilot Study. JMIR Serious Games 2017 Jul 31;5(3):e16. [doi: 10.2196/games.6514] [Medline: 28760730]

XSL•FO
RenderX

37. Berardi AM, Parasuraman R, Haxby JV. Sustained attention in mild Alzheimer's disease. Dev Neuropsychol 2005;28(1):507-537. [doi: 10.1207/s15326942dn2801_4] [Medline: 15992254]
38. Qiu J, Wu Q, Ding G, Xu Y, Feng S. A survey of machine learning for big data processing. EURASIP J Adv Signal Process 2016 Dec;2016(1):67. [doi: 10.1186/s13634-016-0355-x] [Medline: 27594883]
39. Abd-Alrazaq A, Abuelezz I, AlSaad R, et al. Serious Games for Learning Among Older Adults With Cognitive Impairment: Systematic Review and Meta-analysis. J Med Internet Res 2023 Apr 12;25:e43607. [doi: 10.2196/43607] [Medline: 37043277]

Abbreviations:

AD: Alzheimer disease
AUC: area under the receiver operating curve
ML: machine learning
RF: random forest

Edited by N Cahill; submitted 22.11.23; peer-reviewed by J Brown, T Huang; revised version received 08.12.24; accepted 11.12.24; published 03.03.25.

Please cite as:
Torabgar M, Figeys M, Esmail S, Stroulia E, Ríos Rincón AM
Machine Learning Analysis of Engagement Behaviors in Older Adults With Dementia Playing Mobile Games: Exploratory Study
JMIR Serious Games 2025;13:e54797
URL: https://games.jmir.org/2025/1/e54797
doi:10.2196/54797

© Melika Torabgar, Mathieu Figeys, Shaniff Esmail, Eleni Stroulia, Adriana M Ríos Rincón. Originally published in JMIR Serious Games (https://games.jmir.org), 3.3.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Digital Ergonomics of NavegApp, a Novel Serious Game for Spatial Cognition Assessment: Content Validity and Usability Study

Juan Pablo Sanchez-Escudero[1], PhD; David Aguillon[2], MD, PhD; Stella Valencia[3], MSc; Mauricio A Garcia-Barrera[4], PhD; Daniel Camilo Aguirre-Acevedo[5], PhD; Natalia Trujillo[3,6,7], PhD

[1]Group of Epidemiology, Universidad de Antioquia, Medellín, Colombia

[2]Grupo de Neurociencias de Antioquia, University of Antioquia, Medellín, Colombia

[3]Mental Health Research Group, Universidad de Antioquia, Medellín, Colombia

[4]Department of Psychology & Institute on Aging and Lifelong Health, University of Victoria, Victoria, BC, Canada

[5]School of medicine, Universidad de Antioquia, Medellín, Colombia

[6]Stempel College of Public Health and Social Work, Florida International University, Miami, FL, United States

[7]Global Brain Health Institute, University of California, San Francisco, CA, United States

Corresponding Author:
Juan Pablo Sanchez-Escudero, PhD
Group of Epidemiology
Universidad de Antioquia
Calle 70 No. 52-21
Medellín
Colombia
Phone: 57 6042198332
Email: juanp.sanchez@udea.edu.co

Abstract

Background: Alzheimer disease (AD) is the leading cause of dementia worldwide. With aging populations and limited access to effective treatments, there is an urgent need for innovative markers to support timely preventive interventions. Emerging evidence highlights spatial cognition (SC) as a valuable source of cognitive markers for AD. This study presents NavegApp, a serious game (SG) designed to assess 3 key components of SC, which show potential as cognitive markers for the early detection of AD.

Objective: This study aimed to determine the content validity and usability perception of NavegApp across multiple groups of interest.

Methods: A multistep process integrating methodologies from software engineering, psychometrics, and health measurement was implemented to validate the software. Our approach was structured into 3 stages, guided by the software life cycle for health and the Consensus-Based Standards for the Selection of Health Status Measurement Instruments (COSMIN) recommendations for evaluating the psychometric quality of health instruments. To assess content validity, a panel of 8 experts evaluated the relevance and representativeness of tasks included in the app. In addition, 212 participants, categorized into 5 groups based on their clinical status and risk level for AD, were recruited to evaluate the app's digital ergonomics and usability at various stages of development. Complementary analyses were performed to identify group differences and to explore the association between task difficulty and user agreeableness.

Results: NavegApp was validated as a highly usable tool by both experts and users. The expert panel confirmed that the tasks included in the game were representative (Aiken V=0.96-1.00) and relevant (Aiken V=0.96-1.00) for measuring SC components. Both experts and nonexperts rated NavegApp's digital ergonomics positively, with minimal differences between groups (r_{rb} 0.08-0.29). Differences in usability perceptions were observed among participants with sporadic mild cognitive impairment compared to cognitively healthy individuals (r_{rb} 0.26-0.29). A moderate association was also identified between task difficulty and user agreeableness (Cramér V=0.37, 95% CI 0.28-0.54).

Conclusions: NavegApp is a valid and user-friendly SG designed for SC assessment, developed by integrating software engineering and psychometric evaluation methodologies. While the results are promising, further studies are warranted to evaluate

XSL•FO
RenderX

its diagnostic accuracy and construct validity. This work outlines a comprehensive framework for SG development in cognitive assessment, emphasizing the importance of incorporating psychometric validity measures from the outset of the design process.

(JMIR Serious Games 2025;13:e66167) doi:10.2196/66167

KEYWORDS

serious games; spatial cognition; digital neuropsychology; Alzheimer disease; content validity; usability

Introduction

Background

Alzheimer disease (AD) is the leading cause of dementia, accounting for 60% to 80% of cases globally [1]. The rising number of AD cases represents a growing economic and social burden, challenging global health care and social systems [1]. Due to the absence of effective pharmacological treatments, preventive strategies targeting the preclinical stages of the disease have become the forefront approach to delaying the onset of symptoms [2]. Currently, the detection of preclinical AD relies on biomarkers obtained through techniques such as lumbar puncture and positron emission tomography [3]. Despite its accuracy, these methods describe limited scalability and utility in longitudinal population studies due to the associated costs, invasiveness, and potential side effects [4,5]. This is particularly problematic in low- and middle-income countries, where access to expensive diagnostic technologies is limited. Consequently, there is an urgent need for innovative, scalable, and noninvasive markers capable of detecting early cognitive or behavioral changes indicative of AD neuropathology [5,6].

Over the last decade, advances in cognitive neuroscience and software development have led to the emergence of multiple digital tools focused on detecting the earliest cognitive signatures of AD [7]. In this regard, recent evidence indicates that alterations in spatial cognition (SC) processes are present since the preclinical stage of AD. These deficits serve as promising indicators of the prodromal stages of AD, distinguishing affected individuals from healthy controls and marking disease progression [7]. Particularly, tests of spatial navigation applied using immersive virtual reality (VR) interfaces have been demonstrated to be sensitive and specific to the prodromic stage of the disease [7,8].

Although VR-based tests hold major potential, several limitations hinder their widespread adoption in clinical and primary care settings. Key barriers include the high costs of VR devices, the need for highly trained personnel to configure and operate these systems, and logistical challenges such as the time and physical space required for evaluations. For instance, many VR-based tasks necessitate the examinee walking several meters, which can be impractical in clinical environments [7]. The complex interfaces and reliance on unfamiliar control peripherals also present accessibility issues, particularly for older adults who may lack previous experience with such technologies. Together, these factors limit the scalability and feasibility of VR-based assessments for widespread, population-level implementation [4].

In response to these limitations, we have developed NavegApp, a serious game (SG)–based platform encompassing gamified tasks for assessing SC components [9]. This platform was developed by integrating game design principles and measurement properties analyses into the software development life cycle (SDLC), aiming to simplify digital interfaces while preserving adequate psychometric performance for the neuropsychological assessment of individuals at risk of developing neurodegenerative conditions, such as AD. The versatility and compatibility of NavegApp with various devices make it suitable for integration into different settings beyond research, such as primary health care services, nursing homes, or remote assessments. Such flexibility is essential in low- and middle-income countries, where access to specialized cognitive assessment or follow-up services may be limited [10].

Despite the foreseeable advantages, the development of a valid and reliable digital tool that integrates cutting-edge cognitive neuroscience with gamified interfaces into highly usable and ergonomic apps tailored to low- and middle-income countries face several challenges [11]. First, SGs must be precisely tailored for cognitive assessment, requiring a robust theoretical foundation for designing or adapting tasks and interfaces. Second, the SDLC must incorporate psychometric analysis, necessitating a strategic integration of software engineering and health measurement validation best practices. Third, creating an SG for individuals with varying levels of functionality demands careful consideration of digital ergonomics and usability [11]. The following sections present the key considerations and insights we used to address these challenges and implement NavegApp.

SC as a Source of Cognitive Markers for AD

SC involves a range of cognitive functions, including the perception, organization, and use of location and object–based information to understand and navigate through physical or mental spaces [12-14]. When integrated, these processes facilitate complex spatial behaviors, such as solving mazes or tracing routes from point A to point B using landmarks or self-positioning as a reference [13]. Due to the complexity and extent of the brain areas supporting SC, neurodegenerative diseases, such as AD, affect this cognitive process in distinct and characteristic ways [12].

Accrued evidence suggests that subtle cognitive and behavioral deficits in SC can be detected during the preclinical and prodromal stages of AD, stemming from the effects of AD pathology in specific brain regions, such as the medial temporal lobe and parietal cortex [2,8,15-17]. People living with AD exhibit spatial impairments that intensify as AD pathology progressively spreads [12,15]. In the preclinical stage of AD, metabolic alterations and diminished connectivity in the entorhinal cortex, retrosplenial cortex, and precuneus contribute to difficulties with allocentric spatial navigation [8,15,18,19].

XSL•FO
RenderX

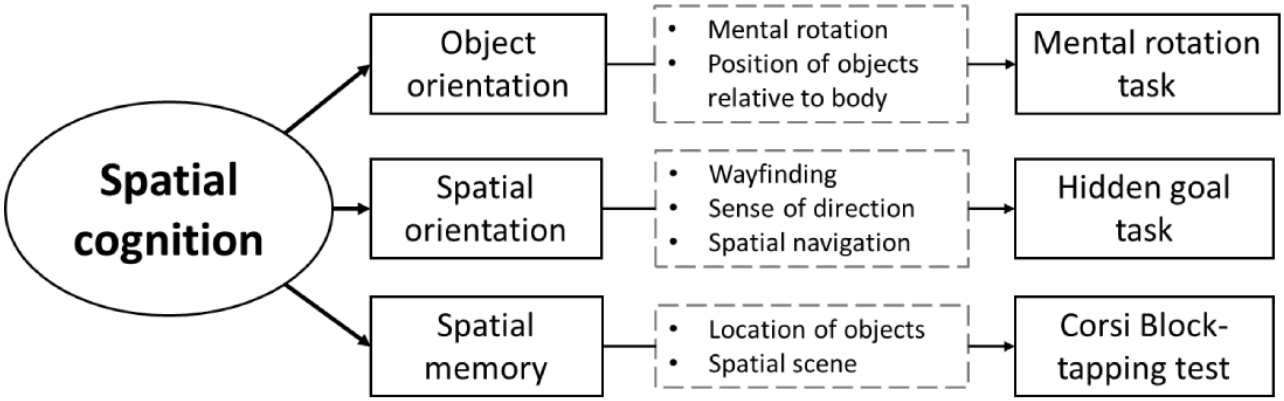
During the prodromal phase, increased neuronal loss and shrinkage of the hippocampus and inferior parietal cortex lead to impairments in path integration, route learning, egocentric navigation, and mental rotation tasks [12,20]. Furthermore, disruptions in glucose metabolism and gray matter loss, resulting from amyloid accumulation in the parietal lobe and posterior parietal cortex, affect spatial working memory and the integration of visual information [8,15].

Mentioned impairments can be detected by using different neuropsychological tasks, as has been previously outlined in the literature [15,16,21-24]. A decreased performance in allocentric spatial navigation tasks, such as the hidden goal task, has been described as an early potential cognitive marker for AD [25-27]. Additional evidence also suggests deficits in motor representation since the prodromal stage of AD, which leads to impairments in mental rotation tasks when compared to healthy, age-matched individuals [28,29]. Complementarily, individuals with mild cognitive impairment (MCI) exhibit decreased performance on the Corsi Block-Tapping test when compared to healthy control individuals, with more deficits observed in those with amnestic MCI and early AD, which suggests a decreased short-term spatial working memory [16,22-24].

To develop a valid and reliable instrument for cognitive assessment, a robust theoretical framework is essential [30]. Such a framework guides the formulation of hypotheses, interpretation of results, and design of tasks during the development of the assessment tool [31]. The SC model, as outlined by Jacobs [13], aligns with previous evidence by emphasizing the role of 3 key processes consistent with empirical evidence of SC deficits: the orientation of individual objects, as seen in mental rotation tasks; spatial orientation for self-navigation within large-scale environments, as seen in allocentric and egocentric spatial navigation tasks; and spatial memory, involving recalling information about object locations and constructing spatial scenes [13].

According to the model by Jacobs [13], these 3 components are interrelated, forming a dynamic network of cognitive abilities that can work synergistically to generate spatial representations or function independently to address specific tasks. This multifaceted nature of SC underscores that it is not a unitary construct and cannot be fully assessed through a single test [13]. Figure 1 illustrates the SC framework by Jacobs [13], and the corresponding neuropsychological tasks included in NavegApp.

Figure 1. Spatial cognition framework and experimental tasks.



Due to the alignment of the SC model by Jacobs [13] with the documented progression of SC deficits in AD, it provides a robust theoretical foundation for selecting and adapting tasks in NavegApp. By offering a comprehensive theoretical definition, the model enables the identification of key task components necessary for accurate construct measurement in psychometric terms. This approach ensures that core elements suitable for gamification are selected without compromising the measurement's validity. Furthermore, while much research has focused on spatial navigation, the model by Jacobs [13] highlights the interconnected and multidimensional nature of SC components in forming spatial representations. This perspective underscores the importance of incorporating multiple metrics to ensure the instrument's validity.

Although the use of SC measures in the detection and staging of AD holds promise, currently, there are no established protocols or tools for its application in clinical or research settings [12]. This gap presents an opportunity to develop a comprehensive evaluation system. However, creating such a tool requires carefully considering the adequacy of the SDLC to include relevant metrics for measurement purposes.

Developing SGs for Cognitive Assessment

The SGs are games developed for purposes beyond entertainment [32]. These games integrate features of computerized tests and video games to create a relaxing and engaging experience for examinees while performing cognitive tasks designed to measure specific aspects of their cognition [33,34]. In recent years, SGs have proven valuable resources for cognitive and behavioral assessment in the context of neurodegenerative diseases [7,35]. Particularly, SGs enhance cognitive assessments by promoting optimal performance through increased engagement and motivation [35]. Furthermore, they enable precise and timely data collection and management, reducing the time required to derive insights from assessments and supporting the clinical decision-making process [33]. Consequently, specific recommendations for designing and adapting SGs for cognitive assessment tailored to individuals living with neurodegenerative diseases such as AD have been suggested [11]. These recommendations emphasize the importance of ergonomic criteria, usability, and measurement properties analysis during the SG development process [7,11,33].

XSL•FO
RenderX

Digital ergonomics involves design principles that create efficient, user-friendly interfaces [9,11]. Usability refers to the perceived ease and quality of use of a system, interface, or product and is a crucial aspect of the user experience [36]. Regarding the measurement properties, ensuring content validity—accurately representing the intended construct from the early development phase—is crucial for capturing essential and relevant information on cognitive performance [7,37].

Although the potential advantages of using video games for cognitive assessment at the earliest stages of AD are noteworthy, this approach remains underexplored and constitutes only a small proportion of the available tools [7]. One critical limitation is the lack of standardized frameworks incorporating evidence–based recommendations for SG development while integrating psychometric principles into the design process.

Most existing discussions and recommendations emphasize the benefits of designing immersive interfaces [38], improving user-interface interaction [39], and enhancing the overall user experience and engagement [40]. While these aspects are undoubtedly important, they do not guarantee the validity of the metrics produced by such tools [37]. Moreover, many instruments designed to measure core components of SC, such as spatial navigation and spatial memory, fail to provide detailed descriptions of their development processes, usability metrics, or feasibility analyses in at-risk populations or individuals in the earliest stages of AD [7]. This gap raises substantially concerns about potential spectrum bias, as the validity and generalizability of these tools may be compromised when applied to diverse populations [7].

The absence of a robust framework for developing SGs specifically tailored for cognitive assessment underscores the need for a comprehensive, transparent, and literature-supported approach. Such a framework should draw on principles from both software engineering and psychometrics to ensure the reliability, validity, and usability of these tools. Thus, developing an SG for cognitive assessment must adhere to software engineering best practices and design principles, particularly the SDLC [36]. This structured process outlines the step-by-step procedures for software development, covering design, development, testing, and maintenance stages, aiming to ensure efficient software development and deliver a high-quality product [36,41]. The process begins with design and development, enabling informed decision-making on system requirements and software architecture. The testing phase aims to ensure software functionality through iterative testing to address anomalies and incorporate user-desired features. This phase also allows collecting user experience data, including usability, digital ergonomics, and critical measurement properties. Once testing and verification are complete, the software is ready for deployment. In the context of cognitive assessment, software deployment typically marks the beginning of the clinical utility validation process [37].

In addition to considering the SDLC, incorporating gamification into SGs for cognitive assessment is equally important [40]. Gamification involves integrating game elements such as badges, counters, or leaderboards into tasks to enhance user engagement without compromising the validity of the assessment [32,37]. For example, graphic elements can be adapted in SC tasks to support a narrative and increase user engagement. However, the spatial relationships among these elements must be preserved to avoid altering task protocols or introducing performance biases.

Given the potential of SC measurements for early identification of individuals at risk for neurodegenerative diseases and the current lack of valid and user-friendly tools for scalable health care cognitive assessment services [9,11,38], this study introduces NavegApp, an SG-based platform designed to assess SC components. By leveraging the SC model by Jacobs [13] and incorporating software engineering best practices, this research seeks to produce a reliable and valid assessment tool that enhances early diagnosis and supports preventive strategies in diverse populations. For this, we focus on 3 critical metrics for SG implementation: content validity, digital ergonomics, and usability perception. Furthermore, this research addresses the gap in standardized frameworks for SG development by offering a detailed account of the development process, serving as a model for future work in the field. Using an iterative SDLC approach, we integrated evidence from cognitive neuroscience and psychometrics to create a fully functional, ready-to-use SG. Thus, the overall objective of this study was to evaluate the content validity, usability, and digital ergonomics of NavegApp as an innovative SG for SC assessment in the context of AD research. This contribution aims to advance tools for early AD detection and improve clinical outcomes through timely intervention.

Methods

Overview

We conducted a prospective observational study implementing a structured, multistep process integrating psychometric, usability, and digital ergonomics analysis into the SDLC [36,37]. This process consisted of 3 stages. The first stage, the prealpha phase, involved software requirement analysis, cognitive tasks gamification, software architecture design, documentation, and coding. The second stage, alpha testing, was focused on identifying and resolving general usability issues, ensuring data transfer integrity, collecting evidence on digital ergonomics and usability perception, and evaluating the app's content validity through an expert panel [37]. The third stage, beta testing, aimed to assess the app's usability by comparing the perceptions of different groups of interest, including caregivers (ie, health professionals and relatives of people living with dementia) and potential end users [42]. Integrating psychometric properties and usability considerations into the SDLC offered the flexibility to test various software versions within target populations. This approach allows for the evaluation of specific hypotheses relevant to software development and improvement.

Participants

A total of 212 participants (n=143, 67.5% female participants) were recruited over 10 months using nonprobabilistic sampling methods. The sample included young, healthy adults; asymptomatic PSEN1-E280A mutation carriers; individuals with sporadic MCI or early dementia; and caregivers. Recruitment was conducted through various research projects

XSL•FO
RenderX

involving individuals at risk of developing AD and their relatives. The cognitive status of participants suspected of having MCI or dementia due to AD was determined through comprehensive clinical examinations, while genetic status was confirmed via genotyping.

The clinical examination protocol encompassed neurological and neuropsychological assessments conducted by trained personnel with extensive experience evaluating this cohort. Examiners were blinded to participants' genetic status during the evaluations to minimize bias. The classification of participants was based on the criteria established by the International Working Group on MCI. This classification incorporated the clinical dementia rating, the subjective cognitive complaints scale, and the Lawton-Brody Instrumental Activities of Daily Living scale. Cognitive impairments were evaluated through a neuropsychological examination that included the Consortium to Establish a Registry for Alzheimer's Disease-Colombian version protocol, validated for use in the country [43,44], alongside assessments of memory, language, visuospatial skills, processing speed, and executive functions.

In addition, a panel of 8 experts with expertise in neurology, experimental psychology, and psychometrics was invited to conduct content validity analyses. The panel was selected based on their accomplishment of eligibility criteria (ie, experience in the field, graduate level, and disposition for participation). Young, healthy adults were enrolled from the university campus. Given the qualitative nature of the data required for testing feasibility and usability, nonprobabilistic sampling was deemed appropriate for developing NavegApp.

Ethics Approval

This study was conducted in accordance with national ethical guidelines for research involving human participants, as outlined in resolution 008430 of 1993, which governs human research in Colombia. The study was classified as minimal risk, as it involved no invasive procedures or collecting sensitive information that could adversely affect participants. The protocol was reviewed and approved by the Ethics Committee of the School of Public Health (Act 21030002-00153-2022).

All participants provided informed consent before their involvement. The process included a comprehensive explanation of the study's objectives and methods, presented in accessible language tailored to the participant's understanding. For individuals with MCI or early dementia, relatives or caregivers were present during the consent process to ensure full comprehension. Their understanding and permission for the individual's participation were confirmed. Participants' privacy and confidentiality were protected by deidentifying all data before analysis. Data were stored in independent files on web servers equipped with security protocols to prevent unauthorized access. Participants received financial compensation equivalent to one day's labor (approximately $15 USD) as recognition for their time and involvement. The study complied with the ethical principles of the Declaration of Helsinki and all applicable regulations.

Procedure

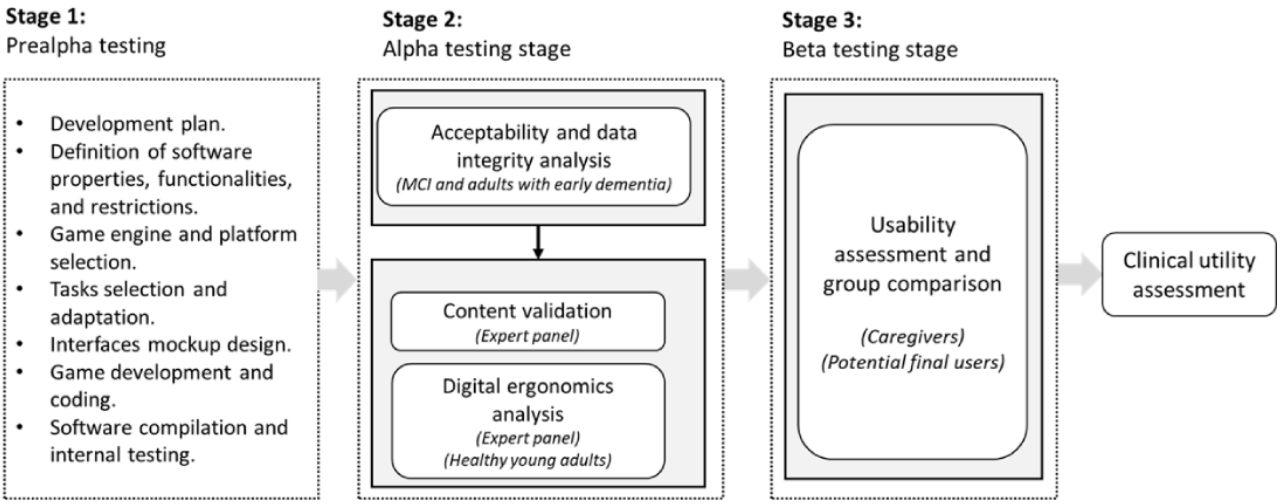
In total, 5 distinct sample groups were enrolled during the validation of NavegApp. Initially, older adults with either MCI or early-stage dementia tested NavegApp as a proof of concept of the interfaces during the alpha-testing phase. The rationale for including such participants was to avoid a potential spectrum bias during the earliest step of the app development by ensuring that the instructions provided, and screen distribution were comprehensible to participants, even when they exhibited signs and symptoms of cognitive impairment [9]. Adjustments were made based on their qualitative feedback and performance.

Next, to analyze content validity, 8 experts were invited to assess the updated version of NavegApp. The information was collected using a 2-round iterative process, consistent with Delphi method recommendations [45,46]. In round 1, experts were asked to give their opinion on the relevance and representativeness of each task included in NavegApp, aligning with the selected SC framework [13]. In addition, their input on the software's digital ergonomics was sought. Subsequently, experts were presented with an adjusted version of NavegApp and an anonymized summary of results from the first round [46]. Qualitative feedback was gathered to improve the software further. Experts were also allowed to evaluate the digital ergonomics of the adjusted version. Concurrently, a healthy sample of undergraduate students was recruited to gather their feedback on digital ergonomics. Furthermore, data on the overall difficulty and user-agreeableness perception of NavegApp were collected.

In the subsequent beta testing stage, expert refinements and healthy young adults' recommendations were incorporated into NavegApp. This phase involved enrolling 2 new groups of participants to evaluate the usability perception of the updated version. As advised by the expert panel, we recruited a separate sample of caregivers, including family members and health professionals affiliated with a Colombian dementia research institution, to test NavegApp and provide feedback on its usability. In addition, a sample of potential users, including individuals in the preclinical stage for familial AD (ie, PSEN1-E280A mutation carriers), individuals with sporadic MCI, and healthy adults, were invited to test NavegApp and provide feedback on its usability. Figure 2 provides a detailed description of NavegApp's development workflow.

XSL•FO
RenderX

Figure 2. The workflow of NavegApp development and validation process. MCI: mild cognitive impairment.



Instruments

Computer System Usability Questionnaire for Videogame

The Computer System Usability Questionnaire for Videogame (VG-CSUQ) is an adapted version of the Spanish Computer System Usability Questionnaire (CSUQ) [42]. This version consists of 15 items rated on a 7-point Likert scale, ranging from "strongly disagree" to "strongly agree." Factor analysis has shown support for a 3-factor structure consistent with previous literature, with factors including system usefulness, information quality, and interface quality [43,44,47-49]. The system usefulness factor assesses task learnability and completion ease, information quality gauges the clarity and user satisfaction with the game's information, and interface quality pertains to the interface's clarity and perceived quality [48]. In total, 4 scores were calculated following the prescribed scoring methodology, one for each usability factor and an overall usability score [43].

Digital Ergonomics Questionnaire

To evaluate NavegApp's digital ergonomics, an ad hoc questionnaire was tailored following the recommendations for SG design in the context of cognitive assessment for neurodegenerative diseases [11]. The questionnaire collects information on 8 ergonomic criteria for evaluating user interfaces. The full version of the questionnaire is present in Multimedia Appendix 1.

NavegApp

NavegApp is an SG designed as a narrative puzzle video game with 3 levels. Each level represents a specific component of the SC model by Jacobs [13], each tailored by gamifying tasks previously described as potentially sensitive to cognitive changes occurring since the preclinical stage of AD [17,22,28,50-55]. The tasks were adapted to a narrative video game format by applying gamification and game design principles aligning with the protocol description available for each one [39,44,48]. Thus, NavegApp includes a gamified version of a hidden goal task (gHGT) for testing allocentric spatial orientation [50], a gamified version of a mental rotation task used for testing the ability to manipulate objects in a 2D space [28], and a gamified version of the Corsi Block-Tapping test for the assessment of visuospatial short-term memory [54,55].

The gHGT comprised 8 trials, each requiring the participant to locate a hidden object (ie, hedgehog) that had been visible for 5 seconds. For orientation, participants could only use 2 distal landmarks. Within each trial, the starting position, spatial landmarks, and hidden goal occupied 1 of the 8 pseudorandom positions while maintaining the same mutual spatial relationships [50]. A tutorial was designed to ensure participants' understanding of the task. An algorithm was implemented to detect unusual performance patterns, such as moving toward the landmark instead of the goal, excessively long paths, or extreme deviations from the goal, prompting a repeat of the tutorial if necessary.

The gamified mental rotation task consisted of 3 conditions (at 0, 90, and 180 degrees) with 16 trials each [28]. In each condition, participants were instructed to mentally rotate the object displayed at the center of the screen in a clockwise direction and subsequently select the correct stimulus from a set of 4 options. The stimuli were presented in the same order and positions as described in the original task, with visual and auditory feedback for correct and incorrect responses. To ensure comprehension of the instructions, a tutorial was implemented using a new stimulus, preserving the original task protocol as closely as possible.

Finally, the gamified Corsi block-tapping test asked the participants to repeat a predefined sequence of blocks in forward and backward sequence [54,55]. In the forward condition, participants were required to replicate the sequence in the same order as presented. In contrast, the sequence had to be reproduced in reverse order in the backward condition. For the gamified version, the participant must follow a mole through 9 holes located in the same position as the blocks from the standard in-person test [54]. The task's stopping criteria were also preserved, ending after 2 consecutive failed attempts at sequences of the same length [54]. A counter displayed on the screen indicated the number of holes to explore in each trial, providing participants with a clear reference. In addition, visual and auditory feedback informed participants whether they successfully found the mole or failed the trial.

XSL•FO
RenderX

Additional scenes were incorporated into the video game to capture essential data, including demographic information and the VG-CSUQ questionnaire. The Unity game engine (version 2022.1.20f1) [47] was used for developing NavegApp. Graphic material was obtained through memberships of 2D game assets in online stores. All the graphic material is protected by the Creative Commons (version 4.0) license. Figure 3 depicts example screenshots of NavegApp's interfaces. The screenshots are presented in the English language for communication purposes. The original content was provided in the Spanish language, as it is the mother tongue of participants involved in the study.

Figure 3. NavegApp's interface screen captures.



Statistical Analysis

Descriptive statistics were used to summarize the performance of formative and summative usability assessments in accordance with the recommendations for quantifying user experience [56]. Our approach adhered to established recommendations for usability and digital ergonomics testing for cognitive games [9,11,56,57], and content validation for measurement instruments in health settings [58-61]. Therefore, for content validity analysis, the experts' opinion consistency was determined following the recommendations and procedures described by Yusoff [58] and Polit et al [59] during the alpha testing stage.

To account for the interrater chance of agreement, the modified κ statistic, based on the content validity index, was calculated [62]. This involved first computing the probability of chance agreement using the formula for a binomial random variable. The modified κ statistic was then derived using the proportion of agreements on relevance and the calculated probability of chance agreement [62]. Results were interpreted based on recommendations from the literature regarding the number of experts. Values >0.72 were classified as indicating good agreement, while values >0.88 were considered evidence of excellent agreement [62].

In addition, Aiken V coefficient [60] was calculated as a complementary measure of the experts' degree of agreement regarding the relevance of the instrument elements using a Likert scale [60,63]. Unlike the content validity index, Aiken V accounts for the ordinal nature of ratings, making it particularly suitable for small samples of expert raters. This characteristic makes Aiken V an ideal measure for evaluating the content validity of instruments during early development stages [60,61].

The rank biserial correlation was used to estimate the magnitude of differences in digital ergonomics perception between experts and young, healthy adults testing the alpha version of NavegApp. In addition, the relationship between agreeableness and perceptions of difficulty was evaluated using the Cramér V coefficient.

The rank biserial correlation was calculated in the beta testing stage to evaluate differences in usability perceptions across participants. To evaluate differences in usability perception due to cognitive decline, the perception of sporadic MCI participants was compared against those of healthy adults in the remaining groups. A 95% CI was computed to estimate the precision of the effect size measures. All statistical analyses were performed using R (version 4.3.1; R Core Team) [62] and RStudio (version 2023.06.1; RStudio Inc) [64].

Results

Sociodemographics

The sociodemographic characteristics of the participants are presented in Table 1. All groups' median years of education were more than 11 years, indicating that at least half of the

XSL•FO
RenderX

participants had finished secondary school. According to the Colombian National Statistics Department, 87.2% (185/212) of the participants belonged to a low socioeconomic status, whereas 87.7% (186/212) were raised in urban areas [65].

Table 1. The sociodemographic characteristics of the participants.

	Alpha testing		Beta testing			
	MCI[a] (n=16)	Healthy young adults (n=20)	Caregivers (n=22)	PSEN1-E280A mutation carriers (n=81)	Healthy controls (n=59)	MCI (n=14)
Sex, n (%)						
Female	11 (69)	10 (50)	15 (68)	58 (72)	39 (66)	10 (71)
Male	5 (31)	10 (50)	7 (32)	23 (28)	20 (34)	4 (29)
Age (y)						
Mean (SD)	72.8 (6.16)	24.2 (6.31)	64.7 (12.7)	31.3 (8.36)	33.7 (10.5)	64.9 (7.59)
Median (IQR)	75.5 (8.3)	22.0 (2.3)	71.5 (19.5)	31.0 (15.0)	34.0 (17.0)	65.0 (8.3)
Years of education (y)						
Mean (SD)	11.0 (5.34)	12.5 (2.31)	12.2 (4.82)	11.6 (3.31)	11.5 (3.66)	12.9 (4.34)
Median (IQR)	13.0 (8.0)	12.0 (2.3)	12.0 (5.0)	11.0 (2.0)	12.0 (2.0)	13.5 (4.8)
Upbringing area, n (%)						
Rural	1 (6)	2 (10)	1 (4)	15 (18)	7 (12)	0 (0)
Urban	15 (94)	18 (90)	21 (96)	66 (82)	52 (88)	14 (100)
Socioeconomic status, n (%)						
High	0 (0)	0 (0)	8 (36)	2 (3)	1 (2)	4 (29)
Low	16 (100)	20 (100)	10 (46)	78 (96)	55 (93)	6 (43)
Medium	0 (0)	0 (0)	4 (18)	1 (1)	3 (5)	4 (29)

[a]MCI: mild cognitive impairment.

Alpha Testing Stage: Data Integrity and General Acceptability

During the alpha testing stage, several issues were identified concerning the integrity of the data uploaded to the server. In total, 3% (1/36) records of task performance was not uploaded due to connectivity issues. Solutions, such as creating a backup of the information within the device and improved communication protocols, were integrated into NavegApp's beta version to prevent future data loss. All the participants could finish at least the first 3 trials for each task, which indicated that the wording of the instructions was clear so the participants with MCI could tackle the cognitive tasks.

Alpha Testing Stage: Content Validity

The expert panel assessed the relevance and representativeness of gamified cognitive tasks in NavegApp in alignment with the SC model by Jacobs [13]. Table 2 displays Aiken V and modified κ indexes, gauging expert consensus. The results show unanimous agreement, with a 100% consensus per the κ index, supporting task relevance and representativeness. The Aiken V index indicated high agreement (ie, between 96% and 100%) among experts regarding task relevance and representativeness.

Table 2. Content validity measures.

Task	Relevance		Representativeness	
	κ	Aiken V (95% CI)	κ	Aiken V (95% CI)
Allocentric spatial navigation (gHGT[a])	1.00	0.96 (0.80-0.99)	1.00	0.96 (0.80-0.99)
Mental rotation (gMRT[b])	1.00	0.96 (0.80-0.99)	1.00	0.96 (0.80-0.99)
Short-term visuospatial working memory (gCorsi[c])	1.00	1.00 (0.86-1.00)	1.00	1.00 (0.86-1.00)

[a]gHGT: gamified version of a hidden goal task.

[b]gMRT: gamified mental rotation task.

[c]gCorsi: gamified Corsi Block-Tapping test.

XSL•FO
RenderX

Alpha Testing Stage: Digital Ergonomics

Following the recommendations for conducting and analyzing Delphi studies [46], an anonymous report was crafted for each expert, and new data regarding their opinion on the digital ergonomics criteria were collected. No differences were found between the first- and second-round scores on any ergonomic criteria. Scores of digital ergonomics for rounds 1 and 2 are provided in Multimedia Appendix 2. Qualitative recommendations such as changes in the wording of instructions, adequations in the interfaces, gamification elements, and a tutorial for the gHGT were included in the beta version of NavegApp.

NavegApp's digital ergonomics were measured using a Likert scale of 7 points. Scores for each criterion were determined by averaging the related items. To summarize the perceptions of the expert panel and healthy young adults, median and IQRs were calculated. The median scores for all ergonomic criteria indicated a perception exceeding 75% (ie, >5.25), reflecting a positive evaluation of NavegApp's digital ergonomics. Table 3 shows minor differences (r_{rb} 0.08-0.29) between the expert panel and healthy young adults regarding their perceptions of digital ergonomics. The experts rated NavegApp's ergonomics slightly higher than healthy young adults, except for the codes significance and orientation criteria. Details of statistical tests are provided in Multimedia Appendix 2.

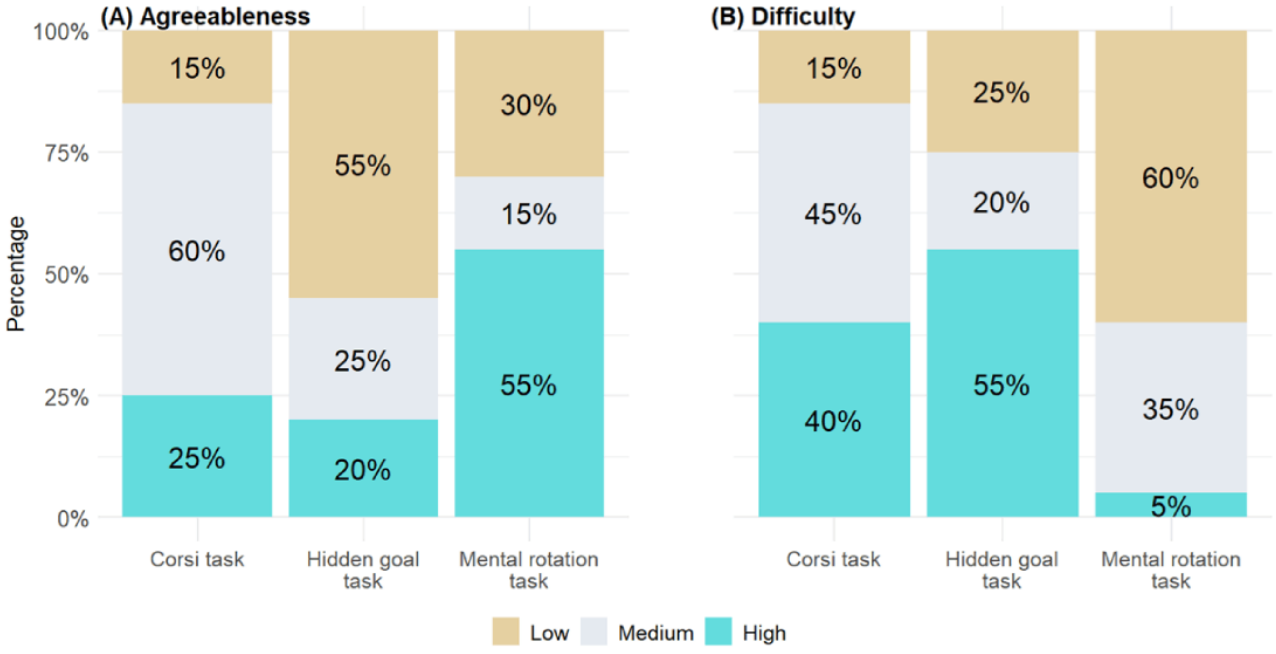
Table 3. Digital ergonomics perception comparison.

Ergonomic criteria	Experts (n=8), median (IQR)	Healthy young adults (n=20), median (IQR)	Rank biserial correlation, r_{rb} (95% CI)
Compatibility	6.75 (1.13)	6.00 (2.00)	0.16 (0.01 to 0.48)
Orientation	6.75 (0.13)	7.00 (0.31)	0.25 (0.02 to 0.59)
Load	6.67 (1.84)	6.33 (0.92)	0.08 (0.01 to 0.52)
Adaptability	7.00 (0.00)	7.00 (0.50)	0.19 (0.01 to 0.46)
Consistency	7.00 (0.17)	6.67 (1.08)	0.29 (0.03 to 0.57)
Codes significance	5.50 (0.25)	5.63 (1.75)	0.26 (0.02 to 0.54)
Control	7.00 (0.00)	7.00 (0.25)	0.13 (–0.36 to 0.56)

Alpha Testing Stage: Difficulty and Agreeableness

Following experts' suggestions, the healthy young adult sample was asked to order the tasks according to agreeableness and difficulty. Analysis using Cramér V indicated a moderate association between the perception of agreeableness and difficulty for each task (χ^2_4=16.5; *P*=.002; V=0.37, 95% CI 0.28-0.54). Further examination of the frequency distribution indicated that as tasks were perceived as more difficult, participants found them less agreeable. Figure 4 displays the difficulty (Figure 4A) and agreeableness (Figure 4B) perception by task.

Figure 4. Agreeableness and difficulty perception by task.



XSL•FO
RenderX

Beta Testing: Usability Assessment

Table 4 details descriptive statistics and effect size measures used to quantify the differences in participants' perceptions of usability, as measured by the VG-CSUQ questionnaire components, for the beta version of NavegApp.

Table 4. Usability perception.

Usability scores	Caregivers (n=22), median (IQR)	PSEN1-E280A mutation carriers (n=81), median (IQR)	Healthy controls (n=59), median (IQR)	MCI[a] (n=14), median (IQR)
Usability total score	93.3 (13.6)	94.4 (11.1)	94.4 (10.0)	83.9 (32.2)
System usefulness	91.7 (10.4)	91.7 (16.7)	91.7 (16.7)	70.8 (33.3)
Information quality	96.7 (12.5)	96.7 (10.0)	96.7 (10.0)	88.3 (25.0)
Interface quality	100 (16.7)	100 (0.0)	100 (4.2)	95.8 (25.0)

[a]MCI: mild cognitive impairment.

Rank biserial correlation was used to assess the magnitude of differences in usability perceptions among the groups testing the beta version of NavegApp. The analysis revealed that the sporadic MCI group showed moderate to large differences, with lower scores in overall usability perception (r_{rb} 0.26-0.29), system usefulness (r_{rb} 0.25-0.37), information quality (r_{rb} 0.27-0.34), and interface quality (r_{rb} 0.13-0.23) when compared to other groups. In contrast, participants from the caregivers, asymptomatic PSEN1-E280A mutation carriers, and healthy control groups exhibited minor differences (r_{rb} 0.00-0.20), indicating a more consistent usability perception. A table detailing the magnitude of these differences for each pair of groups is included in Multimedia Appendix 3.

Discussion

Principal Findings

In this study, we detail the development and validation of NavegApp, a serious video game designed for assessing SC components potentially sensitive to early cognitive change in the AD continuum. Our results confirm that NavegApp is a usable and ergonomic SG encompassing cognitive tasks that are representative and relevant for measuring SC. The consensus among experts and positive feedback on digital ergonomics perception and usability from healthy young adults, caregivers, and final users demonstrates that NavegApp offers a user-friendly interface for executing gamified cognitive tasks effectively. Central to the performance of NavegApp is our approach, which integrates recommendations for usability and digital ergonomics assessment, along with validity measurements directly into the SDLC, particularly during the testing phase [33,36,57].

Our results from the alpha testing stage indicated that NavegApp consistently performed well across multiple devices, accurately displaying all stimuli and instructions with minimal data loss. Notably, participants diagnosed with MCI and early-stage dementia could comprehend and complete at least the initial 3 trials for each task, meeting the primary criteria for task comprehension [11]. These findings align with other studies of SGs developed under similar frameworks, which report positive usability metrics from the early stages of software development [35,66,67]. In addition, our results support using the Unity Game Engine as a suitable platform for developing 2D interfaces for cognitive assessment [68]. Although most research using Unity for cognitive assessment has focused on 3D spatial navigation tasks [18,69], its versatility makes it ideal for creating 2D interfaces compatible with various devices, such as tablets and mobile phones [47,68]. This versatility is crucial for deploying digital tasks in underrepresented populations with limited access to advanced technologies or expensive devices.

Designing reliable and valid SGs for cognitive assessment requires a strong foundation in validity and reliability metrics [7]. Analyzing content validity is crucial in this context, as it assesses whether an instrument effectively measures its intended construct [37,61]. For NavegApp, the expert panel demonstrated a high level of consensus, confirming that all tasks included in the SG were relevant and representative of the SC model by Jacobs [13]. Furthermore, qualitative feedback from the panel provided insights into improving the SG's theoretical foundations and user-interface elements [30,45,46]. This study is the first to include a content validity analysis within the SG development process, setting it apart from other games, primarily focusing on usability [35] and acceptability [67].

Key components of the content validity analysis presented here involved conducting a thorough critical literature review and following the original task protocols to guide the gamification process from the beginning of the SDLC [9,11,46]. Content validity indexes were computed to quantify expert consensus on the relevance and representativeness of various components within the instrument [45,46]. This process aimed to avoid discrepancies in determining the tasks to be included in the SG. In addition, the Delphi method, a structured approach facilitating expert dialogue on complex topics, has proven especially beneficial for gaining insights into content validity and digital ergonomics [46].

Additional considerations, such as potential biases stemming from perceptions of visual stimuli and instructions, were examined using statistical methods during NavegApp's development alpha testing phase. Our findings revealed an association between agreeableness and difficulty perception, prompting us to randomize level presentations to participants in the SG final version. This ensures unbiased performance and prevents overarching perceptions of the test as difficult or unpleasant [70,71]. According to our findings, design decisions can benefit from a phased evaluation approach, such as the one

XSL•FO
RenderX

used in this study, wherein earlier test versions establish a foundational benchmark for assessing subsequent iterations.

While including an expert panel primarily facilitates content validity assessment, their expertise can provide valuable insights into important aspects such as digital ergonomics for building user-friendly interfaces. Integrating digital ergonomics into the SDLC from the alpha stage provides critical information on how design decisions impact the end-user experience [11,44]. Although these principles have been used for developing SG in the context of assessment and training [44,64], there is currently no systematic framework for their inclusion in the SG development process.

To address this gap, our study introduces a self-reported questionnaire that converts digital ergonomics heuristics by Bastien and Scapin [57] into Likert scale items to evaluate 8 fundamental ergonomics criteria recognized as relevant for the design of SG in the context of neurodegenerative diseases [11]. The median scores for these criteria indicated >90% (ie, all scores >6.30) acceptability for NavegApp's interface, except the codes significance criterion, which scored 79% (ie, 5.50). Code significance refers to the semantic relationship between a term and the actions it represents. Scores obtained from experts and young, healthy adults underscore the need to enhance icons and vocabulary to improve user experience and interactions in the subsequent beta phase testing stage. This multistage approach, incorporating multiple sources of evidence, enriches the current literature on the development of SGs for cognitive assessment in neurodegenerative diseases, which often prioritize aspects such as immersive [38], user-interface interactions [39], and gamification [35].

Moving into the beta testing stage, usability assessment was integrated into the SDLC. Usability questionnaires are particularly valuable during this phase [56], as the qualities of the interface, system, and information provided by the software are crucial for enhancing user interaction efficiency [11,42]. While some SGs for cognitive assessment have demonstrated high acceptance [67,69], gaps remain in examining these aspects comprehensively. To address this, we used the VG-CSUQ, an adaptation of the widely used CSUQ, to gather multifaceted usability data in the SG context. This approach contributed to a nuanced assessment and subsequent adjustments of system features, thereby refining users' usability perception.

Given the lack of country- or setting-specific normative data, usability percentages offer a proxy for user perceptions [42]. Median scores exceeding 75% across all usability factors indicate that NavegApp is consistently perceived as usable across all tested groups. All groups rated the interface quality and system usefulness highly, with perception rates exceeding 75%, and no differences were found among the groups.

The differences in information quality perception between sporadic MCI and healthy adult groups can be attributed to the higher perceived difficulty of the tasks by those with decreased cognitive functionality. Despite this, the perception remained highly positive (ie, >80%). This finding suggests that digital tools may be more challenging to implement for individuals in the prodromal stage of AD, necessitating additional efforts to enhance their user experience [11]. It underscores the importance of including participants across the A spectrum during development to avoid spectrum bias and ensure the SG is tested beyond a healthy population [7,11].

Our findings have important implications for cognitive assessment theories related to AD, particularly in the context of screening and diagnosis. Cognitive assessment is a cornerstone in the staging of individuals along the AD continuum, as the presence or absence of subjective and objective cognitive decline distinguishes preclinical and prodromal stages of the disease [3]. Traditionally, subjective cognitive concerns and episodic memory are assessed through person-to-person paper-and-pencil neuropsychological tests, which remain the standard for follow-up, screening, and assessment in at-risk populations [3]. However, the emergence of innovative digital tools, such as SGs, presents an opportunity to rethink these methods. SGs can conduct cognitive assessments using engaging and cost-effective methodologies while maintaining psychometric rigor [9]. Although challenges remain in achieving fully remote cognitive screening programs, our results demonstrate the feasibility of developing user-friendly and highly usable interfaces capable of implementing cognitive assessment protocols in a standardized manner. These findings align with and extend existing theories by demonstrating how digital tools can complement traditional approaches, potentially improving accessibility and scalability in cognitive assessments for AD [9,11].

The development of innovative SGs for cognitive assessment, such as NavegApp, offers unique opportunities for early screening and intervention in AD. These tools bring several advantages to clinical practice. First, SGs eliminate the need for specialized, trained personnel during task execution, enabling remote cognitive assessments and expanding access to at-risk populations. Second, by fostering enhanced engagement and motivation, SGs improve data quality in repetitive cognitive tasks compared to traditional paper-and-pencil assessments [35]. Third, SGs provide greater ecological validity by simulating real-world tasks, offering insights into functional cognition in everyday contexts [7]. In addition, these tools reduce costs and mitigate biases arising from individual differences among evaluators, ensuring consistency in data collection [9]. Together, these features position SGs as powerful instruments for large-scale cognitive screening and monitoring, with the potential to support early diagnosis and intervention in clinical settings.

Limitations

While our results favor NavegApp as a usable tool, it is essential to acknowledge certain limitations. Due to the exploratory nature of the feasibility assessment and the need to collect qualitative data on the performance and perception of potential users, a nonprobabilistic sampling method was preferred for the analyses presented here. However, these sampling methods limit the generalizability of our findings, necessitating further studies to align our results with broader population characteristics [72]. In addition, given the lack of standardized tools for assessing digital ergonomic criteria, we introduced a tailored self-reporting questionnaire grounded in previous recommendations for SG design in neurocognitive disease contexts [11].

XSL•FO
RenderX

However, the psychometric robustness of this tool warrants future evaluation.

Our content validity assessment hinges on a select group of experts meeting specific criteria. Different experts might offer varying degrees of agreement. Finally, while our CSUQ questionnaire adaptation is valuable and represents an effort to develop standardized instruments for assessing the usability of SGs adapted to the Spanish-speaking populations from low- and middle-income countries, it is not exclusive. Alternative instruments, such as the System Usability Scale [44] or Gamefulquest [73], are also suitable options for gauging software and SG usability and user experience.

As group comparisons were a central focus of this study, the imbalance in group sizes, stemming from the nonprobabilistic sampling approach, represents a limitation that warrants consideration. Uneven sample sizes can affect the precision of effect size estimates and the reliability of CIs, potentially introducing bias into group comparisons [72]. Future studies should aim to address this limitation by ensuring balanced sample sizes across groups. Conducting a priori power analyses can help determine the appropriate sample size needed to achieve sufficient statistical power for group comparisons [72]. In addition, using matched sampling methods could further enhance the quality of comparisons by reducing variability and improving the robustness of findings.

Future studies should incorporate additional levels of analysis to refine the development and evaluation of SGs for cognitive assessment. For instance, comparing perceptions and performance across distinct groups using the same software version during the early development phase can help identify potential biases introduced during the gamification process. Conversely, contrasting different software versions at various stages of development can reveal biases stemming from design decisions and modifications to the interface and learning effects associated with repeated use of the SG. To achieve this, longitudinal usability assessments are recommended, enabling the systematic evaluation of changes over time [39].

The preexisting validity evidence of cognitive tasks plays a critical role in shaping the development process, influencing the steps and metrics required for validation and usability. However, a phased approach that collects data on multiple aspects of validity and usability at different stages of development provides richer insights than a singular pilot study. Such an approach facilitates a more comprehensive understanding of task performance and usability dynamics. Finally, future research should investigate how enhanced game design can more effectively target specific cognitive functions. By tailoring tasks to assess nuanced aspects of SC, they could improve their sensitivity for detecting subtle cognitive impairments. This targeted approach may help advance the utility of SGs as scalable tools for early screening and diagnosis of neurodegenerative conditions such as AD.

Conclusions

As the need for scalable cognitive screening programs and early detection of cognitive decline intensifies, digital neuropsychological alternatives, such as SGs, are emerging as a suitable alternative to traditional methods based on molecular biomarkers. These games offer a viable solution for remote, repeated assessments on a population scale [4,5,33]. Despite advances in the field, there is an urgent need to create digital tools that balance both validity and usability for the cognitive testing of people at risk for developing or living with neurodegenerative diseases, such as AD [4]. In this context, we introduce NavegApp, a valid and user-friendly SG designed for SC assessment, crafted by merging software development principles and measurement properties evaluation [37].

The adoption of SGs in health informatics signifies a strategic shift, addressing the limitations inherent in traditional neuropsychological assessment [5,9,11,33]. This study was conducted in response to the growing call for tools that prioritize user experience and validity and are primed for broad-scale cognitive monitoring. While innovative, our approach emphasizes the crucial role of integrating psychometric validity measures from the outset of SG design. Using a blend of usability evaluations, content validity assessments, and digital ergonomic principles, we aim to set the stage for evolving standardized developmental frameworks. Such frameworks are poised to capitalize on cutting-edge health informatics and psychometrics advancements in digital neuropsychology.

Acknowledgments

The authors extend their gratitude to the Group of Neurosciences of Antioquia for their assistance in recruiting participants with mild cognitive impairment during the initial phase of the research and the caregivers for testing the beta version of the serious game. The authors reserved a special acknowledgment for the experts who generously contributed to the content validity analyses. Their valuable insights and suggestions played a crucial role in enhancing the usability of NavegApp. The University of Antioquia and the Colombian Ministry of Science and Innovation supported this study (918-111591891726).

Data Availability

The data supporting this study’s findings are not publicly available due to privacy and confidentiality concerns. However, the data can be accessed from the corresponding author upon reasonable request.

Authors' Contributions

JPS-E conceived and drafted the manuscript, collected research data, and conducted statistical analyses. DA provided support in participant recruitment and clinical neurological assessment. SV participated in recruiting experts to analyze content validity and

XSL•FO
RenderX

the writing process. MAG-B and DCA-A provided theoretical and methodological guidance and feedback throughout the research process. NT provided theoretical guidance and insights into the manuscript. All authors have reviewed and approved the study's content.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Digital ergonomic questionnaire for serious games design.
[DOCX File , 25 KB - games_v13i1e66167_app1.docx]

Multimedia Appendix 2
Digital ergonomics scores and comparisons from Delphi rounds.
[DOCX File , 17 KB - games_v13i1e66167_app2.docx]

Multimedia Appendix 3
Usability perception comparison among groups.
[DOCX File , 17 KB - games_v13i1e66167_app3.docx]

References

1. No authors listed. 2023 Alzheimer's disease facts and figures. Alzheimers Dement 2023 Apr;19(4):1598-1695. [doi: 10.1002/alz.13016] [Medline: 36918389]
2. Rentz DM, Parra Rodriguez MA, Amariglio R, Stern Y, Sperling R, Ferris S. Promising developments in neuropsychological approaches for the detection of preclinical Alzheimer's disease: a selective review. Alzheimers Res Ther 2013 Nov 21;5(6):58 [FREE Full text] [doi: 10.1186/alzrt222] [Medline: 24257331]
3. Jack CRJ, Bennett DA, Blennow K, Carrillo MC, Dunn B, Haeberlein SB, et al. NIA-AA research framework: toward a biological definition of Alzheimer's disease. Alzheimers Dement 2018 Apr;14(4):535-562 [FREE Full text] [doi: 10.1016/j.jalz.2018.02.018] [Medline: 29653606]
4. Levy B, Tsoy E, Gable S. Developing cognitive markers of Alzheimer's disease for primary care: implications for behavioral and global prevention. J Alzheimers Dis 2016 Oct 18;54(4):1259-1272. [doi: 10.3233/JAD-160309] [Medline: 27567831]
5. Whelan R, Barbey FM, Cominetti MR, Gillan CM, Rosická AM. Developments in scalable strategies for detecting early markers of cognitive decline. Transl Psychiatry 2022 Nov 09;12(1):473 [FREE Full text] [doi: 10.1038/s41398-022-02237-w] [Medline: 36351888]
6. Cuffaro L, Di Lorenzo F, Bonavita S, Tedeschi G, Leocani L, Lavorgna L. Dementia care and COVID-19 pandemic: a necessary digital revolution. Neurol Sci 2020 Aug;41(8):1977-1979 [FREE Full text] [doi: 10.1007/s10072-020-04512-4] [Medline: 32556746]
7. Sánchez-Escudero JP, Galvis-Herrera AM, Sánchez-Trujillo D, Torres-López LC, Kennedy CJ, Aguirre-Acevedo DC, et al. Virtual reality and serious videogame-based instruments for assessing spatial navigation in Alzheimer's disease: a systematic review of psychometric properties. Neuropsychol Rev 2024 Feb 26 (forthcoming). [doi: 10.1007/s11065-024-09633-7] [Medline: 38403731]
8. Coughlan G, Coutrot A, Khondoker M, Minihane AM, Spiers H, Hornberger M. Toward personalized cognitive diagnostics of at-genetic-risk Alzheimer's disease. Proc Natl Acad Sci U S A 2019 May 07;116(19):9285-9292 [FREE Full text] [doi: 10.1073/pnas.1901600116] [Medline: 31015296]
9. Manera V, Ben-Sadoun G, Aalbers T, Agopyan H, Askenazy F, Benoit M, et al. Recommendations for the use of serious games in neurodegenerative disorders: 2016 Delphi panel. Front Psychol 2017 Jul 25;8:1243 [FREE Full text] [doi: 10.3389/fpsyg.2017.01243] [Medline: 28790945]
10. Watermeyer T, Calia C. Neuropsychological assessment in preclinical and prodromal Alzheimer disease: a global perspective. J Glob Health 2019 Jun;9(1):010317 [FREE Full text] [doi: 10.7189/jogh.09.010317] [Medline: 31073397]
11. Ben-Sadoun G, Manera V, Alvarez J, Sacco G, Robert P. Recommendations for the design of serious games in neurodegenerative diseases. Front Aging Neurosci 2018 Feb 02;10:13 [FREE Full text] [doi: 10.3389/fnagi.2018.00013] [Medline: 29456501]
12. Possin KL. Visual spatial cognition in neurodegenerative disease. Neurocase 2010 Dec;16(6):466-487 [FREE Full text] [doi: 10.1080/13554791003730600] [Medline: 20526954]
13. Jacobs L. Spatial memory. In: Daroff RB, Aminoff MJ, editors. Encyclopedia of the Neurological Sciences. Cambridge, MA: Academic Press; 2003.
14. Cona G, Scarpazza C. Where is the "where" in the brain? A meta-analysis of neuroimaging studies on spatial cognition. Hum Brain Mapp 2019 Apr 15;40(6):1867-1886 [FREE Full text] [doi: 10.1002/hbm.24496] [Medline: 30600568]

XSL•FO
RenderX

15. Coughlan G, Laczó J, Hort J, Minihane AM, Hornberger M. Spatial navigation deficits - overlooked cognitive marker for preclinical Alzheimer disease? Nat Rev Neurol 2018 Aug;14(8):496-506. [doi: 10.1038/s41582-018-0031-x] [Medline: 29980763]
16. Serino S, Morganti F, Colombo D, Riva G. The contribution of allocentric impairments to the cognitive decline in Alzheimer's disease. In: Proceedings of the 7th International Conference on Pervasive Computing Paradigms for Mental Health. 2018 Presented at: MindCare 2018; January 9-10, 2018; Boston, MA. [doi: 10.1007/978-3-030-01093-5_11]
17. Serino S, Cipresso P, Morganti F, Riva G. The role of egocentric and allocentric abilities in Alzheimer's disease: a systematic review. Ageing Res Rev 2014 Jul;16:32-44. [doi: 10.1016/j.arr.2014.04.004] [Medline: 24943907]
18. Castegnaro A, Howett D, Li A, Harding E, Chan D, Burgess N, et al. Assessing mild cognitive impairment using object-location memory in immersive virtual environments. Hippocampus 2022 Sep;32(9):660-678 [FREE Full text] [doi: 10.1002/hipo.23458] [Medline: 35916343]
19. Howett D, Castegnaro A, Krzywicka K, Hagman J, Marchment D, Henson R, et al. Differentiation of mild cognitive impairment using an entorhinal cortex-based test of virtual reality navigation. Brain 2019 Jun 01;142(6):1751-1766 [FREE Full text] [doi: 10.1093/brain/awz116] [Medline: 31121601]
20. Jacobs HI, Van Boxtel MP, Jolles J, Verhey FR, Uylings HB. Parietal cortex matters in Alzheimer's disease: an overview of structural, functional and metabolic findings. Neurosci Biobehav Rev 2012 Jan;36(1):297-309. [doi: 10.1016/j.neubiorev.2011.06.009] [Medline: 21741401]
21. Iachini T, Iavarone A, Senese VP, Ruotolo F, Ruggiero G. Visuospatial memory in healthy elderly, AD and MCI: a review. Curr Aging Sci 2009 Mar;2(1):43-59. [doi: 10.2174/1874609810902010043] [Medline: 20021398]
22. Rusconi ML, Suardi A, Zanetti M, Rozzini L. Spatial navigation in elderly healthy subjects, amnestic and non amnestic MCI patients. J Neurol Sci 2015 Dec 15;359(1-2):430-437. [doi: 10.1016/j.jns.2015.10.010] [Medline: 26478129]
23. Mitolo M, Gardini S, Fasano F, Crisi G, Pelosi A, Pazzaglia F, et al. Visuospatial memory and neuroimaging correlates in mild cognitive impairment. J Alzheimers Dis 2013;35(1):75-90. [doi: 10.3233/JAD-121288] [Medline: 23357899]
24. Guariglia CC. Spatial working memory in Alzheimer's disease: a study using the Corsi block-tapping test. Dement Neuropsychol 2007;1(4):392-395 [FREE Full text] [doi: 10.1590/S1980-57642008DN10400011] [Medline: 29213417]
25. Nedelska Z, Andel R, Laczó J, Vlcek K, Horinek D, Lisy J, et al. Spatial navigation impairment is proportional to right hippocampal volume. Proc Natl Acad Sci U S A 2012 Feb 14;109(7):2590-2594 [FREE Full text] [doi: 10.1073/pnas.1121588109] [Medline: 22308496]
26. Allison SL, Fagan AM, Morris JC, Head D. Spatial navigation in preclinical Alzheimer's disease. J Alzheimers Dis 2016 Feb 09;52(1):77-90. [doi: 10.3233/JAD-150855] [Medline: 26967209]
27. Allison SL, Rodebaugh TL, Johnston C, Fagan AM, Morris JC, Head D. Developing a spatial navigation screening tool sensitive to the preclinical Alzheimer disease continuum. Arch Clin Neuropsychol 2019 Oct 24;34(7):1138-1155 [FREE Full text] [doi: 10.1093/arclin/acz019] [Medline: 31197326]
28. Suzuki A, Shinozaki J, Yazawa S, Ueki Y, Matsukawa N, Shimohama S, et al. Establishing a new screening system for mild cognitive impairment and Alzheimer's disease with mental rotation tasks that evaluate visuospatial function. J Alzheimers Dis 2018;61(4):1653-1665. [doi: 10.3233/JAD-170801] [Medline: 29376869]
29. Bourrelier J, Kubicki A, Rouaud O, Crognier L, Mourey F. Mental rotation as an indicator of motor representation in patients with mild cognitive impairment. Front Aging Neurosci 2015 Dec 23;7:238 [FREE Full text] [doi: 10.3389/fnagi.2015.00238] [Medline: 26779010]
30. Varpio L, Paradis E, Uijtdehaage S, Young M. The distinctions between theory, theoretical framework, and conceptual framework. Acad Med 2020 Jul;95(7):989-994. [doi: 10.1097/ACM.0000000000003075] [Medline: 31725464]
31. Deterding S, Dixon D, Khaled R, Nacke L. From game design elements to gamefulness: defining "gamification". In: Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments. 2011 Presented at: MindTrek '11; September 28-30, 2011; Tampere, Finland. [doi: 10.1145/2181037.2181040]
32. Farber I, Fua KC, Gupta S, Pautler D. MoCHA: designing games to monitor cognitive health in elders at risk for Alzheimer's disease. In: Proceedings of the 13th International Conference on Advances in Computer Entertainment Technology. 2016 Presented at: ACE '16; November 9-12, 2016; Osaka, Japan. [doi: 10.1145/3001773.3001818]
33. Lumsden J, Edwards EA, Lawrence NS, Coyle D, Munafò MR. Gamification of cognitive assessment and cognitive training: a systematic review of applications and efficacy. JMIR Serious Games 2016 Jul 15;4(2):e11 [FREE Full text] [doi: 10.2196/games.5888] [Medline: 27421244]
34. Manera V, Petit PD, Derreumaux A, Orvieto I, Romagnoli M, Lyttle G, et al. 'Kitchen and cooking,' a serious game for mild cognitive impairment and Alzheimer's disease: a pilot study. Front Aging Neurosci 2015;7:24. [doi: 10.3389/fnagi.2015.00024] [Medline: 25852542]
35. Tong T, Chignell M, Tierney MC, Lee J. A serious game for clinical assessment of cognitive status: validation study. JMIR Serious Games 2016 May 27;4(1):e7 [FREE Full text] [doi: 10.2196/games.5006] [Medline: 27234145]
36. Hauschild AC, Martin R, Holst SC, Wienbeck J, Heider D. Guideline for software life cycle in health informatics. iScience 2022 Dec 22;25(12):105534 [FREE Full text] [doi: 10.1016/j.isci.2022.105534] [Medline: 36437879]

XSL•FO
RenderX

37. Terwee CB, Prinsen CA, Chiarotto A, Westerman MJ, Patrick DL, Alonso J, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. Qual Life Res 2018 May 17;27(5):1159-1170 [FREE Full text] [doi: 10.1007/s11136-018-1829-0] [Medline: 29550964]
38. Tong T, Chignell M, Lam P, Tierney MC, Lee J. Designing serious games for cognitive assessment of the elderly. Proc Int Symp Hum Factors Ergon Healthc 2014 Jul 22;3(1):28-35. [doi: 10.1177/2327857914031004]
39. Graafland M, Dankbaar M, Mert A, Lagro J, De Wit-Zuurendonk L, Schuit S, et al. How to systematically assess serious games applied to health care. JMIR Serious Games 2014 Nov 11;2(2):e11 [FREE Full text] [doi: 10.2196/games.3825] [Medline: 25654163]
40. Mol M, van Schaik A, Dozeman E, Ruwaard J, Vis C, Ebert DD, et al. Dimensionality of the system usability scale among professionals using internet-based interventions for depression: a confirmatory factor analysis. BMC Psychiatry 2020 May 12;20(1):218 [FREE Full text] [doi: 10.1186/s12888-020-02627-8] [Medline: 32398111]
41. Hai-Jew S. Alpha testing, beta testing, and customized testing. In: Designing Instruction For Open Sharing. Cham, Switzerland: Springer; 2019.
42. Hedlefs Aguilar MI, De la Garza González A, Sánchez Miranda MP, Garza Villegas AA. AdaptaciÃ³n al espaÃ±ol del cuestionario de usabilidad de sistemas informÃ¡ticos CSUQ / Spanish language adaptation of the Computer Systems Usability Questionnaire CSUQ. RECI 2016 Jan 14;4(8):84-99. [doi: 10.23913/reci.v4i8.35]
43. Benaida M, Namoun A. An exploratory study of the factors affecting the perceived usability of Algerian educational websites. Turk Online J Educ Technol 2018 Apr;17(2):1-12.
44. Lewis JR. Measuring perceived usability: SUS, UMUX, and CSUQ ratings for four everyday products. Int J Hum Comput Interact 2018 Oct 17;35(15):1404-1419. [doi: 10.1080/10447318.2018.1533152]
45. Grime MM, Wright G. Delphi method. In: Wiley StatsRef: Statistics Reference Online. Hoboken, NJ: John Wiley & Sons; 2014.
46. Beiderbeck D, Frevel N, von der Gracht HA, Schmidt SL, Schweitzer VM. Preparing, conducting, and analyzing Delphi surveys: cross-disciplinary practices, new directions, and advancements. MethodsX 2021;8:101401 [FREE Full text] [doi: 10.1016/j.mex.2021.101401] [Medline: 34430297]
47. Hussain A, Hussain F, Shakeel H, Uddin N, Ghouri TL. Unity game development engine: a technical survey. Univ Sindh J Inf Commun Technol 2020 Jul;4(2):73-81 [FREE Full text]
48. Lewis JR. Measuring perceived usability: the CSUQ, SUS, and UMUX. Int J Hum Comput Interact 2018 Jan 09;34(12):1148-1156. [doi: 10.1080/10447318.2017.1418805]
49. Thirkettle M, Lewis J, Langdridge D, Pike G. A mobile app delivering a gamified battery of cognitive tests designed for repeated play (OU Brainwave): app design and cohort study. JMIR Serious Games 2018 Oct 30;6(4):e10519 [FREE Full text] [doi: 10.2196/10519] [Medline: 30377140]
50. Perrochon A, Kemoun G, Dugué B, Berthoz A. Cognitive impairment assessment through visuospatial memory can be performed with a modified walking Corsi test using the 'magic carpet'. Dement Geriatr Cogn Dis Extra 2014 Jan 30;4(1):1-13 [FREE Full text] [doi: 10.1159/000356727] [Medline: 24575112]
51. Fisk AD, Rogers WA, Charness N, Czaja SJ, Sharit J. Designing for Older Adults: Principles and Creative Human Factors Approaches, Second Edition. Boca Raton, FL: CRC Press; 2009.
52. Zhang BY, Chignell M. A person-centered design framework for serious games for dementia. Proc Hum Factors Ergon Soc Annu Meet 2021 Feb 09;64(1):18-22 [FREE Full text] [doi: 10.1177/1071181320641005]
53. Zichermann G, Cunningham C. Gamification by Design: Implementing Game Mechanics in Web and Mobile Apps. Sebastopol, CA: O'Reilly Media; 2011.
54. Kessels RP, van Zandvoort MJ, Postma A, Kappelle LJ, de Haan EH. The Corsi block-tapping task: standardization and normative data. Appl Neuropsychol 2000 Dec;7(4):252-258. [doi: 10.1207/s15324826an0704_8]
55. Kessels RP, van den Berg E, Ruis C, Brands AM. The backward span of the Corsi Block-Tapping Task and its association with the WAIS-III Digit Span. Assessment 2008 Dec;15(4):426-434. [doi: 10.1177/1073191108315611] [Medline: 18483192]
56. Sauro J, Lewis JR. Quantifying the User Experience: Practical Statistics for User Research. Burlington, MA: Morgan Kaufmann; 2016.
57. Bastien JC, Scapin DL. A validation of ergonomic criteria for the evaluation of human - computer interfaces. Int J Hum Comput Interact 1992 Apr;4(2):183-196. [doi: 10.1080/10447319209526035]
58. Yusoff MS. ABC of content validation and content validity index calculation. Educ Med J 2019 Jun 28;11(2):49-54. [doi: 10.21315/eimj2019.11.2.6]
59. Polit DF, Beck CT, Owen SV. Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. Res Nurs Health 2007 Aug;30(4):459-467. [doi: 10.1002/nur.20199] [Medline: 17654487]
60. Aiken LR. Content validity and reliability of single items or questionnaires. Educ Psychol Meas 1980 Dec 01;40(4):955-959. [doi: 10.1177/001316448004000419]
61. Zamanzadeh V, Ghahramanian A, Rassouli M, Abbaszadeh A, Alavi-Majd H, Nikanfar AR. Design and implementation content validity study: development of an instrument for measuring patient-centered communication. J Caring Sci 2015 Jun;4(2):165-178 [FREE Full text] [doi: 10.15171/jcs.2015.017] [Medline: 26161370]

XSL•FO
RenderX

62. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing. 2021. URL: https://www.R-project.org [accessed 2025-03-18]
63. Aiken LR. Three coefficients for analyzing the reliability and validity of ratings. Educ Psychol Meas 1985 Mar 01;45(1):131-142. [doi: 10.1177/0013164485451012]
64. Khaleghi A, Aghaei Z, Mahdavi MA. A gamification framework for cognitive assessment and cognitive training: qualitative study. JMIR Serious Games 2021 May 18;9(2):e21900 [FREE Full text] [doi: 10.2196/21900] [Medline: 33819164]
65. Censo nacional de población y vivienda 2018. Departamento Administrativo Nacional de Estadística. 2018. URL: https://www.dane.gov.co/index.php/estadisticas-por-tema/demografia-y-poblacion/censo-nacional-de-poblacion-y-vivenda-2018 [accessed 2025-03-18]
66. Valladares-Rodriguez S, Perez-Rodriguez R, Facal D, Fernandez-Iglesias MJ, Anido-Rifon L, Mouriño-Garcia M. Design process and preliminary psychometric study of a video game to detect cognitive impairment in senior adults. PeerJ 2017;5:e3508 [FREE Full text] [doi: 10.7717/peerj.3508] [Medline: 28674661]
67. Valladares-Rodriguez S, Fernández-Iglesias MJ, Anido-Rifón L, Facal D, Rivas-Costa C, Pérez-Rodríguez R. Touchscreen games to detect cognitive impairment in senior adults. A user-interaction pilot study. Int J Med Inform 2019 Jul;127:52-62. [doi: 10.1016/j.ijmedinf.2019.04.012] [Medline: 31128832]
68. Kumar KV, Saiteja C, Anjali M, Reddy KR, Gothane GS. Cognitive game using unity. Int J Creat Res Thoughts 2021 Jun;9(6):b8-13 [FREE Full text]
69. Starrett MJ, McAvan AS, Huffman DJ, Stokes JD, Kyle CT, Smuda DN, et al. Landmarks: a solution for spatial navigation and memory experiments in virtual reality. Behav Res Methods 2021 Jun 16;53(3):1046-1059 [FREE Full text] [doi: 10.3758/s13428-020-01481-6] [Medline: 32939682]
70. Galant-Gołębiewska M, Mika B, Maciejewska M. Analysis of the impact of task difficulty on the operator's workload level. Aviation 2022;26(2) [FREE Full text] [doi: 10.3846/aviation.2022.16808]
71. Stein C, O'Keeffe F, McManus C, Tubridy N, Gaughan M, McGuigan C, et al. Premorbid cognitive functioning influences differences between self-reported cognitive difficulties and cognitive assessment in multiple sclerosis. J Neuropsychol 2024 Mar 22;18(1):47-65. [doi: 10.1111/jnp.12327] [Medline: 37212461]
72. Coolican H. Research Methods and Statistics in Psychology, Sixth edition. London, UK: Psychology Press; 2014.
73. Högberg J, Hamari J, Wästlund E. Gameful Experience Questionnaire (GAMEFULQUEST): an instrument for measuring the perceived gamefulness of system use. User Model User Adapt Interact 2019 Feb 28;29(3):619-660. [doi: 10.1007/s11257-019-09223-w]

Abbreviations

AD: Alzheimer disease
CSUQ: Computer System Usability Questionnaire
gHGT: gamified version of a hidden goal task
MCI: mild cognitive impairment
SC: spatial cognition
SDLC: software development life cycle
SG: serious game
VG-CSUQ: Computer System Usability Questionnaire for videogame
VR: virtual reality

Edited by A Coristine; submitted 05.09.24; peer-reviewed by R Zambrano, W Qi; comments to author 19.10.24; revised version received 20.12.24; accepted 23.12.24; published 02.04.25.

Please cite as:
Sanchez-Escudero JP, Aguillon D, Valencia S, Garcia-Barrera MA, Aguirre-Acevedo DC, Trujillo N
Digital Ergonomics of NavegApp, a Novel Serious Game for Spatial Cognition Assessment: Content Validity and Usability Study
JMIR Serious Games 2025;13:e66167
URL: https://games.jmir.org/2025/1/e66167
doi:10.2196/66167
PMID:

©Juan Pablo Sanchez-Escudero, David Aguillon, Stella Valencia, Mauricio A Garcia-Barrera, Daniel Camilo Aguirre-Acevedo, Natalia Trujillo. Originally published in JMIR Serious Games (https://games.jmir.org), 02.04.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious

XSL•FO
RenderX

Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Effects of Computerized Cognitive Training on Vesicular Acetylcholine Transporter Levels using [18F]Fluoroethoxybenzovesamicol Positron Emission Tomography in Healthy Older Adults: Results from the Improving Neurological Health in Aging via Neuroplasticity-based Computerized Exercise (INHANCE) Randomized Clinical Trial

Mouna Attarha[1], PhD; Ana de Figueiredo Pelegrino[2], MSc; Lydia Ouellet[2], MS; Paule-Joanne Toussaint[2], PhD; Sarah-Jane Grant[1], MA; Thomas Van Vleet[1], PhD; Etienne de Villers-Sidani[2], MDCM

[1]Posit Science Corporation, San Francisco, CA, United States

[2]Department of Neurology and Neurosurgery, McGill University, Montreal, QC, Canada

Corresponding Author:
Mouna Attarha, PhD
Posit Science Corporation
160 Pine Street Suite 200
San Francisco, CA, 94111
United States
Phone: 1 415 394 3100 ext 505
Email: mouna.attarha@positscience.com

Abstract

Background: The cholinergic system mediates essential aspects of cognitive function, yet its structure and function decline progressively with age, by an estimated 2.5% per decade across the lifespan. Cognitive training may help counteract age-related declines in cholinergic functioning and slow associated deficits in cognitive performance.

Objective: This study aims to evaluate whether cognitive training modifies cholinergic binding in older adults.

Methods: The Improving Neurological Health in Aging via Neuroplasticity-based Computerized Exercise (INHANCE) trial is a double-blind randomized controlled trial assessing whether 2 computerized cognitive training programs modify cholinergic expression. The intent-to-treat (ITT) population included 92 community-dwelling healthy older adults aged 65 and above (enrolled July 2021-December 2023; final follow-up June 2024). Participants were randomized at McGill University to either an intervention of speed-based cognitive training exercises designed to improve the speed and accuracy of information processing or an active control of nonspeeded games designed for entertainment (eg, similar in design to Solitaire). Participants completed 35 hours of training on their assigned program at home over a 10-week period using a loaned or personal internet-connected device. Cholinergic binding was measured with the vesicular acetylcholine transporter ligand [^{18}F]fluoroethoxybenzovesamicol (FEOBV) and positron emission tomography (PET). The primary outcome was mean FEOBV binding (standard uptake value ratios [SUVRs]) within the anterior cingulate cortex from baseline to posttest in the ITT population. All other end points were exploratory.

Results: Among the 92 participants in the ITT population (mean age 71.9 years; mean education 16.5 years; 61/92, 66%, women; 88/92, 96%, White), 82 (89%) completed all study activities. The speed-based intervention showed a significant within-group increase in FEOBV binding in the primary region of interest, the anterior cingulate cortex (SUVR change mean +0.044, 95% CI 0.006-0.082, P=.03, medium effect size, ω^2=0.09). The p24c subregion demonstrated a significant between-groups effect favoring speed training (speeded vs nonspeeded SUVR change difference +0.058, 95% CI 0.007-0.110, P=.03, small effect size, ω^2=0.05). Prespecified exploratory analyses revealed significant within-group effects for speed training in the hippocampus (P=.02) and parahippocampal gyrus (P=.04). No effects on FEOBV binding were observed in the active control group.

Conclusions: INHANCE is the largest FEOBV-PET trial to date and demonstrates, for the first time in humans, that speed training can reverse losses in cholinergic terminal densities in brain regions vulnerable to age-related cognitive decline. The 2.3% gain in FEOBV binding in the anterior cingulate achieved over a 10-week intervention may offset the estimated 2.5% decline

XSL•FO
RenderX

typically observed over a decade of natural aging. These findings clarify the neurochemical basis of cognitive training benefits, showing that speed training upregulates binding in networks that support attention, memory, and executive function.

Trial Registration: ClinicalTrials.gov NCT04149457; https://clinicaltrials.gov/study/NCT04149457

International Registered Report Identifier (IRRID): RR2-10.2196/59705

(JMIR Serious Games 2025;13:e75161) doi:10.2196/75161

KEYWORDS

fluoroethoxybenzovesamicol; positron emission tomography; PET; FEOBV-PET; cognitive training; brain training; neuroplasticity; neuromodulatory signaling; aging; anterior cingulate

Introduction

Background

Aging is the strongest known risk factor for dementia, with the risk doubling approximately every 5 years beyond the age of 65 [1]. At the neurological level, aging induces significant system-wide alterations, including decreased or impaired signaling from neuromodulatory centers that gate synaptic plasticity [2-6], expansion of spatiotemporal receptive-field size [7], increased temporal integration windows [8], and elevated neuronal noise [9]. These changes hinder the speed and accuracy of information processing across sensory modalities and compromise the precision and reliability of representational input, which in turn affects the operation of higher-order cognitive functions [10]. Deficits in learning, memory, and executive function are the end products of neurochemical downregulation and the slowed, imprecise functioning of sensory networks that have lost the ability or plasticity to adapt to environmental demands [11].

Research in both animal and human models has consistently highlighted acetylcholine as a pivotal mediator of synaptic plasticity essential for cognition [5,12-14]. Experimental evidence shows that stimulating the release of acetylcholine, such as through the nucleus basalis, improves cognitive processes such as attention, making it an important regulator of learning rate, memory formation, and memory retrieval across sensory modalities [2,3]. Conversely, lesioning or inhibiting acetylcholine impairs plasticity and cognitive performance [4]. Disruptions in cholinergic signaling are linked to age-related memory deficits that precede the early stages of pathological cognitive decline. These disruptions initiate a signaling cascade that interacts with neuropathologies such as amyloidogenesis and tau phosphorylation, ultimately contributing to cholinergic network atrophy, a hallmark of neurodegenerative diseases [15-21].

Clinically, therapeutic approaches aimed at augmenting cholinergic function with cholinesterase inhibitors (donepezil, rivastigmine, and galantamine) have been used to treat mild cognitive impairment (MCI), Alzheimer disease, and related dementias [22], underscoring the significance of neuromodulators such as acetylcholine in preserving the chemical integrity of the brain [6,23-25]. Recent advancements now enable noninvasive, in vivo measurement of cholinergic network integrity in the human brain using positron emission tomography (PET). The radioligand [^{18}F]fluoroethoxybenzovesamicol (FEOBV) selectively binds to the vesicular acetylcholine transporter (VAChT), a protein uniquely expressed by cholinergic neurons [26,27]. The binding pattern across cortical and subcortical areas is consistent with the organization of the cholinergic system and aligns with immunohistology from prior postmortem human studies [28]. Declines in binding are detectable in brain regions known to have reduced cholinergic terminal densities due to aging [29], MCI [30], and dementia [31].

Aging is associated with significant declines in FEOBV binding, with reductions estimated at 2.5% per decade between the ages of 20 and 80 in the anterior cingulate cortex [29], a region that plays a critical role in selective attention [32], learning and memory [33], and executive function [34]. Metabolic dysfunction in this area frequently precedes the earliest cognitive changes observed across the lifespan, and the degree of hypometabolism has been correlated with greater cognitive decline in otherwise healthy adults [35]. Moreover, increased anterior cingulate thickness has been linked to successful cognitive aging [36,37], whereas atrophy in this region serves as a predictor of future dementia development [38].

Decades of research in animal models have shown that cholinergic function can be upregulated by specific behavioral strategies such as cognitive training. Identifying effective behavioral interventions that support cognition in older adults has become an important area of study in recent years [39]. The National Academies of Sciences, Engineering, and Medicine (NASEM) published a review on primary prevention factors and interventions that may independently delay, slow, or prevent cognitive decline. Their report arrived at 3 recommendations (cognitive training, maintaining normal blood pressure, and physical exercise) as having sufficient evidence that “the public should at least have access to these results to help inform their decisions about how they can invest their time and resources to maintain brain health with aging” [40]. In the same year, the American Academy of Neurology (AAN) convened an expert panel and updated its guidelines, stating that clinicians may recommend cognitive training for those with MCI [41]. Most recently, the World Health Organization (WHO) published its 2024 practice guidelines recommending cognitive training as an evidence-based intervention for individuals with dementia [42].

The collective results of the largest trial in cognitive training to date, the Advanced Cognitive Training for Independent and Vital Elderly (ACTIVE), provide strong evidence that a specific type of cognitive training yields sustained cognitive and functional benefits that generalize beyond the trained task. A

XSL•FO
RenderX

selective attention visual speed-of-processing training exercise, referred to as "speed training" in the original publications (now Double Decision, part of the BrainHQ program), reduced dementia incidence by 29%-48% over a decade-long follow-up, depending on the number of training hours completed [43]. Additional studies of speed training have reported improved driving safety [44] and a 48% reduction in at-fault motor vehicle collisions [45], improved balance [46] and a 31% reduction in falls among those at high risk [47], slowed decline in instrumental activities of daily living maintained for 10 years postintervention [48-52], a 68% greater likelihood of improved locus of control maintained over 5 years [53], decreased predicted health care payer–related costs [54], a 38% reduction in the onset of age-related depression [55], a 30% reduction in depressive symptoms [56], and a reduced risk of global decline in health-related quality of life maintained at the 5-year follow-up [57,58].

Another form of training, called Freeze Frame (BrainHQ), was designed to engage tonic and phasic alertness to naturally upregulate neuromodulatory control and improve the accuracy of information processing under speeded conditions. Initial studies demonstrated benefits in executive function, skill acquisition, and spatial and nonspatial attention [59-63], as well as enhancement of the training gains achieved with Double Decision [59]. In a pilot substudy of the phase II ALERT trial, 5 healthy older adults training on Freeze Frame showed a 16%-24% increase in forebrain cholinergic neurotransmission, as measured by FEOBV-PET. These increases in binding paralleled behavioral gains on a sustained vigilance assessment [64].

The design principles of these speeded cognitive training exercises are grounded in a theoretical framework discussed previously [11,65,66]. The neurobiological mechanisms underlying the generalized benefits of these forms of cognitive training, however, remain largely unknown. Several pilot studies have provided initial insights, suggesting that potential mechanisms may include improved functional connectivity within the default mode and central executive networks [67-69], enhanced brain synchronization [70], improved functional magnetic resonance imaging (MRI) and electroencephalography measures of brain activation [69-72], increased hippocampal activation [73], and improved diffusion tensor imaging measures of white matter insulation between brain regions involved in visual and attentional processing [74].

At the core of these network-based changes is synaptic plasticity, a fundamental process regulated by multiple neuromodulatory centers in the brain, including the basal forebrain, which sends cholinergic projections throughout the brain [4,5]. These neuromodulatory systems in general, and cholinergic systems in particular, have long been considered potential drivers of the benefits of computerized training programs and a novel mechanism through which such programs may support brain health and cognition. With the recent development of the FEOBV-PET method for quantitatively assessing cholinergic system integrity in humans, and evidence that analogous forms of cognitive training in rodents alter cholinergic function [75], we hypothesized that cognitive training exercises in humans (Double Decision and Freeze Frame) would affect cholinergic systems. Such findings would inform the neural mechanisms underlying speed-based cognitive training programs and open new avenues of research for developing effective cognitive training interventions.

The Proposed Study

The Improving Neurological Health in Aging via Neuroplasticity-based Computerized Exercise (INHANCE) trial is a randomized controlled study of healthy, community-dwelling older adults aged 65 and above. Its objective was to advance understanding of the mechanisms underlying cognitive training–related benefits in aging by examining neurochemical dynamics associated with cognition. Specifically, INHANCE aims to (1) evaluate, on a primary basis, the effects of 2 computerized training programs on cholinergic binding using FEOBV-PET; (2) assess cognitive and behavioral performance on an exploratory basis; and (3) determine the maintenance of training effects.

Participants were assigned either to speeded cognitive training (Double Decision and Freeze Frame exercises) or to an active control condition involving nonspeeded games designed for entertainment (eg, spin-offs of Solitaire and Candy Crush, titled Double Klondike Solitaire and Bricks Breaking Hex, respectively). The primary end point was FEOBV binding (standard uptake value ratios [SUVRs]) in the anterior cingulate cortex from baseline to posttest for the intent-to-treat (ITT) population. We hypothesized that speed training would result in greater FEOBV binding at posttest compared with baseline.

All other end points were exploratory. Given the extensive projections of cholinergic neurons from the basal forebrain to the limbic system and neocortex, FEOBV SUVRs were evaluated across additional regions of interest: global cortex, frontal lobe, parietal lobe, occipital lobe, temporal lobe, hippocampus, parahippocampal gyrus, striatum, putamen, caudate, posterior cingulate cortex, primary auditory cortex, primary sensorimotor cortex, and the nucleus basalis of Meynert. Participants also completed 3 subtests from the National Institutes of Health The Executive Abilities: Measures and Instruments for Neurobehavioral Evaluation and Research (NIH EXAMINER) cognitive battery to evaluate cognitive control, 2 train-to-task assessments to evaluate target engagement [76,77], and 2 behavioral assessments sensitive to cholinergic function (heart rate variability and pupillometry). All cognitive and behavioral assessments were administered at baseline, posttest, and at a no-contact 3-month follow-up to measure the maintenance of observed effects.

Methods

Ethical Considerations

The study was conducted in accordance with the Declaration of Helsinki and approved by the Western Institutional Review Board (IRB00000533) and the Research Ethics Board of McGill University Health Centre (2020-6474). The FEOBV radioligand was approved by Health Canada (Control #252085). All participants provided written informed consent (see Multimedia Appendix 1).

XSL•FO
RenderX

The consent form described the study as evaluating the efficacy of 2 cognitive training interventions: one focused on "speed" and the other on "executive function." To maintain blinding, both programs were accessed by participants through the same commercial website (web) and commercial app (mobile).

Study data were recorded into a secure, web-based electronic case report form at the study site through the Longitudinal Online Research and Imaging System (LORIS). This system complies with relevant privacy and security standards for electronic trial data entry and storage, as well as the Health Insurance Portability and Accountability Act and the Personal Information Protection and Electronic Documents Act standards for confidentiality and privacy [78]. Following consent, each participant was assigned a standardized Participant Identification Number, composed of digits identifying the study and digits identifying the participant. All electronic case report form entries were deidentified, using the Participant Identification Number rather than the participant's name.

Participants received US $30 (CAD $40) for completing the baseline visit, which included PET and MRI (visit 1); US $10 (CAD $13) for every 10 training sessions completed during the intervention period (up to US $70 [CAD $91] for all 70 sessions); US $30 (CAD $40) for completing the posttest visit, which included PET and MRI (visit 2); and US $30 (CAD $40) for the 3-month end-of-study follow-up visit (visit 3). Participants who completed all study activities were reimbursed a total of US $160 (CAD $211). Payments were provided following completion of each visit. Additional details are available in the study protocol [79].

Study Design

INHANCE is a prospective, double-blind, parallel-arm, active-controlled randomized clinical trial in healthy adults aged 65 and above. Participants were randomized to receive 35 hours of either a computerized visual speed-based cognitive training program (BrainHQ) or an active control consisting of visual nonspeeded computerized games over a 10-week period. A detailed description of the outcome measures, training programs, and statistical analysis plan is provided in the study protocol [79] (see also Multimedia Appendix 2).

Screening

Following informed consent, potential participants completed an in-person structured interview and neuropsychological assessments to determine study eligibility. The structured interview (approximately 20 minutes, in person) was conducted by a research coordinator and included demographic information (eg, year of birth, age, education), medical diagnoses, and current medications, guided by administrator-facing source documents. Neuropsychological assessments were participant-facing. The Montreal Cognitive Assessment (MoCA; about 10 minutes) evaluated global cognitive function across visuospatial processing, executive functioning, naming, memory, attention, language, abstraction, delayed recall, and orientation, yielding a total score of 30, with higher scores indicating better performance [80]. The Geriatric Depression Scale—Short Form (GDS-SF; about 7 minutes) assessed self-reported depressive symptoms, with participants responding yes or no to 15 items about how they felt during the past week [81,82]. The Columbia Suicide Severity Rating Scale (about 10 minutes) assessed the severity and intensity of suicidal ideation, as well as suicidal behavior and lethality [83] (see Multimedia Appendix 3). For additional details on the validation, administration, and adjudication of screening measures, see the study protocol [79].

Participants and Inclusion and Exclusion Criteria

Participants were community-dwelling healthy older adults. Inclusion criteria required individuals to be aged 65 or older, proficient in English or French, capable of fulfilling study requirements, and cognitively intact, defined as an MoCA total score of ≥23, a cut-off that optimizes diagnostic sensitivity and specificity [84]. Exclusion criteria included neurocognitive disorders, suicidal ideation, major depression indicated by a GDS-SF total score >10, prior experience using BrainHQ within the past 5 years, pregnancy, substance abuse, or contraindications to neuroimaging. Concurrent participation in clinical trials involving investigational devices, use of medications with established cholinergic effects [85], or medical conditions that could hinder study engagement were also exclusionary at the discretion of the site investigator (EDVS). Full details are provided in the study protocol [79].

Protocol Changes

Enrollment spanned from July 2021 to December 2023. In October 2022, the protocol and recruitment flyer were revised to state, "Potential participant must be able to communicate in either English or French" (originally, "Potential participant must be a fluent English or French speaker from the age of 12"), as the original wording unintentionally excluded minority groups. This modification to the inclusion criteria was approved by the study's Data Safety Monitoring Board, the Western Institutional Review Board, and the Research Ethics Board of McGill University Health Centre in January 2023.

Recruitment

All participants were recruited near McGill University, Canada, where the FEOBV radiotracer was synthesized and administered. Recruitment methods included public presentations (television, radio, conferences), word of mouth, and flyers. Flyers were posted in churches, community centers, Facebook groups, local libraries, and the neurology clinic at the Montreal Neurological Hospital. They described the study and provided information on inclusion and exclusion criteria, along with the study team's phone number and email address. All study flyers received ethics approval before distribution (see Multimedia Appendix 4).

Randomization

We employed a minimization method of adaptive stratified randomization with a 1:1 allocation to either the intervention or active control group. Groups were stratified based on baseline FEOBV and the baseline NIH EXAMINER composite score. After completion of baseline assessments, a blinded team member at the study site initiated a randomization request by entering both scores and setting the randomization status to "Awaiting Review" in LORIS, a data collection and management system with strict data access controls [86]. The request was then directed to an unblinded team member at the

XSL•FO
RenderX

coordinating center (SJG), who reviewed and confirmed the accuracy of the participant's scores and set the randomization status to "Reviewed," triggering an automated script. The script cumulatively assessed overall study imbalance, and the final group assignment was made at random with a strong probability weighting (0.8) favoring the group that would minimize imbalance. Once the assignment was determined, an automated email containing the participant's group allocation was sent from LORIS to an unblinded site team member, who then conducted the program orientation visit. All randomizations were completed before the participant's first day of program use.

Intervention and Active Control Training Programs

Participants received either visual speed training (Double Decision and Freeze Frame exercises) or an active control consisting of visual nonspeeded computer games designed for entertainment (Double Klondike Solitaire and Bricks Breaking Hex). Both training programs involved 35 hours of computerized training delivered over a 10-week period (approximately 30 minutes per session, 7 sessions per week, for a total of 70 sessions). Training was completed remotely at participants' homes using either a loaned Android tablet or a personal internet-connected device. Participants accessed their assigned program via an identical website or app and were blinded to group assignment (see Figure 1).

Figure 1. Intervention and active control programs. The intervention included 2 tasks: Double Decision, a dual-task paradigm in which participants discriminate a centrally presented visual stimulus while simultaneously locating a target in the peripheral visual field across 40 unique levels, with display exposure duration as the adaptive dimension; and Freeze Frame, a speeded reverse go/no-go paradigm in which participants remember a target image presented at trial onset, followed by a continuous stream of targets and foils with unequal probability, withholding responses to targets and responding quickly to foils, with difficulty as the adaptive dimension. The active control included spin-offs of casual games: Double Klondike Solitaire, in which participants move cards to 8 foundations by suit from Ace to King, and Bricks Breaking Hex, in which participants eliminate groups of same-colored bricks by clicking them.



The active control was designed to account for nonspecific treatment effects, including placebo and practice effects, face validity, interactions with research personnel, computer experience, and was matched to the intervention in overall program use intensity, visual modality, platform delivery, reward delivery, and remuneration of time [87-89]. The control program was strategy-based and did not include speeded elements hypothesized to drive positive neuroplastic changes in the intervention. Control games were developed by an independent gaming company and rated E (for Everyone) by the Entertainment Software Rating Board.

Outcomes

All outcomes were measured in person by blinded assessors. The primary outcome was FEOBV binding, quantified using mean SUVRs in the primary region of interest (ROI), the

XSL•FO
RenderX

anterior cingulate cortex, between baseline and posttest for the ITT population.

FEOBV binding was also assessed in prespecified exploratory regions of interest, including the global cortex (frontal, parietal, occipital, and temporal lobes), hippocampus, parahippocampal gyrus, striatum (putamen and caudate), posterior cingulate cortex, primary auditory cortex, primary sensorimotor cortex, and the nucleus basalis of Meynert, which provides the major source of cholinergic innervation to the cerebral cortex. In the nucleus basalis of Meynert, FEOBV SUVR likely reflects a combination of tracer binding to cholinergic cell bodies and cholinergic brain stem afferents terminating in this region.

A structural MRI (3T Siemens MAGNETOM Prisma; 20 minutes) was acquired during the first scanning session to coregister the PET images (Siemens High-Resolution Research Tomograph; see Multimedia Appendix 2). Cognitive control was assessed using 3 computerized subtests (Flanker, Set-Shifting, and Anti-Saccades) from a validated neuropsychological battery, with performance summarized as the *z* score of the executive composite from NIH EXAMINER [90] (20 minutes). Participants also completed 2 train-to-task cognitive assessments (the Double Decision assessment and the Freeze Frame assessment [15 minutes]) modeled directly on the 2 intervention exercises, with performance summarized as *z* scores to evaluate target engagement. Behavioral measures sensitive to acetylcholine function were also collected, including heart rate variability (using a Shimmer 4-lead Consensys ECG Development Kit) and pupillometry (using Tobii Pro Glasses 2). All cognitive and behavioral assessments were conducted at baseline, posttest, and at a 3-month no-contact follow-up. Additional details are provided in the study protocol [79].

Power and Sample Size Calculations

Based on the 0.27 SD observed in anterior cingulate FEOBV binding in a single-arm, open-label pilot study [64] (N=5), detecting a small-to-moderate effect size (17% increase in anterior cingulate FEOBV binding postintervention) with 80% power and a significance level of .05 required 40 participants per group (80 total). Assuming 15% attrition, at least 92 participants were enrolled to ensure a minimum of 80 completers to adequately power the primary outcome. No interim analyses or stopping guidelines were planned.

Statistical Analyses

A detailed scan protocol, including data acquisition and processing, is provided in the published protocol [79] (see also Multimedia Appendix 2). Briefly, the statistical analysis plan defined a primary ITT population, a single primary outcome measure, a set of exploratory outcome measures, a primary evaluation time point, an exploratory evaluation time point, a primary statistical analysis methodology, criteria for statistical significance, and guidance for result interpretation.

The ITT population was defined a priori as participants who were randomized and completed at least one training session postrandomization (30 minutes).

The primary statistical analysis employed a linear mixed-effects model. Intervention and active control groups in the ITT population were first compared at baseline to identify potential covariates [91]. Each outcome measure was then analyzed using a linear mixed-effects model with treatment group and time as fixed factors, incorporating covariates as needed based on the baseline analysis. Within-group effects at each time point (posttraining and follow-up) were calculated using data from each group, while between-group effects were assessed by including an interaction term (training group × time) to estimate the effect of cognitive training on outcome changes.

Analyses in this study were conducted in accordance with the published protocol, with the exception of additional analyses suggested by anonymous reviewers of related INHANCE trial publications, including a baseline cognition subgroup analysis using the MoCA.

Results

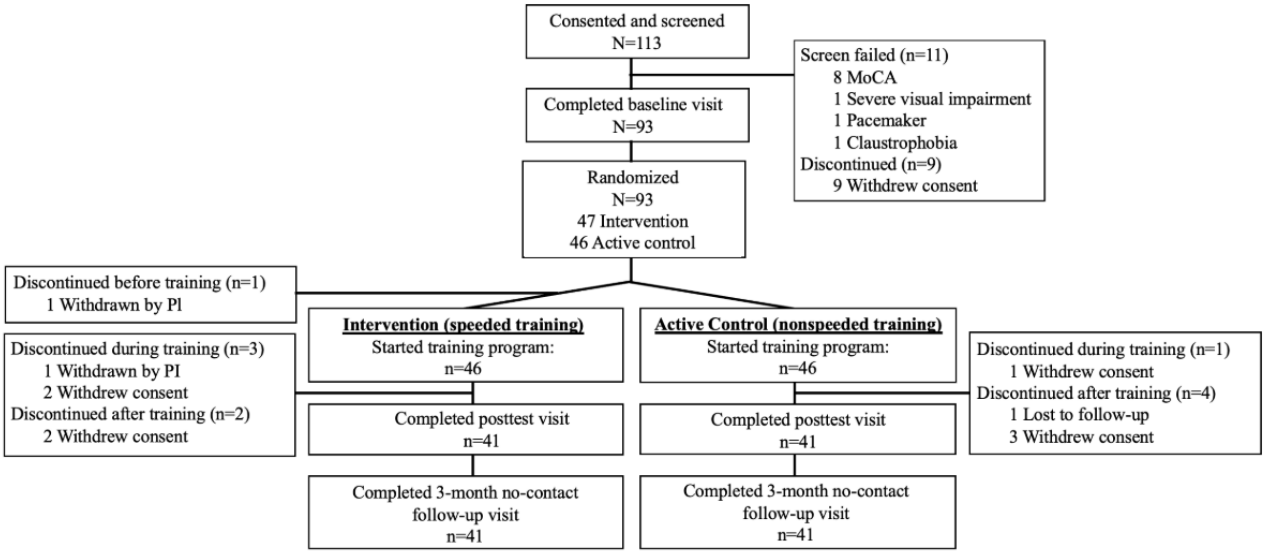
Enrollment

The study was conducted from 2019 to 2024 at McGill University, Canada. Recruitment occurred from July 2021 to December 2023, and the final follow-up visit took place on June 7, 2024.

A total of 113 individuals were screened and provided written informed consent; of these, 20 did not qualify or chose not to participate further. Ninety-three participants completed baseline assessments and were randomized to either the intervention group (speeded training, n=47) or the active control group (nonspeeded training, n=46), with 1 participant withdrawn by the site principal investigator (EDVS) before engaging with the intervention. The ITT population (N=92) included 46 participants in each arm and was analyzed (mean age 71.9 years; mean education 16.5 years; 61/92, 66%, women; and 88/92, 96% White). Overall, 41 out of 46 participants in each arm (or 82/92 combined, 89%) completed both the posttest assessments and the 3-month no-contact follow-up visit. See Figure 2 for the CONSORT (Consolidated Standards of Reporting Trials) flowchart and Multimedia Appendix 5 for the CONSORT checklist.

XSL•FO
RenderX

Figure 2. CONSORT (Consolidated Standards of Reporting Trials) flowchart. Participant flow through the trial. The intention-to-treat (ITT; N=92) population was defined a priori as all randomized participants who completed at least one training session. MoCA: Montreal Cognitive Assessment; PI: principal investigator.



Baseline Characteristics of the ITT Population

Participant characteristics for the ITT population are presented in Table 1. Baseline characteristics were well-balanced between the intervention and active control groups, differing statistically only for the baseline Double Decision assessment, which was included as a covariate in the model (t_{90}=–2.073, *P*=.04). Overall, participants were cognitively normal, without depression, college-educated, and predominantly female and White. Additional baseline characteristics of the sample have been reported previously [92].

Participants randomized to the intervention group (n=46) had a mean NIH EXAMINER executive composite score of 0.38 (SD 0.64). On the Double Decision train-to-task assessment, the mean raw score was 1632.46 (SD 883.24) ms, with lower scores indicating better performance. On the Freeze Frame train-to-task assessment, the mean raw score was 4.22 (SD 1.99), with higher scores indicating better performance.

Participants randomized to the active control group (n=46) had a mean NIH EXAMINER executive composite score of 0.49 (SD 0.50). On the Double Decision train-to-task assessment, the mean raw score was 1320.93 (SD 712.12) ms, and on the Freeze Frame train-to-task assessment, the mean raw score was 5.11 (SD 2.09).

XSL•FO
RenderX

Table 1. Baseline demographic, inclusion, and outcome measures of the ITT[a] population (N=92) by treatment group.

Baseline characteristics	Intervention (n=46)	Active control (n=46)	*P* value
Demographics			
Age, mean (SD), years	72.76 (5.25)	71.04 (4.32)	.09
Education, mean (SD), years	16.28 (3.53)	16.63 (3.30)	.63
Gender (female), n (%)	29 (63)	32 (70)	.66
Race (White), n (%)	44 (96)	44 (96)	.26
Ethnicity (Hispanic or Latino), n (%)	0 (0)	0 (0)	.99
Native language (English), n (%)	17 (37)	16 (35)	.62
Inclusion			
MoCA[b] (total score), mean (SD)	26.37 (1.83)	25.98 (1.84)	.31
GDS-SF[c] (total score), mean (SD)	1.24 (1.85)	1.52 (1.91)	.47
Primary measure			
FEOBV[d] anterior cingulate cortex, SUVR[e] (SD)	1.896 (0.202)	1.911 (0.186)	.72
Exploratory measures			
FEOBV global cortex (SUVR), mean (SD)	1.416 (0.164)	1.459 (0.122)	N/A[f]
FEOBV frontal lobe (SUVR), mean (SD)	1.648 (0.203)	1.693 (0.139)	N/A
FEOBV parietal lobe (SUVR), mean (SD)	1.450 (0.177)	1.465 (0.149)	N/A
FEOBV occipital lobe (SUVR), mean (SD)	1.172 (0.187)	1.238 (0.149)	N/A
FEOBV temporal lobe (SUVR), mean (SD)	1.395 (0.194)	1.440 (0.152)	N/A
FEOBV hippocampus (SUVR), mean (SD)	2.094 (0.315)	2.165 (0.255)	N/A
FEOBV parahippocampal gyrus (SUVR), mean (SD)	1.361 (0.235)	1.413 (0.179)	N/A
FEOBV striatum (SUVR), mean (SD)	5.972 (0.850)	6.220 (0.770)	N/A
FEOBV putamen (SUVR), mean (SD)	6.938 (1.043)	7.207 (0.905)	N/A
FEOBV caudate (SUVR), mean (SD)	5.006 (0.801)	5.232 (0.696)	N/A
FEOBV posterior cingulate (SUVR), mean (SD)	1.942 (0.192)	1.939 (0.205)	N/A
FEOBV primary auditory cortex (SUVR), mean (SD)	1.958 (0.317)	2.071 (0.284)	N/A
FEOBV primary sensorimotor cortex (SUVR), mean (SD)	1.737 (0.227)	1.717 (0.201)	N/A
FEOBV nucleus basalis of Meynert (SUVR), mean (SD)	3.281 (0.568)	3.391 (0.453)	N/A
NIH EXAMINER[g,h] (composite *z* score), mean (SD)	–0.10 (1.12)	0.10 (0.87)	.34
Double decision train-to-task assessment (*z* score), mean (SD)	–0.19 (1.09)	0.19 (0.88)	.04
Freeze frame train-to-task assessment (*z* score), mean (SD)	–0.22 (0.96)	0.21 (1.01)	.07

[a]ITT: intent to treat.

[b]MoCA: Montreal Cognitive Assessment.

[c]GDS-SF: Geriatric Depression Scale—Short Form.

[d]FEOBV: [^{18}F]fluoroethoxybenzovesamicol.

[e]SUVR: standard uptake value ratio.

[f]N/A: not applicable.

[g]NIH EXAMINER: National Institutes of Health The Executive Abilities: Measures and Instruments for Neurobehavioral Evaluation and Research.

[h]Higher scores are better.

Adherence to the Training Programs

Adherence was high throughout the trial, with 89 (97%) participants of the ITT population (N=92) completing the minimum of 10 hours of training, as informed by the ACTIVE study [43,49,93]. Among those randomized to the intervention group (n=46), 38 (83%) completed the prescribed 35 or more hours of training, with a mean of 40.81 hours. In the active control group (n=46), 35 (76%) completed the prescribed 35 or more hours, with a mean of 65.80 hours.

XSL•FO
RenderX

Missing Data

A total of 10 (11%) participants from the ITT population (N=92) had missing MRI/PET, NIH EXAMINER, and train-to-task assessment data at posttest and follow-up due to study withdrawal before the posttest visit (5 from the intervention group and 5 from the active control group).

In addition to the participants who withdrew, data from a small number of participants were missing or invalid for technical reasons. Of note, 3 participants (3%) who completed the study had exceptionally poor atlas alignment, preventing valid calculation of FEOBV SUVRs (1 intervention participant at baseline and posttest; 2 controls at posttest only), 2 participants (2%) had corrupted MRI files at posttest (1 intervention and 1 control), 3 participants (3%) had invalid NIH EXAMINER scores at the 3-month follow-up due to administration of the incorrect form (2 intervention and 1 control), and 2 participants (2%) had a missing Freeze Frame train-to-task assessment due to technical issues (1 intervention at baseline and 1 control at post-test). The statistical model used iterative full-information maximum likelihood estimation to account for missing data.

Training Effects on the Primary Outcome Measure

Table 2 presents FEOBV-PET SUVRs in the primary ROI, the anterior cingulate cortex, for both the intervention and active control groups at baseline and posttest. A significant within-group increase in SUVR was observed for the speeded intervention (mean change +0.044, 95% CI 0.006-0.082, *P*=.03, medium effect size ω^2=0.09). No significant within-group change was observed for the nonspeeded active control (mean change +0.014, 95% CI –0.026 to 0.054, *P*=.50, effect size ω^2<0.01). Tracer uptake in the white matter reference region did not differ between conditions (mean change –2.953, 95% CI –9.636 to 3.718, *P*=.39, effect size ω^2<0.01).

Figure 3 illustrates the change in SUVR from baseline to posttest for the intervention and active control groups. Separate SUVR images for each time point are provided in Figure S1 in Multimedia Appendix 2.

XSL•FO
RenderX

Table 2. Outcome measure analysis. Within- and between-group analyses for the ITT[a] population (N=92) at posttest (visit 2) and follow-up (visit 3) for the intervention and active control groups.

Measures	Intervention within-group differences			Active Control within-group differences			Between groups differences	
	Visit 2 baseline, mean (SD)	Visit 2-visit 1 change, mean (95% CI), *P* value, effect size (ω^2)[b]	Visit 3-visit 1 change, mean (95% CI), *P* value, effect size (ω^2)	Visit 1 baseline, mean (SD)	Visit 2-visit 1 change, mean (95% CI), *P* value, effect size (ω^2)	Visit 3-visit 1 change, mean (95% CI), *P* value, effect size	Visit 2-visit 1 change difference (95% CI), *P* value, effect size	Visit 3-visit 1 change difference (95% CI), *P* value, effect size
Primary measure								
FEOBV[c] anterior cingulate cortex (SUVR[d])	1.896 (0.202)	+0.044 (0.006 to 0.082), .03[e], 0.09	N/A[f]	1.911 (0.186)	+0.014 (–0.026 to 0.054), .49, <0.01	N/A	+0.030 (–0.025 to 0.085), .28, <0.01	N/A
Exploratory measures								
FEOBV global cortex (SUVR)	1.416 (0.164)	+0.026 (–0.007 to 0.059), .13, 0.03	N/A	1.459 (0.122)	+0.011 (–0.020 to 0.042), .48, <0.01	N/A	+0.015 (–0.0230 to 0.060), .52, <0.01	N/A
FEOBV frontal lobe (SUVR)	1.648 (0.203)	+0.033 (–0.013 to 0.080), .17, 0.02	N/A	1.693 (0.139)	+0.009 (–0.023 to 0.040), .58, <0.01	N/A	+0.024 (–0.031 to 0.081), .40, <0.01	N/A
FEOBV parietal lobe (SUVR)	1.450 (0.177)	–0.003 (–0.035 to 0.028), .83, <0.01	N/A	1.465 (0.149)	+0.004 (–0.017 to 0.025), .71, <0.01	N/A	–0.008 (–0.045 to 0.030), .70, <0.01	N/A
FEOBV occipital lobe (SUVR)	1.172 (0.187)	+0.033 (–0.005 to 0.072), .10, 0.04	N/A	1.238 (0.149)	+0.012 (–0.033 to 0.057), .59, <0.01	N/A	+0.021 (–0.037 to 0.080), .49, <0.01	N/A
FEOBV temporal lobe (SUVR)	1.395 (0.194)	+0.043 (–0.002 to 0.089), .07, 0.06	N/A	1.440 (0.152)	+0.016 (–0.030 to 0.062), .49, <0.01	N/A	+0.027 (–0.037 to 0.092), .42, <0.01	N/A
FEOBV hippocampus (SUVR)	2.094 (0.315)	+0.089 (0.015 to 0.163), .02[g], 0.10	N/A	2.165 (0.255)	+0.026 (–0.051 to 0.101), .50, <0.01	N/A	+0.061 (–0.044 to 0.167), .26, <0.01	N/A
FEOBV parahippocampal gyrus (SUVR)	1.361 (0.235)	+0.059 (0.005 to 0.114), .04[g], 0.08	N/A	1.413 (0.179)	+0.021 (–0.032 to 0.072), .43, <0.01	N/A	+0.037 (–0.037 to 0.112), .33, <0.01	N/A
FEOBV striatum (SUVR)	5.972 (0.850)	+0.082 (–0.088 to 0.250), .34, <0.01	N/A	6.220 (0.770)	+0.050 (–0.088 to 0.188), .47, <0.01	N/A	+0.035 (–0.181 to 0.250), .75, <0.01	N/A
FEOBV putamen (SUVR)	6.938 (1.043)	+0.108 (–0.105 to 0.319), .32, <0.01	N/A	7.207 (0.905)	+0.067 (–0.097 to 0.231), .42, <0.01	N/A	+0.044 (–0.221 to 0.308), .75, <0.01	N/A
FEOBV caudate (SUVR)	5.006 (0.801)	+0.061 (–0.082 to 0.201), .40, <0.01	N/A	5.232 (0.696)	+0.032 (–0.089 to 0.153), .60, <0.01	N/A	+0.029 (–0.156 to 0.213), .76, <0.01	N/A

XSL•FO
RenderX

Measures	Intervention within-group differences			Active Control within-group differences			Between groups differences	
	Visit 2 baseline, mean (SD)	Visit 2-visit 1 change, mean (95% CI), *P* value, effect size (ω^2)[b]	Visit 3-visit 1 change, mean (95% CI), *P* value, effect size (ω^2)	Visit 1 baseline, mean (SD)	Visit 2-visit 1 change, mean (95% CI), *P* value, effect size (ω^2)	Visit 3-visit 1 change, mean (95% CI), *P* value, effect size	Visit 2-visit 1 change difference (95% CI), *P* value, effect size	Visit 3-visit 1 change difference (95% CI), *P* value, effect size
FEOBV posterior cingulate (SUVR)	1.942 (0.192)	–0.002 (–0.031 to 0.027), .89, <0.01	N/A	1.939 (0.205)	+0.010 (–0.022 to 0.041), .55, <0.01	N/A	–0.012 (–0.054 to 0.031), .59, <0.01	N/A
FEOBV primary auditory cortex (SUVR)	1.958 (0.317)	+0.005 (–0.049 to 0.059), .86, <0.01	N/A	2.071 (0.284)	+0.026 (–0.021 to 0.073), .28, <0.01	N/A	–0.021 (–0.092 to 0.050), .56, <0.01	N/A
FEOBV primary sensorimotor cortex (SUVR)	1.737 (0.227)	–0.002 (–0.032 to 0.027), .88, <0.01	N/A	1.717 (0.201)	+0.012 (–0.015 to 0.040), .38, <0.01	N/A	–0.014 (–0.054 to 0.026), .48, <.01	N/A
FEOBV nucleus basalis of Meynert (SUVR)	3.281 (0.568)	+0.086 (–0.026 to 0.197), .14, 0.03	N/A	3.391 (0.453)	+0.044 (–0.064 to 0.153), .42, <0.01	N/A	+0.041 (–0.112 to 0.194), .60, <0.01	N/A
NIH EXAMINER[h,i] (composite *z* score)	–0.10 (1.12)	+0.132 (–0.091 to 0.358), .25, <0.01	+ 0.166 (–0.075 to 0.411), .18, 0.02	0.10 (0.87)	+0.091 (–0.131 to 0.314), .42, <0.01	+0.163 (–0.058 to 0.385), .15, 0.03	+0.049 (–0.264 to 0.364), .76, <0.01	+0.006 (–0.319 to 0.333), .97, <0.01
Double Decision train-to-task assessment (*z* score)	–0.19 (1.09)	+1.703 (1.390 to 2.021), <.001[g], 0.73	+1.666 (1.352 to 1.986), <.001[g], 0.72	0.19 (0.88)	+0.174 (–0.166 to 0.512), .32, <0.01	+0.486 (0.245 to 0.726), <.01[g], 0.25	+1.53 (1.073 to 1.989), <.001[g], 0.32	+1.177 (0.787 to 1.570), <.001[g], 0.28
Freeze Frame train-to-task assessment (*z* score)	–0.22 (0.96)	+1.066 (0.763 to 1.363), <.001[g], 0.54	+1.179 (0.918 to 1.435), <.001[g], 0.67	0.21 (1.01)	+0.171 (–0.140 to 0.481), .28, <0.01	+0.294 (–0.002 to 0.585), .05, 0.06	+0.903 (0.477 to 1.327), <.001[g], 0.16	+0.893 (0.508 to 1.277), <.001[g], 0.19

[a]ITT: intent to treat.

[b]Effect sizes are ω^2 (partial) with effect size<0.01 considered "very small," 0.01≤effect size<0.06 considered "small," 0.06≤effect size<0.14 considered "medium," and effect size≥0.14 considered "large."

[c]FEOBV: [^{18}F]fluoroethoxybenzovesamicol.

[d]SUVR: standard uptake value ratio.

[e]*P*<.05.

[f]N/A: not applicable.

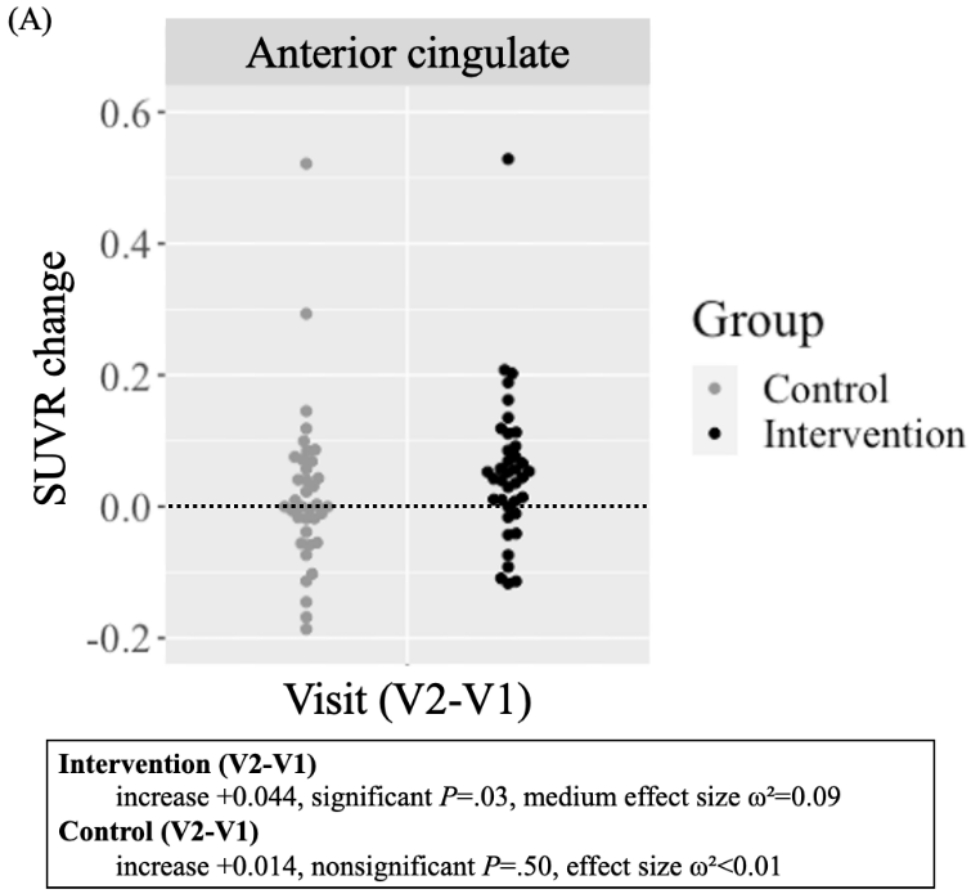
[g]*P*<.05.

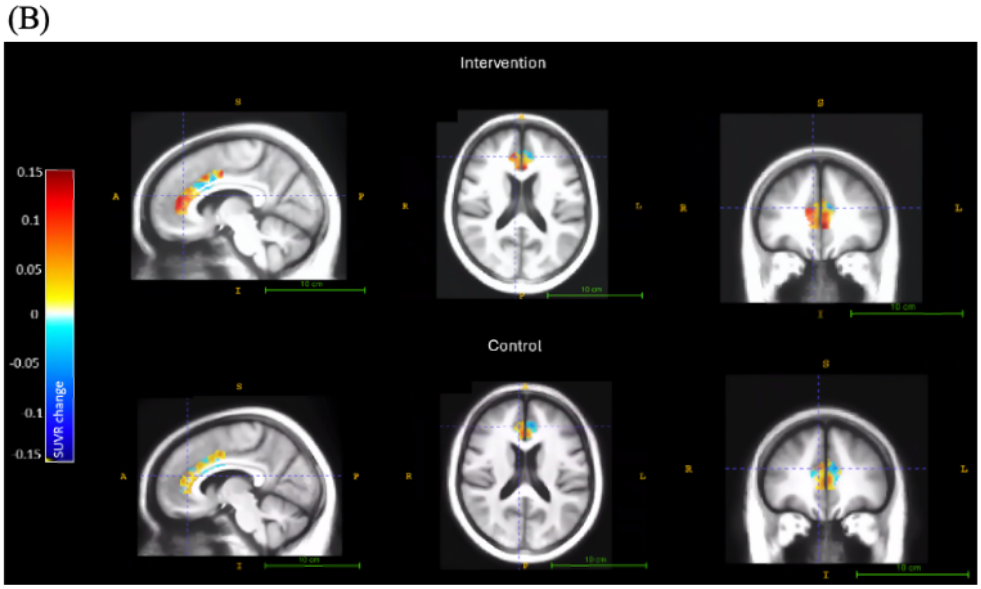
[h]NIH EXAMINER: National Institutes of Health The Executive Abilities: Measures and Instruments for Neurobehavioral Evaluation and Research.

[i]Higher scores are better for all measures.

XSL•FO
RenderX

Figure 3. Standard uptake value ratio (SUVR) change within the primary region of interest (ROI). (A) SUVR binding change in the anterior cingulate at posttest (visit 2) relative to baseline (visit 1), with each icon representing a single participant's change score. Higher scores indicate greater increases in binding. (B) SUVR binding change in the anterior cingulate cortex at posttest relative to baseline, averaged across participants in the intervention (top) and active control (bottom). Warmer colors indicate greater increases in binding. A: anterior; I: inferior; L: left; P: posterior; R: right; S: superior.





The anterior cingulate cortex is a large, functionally heterogeneous region involved in a range of cognitive functions, including attention, action, and social cognition. We evaluated 3 established subregions: p24c, p24ab, and p32 [94]. The p24c and p24ab subregions share a similar functional profile, particularly in reward-related tasks, with p24c specifically linked to action inhibition. By contrast, p32 is primarily engaged during tasks involving emotion regulation and theory of mind. The p24c subregion showed a significant between-group effect favoring the intervention, indicating that speed training increased cholinergic binding more than the active control (speeded vs nonspeeded SUVR change difference +0.058, 95% CI 0.007-0.110, P=.03, small effect size ω^2=0.05). A significant within-group effect was observed for the speeded intervention

XSL•FO
RenderX

(mean change +0.061, 95% CI 0.019-0.103, *P*=.006, large effect size ω^2=0.15) but not for the active control (mean change +0.002, 95% CI –0.028 to 0.033, *P*=.90, effect size ω^2<0.01). For the p24ab subregion, a significant within-group effect was observed for the intervention (mean change +0.070, 95% CI 0.020-0.119, *P*=.008, large effect size ω^2=0.14) but not for the control group (mean change +0.022, 95% CI –0.030 to 0.073, *P*=.41, effect size ω^2<0.01). The p32 subregion showed no significant within- or between-group effects.

Training Effects on Exploratory Outcomes Measures

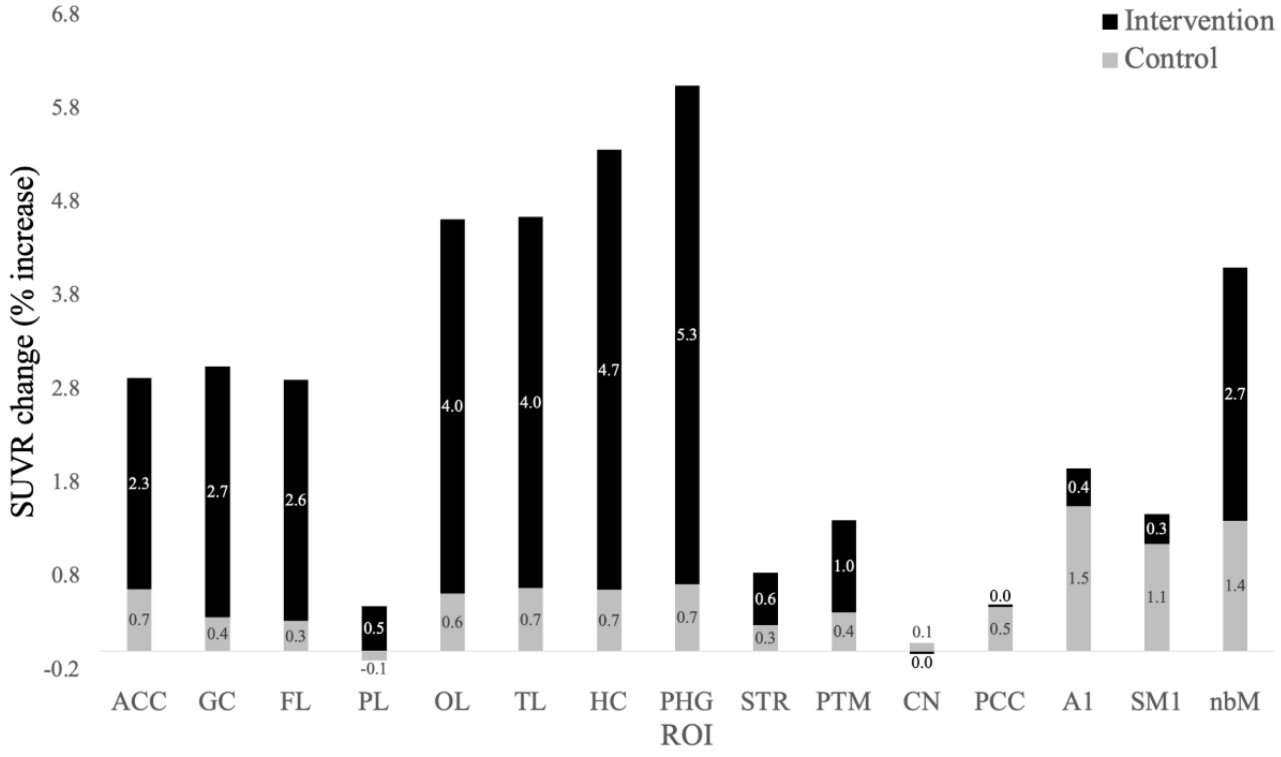
Table 2 presents FEOBV-PET SUVRs in exploratory ROIs for the intervention and active control groups at baseline, baseline versus posttest, and baseline versus follow-up. Significant within-group effects were observed for the speeded intervention in brain regions associated with memory, including the hippocampus (mean change +0.089, 95% CI 0.015-0.163, *P*=.02, medium effect size ω^2=0.10) and the parahippocampal gyrus (mean change +0.059, 95% CI 0.005-0.114, *P*=.04, medium effect size ω^2=0.08). A nonsignificant medium effect was observed for the temporal lobe (mean change +0.043, 95% CI –0.002 to 0.089, *P*=.07, effect size ω^2=0.06). Nonsignificant small effect sizes were observed for the global cortex (mean change +0.026, 95% CI –0.007 to 0.059, *P*=.13, effect size ω^2=0.03), frontal lobe (mean change +0.033, 95% CI –0.013 to 0.080, *P*=.17, effect size ω^2=0.02), occipital lobe (mean change +0.033, 95% CI –0.005 to 0.072, *P*=.10, effect size ω^2=0.04), and another region (mean change +0.086, 95% CI –0.026 to 0.197, *P*=.14, effect size ω^2=0.03).

There were no significant within-group changes in FEOBV SUVR for the active control across any of the exploratory ROIs. For the speeded intervention, no within-group effects were observed in the parietal lobe, striatum, putamen, caudate, posterior cingulate cortex, primary auditory cortex, or primary sensorimotor cortex. Additionally, no significant between-group effects were detected across the exploratory ROIs.

Change scores, expressed as percentages, were calculated across the primary and exploratory ROIs for both the intervention and active control groups (Figure 4). The speeded intervention produced a 2.3% increase in FEOBV SUVR in the anterior cingulate cortex (vs 0.7% in the active control), a 4.7% increase in the hippocampus (vs 0.7% in the control), and a 5.3% increase in the parahippocampal gyrus (vs 0.7% in the control).

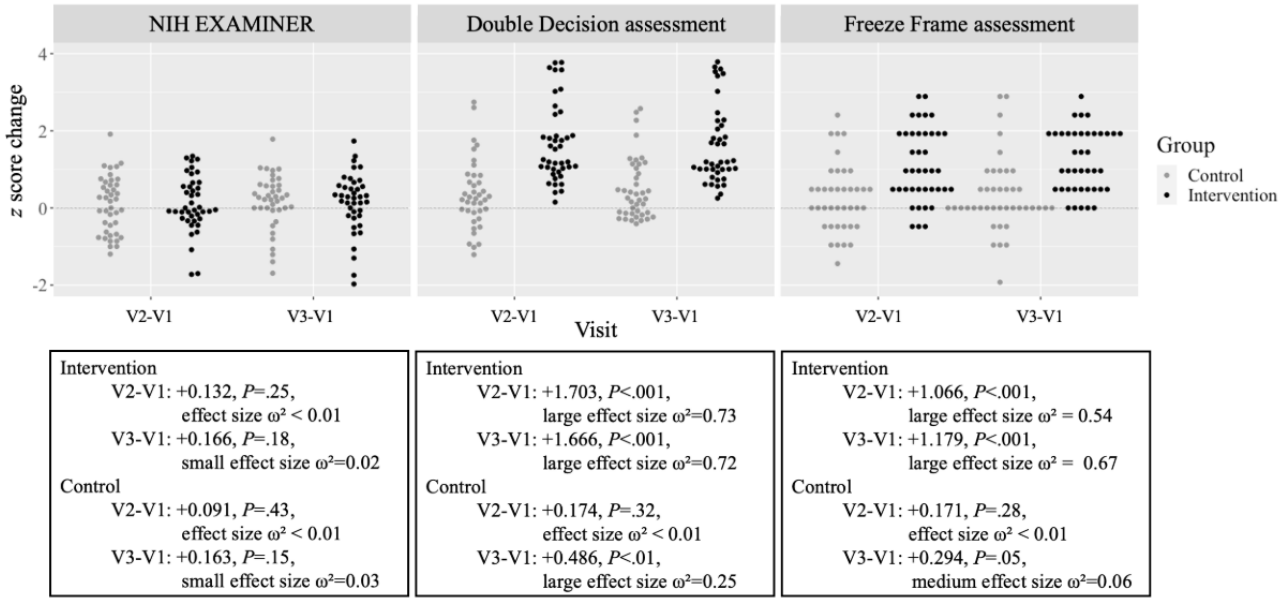
Three exploratory cognitive measures were administered at baseline, posttest, and follow-up (Figure 5). Performance on the cognitive control subtests of the NIH EXAMINER battery (Flanker, Set-Shifting, and Anti-Saccades) showed no significant within- or between-group effects at either posttest or follow-up.

Figure 4. Standard uptake value ratio (SUVR) binding percent increases. Percent change in SUVR binding from baseline to posttest across regions of interest (ROIs) for the intervention and active control groups. Higher scores indicate greater increases in binding. A1: primary auditory cortex; ACC: anterior cingulate cortex; CN: caudate; FL: frontal lobe; GC: global cortex; HC: hippocampus; nbM: nucleus basalis of Meynert; OL: occipital lobe; PCC: posterior cingulate cortex; PHG: parahippocampal gyrus; PL: parietal lobe; PTM: putamen; SMI: primary sensorimotor cortex; STR: striatum; TL: temporal lobe.



XSL•FO
RenderX

Figure 5. Effects on cognitive measures. Cognitive outcomes (z score change) at posttest (V2) and follow-up (V3) relative to baseline (V1) for National Institutes of Health The Executive Abilities: Measures and Instruments for Neurobehavioral Evaluation and Research (NIH EXAMINER), the cognitive assessment for Double Decision, and the cognitive assessment for Freeze Frame. Each icon represents a single participant's change score. Higher scores indicate better cognitive performance across all measures. v: visit.

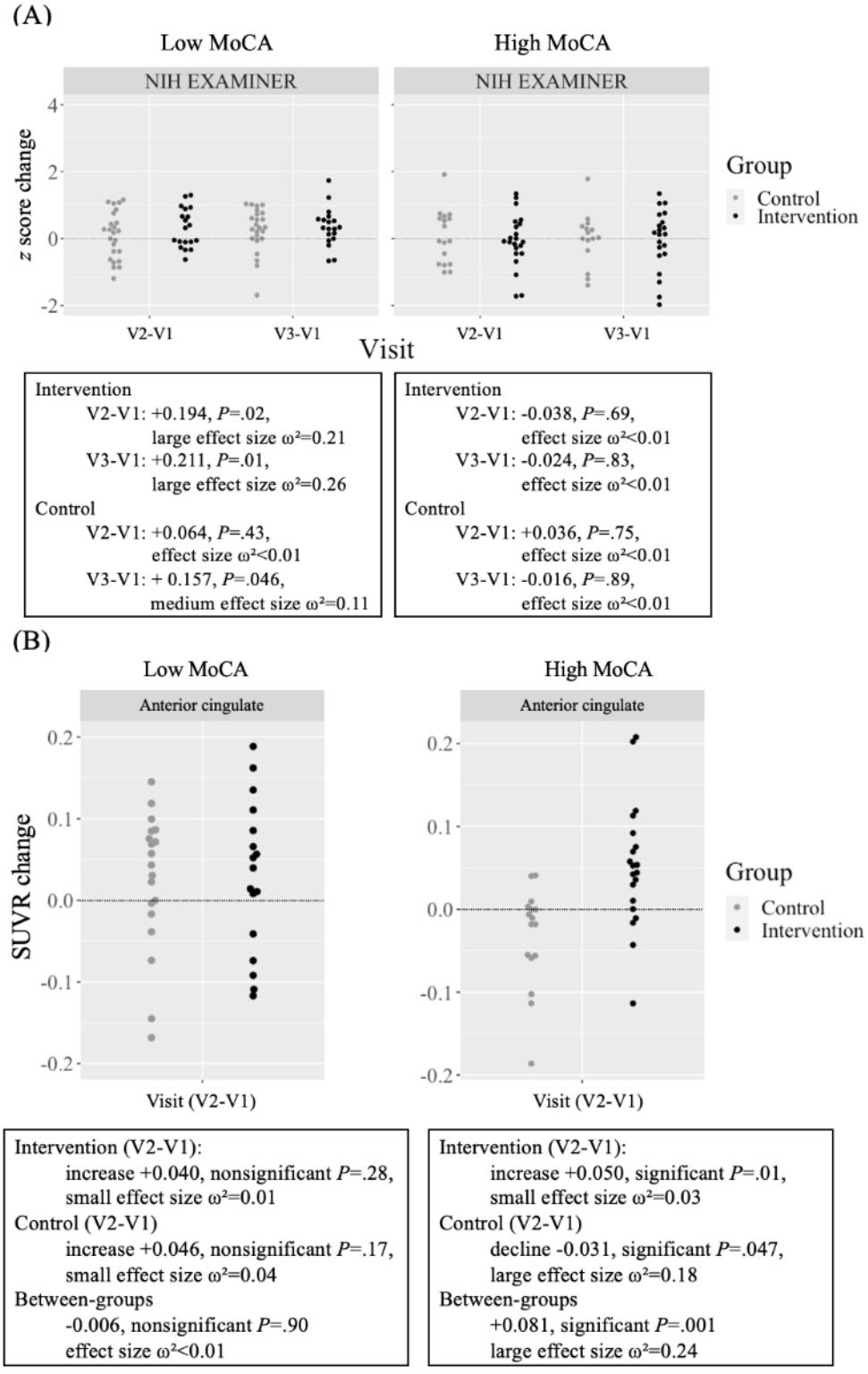


The Double Decision and Freeze Frame assessments, designed to evaluate target engagement, demonstrated significant effects (P<.001). Speeded training produced a significant between-group effect on the Double Decision train-to-task assessment (speeded vs nonspeeded z score change difference +1.53, 95% CI 1.073-1.989, P<.001, large effect size ω^2=0.32) and on the Freeze Frame train-to-task assessment (speeded vs nonspeeded z score change difference +0.903, 95% CI 0.477-1.327, P<.001, large effect size ω^2=0.16) at posttest. These effects were maintained at the 3-month no-contact follow-up for both the Double Decision assessment (speeded vs nonspeeded z score change difference +1.177, 95% CI 0.787-1.570, P<.001, large effect size ω^2=0.28) and the Freeze Frame assessment (speeded vs nonspeeded z score change difference +0.893, 95% CI 0.508-1.277, P<.001, large effect size ω^2=0.19). These results confirm that the training effectively engaged the target and that improvements from a brief, intensive training period were sustained after the intervention concluded.

We explored the effect of baseline cognition using the MoCA, defining "low cognition" as a total score of 23-26 (n=51, 55%, in the ITT) and "high cognition" as a total score of 27-30 (n=41, 45%, in the ITT). At baseline, the low cognition subgroup had a mean NIH EXAMINER executive composite score of 0.09 (SD 0.62), which was notably lower than the mean composite score of 0.65 (SD 0.53) observed in the high cognition subgroup. There was a within-group effect of speed training on the NIH EXAMINER executive composite in the low cognition subgroup (composite z score change mean +0.194, 95% CI 0.042-0.351, P=.02, large effect size ω^2=0.21), but not in the high cognition subgroup (composite z score change mean –0.038, 95% CI –0.230 to 0.151, P=.69, effect size ω^2 < 0.01). No within-group effects were observed for the active control in either the low cognition subgroup (composite z score change mean +0.064, 95% CI –0.092 to 0.223, P=.42, effect size ω^2<0.01) or the high cognition subgroup (composite z score change mean +0.036, 95% CI –0.187 to 0.259, P=.75, effect size ω^2<0.01; see Figure 6A).

XSL•FO
RenderX

Figure 6. Cognitive outcomes and SUVR binding. (A) Cognitive outcomes (z score change) at posttest (visit 2) and follow-up (visit 3) relative to baseline (visit 1) in participants with low (MoCA 23-26, N=51) and high (MoCA 27-30, N=41) baseline cognition, as measured by the NIH EXAMINER. Each icon represents the change score of a single participant. Higher scores indicate better cognitive performance across all measures. (B) Change in SUVR binding at posttest (visit 2) relative to baseline (visit 1) in the anterior cingulate for the low- and high-baseline cognition subgroups. MoCA: Montreal Cognitive Assessment; NIH EXAMINER: National Institutes of Health The Executive Abilities: Measures and Instruments for Neurobehavioral Evaluation and Research; SUVR: standard uptake value ratio; v: visit.



The effect of speed training was maintained at the 3-month no-contact follow-up, demonstrating sustained cognitive benefits after training (composite *z* score change mean +0.211, 95% CI 0.064-0.365, *P*=.01, large effect size ω²=0.26). The active control also showed a modest improvement relative to baseline at the 3-month follow-up (composite *z* score change mean +0.157, 95% CI 0.009-0.308, *P*=.046, medium effect size ω²=0.11). Given that no significant cognitive improvement was observed immediately posttraining (*P*=.43), this change does not reflect a maintenance effect. It may instead represent either

XSL•FO
RenderX

a practice effect on the NIH EXAMINER due to repeated administrations or a delayed cognitive benefit arising from engagement with the control game training.

In the subgroup with high baseline cognition, there was a significant between-group effect favoring the speeded intervention in the primary ROI (speeded vs nonspeeded SUVR change difference +0.081, 95% CI 0.036-0.125, *P*=.001, large effect size ω^2=0.24; see Figure 6B). Within-group analyses showed that the intervention group had a significant increase in binding from baseline to posttest (change mean +0.050; 95% CI 0.016-0.082, *P*=.007, small effect size ω^2=0.03), whereas the active control group demonstrated a significant decline in binding over the same period (change mean –0.031, 95% CI –0.060 to –0.002, *P*=.047, large effect size ω^2=0.18). In the subgroup with low baseline cognition, there were no significant between-group effects (speeded vs nonspeeded SUVR change difference –0.006, 95% CI –0.100 to 0.088, *P*=.90, effect size ω^2<0.01). Likewise, there were no significant within-group effects for either the intervention (change mean +0.040, 95% CI –0.032 to 0.112, *P*=.28, small effect size ω^2=0.01) or the control group (change mean +0.046, 95% CI –0.019 to 0.109, *P*=.17, small effect size ω^2=0.04).

Adverse Events

Participants reported a total of 54 adverse events, of which 4 (7%) were related to the administration of the radioligand. One participant in the active control group reported 2 mild events (dry mouth and unusual sensations in the mouth), and a second participant who was not randomized reported 2 moderate events (nausea and vomiting). Both participants recovered without treatment. No adverse events were attributed to the training programs, and no serious adverse events or deaths occurred.

Discussion

Principal Findings

Age-related declines in neuromodulatory function are closely linked to cognitive decline. In this double-blind, randomized controlled trial, community-dwelling healthy older adults aged 65 and above were assigned to 1 of 2 computerized cognitive training programs: speeded training or nonspeeded training. FEOBV-PET was used to determine whether these training programs could modulate cholinergic function. In the largest FEOBV-PET trial to date, we demonstrate for the first time in humans that speed training significantly increases cholinergic function in the anterior cingulate cortex, our primary ROI. The 2.3% increase in FEOBV SUVR binding observed over the 10-week intervention is comparable in magnitude, but opposite in direction, to the estimated 2.5% decline in anterior cingulate FEOBV SUVR binding [29] per decade of aging.

The speeded intervention showed clear task-based functional associations with the p24c and p24ab subregions of the anterior cingulate. Both exercises involved reward-based tasks performed under conditions of selective attention. One task, in particular, relied heavily on action inhibition, requiring participants to make rapid motor responses to visual foils while withholding responses to specific targets. The p24c and p24ab subregions have been shown to coactivate with elements of the salience and central executive networks, potentially providing a mechanistic explanation for prior findings that speed training enhances functional connectivity within these networks [68]. As expected, no effects were observed in the p32 subregion, consistent with the fact that the training did not involve social theory-of-mind tasks.

The cholinergic system has widespread projections throughout the brain. Exploratory analyses revealed SUVR increases in the hippocampus and parahippocampal gyrus, regions particularly susceptible to age-related declines in cholinergic terminal integrity [95]. These findings suggest potential improvements in memory-related processes and may help explain the memory benefits of speed training reported in previous studies [96]. For example, the well-known ACTIVE trial [97] demonstrated that healthy older adults who completed 10-18 hours of Double Decision training experienced a 29%-48% reduction in dementia risk a decade later [43]. This reduction in dementia risk may reflect the long-term benefits of increased cholinergic function through speed training.

Previous studies using FEOBV-PET have shown that healthy older adults experiencing natural age-related cognitive decline [29], as well as individuals with MCI [30] or dementia [31], share a downregulation of the cholinergic system, albeit to varying degrees. From the perspective of the brain as an information-processing system, this shared neural phenotype suggests that interventions aimed at improving cholinergic neurotransmission could support cognitive health in these populations, regardless of differences in severity or diagnostic classification. Such interventions may help preserve function in healthy aging, slow cognitive decline in MCI, or potentially attenuate progression in dementia. By targeting fundamental neural systems, this approach provides a framework for understanding and addressing cognitive decline across different stages of aging and neurodegeneration.

Our understanding of how computerized cognitive training produces cognitive and functional benefits remains limited. This trial provides direct evidence linking cognitive training to specific changes in brain physiology. Speed training may drive the cognitive and functional improvements observed in the ACTIVE trial [43,97] and other studies in healthy aging [96], MCI [67], and dementia [98,99] by supporting cholinergic network health. This perspective is largely absent from the cognitive training literature, which predominantly emphasizes a neuropsychological framework, focusing on whether training improves specific cognitive functions, such as memory. Incorporating a biological brain health perspective by assessing whether training improves neurological signatures of brain health is essential to align the field of cognitive training with the standards of pharmaceutical drug development.

Improving endogenous cholinergic signaling represents a novel treatment strategy that may offer greater clinical benefits than approaches that rely on exogenously enhancing cholinergic signaling via pharmaceutical therapies. Current treatments for MCI and Alzheimer disease and related dementias include 3 Food and Drug Administration (FDA)–approved cholinesterase inhibitors [100], which increase cholinergic signaling by slowing the breakdown of acetylcholine in and around the synaptic cleft.

XSL•FO
RenderX

These drugs provide benefits that are generally modest and temporary at best, and their efficacy may be limited by their mechanism of action: they prolong the presence of acetylcholine in the synapse but do not restore the amount of acetylcholine released during normal signaling.

The present findings demonstrate increased endogenous VAChT binding in cholinergic terminals of the anterior cingulate, hippocampus, and parahippocampal gyrus following speed training. While VAChT binding does not directly measure acetylcholine concentrations in the synaptic cleft, it indicates an enhanced presynaptic capacity for acetylcholine packaging and release. Several mechanisms may underlie this effect. One potential mechanism is activity-dependent upregulation of VAChT driven by increased neuronal firing and cholinergic demand, consistent with preclinical evidence that heightened cholinergic activity enhances VAChT expression [101]. Another is the induction of trophic signaling: cognitive stimulation in both animal models and humans increases neurotrophic factors such as nerve growth factor [102-108] and brain-derived neurotrophic factor [109-111], which support cholinergic neuron health and can upregulate choline acetyltransferase and VAChT expression. These adaptive changes likely reflect remodeling of existing synapses, including improved vesicle mobilization and transporter availability, rather than synaptogenesis [112-114]. Regardless of the precise biological mechanism underlying the observed increase in binding, these results align with animal studies showing region-specific changes detected by FEOBV-PET, which were subsequently validated through immunohistochemistry following training [75]. Collectively, these findings suggest that cognitive training increases cholinergic signaling capacity, counteracting age-related decline by restoring presynaptic efficiency in circuits supporting attention, memory, and executive function.

The between-group effect size observed in the p24c subregion of the anterior cingulate is consistent with effect sizes reported for cognitive and functional outcomes in meta-analyses. For instance, Basak et al [115] (215 studies) reported benefits in healthy aging (near transfer g=0.38, far transfer g=0.22) and in MCI (near g=0.27, far g=0.18), aligning with prior meta-analyses by Mewborn et al [116] (97 studies; near transfer g=0.44, far transfer g=0.15) and Lampit et al [117] (52 studies; near transfer g=0.22). In patients with dementia, Bahar-Fuchs et al [118] (33 studies) reported significant near transfer benefits (g=0.84), as did Hill et al [119] (17 studies; near transfer g=0.26). Organizations such as NASEM [40], WHO [42], AAN [41], and the Alzheimer's Association [120] have begun to incorporate cognitive training into discussions of brain health, based on meta-analytic evidence in healthy aging, MCI, and dementia [121,122]. Furthermore, given the unmet need for effective interventions across clinical populations, the FDA has indicated that effect sizes of 0.20-0.30 (Cohen d) for a low-risk cognitive training program may provide sufficient efficacy signals to support approval [123]. The speed training intervention used in the ACTIVE trial demonstrated effect sizes within this range, with subsequent analyses showing clinically relevant reductions in dementia risk. In this context, the between-group effect size observed in our study can be considered clinically meaningful.

Exploratory subgroup analyses based on baseline MoCA scores showed that the lack of a within-group effect on NIH EXAMINER (contrary to previous findings [91]) was attributable to differences in baseline cognition. Participants with low baseline cognition demonstrated significant improvements on the cognitive control subtests of NIH EXAMINER (Flanker, Set-Shifting, and Anti-Saccades) following speed training, and these gains were maintained at follow-up. However, the smaller size of this subgroup likely limited statistical power to detect a between-group effect. By contrast, participants with high baseline cognition exhibited a ceiling effect, limiting the potential for measurable improvement on these subtests. These specific tasks may be less sensitive to change in high-functioning older adults [90]. While the low cognition group demonstrated a cognitive benefit, the high cognition group showed between-group differences in FEOBV SUVR within the primary ROI, favoring the speed training intervention. Previous research has documented distinct patterns of FEOBV-PET SUVR decline across healthy and clinical populations, with the most pronounced reductions in healthy adults observed in the anterior cingulate, striatum, and primary sensorimotor cortex. By contrast, individuals with MCI exhibit the greatest declines in the global cortex, occipital lobe, and parietal lobe, whereas those with dementia show the most pronounced reductions in the temporal cortex [29-31,34,35]. The distinct pattern of cholinergic changes observed across the primary and exploratory ROIs between the low and high cognition subgroups aligns with these prior findings. Although further research is needed to fully elucidate this dissociation, we interpret the cognitive gains in the low cognition group and the SUVR changes in the high cognition group as reflecting regional differences in cerebral SUVR topography between these subpopulations, as well as the limited sensitivity of the 3 selected NIH EXAMINER subtests. To evaluate the relationship between changes in FEOBV-PET SUVRs and cognitive performance, future studies should employ a validated, comprehensive cognitive assessment battery (with measures of processing speed, attention, memory, and executive function) and recruit healthy older adults within a narrow high-MoCA range (27-30) when designating the anterior cingulate as the primary ROI.

Strengths

This is the largest FEOBV-PET study to date, conducted within a rigorously designed randomized clinical trial, featuring a well-matched active control and an a priori statistical analysis plan with prespecified ROIs.

In vivo FEOBV-PET provides a novel biomarker for assessing cholinergic function. Our trial not only replicates findings from smaller pilot studies demonstrating that FEOBV-PET can safely localize cholinergic terminals in humans, but also shows that terminal densities can change following a 10-week cognitive training intervention. As an imaging tool specific to neuromodulatory function, FEOBV-PET offers a way to quantify cholinergic status across disease states and may ultimately serve as a valuable method to monitor treatment response.

XSL•FO
RenderX

Limitations

The limited racial and ethnic diversity of trial participants constrains the generalizability of these findings to non-White populations. To address this, we implemented extensive outreach initiatives over the course of the trial, including contacting 20 community centers in Montreal with greater representation, more than 200 churches, and 10 local libraries, as well as participating in multiple media interviews about INHANCE. Despite these and other efforts, most participants identified as White. Barriers to minority representation in clinical trials are well documented and may include mistrust stemming from historical research abuses, time and resource constraints, cultural and language differences, discomfort with the research process, and perceptions of increased personal risk. In the case of INHANCE, the requirement to undergo radioligand imaging and commit to 35 hours of training may have further discouraged participation [124-127]. Future studies should aim to reduce participant burden and implement more equitable research practices [128].

This mechanism paper was motivated by the positive cognitive and functional benefits observed in healthy older adults randomized to speed training in the ACTIVE trial. Although we intended to recruit only healthy older adults, the relatively low MoCA cut-off of 23 [84] may have inadvertently included participants with undiagnosed MCI, thereby introducing additional variance into our trial.

Finally, only known confounders of cognitive training reported in the literature (eg, major depression) were specified in the published protocol. Potential confounders of cholinergic plasticity, such as smoking or systemic estrogen use through hormone replacement therapy, were not accounted for in this study.

Future Directions

Future pivotal trials should evaluate whether speed training increases cholinergic binding in clinically diagnosed MCI populations and determine if FEOBV-PET–measured reductions in binding can predict progression from MCI to dementia, as well as whether this transition can be favorably modified by improving neurotransmission through speed training.

Conclusions

Identifying and implementing effective strategies to support brain health has the potential to reduce health care costs, increase workforce participation and community engagement, and improve quality of life. As neuromodulatory centers of the brain play a causal role in cognitive performance, improving cholinergic signaling may help protect against age-related cognitive decline. This trial demonstrates that an intensive, time-limited period of speed training can enhance neuromodulatory health in community-dwelling adults aged 65 and older, specifically in regions that support selective attention, inhibitory control, and memory. These findings support the use of this promising, low-risk intervention and contribute to a mechanistic understanding of cognitive training that establishes a foundation for future trials.

Acknowledgments

Research reported in this publication was supported by the National Institute on Aging of the National Institutes of Health under Award Numbers R44AG039965 and 3R44AG039965-06S1 (to MA and TVV). The funders had no role in the design or conduct of the study; collection, management, analysis, or interpretation of the data; preparation, review, or approval of the manuscript; or the decision to submit the manuscript for publication. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. We are deeply grateful to the members of McGill's Clinical Research Unit (Abed Kadiry, Angela Genge, Auréline Boucher, Christina Caccese, Cyrena Gerardi, David Costa, Dylan Sembinelli, Farah Naz Sulaiman, Hung-Hsin Hsiao, James Sirhan, Julian Santorelli, Justine Craft, Lindsay Hannah, Luciano Buonamici, Martin Chevrier, Michael Pileggi, Michal Friedman, Niloofar Barzegar, Reda Bouhachi, Robert Le, Romina Perrotti, Ronaldo Lopez, Salma Khalil, Soheil Quchani, Tamim Rezai, and Gabriele Riva); members of Etienne's lab at McGill's The Neuro (Carla Benea, Jonathan Côté, Mike Cisneros-Franco, and Patrice Voss); members of McGill's Longitudinal Online Research and Imaging System (LORIS) team (Laetitia Fesselier, Rida Abou-Haidar, and Samir Das); and members of Bedard's lab at the Université du Québec à Montréal (Rebekah Wickens). At Posit Science, we thank Greg Sabatini, software engineer, for managing the technical implementation and maintenance of the intervention and control programs, and Sawsan Dabit, research coordinator, for overseeing Institutional Review Board–related activities at trial launch. We also thank our Data Safety and Monitoring Board for their oversight: Dr. Feng Vankee Lin (content expert), Dr. Gary Cutter (biostatistician), Dr. Matt Sutterer (National Institutes of Health Program Officer), and Caroline Sferrazza (health specialist). Finally, we are grateful to an anonymous reviewer whose comments inspired the Montreal Cognitive Assessment (MoCA) subgroup analysis. This manuscript was not generated with artificial intelligence.

Data Availability

Beginning May 31, 2026, deidentified participant data from INHANCE will be made available to verified academic and industry researchers. Investigators will be able to conduct both confirmatory and exploratory analyses using the Longitudinal Online Research and Imaging System (LORIS), a freely accessible, open-source data archive with provenance-sharing features. The dataset will be accessible online [129].

XSL•FO
RenderX

Authors' Contributions

Conceptualization: MA (equal), EDVS (equal), TVV (equal)
Data curation: MA (equal), ADFP (equal), SJG (supporting), LO (supporting)
Formal analysis: ADFP (equal), MA (equal), PJT (supporting), EDVS (supporting)
Funding acquisition: MA (equal), TVV (equal)
Investigation: SJG (lead), LO (supporting), PJT (supporting)
Methodology: MA (equal), EDVS (equal), TVV (supporting), ADFP (supporting), PJT (supporting)
Project administration: SJG (lead), LO (supporting), PJT (supporting)
Resources: MA (equal), EDVS (equal), TVV (supporting)
Software: MA (lead), SJG (supporting)
Validation: SJG (lead)
Supervision: MA (equal), EDVS (equal), TVV (supporting)
Visualization: ADFP (equal), MA (equal)
Writing – original draft: MA (lead)
Writing – review and editing: EDVS (equal), ADFP (equal), TVV (supporting)

Conflicts of Interest

MA, TVV, and SJG are employees at Posit Science, the company that develops BrainHQ's cognitive training and assessment programs. They are shareholders of Posit Science stock. EDVS, ADFP, PJT, and LO declare no competing interests.

Multimedia Appendix 1
Participant consent forms in English and French.
[PDF File (Adobe PDF File), 968 KB - games_v13i1e75161_app1.pdf]

Multimedia Appendix 2
Positron emission tomography and magnetic resonance imaging acquisition and data analysis.
[PDF File (Adobe PDF File), 1839 KB - games_v13i1e75161_app2.pdf]

Multimedia Appendix 3
Participant-facing questionnaires in English and French.
[PDF File (Adobe PDF File), 841 KB - games_v13i1e75161_app3.pdf]

Multimedia Appendix 4
Study flyers in English and French.
[PDF File (Adobe PDF File), 2718 KB - games_v13i1e75161_app4.pdf]

Multimedia Appendix 5
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1110 KB - games_v13i1e75161_app5.pdf]

References

1. Jorm AF, Jolley D. The incidence of dementia: a meta-analysis. Neurology 1998 Sep 01;51(3):728-733. [doi: 10.1212/wnl.51.3.728] [Medline: 9748017]
2. Kilgard MP, Merzenich MM. Cortical map reorganization enabled by nucleus basalis activity. Science 1998 Mar 13;279(5357):1714-1718. [doi: 10.1126/science.279.5357.1714] [Medline: 9497289]
3. Kilgard MP, Merzenich MM. Plasticity of temporal information processing in the primary auditory cortex. Nat Neurosci 1998 Dec;1(8):727-731. [doi: 10.1038/3729] [Medline: 10196590]
4. Kunnath AJ, Gifford RH, Wallace MT. Cholinergic modulation of sensory perception and plasticity. Neurosci Biobehav Rev 2023 Sep;152:105323. [doi: 10.1016/j.neubiorev.2023.105323] [Medline: 37467908]
5. Picciotto M, Higley M, Mineur Y. Acetylcholine as a neuromodulator: cholinergic signaling shapes nervous system function and behavior. Neuron 2012 Oct 04;76(1):116-129. [doi: 10.1016/j.neuron.2012.08.036] [Medline: 23040810]
6. Hampel H, Mesulam M, Cuello A, Farlow MR, Giacobini E, Grossberg GT, et al. The cholinergic system in the pathophysiology and treatment of Alzheimer's disease. Brain 2018 Jul 01;141(7):1917-1933 [FREE Full text] [doi: 10.1093/brain/awy132] [Medline: 29850777]
7. Mahncke H, Bronstone A, Merzenich M. Brain plasticity and functional losses in the aged: scientific bases for a novel intervention. Prog Brain Res 2006;157:81-109. [doi: 10.1016/S0079-6123(06)57006-2] [Medline: 17046669]

XSL•FO
RenderX

8. Brooks CJ, Chan YM, Anderson AJ, McKendrick AM. Audiovisual temporal perception in aging: the role of multisensory integration and age-related sensory loss. Front Hum Neurosci 2018 May 9;12:192. [doi: 10.3389/fnhum.2018.00192] [Medline: 29867415]
9. Schmolesky M, Wang Y, Pu M, Leventhal A. Degradation of stimulus selectivity of visual cortical cells in senescent rhesus monkeys. Nat Neurosci 2000 Apr;3(4):384-390 [FREE Full text] [doi: 10.1038/73957] [Medline: 10725929]
10. Merzenich M, Nahum M, Van Vleet TM. Changing Brains: Applying Brain Plasticity to Advance and Recover Human Ability. Amsterdam, The Netherlands: Elsevier; 2013.
11. Merzenich MM, Van Vleet TM, Nahum M. Brain plasticity-based therapeutics. Front Hum Neurosci 2014 Jun 27;8:385. [doi: 10.3389/fnhum.2014.00385] [Medline: 25018719]
12. Micheau J, Marighetto A. Acetylcholine and memory: a long, complex and chaotic but still living relationship. Behav Brain Res 2011 Aug 10;221(2):424-429. [doi: 10.1016/j.bbr.2010.11.052] [Medline: 21130809]
13. Himmelheber AM, Sarter M, Bruno JP. Increases in cortical acetylcholine release during sustained attention performance in rats. Brain Res Cogn Brain Res 2000 Jun;9(3):313-325. [doi: 10.1016/s0926-6410(00)00012-4] [Medline: 10808142]
14. Parikh V, Bangasser D. Cholinergic signaling dynamics and cognitive control of attention. Curr Top Behav Neurosci 2020;45:71-87 [FREE Full text] [doi: 10.1007/7854_2020_133] [Medline: 32447715]
15. Mesulam M. The cholinergic lesion of Alzheimer's disease: pivotal factor or side show? Learn Mem 2004 Jan;11(1):43-49. [doi: 10.1101/lm.69204] [Medline: 14747516]
16. Schliebs R, Arendt T. The significance of the cholinergic system in the brain during aging and in Alzheimer's disease. J Neural Transm (Vienna) 2006 Nov;113(11):1625-1644. [doi: 10.1007/s00702-006-0579-2] [Medline: 17039298]
17. Schliebs R, Arendt T. The cholinergic system in aging and neuronal degeneration. Behav Brain Res 2011 Aug 10;221(2):555-563. [doi: 10.1016/j.bbr.2010.11.058] [Medline: 21145918]
18. Orlando IF, Shine JM, Robbins TW, Rowe JB, O'Callaghan C. Noradrenergic and cholinergic systems take centre stage in neuropsychiatric diseases of ageing. Neurosci Biobehav Rev 2023 Jun;149:105167. [doi: 10.1016/j.neubiorev.2023.105167] [Medline: 37054802]
19. Berry AS, Harrison TM. New perspectives on the basal forebrain cholinergic system in Alzheimer's disease. Neurosci Biobehav Rev 2023 Jul;150:105192 [FREE Full text] [doi: 10.1016/j.neubiorev.2023.105192] [Medline: 37086935]
20. Chaves-Coira I, García-Magro N, Zegarra-Valdivia J, Torres-Alemán I, Núñez Á. Cognitive deficits in aging related to changes in basal forebrain neuronal activity. Cells 2023 May 25;12(11):1477 [FREE Full text] [doi: 10.3390/cells12111477] [Medline: 37296598]
21. Shanbhag NM, Padmanabhan JL, Zhang Z, Harel BT, Jia H, Kangarloo T, et al. An acetylcholine M1 receptor-positive allosteric modulator (TAK-071) in Parkinson disease with cognitive impairment: a phase 2 randomized clinical trial. JAMA Neurol 2025 Feb 01;82(2):152-159. [doi: 10.1001/jamaneurol.2024.4519] [Medline: 39761063]
22. Rockwood K. Size of the treatment effect on cognition of cholinesterase inhibition in Alzheimer's disease. J Neurol Neurosurg Psychiatry 2004 May;75(5):677-685 [FREE Full text] [doi: 10.1136/jnnp.2003.029074] [Medline: 15090558]
23. Singh S, Agrawal N, Goyal A. Role of alpha-7-nicotinic acetylcholine receptor in Alzheimer's disease. CNS Neurol Disord Drug Targets 2024 Mar;23(3):384-394. [doi: 10.2174/1871527322666230627123426] [Medline: 37366362]
24. Muir JL. Acetylcholine, aging, and Alzheimer's disease. Pharmacol Biochem Behav 1997 Apr;56(4):687-696. [doi: 10.1016/s0091-3057(96)00431-5] [Medline: 9130295]
25. Hasselmo ME, Sarter M. Modes and models of forebrain cholinergic neuromodulation of cognition. Neuropsychopharmacology 2011 Jan;36(1):52-73 [FREE Full text] [doi: 10.1038/npp.2010.104] [Medline: 20668433]
26. Kuhl DE, Koeppe RA, Fessler JA, Minoshima S, Ackermann RJ, Carey JE, et al. In vivo mapping of cholinergic neurons in the human brain using SPECT and IBVM. J Nucl Med 1994 Mar;35(3):405-410 [FREE Full text] [Medline: 8113884]
27. Petrou M, Frey KA, Kilbourn MR, Scott PJH, Raffel DM, Bohnen NI, et al. In vivo imaging of human cholinergic nerve terminals with (-)-5-(18)F-fluoroethoxybenzovesamicol: biodistribution, dosimetry, and tracer kinetic analyses. J Nucl Med 2014 Mar;55(3):396-404 [FREE Full text] [doi: 10.2967/jnumed.113.124792] [Medline: 24481024]
28. Okkels N, Horsager J, Labrador-Espinosa MA, Hansen FO, Andersen KB, Just MK, et al. Distribution of cholinergic nerve terminals in the aged human brain measured with [F]FEOBV PET and its correlation with histological data. Neuroimage 2023 Apr 01;269:119908. [doi: 10.1016/j.neuroimage.2023.119908] [Medline: 36720436]
29. Albin RL, Bohnen NI, Muller MLTM, Dauer WT, Sarter M, Frey KA, et al. Regional vesicular acetylcholine transporter distribution in human brain: a [18F]fluoroethoxybenzovesamicol positron emission tomography study. J Comp Neurol 2018 Dec 01;526(17):2884-2897 [FREE Full text] [doi: 10.1002/cne.24541] [Medline: 30255936]
30. Xia Y, Eeles E, Fripp J, Pinsker D, Thomas P, Latter M, et al. Reduced cortical cholinergic innervation measured using [F]-FEOBV PET imaging correlates with cognitive decline in mild cognitive impairment. Neuroimage Clin 2022;34:102992 [FREE Full text] [doi: 10.1016/j.nicl.2022.102992] [Medline: 35344804]
31. Aghourian M, Legault-Denis C, Soucy JP, Rosa-Neto P, Gauthier S, Kostikov A, et al. Quantification of brain cholinergic denervation in Alzheimer's disease using PET imaging with [F]-FEOBV. Mol Psychiatry 2017 Nov;22(11):1531-1538. [doi: 10.1038/mp.2017.183] [Medline: 28894304]
32. Posner MI, Rothbart MK. Attention, self-regulation and consciousness. Philos Trans R Soc Lond B Biol Sci 1998 Nov 29;353(1377):1915-1927 [FREE Full text] [doi: 10.1098/rstb.1998.0344] [Medline: 9854264]

XSL•FO
RenderX

33. Martinelli P, Sperduti M, Devauchelle AD, Kalenzaga S, Gallarda T, Lion S, et al. Age-related changes in the functional network underlying specific and general autobiographical memory retrieval: a pivotal role for the anterior cingulate cortex. PLoS One 2013 Dec;8(12):e82385 [FREE Full text] [doi: 10.1371/journal.pone.0082385] [Medline: 24367516]
34. Pardo J, Nyabwari S, Lee J, Alzheimer's Disease Neuroimaging Initiative. Aging-related hypometabolism in the anterior cingulate cortex of cognitively intact, amyloid-negative seniors at rest mediates the relationship between age and executive function but not memory. Cereb Cortex Commun 2020;1(1):tgaa020 [FREE Full text] [doi: 10.1093/texcom/tgaa020] [Medline: 34296097]
35. Pardo JV, Lee JT, Sheikh SA, Surerus-Johnson C, Shah H, Munch KR, et al. Where the brain grows old: decline in anterior cingulate and medial prefrontal function with normal aging. Neuroimage 2007 Apr 15;35(3):1231-1237 [FREE Full text] [doi: 10.1016/j.neuroimage.2006.12.044] [Medline: 17321756]
36. Pezzoli S, Giorgio J, Martersteck A, Dobyns L, Harrison TM, Jagust WJ. Successful cognitive aging is associated with thicker anterior cingulate cortex and lower tau deposition compared to typical aging. Alzheimers Dement 2024 Jan;20(1):341-355 [FREE Full text] [doi: 10.1002/alz.13438] [Medline: 37614157]
37. Touroutoglou A, Wong B, Andreano JM. What is so super about ageing? Lancet Healthy Longev 2023 Aug;4(8):e358-e359 [FREE Full text] [doi: 10.1016/S2666-7568(23)00103-4] [Medline: 37454674]
38. Tan RH, Pok K, Wong S, Brooks D, Halliday GM, Kril JJ. The pathogenesis of cingulate atrophy in behavioral variant frontotemporal dementia and Alzheimer's disease. Acta Neuropathol Commun 2013 Jul 05;1:30 [FREE Full text] [doi: 10.1186/2051-5960-1-30] [Medline: 24252534]
39. Baker LD, Espeland MA, Whitmer RA, Snyder HM, Leng X, Lovato L, et al. Structured vs self-guided multidomain lifestyle interventions for global cognitive function: the US POINTER randomized clinical trial. JAMA 2025 Aug 26;334(8):681-691. [doi: 10.1001/jama.2025.12923] [Medline: 40720610]
40. National Academies of Sciences, Engineering, and Medicine, Health and Medicine Division, Board on Health Sciences Policy, Committee on Preventing Dementia and Cognitive Impairment. In: Downey A, Stroud C, Landis S, Leshner AI, editors. Preventing Cognitive Decline and Dementia: A Way Forward. Washington, DC: National Academies Press (US); Jul 22, 2017.
41. Petersen RC, Lopez O, Armstrong MJ, Getchius TSD, Ganguli M, Gloss D, et al. Practice guideline update summary: mild cognitive impairment [RETIRED]: Report of the Guideline Development, Dissemination, and Implementation Subcommittee of the American Academy of Neurology. Neurology 2018 Jan 16;90(3):126-135 [FREE Full text] [doi: 10.1212/WNL.0000000000004826] [Medline: 29282327]
42. World Health Organization (WHO). Psychological Interventions Implementation Manual: Integrating Evidence-Based Psychological Interventions into Existing Services. Geneva, Switzerland: World Health Organization; 2024.
43. Edwards JD, Xu H, Clark DO, Guey LT, Ross LA, Unverzagt FW. Speed of processing training results in lower risk of dementia. Alzheimers Dement (N Y) 2017 Nov;3(4):603-611 [FREE Full text] [doi: 10.1016/j.trci.2017.09.002] [Medline: 29201994]
44. Roenker DL, Cissell GM, Ball KK, Wadley VG, Edwards JD. Speed-of-processing and driving simulator training result in improved driving performance. Hum Factors 2003;45(2):218-233. [doi: 10.1518/hfes.45.2.218.27241] [Medline: 14529195]
45. Ball K, Edwards JD, Ross LA, McGwin G. Cognitive training decreases motor vehicle collision involvement of older drivers. J Am Geriatr Soc 2010 Nov;58(11):2107-2113 [FREE Full text] [doi: 10.1111/j.1532-5415.2010.03138.x] [Medline: 21054291]
46. Smith-Ray RL, Makowski-Woidan B, Hughes SL. A randomized trial to measure the impact of a community-based cognitive training intervention on balance and gait in cognitively intact Black older adults. Health Educ Behav 2014 Oct;41(1 Suppl):62S-69S [FREE Full text] [doi: 10.1177/1090198114537068] [Medline: 25274713]
47. Sprague BN, Ross LA, Ball KK. Does cognitive training reduce falls across ten years?: Data from the ACTIVE trial. Int J Environ Res Public Health 2023 Mar 11;20(6):4941 [FREE Full text] [doi: 10.3390/ijerph20064941] [Medline: 36981850]
48. Rebok GW, Ball K, Guey LT, Jones RN, Kim HY, King JW, ACTIVE Study Group. Ten-year effects of the advanced cognitive training for independent and vital elderly cognitive training trial on cognition and everyday functioning in older adults. J Am Geriatr Soc 2014 Jan;62(1):16-24 [FREE Full text] [doi: 10.1111/jgs.12607] [Medline: 24417410]
49. Ball K, Berch DB, Helmers KF, Jobe JB, Leveck MD, Marsiske M, Advanced Cognitive Training for IndependentVital Elderly Study Group. Effects of cognitive training interventions with older adults: a randomized controlled trial. JAMA 2002 Nov 13;288(18):2271-2281 [FREE Full text] [doi: 10.1001/jama.288.18.2271] [Medline: 12425704]
50. Edwards JD, Wadley VG, Myers RES, Roenker DL, Cissell GM, Ball KK. Transfer of a speed of processing intervention to near and far cognitive functions. Gerontology 2002;48(5):329-340. [doi: 10.1159/000065259] [Medline: 12169801]
51. Edwards JD, Wadley VG, Vance DE, Wood K, Roenker DL, Ball KK. The impact of speed of processing training on cognitive and everyday performance. Aging Ment Health 2005 May;9(3):262-271. [doi: 10.1080/13607860412331336788] [Medline: 16019280]
52. Ball K, Edwards J, Ross L. The impact of speed of processing training on cognitive and everyday functions. J Gerontol B Psychol Sci Soc Sci 2007 Jun;62 Spec No 1:19-31. [doi: 10.1093/geronb/62.special_issue_1.19] [Medline: 17565162]

XSL•FO
RenderX

53. Wolinsky FD, Vander Weg MW, Martin R, Unverzagt FW, Willis SL, Marsiske M, et al. Does cognitive training improve internal locus of control among older adults? J Gerontol B Psychol Sci Soc Sci 2010 Sep;65(5):591-598 [FREE Full text] [doi: 10.1093/geronb/gbp117] [Medline: 20008028]
54. Wolinsky FD, Mahncke HW, Kosinski M, Unverzagt FW, Smith DM, Jones RN, et al. The ACTIVE cognitive training trial and predicted medical expenditures. BMC Health Serv Res 2009 Jun 29;9:109 [FREE Full text] [doi: 10.1186/1472-6963-9-109] [Medline: 19558724]
55. Wolinsky FD, Mahncke HW, Weg MWV, Martin R, Unverzagt FW, Ball KK, et al. The ACTIVE cognitive training interventions and the onset of and recovery from suspected clinical depression. J Gerontol B Psychol Sci Soc Sci 2009 Sep;64(5):577-585 [FREE Full text] [doi: 10.1093/geronb/gbp061] [Medline: 19617456]
56. Wolinsky FD, Vander Weg MW, Martin R, Unverzagt FW, Ball KK, Jones RN, et al. The effect of speed-of-processing training on depressive symptoms in ACTIVE. J Gerontol A Biol Sci Med Sci 2009 Apr;64(4):468-472 [FREE Full text] [doi: 10.1093/gerona/gln044] [Medline: 19181719]
57. Wolinsky FD, Unverzagt FW, Smith DM, Jones R, Wright E, Tennstedt SL. The effects of the ACTIVE cognitive training trial on clinically relevant declines in health-related quality of life. J Gerontol B Psychol Sci Soc Sci 2006 Sep;61(5):S281-S287. [doi: 10.1093/geronb/61.5.s281] [Medline: 16960242]
58. Wolinsky F, Unverzagt F, Smith D, Jones R, Stoddard A, Tennstedt S. The ACTIVE cognitive training trial and health-related quality of life: protection that lasts for 5 years. J Gerontol A Biol Sci Med Sci 2006 Dec;61(12):1324-1329. [doi: 10.1093/gerona/61.12.1324] [Medline: 17234829]
59. Van Vleet TM, DeGutis JM, Merzenich MM, Simpson GV, Zomet A, Dabit S. Targeting alertness to improve cognition in older adults: a preliminary report of benefits in executive function and skill acquisition. Cortex 2016 Sep;82:100-118 [FREE Full text] [doi: 10.1016/j.cortex.2016.05.015] [Medline: 27372902]
60. Degutis JM, Van Vleet TM. Tonic and phasic alertness training: a novel behavioral therapy to improve spatial and non-spatial attention in patients with hemispatial neglect. Front Hum Neurosci 2010;4(60):1-17 [FREE Full text] [doi: 10.3389/fnhum.2010.00060] [Medline: 20838474]
61. Van VT, DeGutis J. The nonspatial side of spatial neglect and related approaches to treatment. Prog Brain Res 2013;207:327-349. [doi: 10.1016/b978-0-444-63327-9.00012-6]
62. DeGutis J, Grosso M, VanVleet T, Esterman M, Pistorino L, Cronin-Golomb A. Sustained attention training reduces spatial bias in Parkinson's disease: a pilot case series. Neurocase 2016 Sep;22(2):179-186 [FREE Full text] [doi: 10.1080/13554794.2015.1088035] [Medline: 26360648]
63. Van Vleet TM, Chen A, Vernon A, Novakovic-Agopian T, D'Esposito MT. Tonic and phasic alertness training: a novel treatment for executive control dysfunction following mild traumatic brain injury. Neurocase 2015 Jul;21(4):489-498. [doi: 10.1080/13554794.2014.928329] [Medline: 24984231]
64. Van Vleet T, de Villers-Sidani E, Cote J, Merzenich M, Rosa-Neto P, Kang MS. P2 - 005: Neuroplasticity - based visual attention training and the expression of acetylcholine in healthy older adults. Alzheimer's & Dementia 2018 Jul;14(7S_Part_12):P666. [doi: 10.1016/j.jalz.2018.06.689]
65. Delahunt PB, Hardy JL, Brenner D, Chan SC, Dewey JA, Mahncke HW. InSight: scientific principles of a brain-plasticity-based visual training program. Google Scholar. San Francisco, CA: Posit Science Corporation; 2008. URL: https://tinyurl.com/brkrp868 [accessed 2025-01-31]
66. Attarha M, Mahncke H, Merzenich M. The real-world usability, feasibility, and performance distributions of deploying a digital toolbox of computerized assessments to remotely evaluate brain health: development and usability study. JMIR Form Res 2024 May 13;8:e53623 [FREE Full text] [doi: 10.2196/53623] [Medline: 38739916]
67. Lin F, Heffner KL, Ren P, Tivarus ME, Brasch J, Chen DG, et al. Cognitive and neural effects of vision-based speed-of-processing training in older adults with amnestic mild cognitive impairment: a pilot study. J Am Geriatr Soc 2016 Jun;64(6):1293-1298 [FREE Full text] [doi: 10.1111/jgs.14132] [Medline: 27321608]
68. Lin FV, Tao Y, Chen Q, Anthony M, Zhang Z, Tadin D, et al. Processing speed and attention training modifies autonomic flexibility: a mechanistic intervention study. Neuroimage 2020 Jun;213:116730. [doi: 10.1016/j.neuroimage.2020.116730] [Medline: 32165263]
69. Ross L, Webb C, Whitaker C, Hicks JM, Schmidt EL, Samimy S, et al. The effects of useful field of view training on brain activity and connectivity. J Gerontol B Psychol Sci Soc Sci 2019 Sep 15;74(7):1152-1162 [FREE Full text] [doi: 10.1093/geronb/gby041] [Medline: 29757433]
70. Frantzidis CA, Ladas AKI, Vivas AB, Tsolaki M, Bamidis PD. Cognitive and physical training for the elderly: evaluating outcome efficacy by means of neurophysiological synchronization. Int J Psychophysiol 2014 Jul;93(1):1-11. [doi: 10.1016/j.ijpsycho.2014.01.007] [Medline: 24472698]
71. Scalf P, Colcombe S, McCarley J, Erickson KI, Alvarado M, Kim JS, et al. The neural correlates of an expanded functional field of view. J Gerontol B Psychol Sci Soc Sci 2007 Jun;62 Spec No 1:32-44. [doi: 10.1093/geronb/62.special_issue_1.32] [Medline: 17565163]
72. Berry AS, Zanto TP, Clapp WC, Hardy JL, Delahunt PB, Mahncke HW, et al. The influence of perceptual training on working memory in older adults. PLoS One 2010 Jul 14;5(7):e11537 [FREE Full text] [doi: 10.1371/journal.pone.0011537] [Medline: 20644719]

XSL•FO
RenderX

73. Rosen AC, Sugiura L, Kramer JH, Whitfield-Gabrieli S, Gabrieli JD. Cognitive training changes hippocampal function in mild cognitive impairment: a pilot study. J Alzheimers Dis 2011 Oct;26 Suppl 3(Suppl 3):349-357 [FREE Full text] [doi: 10.3233/JAD-2011-0009] [Medline: 21971474]
74. Strenziok M, Parasuraman R, Clarke E, Cisler DS, Thompson JC, Greenwood PM. Neurocognitive enhancement in older adults: comparison of three cognitive training tasks to test a hypothesis of training transfer in brain connectivity. Neuroimage 2014 Jan 15;85 Pt 3:1027-1039. [doi: 10.1016/j.neuroimage.2013.07.069] [Medline: 23933474]
75. Cisneros-Franco JM, Voss P, Kang MS, Thomas ME, Côté J, Ross K, et al. PET imaging of perceptual learning-induced changes in the aged rodent cholinergic system. Front Neurosci 2019;13:1438 [FREE Full text] [doi: 10.3389/fnins.2019.01438] [Medline: 32038142]
76. Attarha M, De Figueiredo Pelegrino A, Ouellet L, Toussaint PJ, Grant SJ, Van Vleet T, et al. Association of a brief computerized cognitive assessment with cholinergic neurotransmission: assessment validation study. JMIR Form Res 2025 Jul 07;9(1):e68374 [FREE Full text] [doi: 10.2196/68374] [Medline: 40622012]
77. Attarha M, Carolina de Figueiredo Pelegrino A, Ouellet L, Grant SJ, de Villers-Sidani E, Van Vleet T. Bringing executive function testing online: assessment validation study. JMIR Form Res 2025 Jul 14;9(1):e75687 [FREE Full text] [doi: 10.2196/75687] [Medline: 40654319]
78. FAQ. LORIS. URL: https://loris.ca/faq.html [accessed 2025-01-31]
79. Attarha M, Pelegrino ACDF, Toussaint PJ, Grant SJ, Vleet TV. Villers-Sidani E de. Improving Neurological Health in Aging Via Neuroplasticity-Based Computerized Exercise: Protocol for a Randomized Controlled Trial. JMIR Research Protocols. 2024 ;13(1):e59705. [doi: 10.2196/59705]
80. Nasreddine ZS, Phillips NA, Bedirian V, Charbonneau S, Whitehead V, Collin I, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. J Am Geriatr Soc 2005 Apr;53(4):695-699. [doi: 10.1111/j.1532-5415.2005.53221.x] [Medline: 15817019]
81. Yesavage JA, Brink TL, Rose TL, Lum O, Huang V, Adey M, et al. Development and validation of a geriatric depression screening scale: a preliminary report. J Psychiatr Res 1982 Jan;17(1):37-49. [doi: 10.1016/0022-3956(82)90033-4] [Medline: 7183759]
82. Brink TL, Yesavage JA, Lum O, Heersema PH, Adey M, Rose TL. Screening tests for geriatric depression. Clinical Gerontologist 2008 Oct 25;1(1):37-43. [doi: 10.1300/J018v01n01_06]
83. Posner K, Brown GK, Stanley B, Brent DA, Yershova KV, Oquendo MA, et al. The Columbia-Suicide Severity Rating Scale: initial validity and internal consistency findings from three multisite studies with adolescents and adults. Am J Psychiatry 2011 Dec;168(12):1266-1277 [FREE Full text] [doi: 10.1176/appi.ajp.2011.10111704] [Medline: 22193671]
84. Carson N, Leach L, Murphy KJ. A re-examination of Montreal Cognitive Assessment (MoCA) cutoff scores. Int J Geriatr Psychiatry 2018 Feb;33(2):379-388. [doi: 10.1002/gps.4756] [Medline: 28731508]
85. Vinogradov S, Fisher M, Warm H, Holland C, Kirshner MA, Pollock BG. The cognitive cost of anticholinergic burden: decreased response to cognitive training in schizophrenia. Am J Psychiatry 2009 Sep;166(9):1055-1062 [FREE Full text] [doi: 10.1176/appi.ajp.2009.09010017] [Medline: 19570929]
86. Home page. LORIS. URL: https://loris.ca/index.html [accessed 2025-01-31]
87. Lee H, Kent J, Wendel C, Wolinsky FD, Foster ED, Merzenich MM, et al. Home-based, adaptive cognitive training for cognitively normal older adults: initial efficacy trial. J Gerontol B Psychol Sci Soc Sci 2020 Jun 02;75(6):1144-1154 [FREE Full text] [doi: 10.1093/geronb/gbz073] [Medline: 31140569]
88. Nahum M, Lee H, Fisher M, Green MF, Hooker CI, Ventura J, et al. Online social cognition training in schizophrenia: a double-blind, randomized, controlled multi-site clinical trial. Schizophr Bull 2021 Jan 23;47(1):108-117 [FREE Full text] [doi: 10.1093/schbul/sbaa085] [Medline: 32614046]
89. Biagianti B, Roach BJ, Fisher M, Loewy R, Ford JM, Vinogradov S, et al. Trait aspects of auditory mismatch negativity predict response to auditory training in individuals with early illness schizophrenia. Neuropsychiatr Electrophysiol 2017 Jun;3 [FREE Full text] [doi: 10.1186/s40810-017-0024-9] [Medline: 28845238]
90. Kramer JH, Mungas D, Possin KL, Rankin KP, Boxer AL, Rosen HJ, et al. NIH EXAMINER: conceptualization and development of an executive function battery. J Int Neuropsychol Soc 2014 Jan;20(1):11-19 [FREE Full text] [doi: 10.1017/S1355617713001094] [Medline: 24103232]
91. Mahncke H, DeGutis J, Levin H, Newsome MR, Bell MD, Grills C, et al. A randomized clinical trial of plasticity-based cognitive training in mild traumatic brain injury. Brain 2021 Aug 17;144(7):1994-2008 [FREE Full text] [doi: 10.1093/brain/awab202] [Medline: 34312662]
92. Pelegrino ADF, Attarha M, Toussaint PJ, Ouellet L, Grant SJ, Van Vleet T, et al. Cholinergic neurotransmission in the anterior cingulate cortex is associated with cognitive performance in healthy older adults: Baseline characteristics of the Improving Neurological Health in Aging via Neuroplasticity-based Computerized Exercise (INHANCE) trial. Neuroimage Rep 2025 Mar;5(1) [FREE Full text] [doi: 10.1016/j.ynirp.2025.100234] [Medline: 40191405]
93. Jobe J, Smith D, Ball K, Tennstedt S, Marsiske M, Willis S, et al. Active. Controlled Clinical Trials 2001 Aug;22(4):453-479 [FREE Full text] [doi: 10.1016/S0197-2456(01)00139-8]

XSL•FO
RenderX

94. Palomero-Gallagher N, Hoffstaedter F, Mohlberg H, Eickhoff S, Amunts K, Zilles K. Human pregenual anterior cingulate cortex: structural, functional, and connectional heterogeneity. Cereb Cortex 2019 Jun 01;29(6):2552-2574 [FREE Full text] [doi: 10.1093/cercor/bhy124] [Medline: 29850806]
95. Kanel P, van der Zee S, Sanchez-Catasus CA, Koeppe RA, Scott PJH, van Laar T, et al. Cerebral topography of vesicular cholinergic transporter changes in neurologically intact adults: a [F]FEOBV PET study. Aging Brain 2022;2:2-18. [doi: 10.1016/j.nbas.2022.100039] [Medline: 35465252]
96. Smith GE, Housen P, Yaffe K, Ruff R, Kennison RF, Mahncke HW, et al. A cognitive training program based on principles of brain plasticity: results from the Improvement in Memory with Plasticity-based Adaptive Cognitive Training (IMPACT) study. J Am Geriatr Soc 2009 Apr;57(4):594-603 [FREE Full text] [doi: 10.1111/j.1532-5415.2008.02167.x] [Medline: 19220558]
97. Tennstedt SL, Unverzagt FW. The ACTIVE study: study overview and major findings. J Aging Health 2013 Dec;25(8 Suppl):3S-20S [FREE Full text] [doi: 10.1177/0898264313518133] [Medline: 24385638]
98. Roach JC, Rapozo MK, Hara J, Glusman G, Lovejoy J, Shankle WR, COCOA Consortium:. A remotely coached multimodal lifestyle intervention for Alzheimer's disease ameliorates functional and cognitive outcomes. J Alzheimers Dis 2023;96(2):591-607. [doi: 10.3233/JAD-230403] [Medline: 37840487]
99. Toups K, Hathaway A, Gordon D, Chung H, Raji C, Boyd A, et al. Precision medicine approach to Alzheimer's disease: successful pilot project. J Alzheimers Dis 2022;88(4):1411-1421 [FREE Full text] [doi: 10.3233/JAD-215707] [Medline: 35811518]
100. Lanctôt KL, Rajaram RD, Herrmann N. Therapy for Alzheimer's disease: how effective are current treatments? Ther Adv Neurol Disord 2009 May;2(3):163-180 [FREE Full text] [doi: 10.1177/1756285609102724] [Medline: 21179526]
101. Liu W, Li J, Yang M, Ke X, Dai Y, Lin H, et al. Chemical genetic activation of the cholinergic basal forebrain hippocampal circuit rescues memory loss in Alzheimer's disease. Alzheimers Res Ther 2022 Apr 13;14(1):53 [FREE Full text] [doi: 10.1186/s13195-022-00994-w] [Medline: 35418161]
102. Aloe L, Rocco ML, Balzamino BO, Micera A. Nerve growth factor: a focus on neuroscience and therapy. Curr Neuropharmacol 2015 Jul;13(3):294-303 [FREE Full text] [doi: 10.2174/1570159x13666150403231920] [Medline: 26411962]
103. Pham TM, Ickes B, Albeck D, Söderström S, Granholm AC, Mohammed AH. Changes in brain nerve growth factor levels and nerve growth factor receptors in rats exposed to environmental enrichment for one year. Neuroscience 1999;94(1):279-286. [doi: 10.1016/s0306-4522(99)00316-4] [Medline: 10613518]
104. Hall JM, Gomez-Pinilla F, Savage LM. Nerve growth factor is responsible for exercise-induced recovery of septohippocampal cholinergic structure and function. Front Neurosci 2018 Nov;12:773 [FREE Full text] [doi: 10.3389/fnins.2018.00773] [Medline: 30443202]
105. Lazarov O, Robinson J, Tang YP, Hairston IS, Korade-Mirnics Z, Lee VMY, et al. Environmental enrichment reduces Abeta levels and amyloid deposition in transgenic mice. Cell 2005 Mar 11;120(5):701-713 [FREE Full text] [doi: 10.1016/j.cell.2005.01.015] [Medline: 15766532]
106. Baroncelli L, Braschi C, Spolidoro M, Begenisic T, Sale A, Maffei L. Nurturing brain plasticity: impact of environmental enrichment. Cell Death Differ 2010 Jul;17(7):1092-1103. [doi: 10.1038/cdd.2009.193] [Medline: 20019745]
107. Dobransky T, Rylett RJ. A model for dynamic regulation of choline acetyltransferase by phosphorylation. J Neurochem 2005 Oct;95(2):305-313 [FREE Full text] [doi: 10.1111/j.1471-4159.2005.03367.x] [Medline: 16135099]
108. Erickson JD, Varoqui H, Schafer MK, Modi W, Diebler MF, Weihe E, et al. Functional identification of a vesicular acetylcholine transporter and its expression from a cholinergic gene locus. J Biol Chem 1994 Sep 02;269(35):21929-21932 [FREE Full text] [Medline: 8071310]
109. Vinogradov S, Fisher M, Holland C, Shelly W, Wolkowitz O, Mellon SH. Is serum brain-derived neurotrophic factor a biomarker for cognitive enhancement in schizophrenia? Biol Psychiatry 2009 Sep 15;66(6):549-553 [FREE Full text] [doi: 10.1016/j.biopsych.2009.02.017] [Medline: 19368899]
110. Fisher M, Mellon SH, Wolkowitz O, Vinogradov S. Neuroscience-informed auditory training in schizophrenia: a final report of the effects on cognition and serum brain-derived neurotrophic factor. Schizophr Res Cogn 2016 Mar 01;3:1-7 [FREE Full text] [doi: 10.1016/j.scog.2015.10.006] [Medline: 26705516]
111. Pressler SJ, Titler M, Koelling TM, Riley PL, Jung M, Hoyland-Domenico L, et al. Nurse-enhanced computerized cognitive training increases serum brain-derived neurotropic factor levels and improves working memory in heart failure. J Card Fail 2015 Aug;21(8):630-641. [doi: 10.1016/j.cardfail.2015.05.004] [Medline: 25982826]
112. Andersen KB, Hansen AK, Knudsen K, Schacht AC, Damholdt MF, Brooks DJ, et al. Healthy brain aging assessed with [F]FDG and [C]UCB-J PET. Nucl Med Biol 2022;112-113:52-58 [FREE Full text] [doi: 10.1016/j.nucmedbio.2022.06.007] [Medline: 35820300]
113. Scheff SW, Price DA, Sparks DL. Quantitative assessment of possible age-related change in synaptic numbers in the human frontal cortex. Neurobiol Aging 2001;22(3):355-365. [doi: 10.1016/s0197-4580(01)00222-6] [Medline: 11378240]
114. Itoh Y, Yamada M, Suematsu N, Matsushita M, Otomo E. An immunohistochemical study of centenarian brains: a comparison. J Neurol Sci 1998 Apr 15;157(1):73-81. [doi: 10.1016/s0022-510x(98)00050-1] [Medline: 9600680]

XSL•FO
RenderX

115. Basak C, Qin S, O'Connell MA. Differential effects of cognitive training modules in healthy aging and mild cognitive impairment: a comprehensive meta-analysis of randomized controlled trials. Psychol Aging 2020 Mar;35(2):220-249 [FREE Full text] [doi: 10.1037/pag0000442] [Medline: 32011155]
116. Mewborn CM, Lindbergh CA, Stephen Miller L. Cognitive interventions for cognitively healthy, mildly impaired, and mixed samples of older adults: a systematic review and meta-analysis of randomized-controlled trials. Neuropsychol Rev 2017 Jul 19;27(4):403-439. [doi: 10.1007/s11065-017-9350-8] [Medline: 28726168]
117. Lampit A, Hallock H, Valenzuela M. Computerized cognitive training in cognitively healthy older adults: a systematic review and meta-analysis of effect modifiers. PLoS Med 2014 Nov;11(11):e1001756. [doi: 10.1371/journal.pmed.1001756] [Medline: 25405755]
118. Bahar-Fuchs A, Martyr A, Goh A, Sabates J, Clare L. Cognitive training for people with mild to moderate dementia. Cochrane Database Syst Rev 2019 Mar 25;3(3):CD013069 [FREE Full text] [doi: 10.1002/14651858.CD013069.pub2] [Medline: 30909318]
119. Hill NTM, Mowszowski L, Naismith SL, Chadwick VL, Valenzuela M, Lampit A. Computerized cognitive training in older adults with mild cognitive impairment or dementia: a systematic review and meta-analysis. Am J Psychiatry 2017 Apr 01;174(4):329-340. [doi: 10.1176/appi.ajp.2016.16030360] [Medline: 27838936]
120. Baumgart M, Snyder HM, Carrillo MC, Fazio S, Kim H, Johns H. Summary of the evidence on modifiable risk factors for cognitive decline and dementia: a population-based perspective. Alzheimers Dement 2015 Jun;11(6):718-726 [FREE Full text] [doi: 10.1016/j.jalz.2015.05.016] [Medline: 26045020]
121. Chan ATC, Ip RTF, Tran JYS, Chan JYC, Tsoi KKF. Computerized cognitive training for memory functions in mild cognitive impairment or dementia: a systematic review and meta-analysis. NPJ Digit Med 2024 Jan 03;7(1):1 [FREE Full text] [doi: 10.1038/s41746-023-00987-5] [Medline: 38172429]
122. Sprague BN, Freed SA, Webb CE, Phillips CB, Hyun J, Ross LA. The impact of behavioral interventions on cognitive function in healthy older adults: a systematic review. Ageing Res Rev 2019 Jul;52:32-52 [FREE Full text] [doi: 10.1016/j.arr.2019.04.002] [Medline: 31002885]
123. Kollins SH, DeLoss DJ, Canadas E, Lutz J, Findling RL, Keefe RSE, et al. A novel digital intervention for actively reducing severity of paediatric ADHD (STARS-ADHD): a randomised controlled trial. Lancet Digit Health 2020 Apr;2(4):e168-e178 [FREE Full text] [doi: 10.1016/S2589-7500(20)30017-0] [Medline: 33334505]
124. Clark LT, Watkins L, Piña IL, Elmer M, Akinboboye O, Gorham M, et al. Increasing diversity in clinical trials: overcoming critical barriers. Curr Probl Cardiol 2019 May;44(5):148-172 [FREE Full text] [doi: 10.1016/j.cpcardiol.2018.11.002] [Medline: 30545650]
125. Bae AS. Key barriers against racial and ethnic minority participation in U.S. clinical trials. Int J Clin Trials 2022 Jul 26;9(3):227-233. [doi: 10.18203/2349-3259.ijct20221876]
126. Durant RW, Wenzel JA, Scarinci IC, Paterniti DA, Fouad MN, Hurd TC, et al. Perspectives on barriers and facilitators to minority recruitment for clinical trials among cancer center leaders, investigators, research staff, and referring clinicians: enhancing minority participation in clinical trials (EMPaCT). Cancer 2014 Apr 01;120 Suppl 7(0 7):1097-1105. [doi: 10.1002/cncr.28574] [Medline: 24643647]
127. Allison K, Patel D, Kaur R. Assessing multiple factors affecting minority participation in clinical trials: development of the clinical trials participation barriers survey. Cureus 2022 Apr;14(4):e24424 [FREE Full text] [doi: 10.7759/cureus.24424] [Medline: 35637812]
128. Diaz V. Encouraging participation of minorities in research studies. Ann Fam Med 2012;10(4):372-373 [FREE Full text] [doi: 10.1370/afm.1426] [Medline: 22778129]
129. ORIS (Longitudinal Online Research and Imaging System). URL: https://inhance.loris.ca/ [accessed 2025-09-15]

Abbreviations

AAN: American Academy of Neurology
ACTIVE: Advanced Cognitive Training for Independent and Vital Elderly
CONSORT: Consolidated Standards of Reporting Trials
FDA: Food and Drug Administration
FEOBV: [^{18}F]fluoroethoxybenzovesamicol
GDS-SF: Geriatric Depression Scale—Short Form
INHANCE: Improving Neurological Health in Aging via Neuroplasticity-based Computerized Exercise
ITT: intent to treat
LORIS: Longitudinal Online Research and Imaging System
MCI: mild cognitive impairment
MoCA: Montreal Cognitive Assessment
MRI: magnetic resonance imaging
NASEM: National Academies of Sciences, Engineering, and Medicine

XSL•FO
RenderX

NIH EXAMINER: National Institutes of Health The Executive Abilities: Measures and Instruments for Neurobehavioral Evaluation and Research
PET: positron emission tomography
ROI: region of interest
SUVR: standard uptake value ratio
VAChT: vesicular acetylcholine transporter
WHO: World Health Organization

Edited by A Coristine; submitted 28.03.25; peer-reviewed by J Horsager, J Russell; comments to author 20.08.25; revised version received 28.08.25; accepted 08.09.25; published 13.10.25.

Please cite as:
Attarha M, de Figueiredo Pelegrino A, Ouellet L, Toussaint PJ, Grant SJ, Van Vleet T, de Villers-Sidani E
Effects of Computerized Cognitive Training on Vesicular Acetylcholine Transporter Levels using [18F]Fluoroethoxybenzovesamicol Positron Emission Tomography in Healthy Older Adults: Results from the Improving Neurological Health in Aging via Neuroplasticity-based Computerized Exercise (INHANCE) Randomized Clinical Trial
JMIR Serious Games 2025;13:e75161
URL: https://games.jmir.org/2025/1/e75161
doi:10.2196/75161
PMID:

©Mouna Attarha, Ana de Figueiredo Pelegrino, Lydia Ouellet, Paule-Joanne Toussaint, Sarah-Jane Grant, Thomas Van Vleet, Etienne de Villers-Sidani. Originally published in JMIR Serious Games (https://games.jmir.org), 13.10.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Effect of Gamification on Task Engagement During an Eye-Tracking Test Battery in 5-Year-Old Children Born Preterm: Observational Study

Andrea De Gobbis[1,2], MEng; Marta Malavolta[2], MEng; Nathalie De Beukelaer[3,4], PhD; Els Ortibus[3,4,5], MD, PhD; Aleksander Sadikov[1,2], PhD; Vida Groznik[1,2], PhD

[1]NEUS Diagnostics d.o.o., Ljubljana, Slovenia

[2]Faculty of Computer and Information Science, University of Ljubljana, Ljubljana, Slovenia

[3]Department of Development and Regeneration, KU Leuven, Leuven, Belgium

[4]Leuven Child and Youth Institute, KU Leuven, Leuven, Belgium

[5]Centre for Developmental Disabilities, Universitair Ziekenhuis Leuven, Leuven, Belgium

Corresponding Author:
Andrea De Gobbis, MEng
NEUS Diagnostics d.o.o.
Ziherlova ulica 40E
Ljubljana, 1000
Slovenia
Phone: 386 41 918 373
Email: andrea.dg@neus-diagnostics.com

Abstract

Background: Recent studies suggest that eye movements during tasks reflect cognitive processes and that analysis of eye movements using eye-tracking devices can identify developmental impairments in young children. Maintaining engagement during eye-tracking assessments in young children is challenging and often results in data loss due to distractions. This leads to incomplete recordings and repeated measurements, which can be taxing for young children. Gamification of eye-tracking procedures for cognitive diagnosis might increase engagement and help mitigate these problems, but its effects should be studied and quantified.

Objective: This study compares a standard eye-tracking test battery designed by us with a gamified cartoon version to evaluate the effectiveness of gamification in reducing data loss. The gamified test incorporated child-friendly visuals to provide context for the stimuli presented. The study has two objectives: (1) to compare the data quality between the two versions of the test and (2) to investigate whether, by applying a dynamic stopping criterion to both tests, the higher data quality of the gamified procedure allows earlier test termination.

Methods: Data were collected in a cohort of 25 children born preterm aged 5 years. We measured data quality using a metric derived from robustness, which we defined as the lost data index (LDI), along with task completion rates and feedback from participants. Data analyses were performed as follows: (1) direct comparison of the LDI for the two tests and (2) demonstrating that, although the base gamified test is longer, applying a stopping criterion results in comparable durations. The stopping criterion was based on the number of tasks with an LDI value below a predefined threshold.

Results: The gamified version demonstrated a significant reduction in average LDI compared with the standard version in the first (P<.001, Mann-Whitney U test) and second (P=.01, U test) quarters of the test. In addition, a lower rate of missing values, concentrated at the beginning of the tests, allowed the cartoon test to be stopped after fewer tasks. This, together with the longer tasks of the cartoon test, resulted in comparable test lengths for all thresholds measured by area under the curve (P=.50, U test) and at the chosen threshold of 0.2 LDI (P=.21, U test). Increased engagement was further supported by positive feedback, with 79% (11/14) of the participants who provided feedback preferring the gamified version.

Conclusions: These findings highlight the potential of serious games in eye-tracking–based cognitive assessments for 5-year-old children born preterm. Specifically, gamification might reduce missing values and increase participant engagement, leading to higher retention rates and more effective tests, without significantly lengthening testing procedures.

XSL•FO
RenderX

(JMIR Serious Games 2025;13:e77109) doi:10.2196/77109

KEYWORDS

child; data accuracy; eye movement data analysis; eye-tracking; gamification; premature birth

Introduction

Overview

Prematurity leads to an increased risk of altered neurodevelopmental outcomes in childhood and adolescence, such as cerebral palsy and impaired cognitive outcomes [1], often resulting from brain lesions [2], with many survivors facing a lifetime of disability [3]. Early diagnosis of these adverse outcomes is an important strategy, as it can lead to prompt treatment and a wider array of therapeutic options [2]. In this context, evidence suggests that oculomotor movements during specific tasks, such as smooth pursuit [4], reaction to stimuli [5], or scene exploration [6], serve as biomarkers of cognitive processes and show differences between children with cognitive impairments and typically developing children [7] that could be used to predict future development [8,9]. Therefore, a computerized neuropsychological test battery based on eye-tracking represents a noninvasive, qualitative assessment for the early detection of cognitive delays in children born preterm. Moreover, young nonverbal children are particularly suited to the eye-tracking approach, making it a desirable tool for early detection [10]. However, one of the main challenges when assessing younger participants is maintaining their engagement [11-13]. In a previous study [14], we suggested that certain paradigms for eye-tracking testing might be uninteresting to children born preterm aged 2-6 years.

Serious games offer a solution to this problem by incorporating dynamic and visually stimulating elements [15]. A gamified procedure reduces the monotony associated with repeated exposure to similar tasks, has been shown to lead to more consistent measurements [16], and can be administered remotely [17]. These characteristics are particularly important in pediatric assessments, where maintaining engagement is critical for data reliability [13]. However, the application of gamification to an eye-tracking paradigm introduces specific challenges, such as the possible addition of distracting elements [18-20]. Therefore, when introducing a gamified version of an existing cognitive assessment tool based on eye-tracking, it is necessary to directly compare the two versions [16,21].

Related Works

Eye-Tracking

Different eye-tracking paradigms have been used to obtain measures related to visuocognitive functions in children, such as memory [22], social orienting [6], and smooth pursuit ability [4,23]. Measures include reaction times and looking times for visual processing [5,24,25], reaction times for processing speed and inhibition [26], gain, time delay, and error in smooth pursuit [4,23], and looking times for memory and social orienting [6,22]. In a recent study, Hokkenet et al [7] showed that the scan paths of children born preterm with cerebral visual impairment differ measurably from those of typically developing children.

The prevalence of missing values when testing young participants has been identified as one of the main challenges when using a remote eye-tracking system [12,27], which is an eye-tracking device fixed to a computer screen. Gredebäck et al [12] and Corbetta et al [11] suggested using engaging stimuli in the test to keep the attention of the child focused on the screen.

In adult and older adult populations, test batteries that combine different tasks have been used for the early diagnosis of dementia [28,29]. Such methodologies, which combine various kinds of stimuli and measure a variety of cognitive functions, provided a blueprint for our combined test battery. The potential carryover effects when combining multiple tasks in younger populations remain unclear, but the positive results obtained in adults [28,29] suggest that these types of tests provide a promising framework.

Gamification

The idea of using serious games to increase participant engagement in clinical tests is not new [15,18,30-32], and guidelines on introducing gamified elements to existing tests are available [33]. In their review, Lumsden et al [15] highlighted that serious games are often used to improve motivation and enhance appeal for a particular age group (either older or younger participants) during cognitive assessments. Conversely, other studies highlight that gamification might increase cognitive loads in adult participants [34], possibly introduce distracting elements that fatigue participants earlier [19,20], and may not influence data quality [35]. Another consideration when substituting a set of stimuli with different ones is the potential effect on measured outcomes [15]. For example, Kooiker et al [24] reported different reaction times to an animated cartoon stimulus versus a static shape. Held et al [19] hypothesized that gamification elements can distract participants from the core task, a view supported by Wiley et al [20], who suggest that gamification might introduce an expectation of an interesting activity that is later unmet. Therefore, when gamifying an existing paradigm, the adapted test may not be directly comparable with the original test. Even considering these drawbacks, in some cases presenting gamified stimuli may be necessary to keep young participants engaged, as they might otherwise get bored with the test. For example, the standard test we designed in a previous work [36] might have been too simplistic for children older than 24 months [14], leading to reduced engagement and data quality in a preliminary trial. Moreover, Wiley et al [16] reported that using game elements might even increase internal consistency and test-retest reliability.

Measuring Engagement

Improvements in eye-tracking technologies have allowed quantitative measurements of participant engagement [37,38]. Different methods can be used to assess engagement during testing from eye-tracking data, such as clustering of gaze

XSL•FO
RenderX

trajectories [37] and inter-subject correlation [39], in conjunction with looking time and feedback. Finally, missing values are also useful for determining data quality [27,40], providing a baseline for the amount of usable data and offering an indication of participants' engagement by measuring looking time at the screen.

Direct comparisons between two variants (ie, simple vs gamified) of the same cognitive assessment test to measure engagement and data quality using eye-tracking have been conducted previously in adult populations, with contrasting results. Scharingen et al [34], for example, suggested that gamification might increase cognitive load without strong evidence of improved attention, corroborating their findings with both eye-tracking measures and EEGs. In contrast, Friehs et al [21] reported that gamification can be beneficial for certain tasks, focusing on reaction times measured by an eye-tracking device, finding no differences in reaction times between versions and an increase in positive feedback with the gamified version. These conflicting findings underline the difficulties of directly comparing different procedures and measuring engagement and interest. A similar study by Gallagher et al [18] investigated the effects of gamification on children aged 8-12 years with attention deficit hyperactivity disorder, finding that game elements can increase intrinsic motivation but also distract participants from the task. However, the magnitude of the increase in intrinsic motivation was unclear because the authors did not directly compare the two versions of the task. To our knowledge, similar comparison studies have not been conducted in a sample of 5-year-old children born preterm. Therefore, we plan to address this research gap in this study by focusing on measures readily available during an eye-tracking experiment.

This Study

The aim of this study is to compare data quality and engagement in a cohort of 5-year-old children born preterm for two variants of an eye-tracking test battery: a standard version [14,36] and a gamified cartoon version. The comparison is independent of task performance and uses, as a comparison measure, an index of lost data inspired by robustness [27,40], representing the ratio of nonmissing gaze points to the total number of eye-tracking samples. We propose that the resulting lost data index (LDI) not only provides an indication of data quality but also reflects test engagement by quantifying the time the participant spends looking at the screen, making it an optimal comparison metric. A secondary objective of this study is to show that, while gamification might increase the base duration of a test, using a dynamic stopping criterion based on data quality results in comparable test lengths for the two variants.

Methods

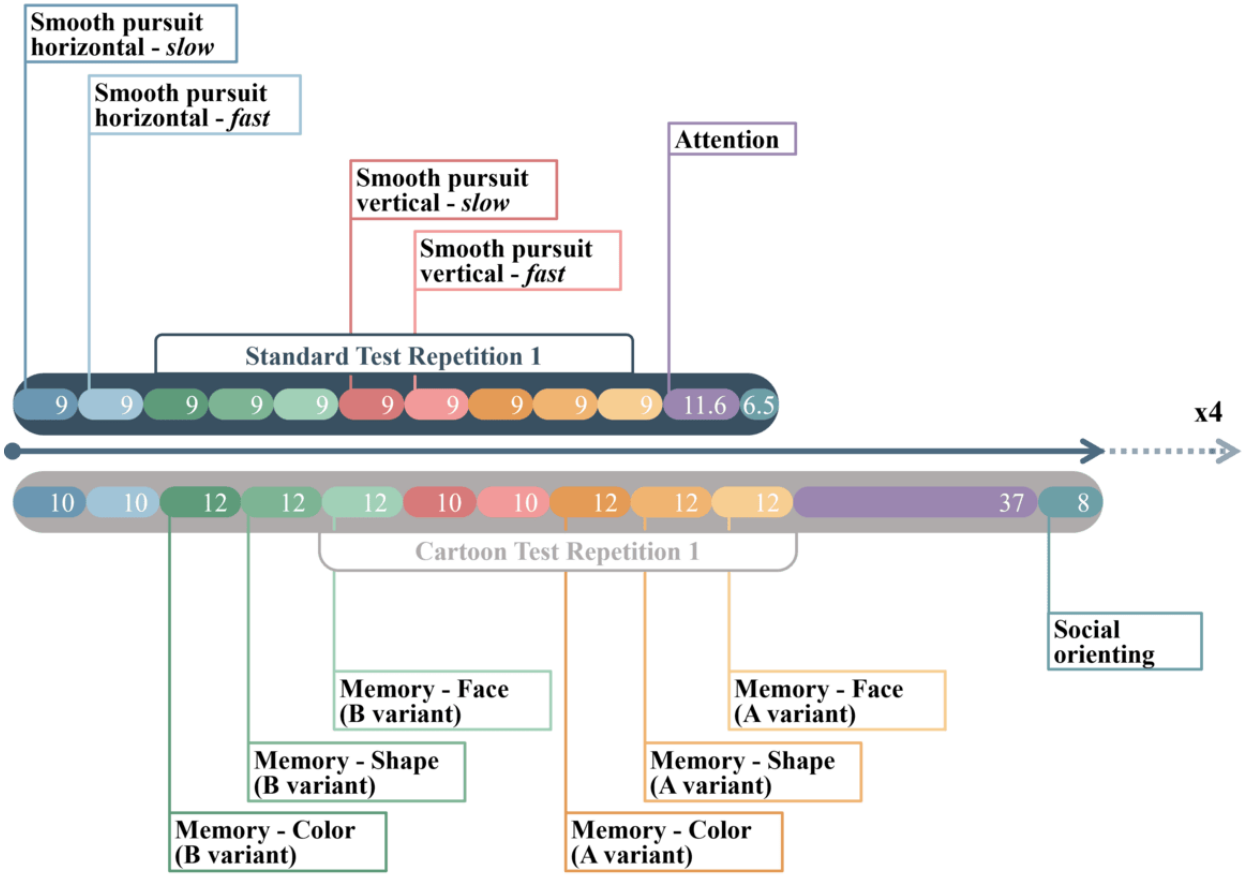
Overview

In this observational cohort study, we conducted a controlled experiment on a sample of 5-year-old children born preterm. The experimental procedure involved presenting each participant with two variants of the test battery, one after the other, in a pseudorandomized order.

Both the standard and the cartoon test consist of a total of 12 tasks, including 4 smooth pursuit tasks, 6 memory tasks, 1 attention task, and 1 social orienting task. During each test, these tasks are presented in a fixed order (Figure 1) and repeated 4 times, for a total of 48 tasks comprising the complete test. The task order was chosen to maintain a sense of variety, with tasks based on scene exploration (memory task and social orienting task) alternated with more dynamic tasks (smooth pursuit task and attention task). The standard test builds upon the paradigm we introduced in previous works; for further details, we refer to De Gobbis et al [36] and Malavolta et al [14].

XSL•FO
RenderX

Figure 1. Basic structure of the two tests. Both tests are composed of 4 repetitions of the same 12 tasks. Task types follow the same order in both tests. Along with the task names, the duration of each task is indicated in seconds.



The Standard and Cartoon Tests

The standard test is an evolution of an eye-tracking–based paradigm we developed in a previous study [14,36] and is composed of 4 types of tasks. Each task starts with the appearance of a yellow smiley face at the center of the screen for 1.5 seconds, accompanied by a sound, to focus the participant's attention, and all stimuli are presented on a black background. The cartoon version was created to address a problem in the standard test identified in previous work [14]: participants aged 2-8 years tended to interrupt the test by standing up or becoming distracted. This lack of engagement suggested that the standard test was insufficiently interesting for this age group. To address this, we redesigned the test using the Unity game engine, incorporating dynamic cartoon elements to create a more engaging experience while preserving the core elements of the task.

To create the cartoon version of the test, we followed the RECIPE guidelines described by Nicholson [33]. Specifically, we aimed to create a context for the stimuli appearing on the screen by translating them into a cartoon setting. To make the comparison as fair as possible between the two versions, we did not introduce any interactive elements in the cartoon version and, similarly, no feedback was given on task performance. As such, we focused on what Nicholson defines as exposition and information, and what other authors refer to as intrinsic motivation [41], which, in short, consists of communicating to the participant the context of the action they are performing without a direct feedback loop, such as points. During the design of this gamified cartoon version, we focused on the following priorities: context and storytelling, visual interest, audio, and fidelity-engagement compromise.

- Context and storytelling: A narrative setting was introduced to provide context for the tasks, making them more engaging for participants. For example, in the social orienting task, the original stimuli, such as faces and houses, were preserved but enhanced with animations, including a swirling vortex transition and nature, without altering the essence of the stimuli.
- Visual interest: Vibrant colors, dynamic animations, and visually appealing transitions were introduced to maintain children's attention throughout the tasks.
- Audio: Sound effects and music were added to create a more engaging experience. The sounds were selected to complement the visuals—for example, a chirping sound when a bird appeared on screen.
- Fidelity-engagement compromise: When adding game elements to a cognitive test there is always a trade-off between game elements and the core task [41]. In this study, we made an effort to maintain a balance between making the tasks engaging and preserving the original test's fidelity. While some tasks, such as the social orienting task, retained their original structure with some adjustments, others, such as the attention task, underwent larger changes to introduce a storytelling element while preserving the main objectives.

XSL•FO
RenderX

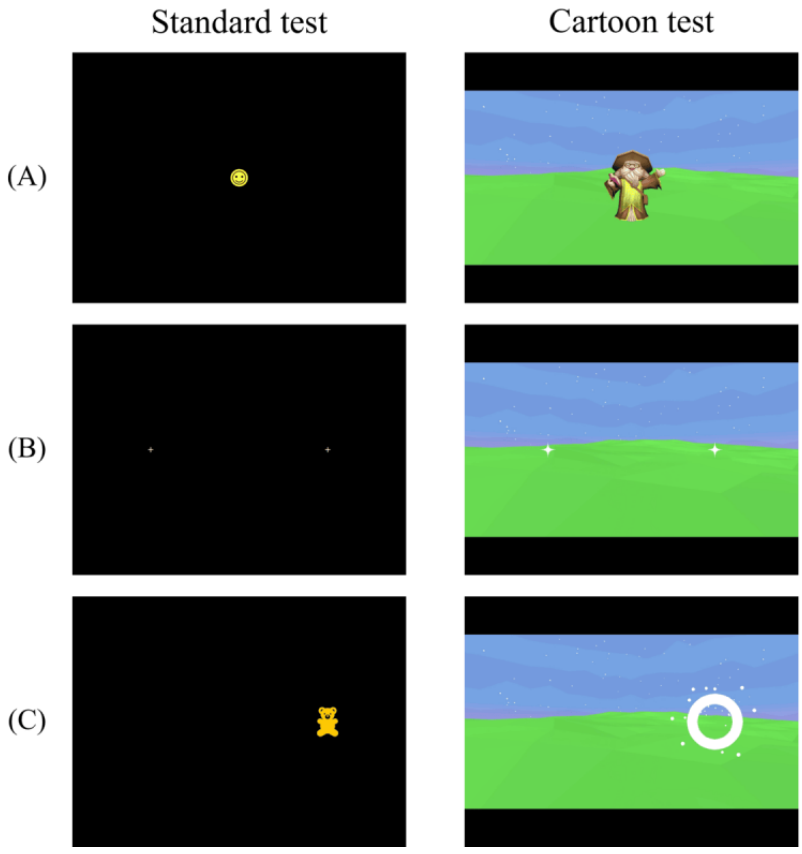
This process resulted in several differences from the standard version. The black background of the standard test was replaced with a computer-generated image of a grass field under a twilight sky, with a dark blue sky and stars. This background was consistent across all tasks (Figure 2). While the smiley face in the standard test serves as the initial fixation for all tasks, in the cartoon version, each task has a different target depending on how it is framed. All stimuli in this test are copyright-free Unity models, and all timings from the standard test were maintained; however, due to animations and transitions, the tasks are uniformly longer.

Figure 2. Examples of visual elements in the cartoon version. The wizard casts a spell in the attention task (left), a treasure chest with visual effects is shown in the memory task (center), and a fairy stimulus is presented for the smooth pursuit task (right). A common background is shown behind the objects across all tasks.



The tasks used in the cartoon version were the same as those in the standard version. These 4 task types are described below, along with the differences introduced in the cartoon version. The differences are also summarized in Table 1, and a side-by-side comparison of the attention task is shown in Figure 3. . A video example of the tests is available in Multimedia Appendices 1 and 2.

Figure 3. Comparison of the two versions of the attention task: the standard version (left) and the cartoon version (right). (A) Initial fixation (smiley face or wizard), (B) cues (crosses or sparkles), and (C) target (animal icon or fireworks).



XSL•FO
RenderX

Table 1. Summary of the differences between the two versions of the test.

Element	Standard test	Cartoon test
Smooth pursuit task		
Target	Yellow smiley face	A bird, a dragon, and a witch (horizontal) and a fairy and a slime (vertical)
Attention task		
Fixation	Yellow smiley face	A magician
Cues	White crosses	Small white sparkle
Target	Colored animal graphic (dog, cat, etc)	Firework explosion
Memory task		
Fixation	Yellow smiley face	Shaking treasure chest
Pictures (color modality)	Colored circles	Colored spheres
Pictures (shape modality)	2D shapes (circle, square, and diamond)	3D shapes (sphere, cube, and diamond)
Pictures (face modality)	Photographs	Photographs in a 3D frame
Social orienting task		
Fixation	Yellow smiley face	Purple vortex
Pictures	Photographs	Photographs in a 3D frame

Smooth pursuit task: In the standard version, a yellow smiley face moves smoothly across the screen in a sinusoidal pattern. There are 4 variants of the task, defined by two speeds along each of the two axes (horizontal and vertical). The periods used are 1600 ms (0.625 Hz) and 2400 ms (≈0.4 Hz). All variants last 7.5 seconds, plus 1.5 seconds for initial fixation with the smiley face. The design of this task was inspired by previous studies using smooth pursuit [4,23]. In the cartoon version, in place of the smiley face, different stimuli move on the screen. The movement is again an oscillation with the same orientation, speed, and amplitude. Before moving, the stimulus remains at the center of the screen for 1.5 seconds to induce fixation. The models used are a bird, a witch, and a dragon for the horizontal variant, and a fairy and a bouncing slime for the vertical one. All of them can be considered characters (no inanimate subjects) to increase engagement and motivate their movements. As they move, a sound is played depending on the model (eg, chirping for the bird and a bell jingle for the fairy). The task is 1 second longer due to a small break before the appearance of the next stimulus (10 seconds).

Memory task: The task starts by showing two pictures for 3 seconds (habituation phase), followed by a blank screen for 0.5 seconds, and then two pictures are displayed for 4 seconds, with one of them replaced by a new image (novelty phase). There are 2 variants of this task: in variant A, based on the work of Ross-Sheehy et al [22], the two initial pictures are different, whereas in the classical visual pair comparison (variant B) [42], the two initial pictures are the same. Furthermore, each variant is presented in 3 distinct modalities, depending on the pictures shown. These include circles of different colors (color modality), different simple shapes (shape modality), and pictures of human faces (face modality). The colors used are red, yellow, and green, while the available shapes are a square, a diamond (a square rotated by 90 degrees), and a circle. The face pictures used in the test are from the London Faces Dataset [43]. In the cartoon version, the initial fixation target is a treasure chest that falls from the top of the screen and stops in the center. The chest then shakes, accompanied by a suitable sound effect, then explodes, and two objects appear in place of the two pictures. Then, after the habituation phase, the two objects leave the field of view by moving off the screen from the bottom. After a short delay, the objects move upward and return to their previous positions, with one of them changed. As in the standard version, the task has 3 modalities: colored spheres (color), different 3D shapes (shape), and framed pictures of faces (face). Due to the animations (eg, treasure chest appearing, objects moving) the task is 3 seconds longer than its standard counterpart, with a total duration of 12 seconds.

Attention task: The task is inspired by the “Infant Orienting With Attention” paradigm [26] and is composed of 4 modalities. In each modality, after the first fixation at the center (the yellow smiley face), little white crosses, called cues, appear to one side or both sides of the screen. The cues stay on the screen for 0.1 seconds with 0.15 seconds blank screens before and after. A sound accompanies the cues. Subsequently, a target (a brightly colored animal graphic, such as a dog or a cat) appears on one side of the screen for a duration of 1 second, and the participant is expected to fixate on this target. The 4 modalities differ on the position of the cues: the valid modality has the cue on the same side where the target will appear, the invalid one has the cue on the opposite position, the double modality has cues on both sides, and finally there is a baseline modality without cues. The 4 modalities are presented in a different order in each repetition. The duration of each modality is 2.9 seconds, resulting in a total duration of the task of 11.6 seconds. In the cartoon version, we framed the entire task to give it a simple

XSL•FO
RenderX

story: a wizard tries to cast a magic spell with varying results. The spell is supposed to create a small sparkle (the cue) and then a firework explosion (the target) in the same position, which would correspond to the valid modality in the standard task. As such, depending on the modality, the casting is framed as successful (valid modality), unsuccessful (invalid), or partially successful (baseline, double). All 4 modalities follow the same structure. The wizard first appears at the center of the screen in a puff of smoke, before casting the spell with a short animation; this first phase corresponds to the initial fixation. The wizard then disappears in another puff of smoke and a white sparkle(s) (not animated) appears as the cue(s), followed by an animated firework explosion as the target. Then the wizard reappears, and a different animation is played depending on how successful his spell was: a happy clapping (successful), shamefully hanging his head (unsuccessful), or an unsure humming (partially successful). The addition of these storytelling elements resulted in a significantly increased time: the total task lasts 37 seconds, compared to the 11.6 seconds of the standard version.

Social orienting task: The task follows a paradigm from Telford et al [6]. After the smiley face at the beginning, the participant is shown two pictures, each filling half the screen, one of a house and the other of a human face taken from the London Faces Dataset [43]. The pictures of faces all look forward with a neutral expression and were chosen to maintain an approximate balance of gender and ethnicity across the 4 repetitions. The two pictures stay on screen for 5 seconds; thus, the total duration of the task is 6.5 seconds. In the cartoon version, at the beginning of the task, a bright purple vortex is shown at the center of the screen to attract the participant's attention, and the two framed pictures appear as if ejected from the vortex, showing the same pictures used in the standard version. Due to the animation of the vortex appearing, disappearing, and the pictures exiting the vortex, the task is slightly longer than the standard version, lasting 8 seconds.

Recruitment

Participants in this study were recruited between February 2024 and February 2025 from children undergoing periodic checks at the Centre for Developmental Disabilities, University Hospitals Leuven, Belgium. Inclusion criteria were age 4-6 years and born preterm (gestational age <32 weeks, birth weight <1500 g, or both). Children with severe visual impairments (eg, acuity <0.2 on a decimal scale), behavioral problems, or any other impairment that might influence the results of the measurements were excluded from the study.

A total of 26 children (mean age 60.2, SD 2.62 months) were recruited for the study, all of whom were born preterm (gestational age <32 weeks). Among these, one left during the break and was thus unable to have a recording for the second test. No test was interrupted due to unrelated conditions unrelated to the participants, such as malfunctions or a request from the parents. As such, we analyzed data from 25 children, of whom 15 (60%) were male and the remaining female. The median birth weight was 1440 (IQR 1220-1580) g, and the median gestational age was 31 (IQR 30-32) weeks. Among the participants, 12 were shown the cartoon version first (5 females, 7 males) and 13 the standard version first (5 females, 8 males).

Data Gathering Procedure

To maximize the effectiveness of the recordings while minimizing distractions, we followed the guidelines outlined by Duchowski [44] when creating the setup. Participants were instructed to sit on a comfortable chair positioned 60 cm from a 21-inch computer screen with a mounted Tobii Pro Nano eye-tracker (60 Hz). The child was separated from the rest of the room using a white folding screen, and a second white screen was positioned behind the computer screen. Behind the two screens, an operator could visualize in real time where the participant was looking on the screen and, if necessary, stop the test. The two test variants were then played sequentially, with a 10-minute break in between; the order of the variants was alternated across participants. During the break, children were encouraged to play with their parents and relax.

The operator was instructed to let the test play to its conclusion and interrupt it only if: (1) the child communicated verbally that they wanted to stop, (2) the child was crying, fussy, or unable to continue, (3) the child got up from the chair and walked out of the test area, and (4) the child interfered with the setup of the test (eg, pushed either of the folding screens causing them to fall). All these conditions were interpreted as a definitive loss of engagement from the participant. In addition, the test could also be interrupted due to software or hardware malfunction or after a request from the child's parents.

After both tests were concluded, the operator asked the participant which version they preferred with the question, "Which video did you like the most?" If the participant preferred the standard version, the operator would ask them the reason.

Statistical Analysis

Measures

For each participant, the recorded gaze trajectories were divided into the 48 tasks composing each test variant. Next, we computed the LDI, a quantity based on robustness, on a task-by-task basis. Robustness is a measure of data quality for eye-tracking data [27,40], which encodes the tendency of the device to incorrectly record a sample and is calculated by dividing the number of nonmissing values by the total number of samples. Robustness offers an upper bound to the amount of data that can be parsed and analyzed in any subsequent study. In this study, we took inspiration from this measure to define the LDI as the ratio between gaze points that were either (1) missing or (2) not positioned inside the screen boundaries, and the total number of samples. Moreover, small gaps in valid samples (ie, <150 ms) were ignored when computing this quantity, as they were probably due to blinking, which represents an unavoidable data loss and can be imputed. An LDI of 0 means no data loss, while a value of 1 signifies that the entire task is unusable; small values are therefore preferable and, depending on the stimuli, tasks with an LDI value above a certain threshold might be discarded [4]. From the LDI value, it is possible to obtain the total time of disruption by multiplying it with task duration; in this study, however, we only used standard LDI to maintain normalization across tasks. The main differences between robustness and LDI are that (1) most disruption due to the device (usually <150 ms) and blinks are ignored by the LDI

XSL•FO
RenderX

because they are unavoidable, and (2) gaze positions that are detected but not usable (outside the screen) are counted. These differences result in a slightly different feature that is less about the eye-tracking device and more about the measure of the avoidable disruption caused by the participant. For example, a participant that continuously looks away from the screen because they are distracted will have a higher LDI than a participant who correctly fixates on the stimuli. As such, the LDI is both an upper bound on unusable data and an indication of engagement in the test. The latter interpretation must be considered with caution, since a participant with a perfect LDI of 0 might not be following the tasks and may be completely uninterested, while still looking at the screen. That is, the LDI contains no indication of task performance. Our main motivation for choosing LDI in place of robustness or another feature describing engagement is that LDI is a general measurement that can be computed for all kinds of eye-tracking tasks, in contrast with other outcomes that are specific to certain kinds of tasks (eg, the number of fixations or saccade length depends heavily on the stimulus chosen). Thus, it can be compared across different tasks and potentially generalized to other kinds of tests. Moreover, it provides an indication of both engagement and data quality, as a bound to recording time that can be analyzed. If a certain task is missing from the test, for example, because the operator interrupted the test, the LDI is set to 1, signifying no data available.

Once we defined LDI as a measure of data quality and engagement, we applied two approaches to compare the standard and cartoon variants of the test.

Experiment 1: Direct Comparison of LDI and Evolution in Time

We directly compared LDI values between the two tests to show that the cartoon test had significantly lower LDI. The comparison was done by dividing the complete test into 4 repetitions of 12 tasks each. Then, for each repetition, the average LDI over all 12 tasks was computed by weighting each task by its respective duration. The comparison consisted of four 1-tailed Mann-Whitney *U* tests to find evidence that the cartoon test had a lower LDI. We chose the *U* test because LDI is bounded between 0 and 1 and therefore nonnormally distributed.

In addition, we examined the evolution of the LDI during the test and visually investigated its general trend in both versions.

For experiment 1, we set the significance level to α=.05 and applied a Bonferroni correction (N_{test}=4) to account for repeated statistical tests (equivalent significance $\alpha_{Bonf}=\alpha/N_{test}$=.0125).

Experiment 2: Admissible Tasks and Test Length

As a second experiment, we investigated the expected length of the two tests if we were to introduce a stopping criterion based on LDI in a simulated clinical setting. In general, there are two broad categories of testing procedures: a static test, which is kept the same for all participants and has a fixed length, and a dynamic test, which is adapted to the participant and can have a variable length. In this study, a dynamic test is defined as a testing procedure in which we can choose, based on real-time measures, which task to present to the participant. Small LDI values in a static test will lead to fewer repetitions being discarded in the cartoon version, which allows features to be aggregated to obtain more stable outcome measurements. In the case of a dynamic test with a stopping criterion based on lower LDI, the cartoon test will lead to shorter tests on average.

To show that a lower LDI results in a shorter testing time, we simulated a dynamic testing procedure by introducing a plausible stopping criterion for our test. If we were to use a similar criterion in a comparable population, testing time between the two versions would be comparable, and the use of the cartoon version would lead to an increase in correctly completed tests. To simulate a dynamic stopping criterion, we used the same dataset and artificially censored data after the simulated stopping time. The criterion used in this study to decide if a task should be discarded is based exclusively on lost data. Additional motivation to discard tasks could be low precision or accuracy [40], but precision and accuracy thresholds usually depend on the specific quantity measured (eg, smaller areas of interest necessitate higher precision). As such, to keep our considerations as general as possible and in line with the preceding considerations, we did not include them in the analysis, focusing on LDI. Specifically, we considered a task admissible if its LDI was below a certain threshold τ. This criterion for discarding sections of recordings is not new in eye-tracking applications, for example, Gredebäck et al [4] use τ=0.2 in their smooth pursuit study.

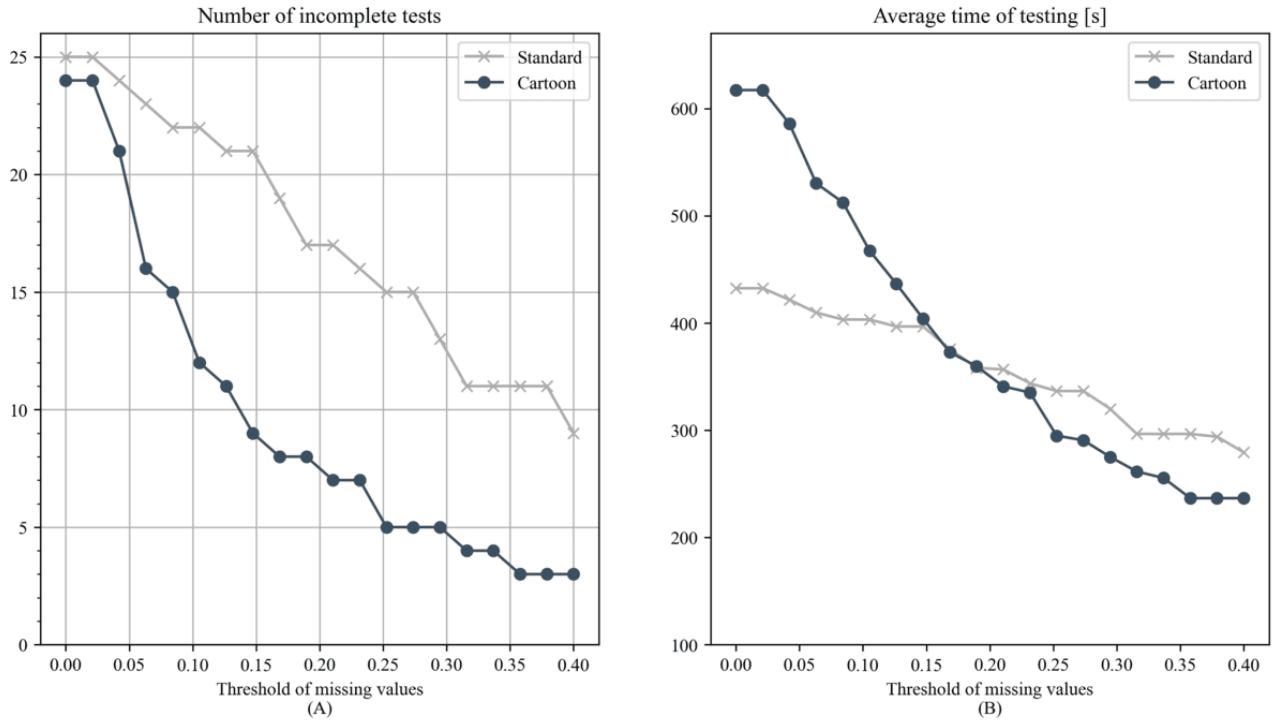
The stopping criterion chosen in this study was intended to be as general as possible and rests on the reasonable assumption that a certain number of tasks is necessary to reach a diagnosis. Thus, the test was run normally and could be stopped prematurely once we record M different admissible (not discarded) tasks, where M ranged between 1 and 12 (in our analysis M=12; thus, all tasks needed to be admissible to obtain a complete test). If the fourth repetition concluded without reaching the desired number, the test was deemed incomplete. It is important to note that, in a real dynamic setting, it would be possible to select the order of the tasks and present only the missing ones to the participant. This approach has some drawbacks: for example, it would lead to test recordings with tasks in different orders, and therefore researchers would need to consider order effects during the trial and analysis. In addition, simulating this approach with our dataset would introduce distortion in the results, since tasks were always presented in the same order. For these reasons, in our simulation we assumed the tasks were always presented in the same order, and every time a certain task was not admissible, it was necessary to run a new repetition.

Since test length and number of admissible tasks depend on the threshold τ, we obtained a curve for each test version ($\tau \rightarrow T_{duration}$), which described test length when varying the threshold. A numerical comparison could be done by fixing a certain τ—that is a point comparison (in this study, we chose the τ=0.2, inspired by previous work [4])—or by computing the area under the curve (AUC) for each participant. It is worth noting that this AUC referred to the curve shown in Figure 4 and should not be confused with the area under the “receiver

XSL•FO
RenderX

operator characteristic" curve or the same term used in pharmacokinetics. For both point comparisons and AUC, we used a 1-tailed Mann-Whitney *U* test (significance level α=.10) to determine whether there was a significant difference in test length between the standard and cartoon tests when using a dynamic stopping criterion (with the alternative hypothesis that the cartoon test was longer).

Figure 4. (A) Number of participants who would have successfully completed the test for each plotted threshold τ. (B) Average total duration of the simulated dynamic test with the stopping criterion for each threshold τ. In panel (B) for τ=0 (no missing value tolerated), both tests run their entire course, and the cartoon test is longer.



Ethical Considerations

This study was conducted with the approval of the Ethical Committee KU Leuven/UZ Leuven (EC number S68019) and on the basis of previous informed consent by the participants' parents or their legally authorized representatives. Eye-tracking data are naturally anonymous, as it is impossible to clearly identify a person from gaze movements alone. Clinical data were partially anonymized by removing names and surnames. All personal data were treated as confidential by the researchers and not shared with anyone external to this study.

Results

Experiment 1: Direct Comparison of LDI and Evolution in Time

After computing the LDI by task, qualitative inspection of its trend over time (Figure 5) shows an approximately linear effect of testing time on the average engagement of the participants. Aggregating the values reveals evidence that the LDI of the entire test is lower in the cartoon version, but not significantly (*P*=.05, Mann-Whitney *U* test, Table 2). This is because the effect is not uniformly distributed across the entire test: the initial repetitions show a larger gap in LDI values between the two versions. In the first repetition, the cartoon version retains a significantly larger percentage of data (*P*<.001; median cartoon LDI 0.11, IQR 0.04-0.12 vs median standard LDI 0.22, IQR 0.12-0.29). This effect decreases over the course of the test, and in the last repetition the LDI is approximately the same between the two versions (*P*=.68). Figure 6 shows the distribution for each repetition, together with the respective *P* values.

We propose two explanations of the observed phenomenon: first, because of the increased test length, the fourth repetition of the cartoon version starts at 471 seconds, more than two minutes after the start of the fourth repetition in the standard version (325 seconds). Thus, it is possible that, due to the effect of testing time on LDI (Figure 5), we have reached the limit of engagement for the participant. Second, this effect might be caused by a form of survival bias. Indeed, by inspecting when tests were interrupted (Figure 7), we observe that the cartoon test is usually stopped later, with only 16% (4/25) of participants interrupted before 432 seconds in the cartoon version versus 56% (14/25) in the standard version, but both of them are stopped in the third and fourth repetition in comparable numbers. Therefore, in the fourth repetition the participants that are still being tested are probably the more attentive ones, independently of the variants shown, and so data loss is comparable.

XSL•FO
RenderX

Figure 5. Average lost data index over time for the two versions of the test. The vertical axis represents the lost data index by task, and the horizontal axis represents time. The cartoon test is longer. Repetitions are separated by gaps in the lines.

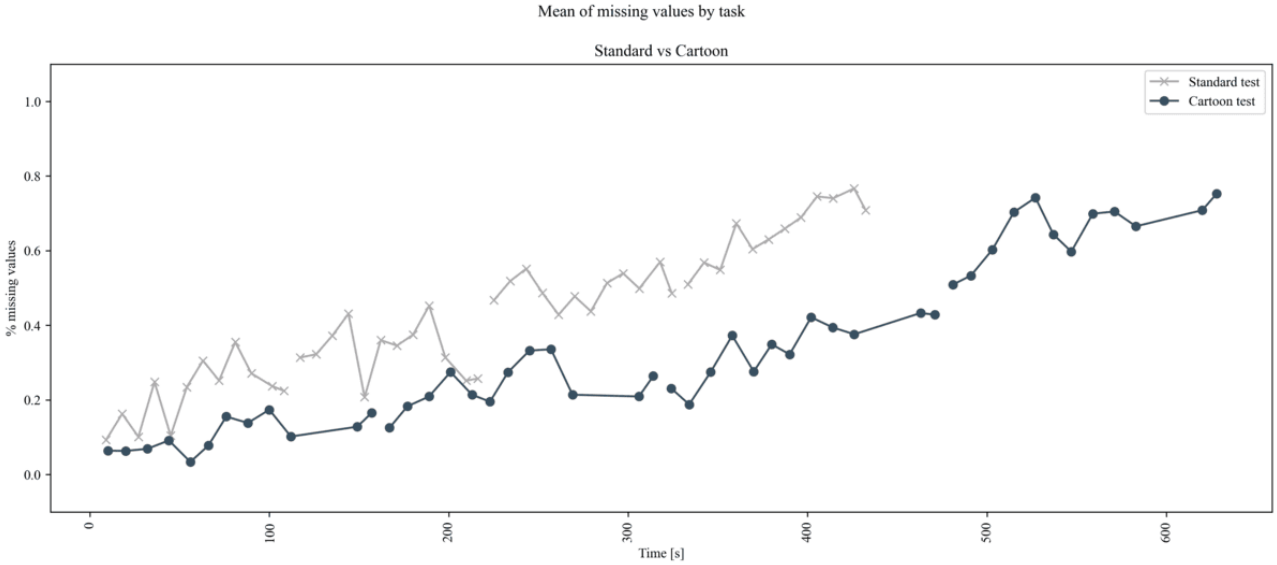


Table 2. Summary of results. For each hypothesis test, the *P* value and average values of the quantity compared are reported. Average values are reported as mean (standard deviation). Separately, the first quartile and third quartile are reported. All tests were performed using the 1-tailed Mann-Whitney U test.

Test	*P* value	Value standard, mean (SD)	Value standard, IQR	Value cartoon, mean (SD)	Value cartoon, IQR
Experiment 1					
Complete test (LDI[a])	.05	0.43 (0.19)	0.26-0.54	0.33 (0.19)	0.19-0.45
Repetition 1 (LDI)	<.001	0.22 (0.17)	0.12-0.29	0.11 (0.11)	0.04-0.12
Repetition 2 (LDI)	.01	0.33 (0.17)	0.22-0.45	0.24 (0.22)	0.07-0.37
Repetition 3 (LDI)	.02	0.50 (0.27)	0.27-0.69	0.34 (0.25)	0.14-0.41
Repetition 4 (LDI)	.68	0.65 (0.32)	0.28-1.00	0.66 (0.37)	0.26-1.00
Experiment 2					
Test time (τ=0.2, in seconds)	.21	357 (117)	270-432	360 (201)	157-628
Test time (AUC[b])	.50	75.6 (17.6)	60.3-91.0	80.7 (27.4)	56.0-99.6

[a]LDI: lost data index.

[b]AUC: area under the curve.

XSL•FO
RenderX

Figure 6. Aggregated lost data index values by repetition with *P* values from the 1-tailed Mann-Whitney univariate *U* test.

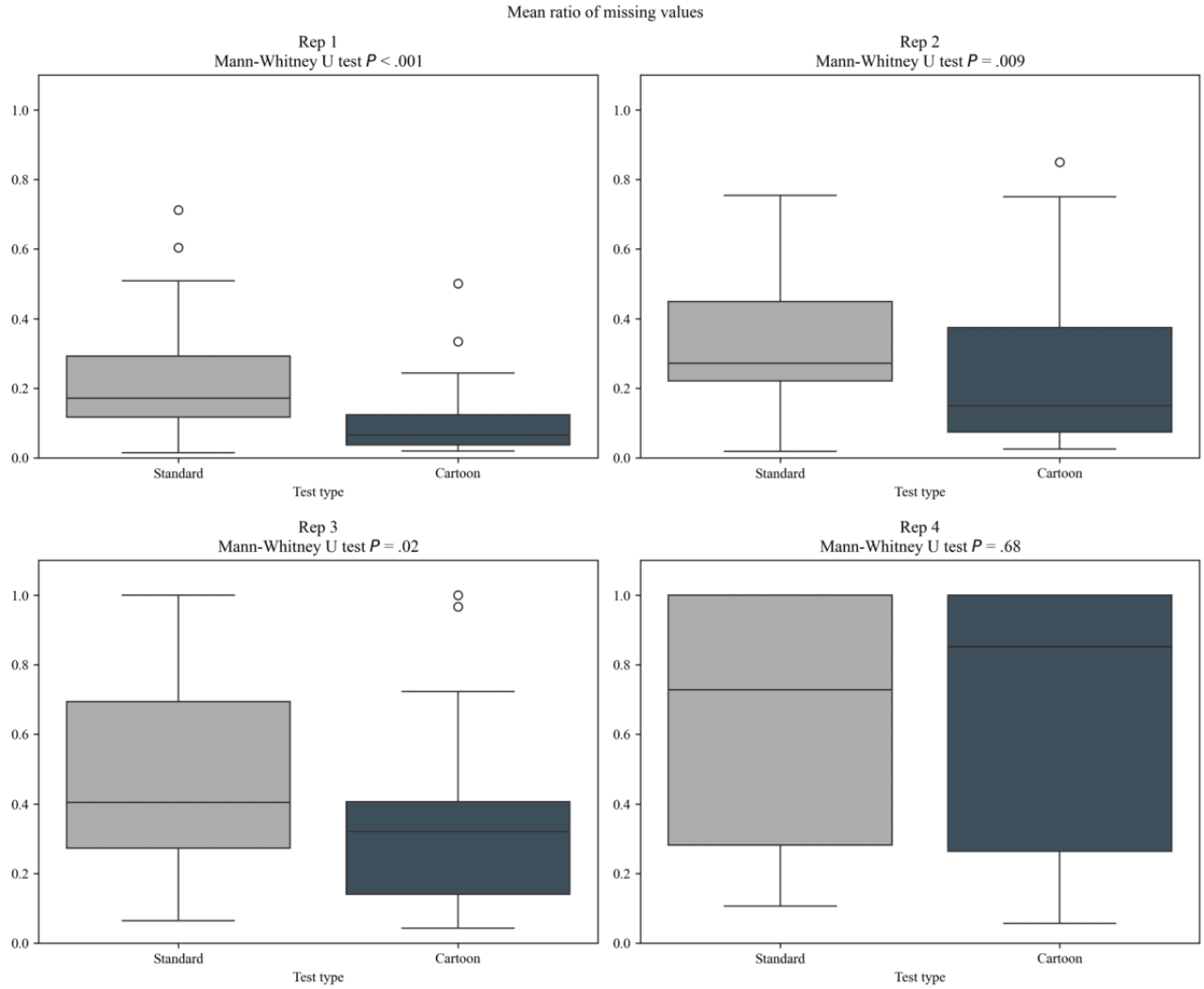
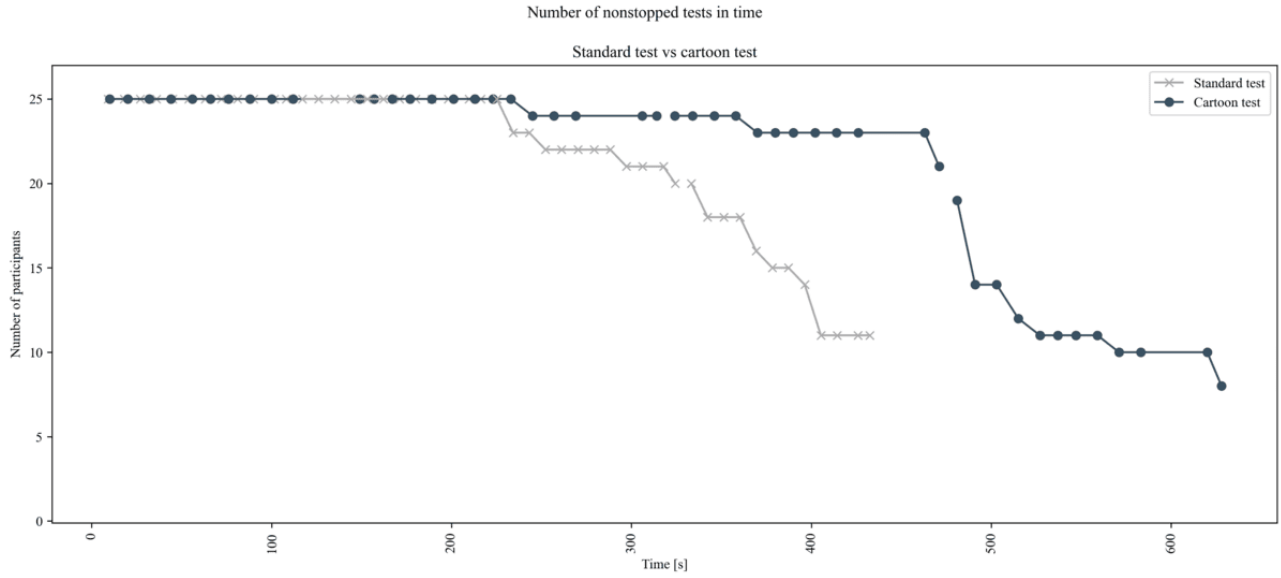


Figure 7. Number of uninterrupted tests over time for the two versions of the test. The vertical axis indicates the total number of participants who had not interrupted the test at a given time point.



These results are also reflected in the stated preferences of the participants: of the 25 children who have completed both tests, a total of 14 answered the follow-up question about which test they preferred. Out of the 14 participants, 11 (79%) showed a preference for the cartoon version over the standard version, while 2 (14%) preferred the standard version, and 1 (7%)

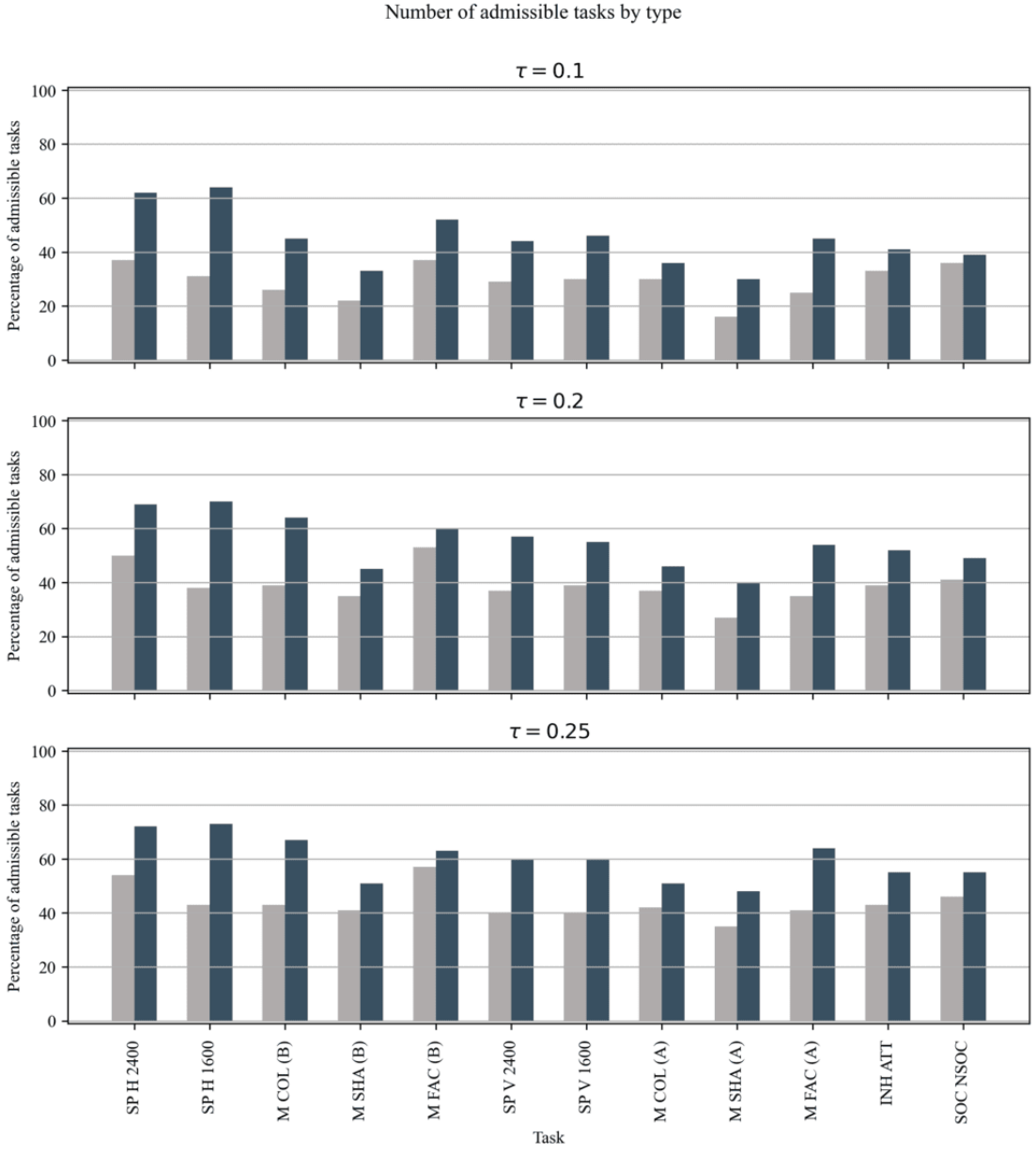
XSL•FO
RenderX

indicated that they did not like either version. The 2 participants who preferred the standard version provided distinct reasons: one participant was scared by the wizard of the cartoon version (Figure 2), while the other liked the standard version because it was shorter. Figure 6 shows aggregated LDI values by repetition, with *P* values of the 1-tailed univariate *U* test.

Experiment 2: Admissible Tasks and Test Length

Given the results of experiment 1 (ie, smaller LDI in the initial repetitions of the cartoon version), a practical consequence is that fewer tasks would be discarded due to data loss. This would result in either more robust measurements when using a static test or a shorter procedure when using a dynamic test. Figure 8 illustrates this effect: for 3 threshold value (τ=0.1, τ=0.2, and τ=0.25) the total number of admissible tasks is consistently higher in the cartoon test. This is expected given the lower LDI demonstrated in the previous section and should not be interpreted as additional evidence. Instead, it serves as an example of expected behavior in a clinical setting. Additionally, Figure 4 shows how the number of incomplete tests and the average testing time vary as the parameter τ is changed. Then we conducted a point comparison using a value of τ=0.2 and another comparison by computing the AUC for each participant. The point comparison shows that at τ=0.2, 68% (17/25) of participants had incomplete standard tests, compared with 32% (8/25) for the cartoon test. In addition, there is no strong evidence that testing time is higher for the cartoon version (*P*=.21 *U* test). When comparing AUCs, the situation is analogous, and a Mann-Whitney test shows that there is no strong evidence that the cartoon test has a higher AUC compared with the standard test (*P*=.50). Table 2 provides a full exposition of the results.

Figure 8. Percentage of total admissible tasks for τ=0.1, τ=0.2, and τ=0.25, divided by task type. A task is considered admissible if the lost data index was ≤τ. INH ATT: attention task; M COL (A/B): memory task-color (variant A/B); M FAC (A/B): memory task-face (variant A/B); M SHA (A/B): memory task-shape (variant A/B); SOC NSOC: social orienting task; SP H 1600: smooth pursuit task horizontal-fast; SP H 2400: smooth pursuit task horizontal-slow; SP V 1600: smooth pursuit task vertical-fast; SP V 2400: smooth pursuit task vertical-slow.



XSL•FO
RenderX

Discussion

Principal Results

In this study, we investigate the effect of gamification on participants' engagement during an eye-tracking test battery in a cohort of 5-year-old children born preterm. This age group was selected based on observations from previous studies [14,36], in which we identified challenges in capturing the attention of older children compared with infants.

The core result of this study is that, compared with a standard testing procedure, the cartoon version better maintained participant's attention on the screen in the considered population. The first 3 repetitions of the tasks displayed a lower LDI in the cartoon version, indicating that participants focused on the screen more often and were less frequently distracted. This result is supported by the subjective feedback of the participants: among those who expressed a preference, the majority (11/14, 79%) preferred the cartoon test. Among the 2 (2/14, 14%) that preferred the standard test, one was scared by an element, and the other felt that the test was too long. This finding underscores the importance of selecting appropriate gamified stimuli and suggests that a dynamic stopping criterion might enhance the experience.

Another observation of this study is the decline in engagement at the fourth repetition across both the cartoon test and standard assessment. This consistent trend suggests that the repetitiveness of the stimuli plays a significant role in reduced motivation, and that gamification cannot fully compensate for this decline. The phenomenon can be attributed to various factors, including boredom or habituation to stimuli, all of which are prevalent in pediatric populations with shorter attention spans [15]. As such, it is essential to ensure high-quality recordings during the initial repetitions of the stimuli. The cartoon test, despite its engaging nature, is not immune to these effects. However, its ability to mitigate the impact of test duration relative to the standard test is noteworthy, with a higher retention rate (21/25, 84% vs 15/25, 60%, an increase of 40%).

By incorporating visually stimulating elements, the cartoon test increases task engagement and reduces the monotony associated with repeated exposure to similar stimuli. Moreover, the capacity for longer testing durations and the inclusion of more admissible tasks, due to a lower LDI, address two limitations of traditional assessment methods. By sustaining engagement through visually appealing and dynamic content, the cartoon test enables researchers to gather a more extensive dataset within a single session. This can be clearly seen in the results summarized in Figure 4, where the two tests, if dynamic stopping is used, reach a similar duration, indicating that with the cartoon test fewer repetitions are sufficient to gather enough data for analysis. Indeed, with the stopping criterion described and τ=0.2, about 40% (10/25) of the cartoon tests required only 1 repetition to obtain enough admissible tasks, compared with 9% (2/25) of the standard tests, thereby minimizing the disruptive effect caused by a large number of repetitions. These results show that gamification of eye-tracking test batteries for 5-year-old children born preterm is a promising research direction to increase data quality and engagement, without necessarily adding to the duration of the procedure.

It is also important to note that the effect of gamification on task performance is outside the scope of this study. In contrast with previous studies [16,18,21], we focused exclusively on the engagement and data quality aspects of gamification.

Comparison With Previous Work

Our results appear to be in contrast with a previous study by Scharinger et al [34], in which they used pupil diameter and fixation duration to measure engagement in a gamified version of the N-block test. By comparing with the standard version of the exercise, they did not find a significant effect of gamification on the measured outcomes. This lack of a difference between the two versions can be explained by the age of their sample, which was composed of university students, or by the choice of measured outcomes: LDI in our case and pupil diameter in theirs. It is worth noting that, even without a measurable difference in gaze behavior, the participants found the gamified version subjectively more interesting. In contrast, our results agree with a similar comparison conducted by Friehs et al [21], in which a cohort of university-aged participants displayed similar performance and increased enjoyment with the gamified paradigm. The main difference with previous studies is that, in our case, the age of the participants makes it challenging to use a questionnaire to measure engagement with the test. Consequently, we have focused on the eye-tracking measures such as LDI to compare the two versions, in addition to short qualitative feedback from the participants. Previous research by Wiley et al [16] also highlights that the increase in engagement brought by gamification can increase consistency in performance of the stop-signal task between multiple tests in an adult population. This would be in accordance with our own results, since low LDI values and a high number of admissible tasks are linked to more stable measurements. However, again, the key dissimilarity is the age difference in the participants. Gallaghar et al [18] have tried to extend the results of Wiley et al [16] to younger participants (aged 8-12 years) with attention deficit hyperactivity disorder, receiving positive feedback, although a direct comparison with the standard version of the task was not included.

Limitations

Finally, while the cartoon version demonstrates considerable promise, this study is not without limitations. First, we had a small sample size, which restricts the generalizability of findings and is the reason we focused on a single quantity of interest and did not take into account the effects of specific task types. Second, our approach of considering only the LDI, while increasing the scope and generalizability of our results, could be complemented by other measures such as pupil size [34]. Finally, while some of these results might translate to different populations, as shown in other studies [16,18,21], the test we used was designed specifically for young children, and we cannot assume that similar results would be found in adults.

Future Work

The results of this study suggest that future work in the development of cartoon versions of existing eye-tracking

XSL•FO
RenderX

paradigms could be game-changing when testing young children, who are otherwise difficult to keep engaged. Future work should focus on further analysis including pupil size [34] and task performance [16]. Moreover, in this study we did not consider carry-over effects due to task order, and this would be a crucial step in creating a dynamic version of the test, where gamification could be used to shorten the procedure.

Conclusions

In this paper we directly compared two versions of the same eye-tracking test battery, a standard version, and a gamified cartoon version. We have shown that within a group of preterm born five-year-old participants, the cartoon version is correlated with a lower level of data loss and distraction by the participant. This difference is especially present in the first repetitions of the tasks, and its decline during the test might be due to the repetitiveness of the stimuli. Nonetheless, due to the lower data loss caused by gamification, we can expect not only to improve data quality in the recordings but, by implementing dynamical stopping criteria, shorten the testing procedure. Our results provide practical guidelines to gamifying eye-tracking tests by offering a direct comparison and, by implementing the recommendations outlined in the discussion, researchers can refine the design of serious games and maximize their potential as a reliable and effective tool in pediatric cognitive research.

Acknowledgments

The authors would like to thank Marieke Geldof and Diete Vandendungen for their assistance in recruiting the participants, performing the tests, and collecting the clinical data.

This research was funded by the European Union Horizon 2020 Marie Skłodowska-Curie Actions-Innovative Training Networks (MSCA-ITN) project “Premature Newborn Motor and Cognitive Impairments: Early diagnosis” (PARENT, grant 956394) and the Slovenian Research and Innovation Agency (ARIS) under the research program Artificial Intelligence and Intelligent Systems (grant P2-0209).

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

Conceptualization of this study was done by ADG, EO, AS, and VG. Methodology was developed by ADG, AS, and VG. Validation was performed by ADG. Formal analysis was conducted by ADG and VG. Investigation was carried out by ADG, AS, and VG. Resources were provided by ADG. Data curation was managed by ADG. The original draft was written by ADG, and review and editing were completed by MM, NDB, EO, AS, and VG. Visualization was handled by ADG and MM. Supervision and project administration were overseen by NDB, EO, AS, and VG. Funding was acquired by EO, AS, and VG. All authors have read and approved the published version of the manuscript.

Conflicts of Interest

VG and AS are co-founders of NEUS Diagnostics d.o.o., VG is the CEO of NEUS Diagnostics d.o.o., and ADG was employed by NEUS Diagnostics d.o.o. as a researcher within the project PARENT. All other authors declare no conflicts of interest.

Multimedia Appendix 1
Example of some of the tasks in the standard test. Tasks: Smooth pursuit task (horizontal), Memory task (colour modality, variant B), Memory task (face modality, variant B), Attention task (valid modality only), Social orienting task.
[MP4 File (MP4 Video), 2348 KB - games_v13i1e77109_app1.mp4]

Multimedia Appendix 2
Example of some of the tasks in the cartoon test. Tasks: Smooth pursuit task (horizontal, bird), Memory task (colour modality, variant A), Memory task (face modality, variant A), Attention task (valid modality only), Social orienting task.
[MP4 File (MP4 Video), 12376 KB - games_v13i1e77109_app2.mp4]

References

1. Ylijoki M, Sentenac M, Pape B, Zeitlin J, Lehtonen L. The aetiology of preterm birth and risks of cerebral palsy and cognitive impairment: a systematic review and meta-analysis. Acta Paediatr 2024;113(4):643-653. [doi: 10.1111/apa.17118] [Medline: 38265113]
2. Axford SB, Burnett AC, Seid AM, Anderson PJ, Waterland JL, Gilchrist CP, et al. Risk factor effects on neurodevelopment at 2 years in very preterm children: a systematic review. Pediatrics 2025;155(6):e2024069565. [doi: 10.1542/peds.2024-069565] [Medline: 40368397]

XSL•FO
RenderX

3. Bhutta AT, Cleves MA, Casey PH, Cradock MM, Anand KJS. Cognitive and behavioral outcomes of school-aged children who were born preterm: a meta-analysis. JAMA 2002;288(6):728-737. [doi: 10.1001/jama.288.6.728] [Medline: 12169077]
4. Gredebäck G, von Hofsten C, Karlsson J, Aus K. The development of two-dimensional tracking: a longitudinal study of circular pursuit. Exp Brain Res 2005;163(2):204-213. [doi: 10.1007/s00221-004-2162-0] [Medline: 15688173]
5. Kooiker MJ, Verbunt HJ, van der Steen J, Pel JJ. Combining visual sensory functions and visuospatial orienting functions in children with visual pathology: a longitudinal study. Brain Dev 2019;41(2):135-149. [doi: 10.1016/j.braindev.2018.09.006] [Medline: 30301589]
6. Telford EJ, Fletcher-Watson S, Gillespie-Smith K, Pataky R, Sparrow S, Murray IC, et al. Preterm birth is associated with atypical social orienting in infancy detected using eye tracking. J Child Psychol Psychiatry 2016;57(7):861-868. [doi: 10.1111/jcpp.12546] [Medline: 26934180]
7. Hokken MJ, Stein N, Pereira RR, Rours IG, Frens MA, van der Steen J, et al. Eyes on CVI: eye movements unveil distinct visual search patterns in cerebral visual impairment compared to ADHD, dyslexia, and neurotypical children. Res Dev Disabil 2024;151:104767 [FREE Full text] [doi: 10.1016/j.ridd.2024.104767] [Medline: 38861794]
8. Beunders VAA, Vermeulen MJ, Roelants JA, Rietema N, Swarte RMC, Reiss IKM, et al. Early visuospatial attention and processing and related neurodevelopmental outcome at 2 years in children born very preterm. Pediatr Res 2021;90(3):608-616. [doi: 10.1038/s41390-020-01206-7] [Medline: 33070166]
9. Chorna O, Corsi G, Del Secco S, Bancale A, Guzzetta A. Correlation between early visual functions and cognitive outcome in infants at risk for cerebral palsy or other neurodevelopmental disorders: a systematic review. Children (Basel) 2024;11(6):747 [FREE Full text] [doi: 10.3390/children11060747] [Medline: 38929326]
10. Oakes L. Advances in eye tracking in infancy research. Infancy 2011;17(1):1-8. [doi: 10.1111/j.1532-7078.2011.00101.x] [Medline: 32693501]
11. Corbetta D, Guan Y, Williams JL. Infant eye‐tracking in the context of goal‐directed actions. Infancy 2011;17(1):102-125. [doi: 10.1111/j.1532-7078.2011.00093.x]
12. Gredebäck G, Johnson S, von Hofsten C. Eye tracking in infancy research. Dev Neuropsychol 2010;35(1):1-19. [doi: 10.1080/87565640903325758] [Medline: 20390589]
13. Venker CE, Kover ST. An open conversation on using eye-gaze methods in studies of neurodevelopmental disorders. J Speech Lang Hear Res 2015;58(6):1719-1732 [FREE Full text] [doi: 10.1044/2015_JSLHR-L-14-0304] [Medline: 26363412]
14. Eye-tracking test battery for newborns: a pilot feasibility study. CEUR Workshop Proceedings. 2022. URL: https://ceur-ws.org/Vol-3363/paper04.pdf [accessed 2025-09-13]
15. Lumsden J, Edwards EA, Lawrence NS, Coyle D, Munafò MR. Gamification of cognitive assessment and cognitive training: a systematic review of applications and efficacy. JMIR Serious Games 2016;4(2):e11 [FREE Full text] [doi: 10.2196/games.5888] [Medline: 27421244]
16. Wiley K, Berger P, Friehs MA, Mandryk RL. Measuring the reliability of a gamified stroop task: quantitative experiment. JMIR Serious Games 2024;12:e50315 [FREE Full text] [doi: 10.2196/50315] [Medline: 38598265]
17. Stein A, Riddle J, Caulfield KA, Dux PE, Friehs MA, Schroeder PA, et al. Safety, feasibility, and tolerability of ten days of at-home, remotely supervised tDCS during gamified attention training in children with acquired brain injury: an open-label, dose-controlled pilot trial. Brain Sci 2025;15(6):561 [FREE Full text] [doi: 10.3390/brainsci15060561] [Medline: 40563733]
18. Gallagher R, Kessler K, Bramham J, Dechant M, Friehs MA. A proof-of-concept study exploring the effects of impulsivity on a gamified version of the stop-signal task in children. Front Psychol 2023;14:1068229 [FREE Full text] [doi: 10.3389/fpsyg.2023.1068229] [Medline: 36844283]
19. Held L, Pannermayr J, Kaufmann A, Scheffer M, Flores P, Dechant M, et al. No impact of story context and avatar power on performance in a stop-signal game. Heliyon 2025;11(1):e41039 [FREE Full text] [doi: 10.1016/j.heliyon.2024.e41039] [Medline: 39802016]
20. Wiley K, Vedress S, Mandryk K. How points and theme affect performance and experience in a gamified cognitive task. 2020 Presented at: CHI '20: 2020 CHI Conference on Human Factors in Computing Systems; April 25-30, 2020; Honolulu, HI p. 1-15. [doi: 10.1145/3313831.3376697]
21. Friehs MA, Dechant M, Vedress S, Frings C, Mandryk RL. Effective gamification of the stop-signal task: two controlled laboratory experiments. JMIR Serious Games 2020 Sep 08;8(3):e17810 [FREE Full text] [doi: 10.2196/17810] [Medline: 32897233]
22. Ross-Sheehy S, Eschman B. Assessing visual STM in infants and adults: eye movements and pupil dynamics reflect memory maintenance. Visual Cognition 2019;27(1):78-92. [doi: 10.1080/13506285.2019.1600089]
23. Kaul YF, Rosander K, von Hofsten C, Strand Brodd K, Holmström G, Hellström-Westas L. Visual tracking at 4 months in preterm infants predicts 6.5-year cognition and attention. Pediatr Res 2022;92(4):1082-1089. [doi: 10.1038/s41390-021-01895-8] [Medline: 34949760]
24. Kooiker MJ, van der Steen J, Pel JJ. Reliability of visual orienting response measures in children with and without visual impairments. J Neurosci Methods 2014;233:54-62. [doi: 10.1016/j.jneumeth.2014.06.005] [Medline: 24931709]
25. Kooiker MJG, Pel JJM, Verbunt HJM, de Wit GC, van Genderen MM, van der Steen J. Quantification of visual function assessment using remote eye tracking in children: validity and applicability. Acta Ophthalmol 2016;94(6):599-608 [FREE Full text] [doi: 10.1111/aos.13038] [Medline: 27130515]

XSL•FO
RenderX

26. Ross-Sheehy S, Schneegans S, Spencer JP. The infant orienting with attention task: assessing the neural basis of spatial attention in infancy. Infancy 2015;20(5):467-506 [FREE Full text] [doi: 10.1111/infa.12087] [Medline: 26273232]
27. Wass SV, Forssman L, Leppänen J. Robustness and precision: how data quality may influence key dependent variables in infant eye‐tracker analyses. Infancy 2014;19(5):427-460. [doi: 10.1111/infa.12055]
28. Groznik V. Gaze behaviour during reading as a predictor of mild cognitive impairment. 2021 Presented at: IEEE EMBS International Conference on Biomedical and Health Informatics (BHI); July 27-30, 2021; Athens, Greece p. 1-4. [doi: 10.1109/bhi50953.2021.9508586]
29. Oyama A, Takeda S, Ito Y, Nakajima T, Takami Y, Takeya Y, et al. Novel method for rapid assessment of cognitive impairment using high-performance eye-tracking technology. Sci Rep 2019;9(1):12932 [FREE Full text] [doi: 10.1038/s41598-019-49275-x] [Medline: 31506486]
30. Damaševičius R, Maskeliūnas R, Blažauskas T. Serious games and gamification in healthcare: A meta-review. Information 2023;14(2):105. [doi: 10.3390/info14020105]
31. Haoran G, Bazakidi E, Zary N. Serious games in health professions education: review of trends and learning efficacy. Yearb Med Inform 2019;28(1):240-248 [FREE Full text] [doi: 10.1055/s-0039-1677904] [Medline: 31022747]
32. McPherson J, Burns NR. Assessing the validity of computer-game-like tests of processing speed and working memory. Behav Res Methods 2008;40(4):969-981. [doi: 10.3758/BRM.40.4.969] [Medline: 19001388]
33. Nicholson S. A RECIPE for meaningful gamification. In: Reiners T, Wood LC, editors. Gamification in Education and Business. Cham: Springer International Publishing; 2015:1-20.
34. Scharinger C, Prislan L, Bernecker K, Ninaus M. Gamification of an n-back working memory task - is it worth the effort? An EEG and eye-tracking study. Biol Psychol 2023;179:108545 [FREE Full text] [doi: 10.1016/j.biopsycho.2023.108545] [Medline: 36965785]
35. Hawkins GE, Rae B, Nesbitt KV, Brown SD. Gamelike features might not improve data. Behav Res Methods 2013;45(2):301-318. [doi: 10.3758/s13428-012-0264-3] [Medline: 23055169]
36. Eye-tracking test battery for detecting cognitive impairments in premature children. CEUR Workshop Proceedings. 2022. URL: https://ceur-ws.org/Vol-3363/paper03.pdf [accessed 2025-09-13]
37. Frutos-Pascual M, Garcia-Zapirain B. Assessing visual attention using eye tracking sensors in intelligent cognitive therapies based on serious games. Sensors (Basel) 2015;15(5):11092-11117 [FREE Full text] [doi: 10.3390/s150511092] [Medline: 25985158]
38. Groznik A, Sadikov A. Gamification in cognitive assessment and cognitive training for mild cognitive impairment. In: Augmented Reality Games II: The Gamification of Education, Medicine and Art Ed by V Geroimenko. Cham: Springer International Publishing; 2019:179-204.
39. Madsen J, Júlio SU, Gucik PJ, Steinberg R, Parra LC. Synchronized eye movements predict test scores in online video education. Proc Natl Acad Sci U S A 2021;118(5):e2016980118 [FREE Full text] [doi: 10.1073/pnas.2016980118] [Medline: 33509928]
40. Holmqvist K, Nyström M, Mulvey F. Eye tracker data quality: what it is and how to measure it. 2012 Presented at: ETRA '12: Symposium on Eye Tracking Research and Applications; March 28-30, 2012; Santa Barbara California p. 45-52. [doi: 10.1145/2168556.2168563]
41. Yeung T. Mapping practices of academic game development. 2025 Presented at: FDG '25: 20th International Conference on the Foundations of Digital Games; April 15-18, 2025; Vienna, Austria p. 1-11. [doi: 10.1145/3723498.3723799]
42. Crutcher MD, Calhoun-Haney R, Manzanares CM, Lah JJ, Levey AI, Zola SM. Eye tracking during a visual paired comparison task as a predictor of early dementia. Am J Alzheimers Dis Other Demen 2009;24(3):258-266 [FREE Full text] [doi: 10.1177/1533317509332093] [Medline: 19246573]
43. DeBruine L, Jones B. Face Research Lab London Set. figshare. 2017. URL: https://figshare.com/articles/dataset/Face_Research_Lab_London_Set/5047666 [accessed 2025-09-13]
44. Duchowski AT. Eye Tracking Methodology. Cham, Switzerland: Springer International Publishing; 2017.

Abbreviations

AUC: area under the curve
LDI: lost data index

XSL•FO
RenderX

Edited by A Coristine; submitted 07.05.25; peer-reviewed by M Friehs, R Garotti; comments to author 29.05.25; revised version received 21.07.25; accepted 28.08.25; published 27.10.25.

Please cite as:
De Gobbis A, Malavolta M, De Beukelaer N, Ortibus E, Sadikov A, Groznik V
Effect of Gamification on Task Engagement During an Eye-Tracking Test Battery in 5-Year-Old Children Born Preterm: Observational Study
JMIR Serious Games 2025;13:e77109
URL: https://games.jmir.org/2025/1/e77109
doi:10.2196/77109
PMID:

©Andrea De Gobbis, Marta Malavolta, Nathalie De Beukelaer, Els Ortibus, Aleksander Sadikov, Vida Groznik. Originally published in JMIR Serious Games (https://games.jmir.org), 27.10.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Measuring Stress and Perceptions for a Virtual Reality–Based Pericardiocentesis Procedure Simulation for Medical Training: Usability Study

Alberto Rubio-López[1,2], MSc, MD; Rodrigo García Carmona[3], PhD; Laura Zarandieta Román[3], MSc; Alejandro Rubio Navas[4], MSc; Angel González-Pinto[5,6], MD, PhD; Pablo Cardinal-Fernández[7,8], MD, PhD

[1]Intensive Care Unit, Hospital Universitario HM Montepríncipe, HM Hospitales, Avda Montepríncipe, Boadilla del Monte, Madrid, Spain

[2]Facultad de Medicina, Universidad San Pablo CEU, Madrid, Spain

[3]Departamento de Tecnologías de la Información, Escuela Politécnica Superior, Universidad San Pablo CEU, Madrid, Spain

[4]Departamento de Fisiología, Facultad de Biología, Universidad Autónoma de Madrid, Madrid, Spain

[5]Jefe de Servicio. Unidad de Cirugía Cardiaca, Hospital General Universitario Gregorio Marañón, Madrid, Spain

[6]Jefe de Servicio de Cirugía Cardiaca, Hospital Universitario HM Montepríncipe, HM Hospitales, Boadilla del Monte, Madrid, Spain

[7]Intensive Care Unit, Hospital Universitario HM Torrelodones, HM Hospitales, Madrid, Spain

[8]Facultad HM de Ciencias de la Salud, Camilo José Cela University, Villanueva de la Cañada, Madrid, Spain

Corresponding Author:
Alberto Rubio-López, MSc, MD
Intensive Care Unit, Hospital Universitario HM Montepríncipe, HM Hospitales, Avda Montepríncipe, Boadilla del Monte, Madrid, Spain

Abstract

Background: Virtual reality (VR) is increasingly used in medical education, providing immersive environments for training in high-risk procedures such as pericardiocentesis. This lifesaving procedure requires technical precision and induces cognitive and physiological stress. Evaluating both usability and stress responses in a VR-based pericardiocentesis simulation is essential. Heart rate variability (HRV) serves as an objective stress marker, while prior VR experience may influence usability and stress perception.

Objectives: This study aimed to assess the usability of a VR-based pericardiocentesis simulation, examine the relationship between usability perceptions and physiological stress (HRV), and determine the impact of prior VR experience on usability scores and stress responses.

Methods: A total of 119 final-year medical students participated in a VR pericardiocentesis simulation. Usability was evaluated using the System Usability Scale (SUS), the Post-Study System Usability Questionnaire, the Presence Questionnaire, and the Simulator Sickness Questionnaire. Physiological stress was assessed through HRV parameters, including the root-mean-square of successive differences (rMSSDs), percentage of differences greater than 50 ms (PNN50), low-frequency to high-frequency ratio, and nonlinear HRV indices (SD1/SD2 ratio, Poincaré area). Statistical analyses included descriptive statistics, Spearman correlations, and Mann-Whitney *U* tests to explore relationships between usability, stress, and prior VR experience.

Results: The VR simulation received a mean SUS score of 75.00 (SD 6.41; 95% CI 73.42 - 76.58), exceeding the general usability threshold of 68 (*P*=.002). The mean Post-Study System Usability Questionnaire score of 2.92 (SD 1.83; 95% CI 2.55 - 3.29) indicated moderate satisfaction, while the mean Presence Questionnaire score of 109.46 (SD 9.12; 95% CI 107.88 - 111.04) reflected strong immersion. Simulator sickness symptoms were mild (mean Simulator Sickness Questionnaire score 12.43, SD 15.41; 95% CI 9.28 - 15.58), although novice users reported significantly higher nausea levels (*P*=.02). Physiological stress analysis revealed a mean rMSSD of 281.27 (SD 98.99; 95% CI 259.45 - 303.09) ms and PNN50 of 56.85% (SD 19.70%; 95% CI 52.23% - 61.47%), indicating moderate autonomic balance. A significant negative correlation was observed between HRV parameters (rMSSD and PNN50) and simulator sickness (*P*=.04; Spearman ρ=−0.23), suggesting that higher physiological stress was associated with increased simulator sickness symptoms. Prior VR experience was linked to higher usability scores (SUS +5.2; 95% CI 3.12-7.28; *P*=.03) and lower simulator sickness symptoms (*P*=.02) but did not significantly affect HRV markers.

Conclusions: VR-based simulations for high-risk medical procedures are effective training tools with high usability (SUS=75) and strong immersion. Simulator sickness correlated with physiological stress, emphasizing the need for design refinements to improve user comfort. Prior VR experience improved usability and reduced simulator sickness but did not significantly impact HRV markers. Future research should refine VR interfaces to balance immersion with minimized cognitive and physical discomfort.

XSL•FO
RenderX

(JMIR Serious Games 2025;13:e68515) doi:10.2196/68515

KEYWORDS

VR; pericardiocentesis simulation; usability assessment; heart rate variability; physiological stress; simulator sickness; System Usability Scale; Presence Questionnaire; virtual reality

Introduction

Virtual reality (VR) has rapidly emerged as a transformative tool in medical education. Its ability to simulate high-risk procedures in a controlled and safe environment has made VR an invaluable resource for developing essential clinical skills [1]. One of these procedures, pericardiocentesis—an urgent, lifesaving intervention that involves inserting a needle into the pericardium to drain fluid around the heart—requires not only technical precision but also the ability to remain calm under extreme pressure [2]. In real clinical settings, this combination of skills is critical, and VR simulations offer a safe and repeatable platform for medical students to develop these critical skills without the risks associated with real-life clinical practice.

The immersive nature of VR is especially effective in recreating the cognitive and emotional challenges clinicians face during real-life medical emergencies. However, this immersion can also induce significant stress in users, particularly when simulating high-stress, high-risk procedures such as pericardiocentesis. To ensure the effectiveness of VR simulations in preparing medical students for real-world scenarios, it is crucial to evaluate both the usability of the system and the physiological stress responses it triggers [3]. Heart rate variability (HRV) is a well-established biomarker used to measure stress, providing objective data on how users manage stress in high-pressure situations. Evaluating both aspects is essential to optimize VR-based training and ensure that students are adequately prepared for high-pressure clinical scenarios [4].

For VR to function as an effective educational tool, it must be intuitive and user-friendly. Standard usability scales, such as the System Usability Scale (SUS) [5] and the Post-Study System Usability Questionnaire (PSSUQ) [6], assess how efficiently users interact with the system and whether the technology facilitates or hinders learning. A high usability score indicates a low-cognitive load, allowing learners to focus on the procedural and clinical aspects of the task rather than on navigating the VR environment [7].

Another key factor in effective VR training is presence, defined as the degree to which users feel immersed in the virtual environment. The Presence Questionnaire (PQ) [8] evaluates this aspect by measuring user engagement and perceived realism. A strong sense of presence is essential for simulating real-world clinical pressure, as it encourages learners to respond as they would in an actual medical emergency [9]. Balancing high usability with a strong presence ensures that VR simulations are both accessible and engaging for medical trainees.

Training for high-risk medical procedures involves not only cognitive demands but also physiological stress. Immersive VR simulations have been shown to trigger autonomic nervous system (ANS) responses, which can influence performance, learning, and emotional regulation [10]. HRV is a well-established physiological marker of stress, providing insights into autonomic regulation during demanding tasks [11]. HRV parameters reflect different aspects of autonomic function:

- Parasympathetic (vagal) tone: indicators such as the root-mean-square of successive differences (rMSSD), proportion of NN50 (pNN50), high-frequency (HF) power, and standard deviation 1 (SD1) are associated with relaxation and adaptive stress responses [12].
- Overall ANS oscillations: metrics including SD of NN intervals, low-frequency (LF) power, and standard deviation 2 (SD2) represent the combined influence of both sympathetic and parasympathetic branches of the ANS, indicating autonomic flexibility and stress regulation.
- Sympathovagal balance: the LF/HF ratio and SD2/SD1 ratio indicate the relative dominance of sympathetic (stress-related) and parasympathetic (relaxation-related) activity, where a higher LF/HF ratio suggests greater sympathetic activation.

By analyzing these HRV parameters, researchers can assess how physiological stress responses interact with cognitive and emotional demands in VR-based medical training [13]. However, the extent to which these markers correlate with usability and immersion remains unclear. User experience with VR varies significantly, and prior exposure to VR technology may affect both usability perceptions and stress responses [14]. Medical students with previous VR experience may find the system more intuitive, report higher usability scores, and experience less stress, as they are already familiar with navigating virtual environments. Conversely, students without prior VR exposure may struggle with interaction mechanics, leading to heightened cognitive load and stress, particularly in high-stakes scenarios such as pericardiocentesis. Understanding how prior VR experience influences usability and physiological stress is crucial for designing adaptive VR training programs that accommodate users with varying levels of familiarity.

While previous research has investigated VR usability and stress responses, few studies have explored their relationship, especially in the context of high-risk medical procedures such as pericardiocentesis. Furthermore, the role of prior VR experience in shaping both usability perceptions and physiological adaptation to stress remains poorly understood. This study aimed to fill this gap by examining the interplay between usability, HRV-based stress responses, and prior VR experience in a cohort of final-year medical students undergoing a pericardiocentesis simulation. Findings from this research will provide valuable insights into how VR training environments can be refined to better prepare students for the cognitive and physiological demands of real-world clinical practice.

XSL•FO
RenderX

Methods

Participants

A total of 119 final-year medical students were recruited through in-class announcements. During lecture sessions, researchers provided an overview of the study's objectives, procedures, risks, and benefits, allowing students to ask questions before deciding to enroll. Participation was voluntary, with no financial or material compensation provided. Interested students signed informed consent forms, which outlined confidentiality procedures and withdrawal rights. Recruitment was completed within 1 month. Participants ranged in age from 22 to 45 (mean 24.03, SD 3.4) years, with 78.4% (93/119) identifying as female. The sample was representative of the student body demographics at the institution. Prior VR experience was assessed as a binary variable (yes/no)—participants reported whether they had previously used VR for any purpose. Those with prior VR exposure (31/119, 26.1%) engaged with VR-based applications (eg, gaming, training, and immersive experiences), whereas the remaining 73.9% (88/119) had no prior VR experience.

Inclusion and Exclusion Criteria

Eligible participants were final-year medical students aged 22 - 45 years without uncorrected visual impairments. Students using corrective eyewear (glasses or contact lenses) were eligible. To ensure the validity of HRV measurements and usability assessments, the following exclusion criteria were applied: (1) cardiac medication use (eg, β-blockers) due to potential effects on autonomic regulation; (2) excessive caffeine consumption (>400 mg/day, equivalent to ~4 cups of coffee), as it can alter physiological stress responses; (3) use of sympathomimetic substances (eg, decongestants and stimulants) due to potential HRV interference; (4) history of severe motion sickness, which could bias usability and simulator sickness assessments; and (5) incomplete data or poor-quality HRV recordings, ensuring data integrity. These criteria minimized confounding factors and ensured a controlled evaluation of usability and physiological responses.

Instruments

Usability Scales

After completing the VR simulation, participants electronically completed four validated questionnaires via Microsoft Forms. They accessed the forms on personal devices in a quiet environment to ensure consistent data collection.

- The System Usability Scale is a 10-item, 5-point Likert scale measuring ease of use (0 - 100, higher scores=better usability). It includes 2 subscales: SUS-Usability (overall ease of interaction) and SUS-Learning (effort required to learn system use).
- The Post-Study System Usability Questionnaire is a 16-item scale measuring user satisfaction in 3 domains: system usefulness, information quality, and interface quality. Responses are recorded on a 7-point Likert scale, where lower scores indicate higher usability perceptions.
- The Presence Questionnaire is a 24-item scale assessing immersion within the VR environment, with subscales measuring involvement, control, and interface quality. Scores range from 1 ("not at all") to 7 ("very much"), with higher scores reflecting stronger engagement with the virtual environment.
- The Simulator Sickness Questionnaire (SSQ) evaluates symptoms of simulator sickness across 3 domains: nausea, oculomotor discomfort, and disorientation. Participants rate their symptoms on a 4-point scale (0="none" to 3="severe"). The SSQ provides both individual subscale scores and a total sickness score.

The complete versions of these questionnaires are available as Multimedia Appendices 1-4.

HRV Parameters

HRV was continuously recorded using the Biosignal Plux system (Figure 1). Three torso-mounted electrodes measured heart rate, with data processed via OpenSignals (Plux Wireles Biosignals) software to extract key HRV parameters:

- rMSSD: reflects parasympathetic activity, indicating relaxation.
- PNN50: percentage of successive heart rate intervals that differ by more than 50 ms, reflecting vagal tone.
- LF/HF ratio: indicator of autonomic balance (higher= sympathetic dominance, lower= parasympathetic dominance).
- SD1/SD2 ratio: reflects short-term versus long-term HRV variability, with larger values indicating greater autonomic flexibility.

XSL•FO
RenderX

Figure 1. Participant interacting with the virtual reality simulation for pericardiocentesis using Oculus Rift S.The simulation replicates the procedural steps required in a real clinical setting. The session lasted approximately 10 minutes and included real-time physiological stress monitoring via heart rate variability measurements.



VR Simulation

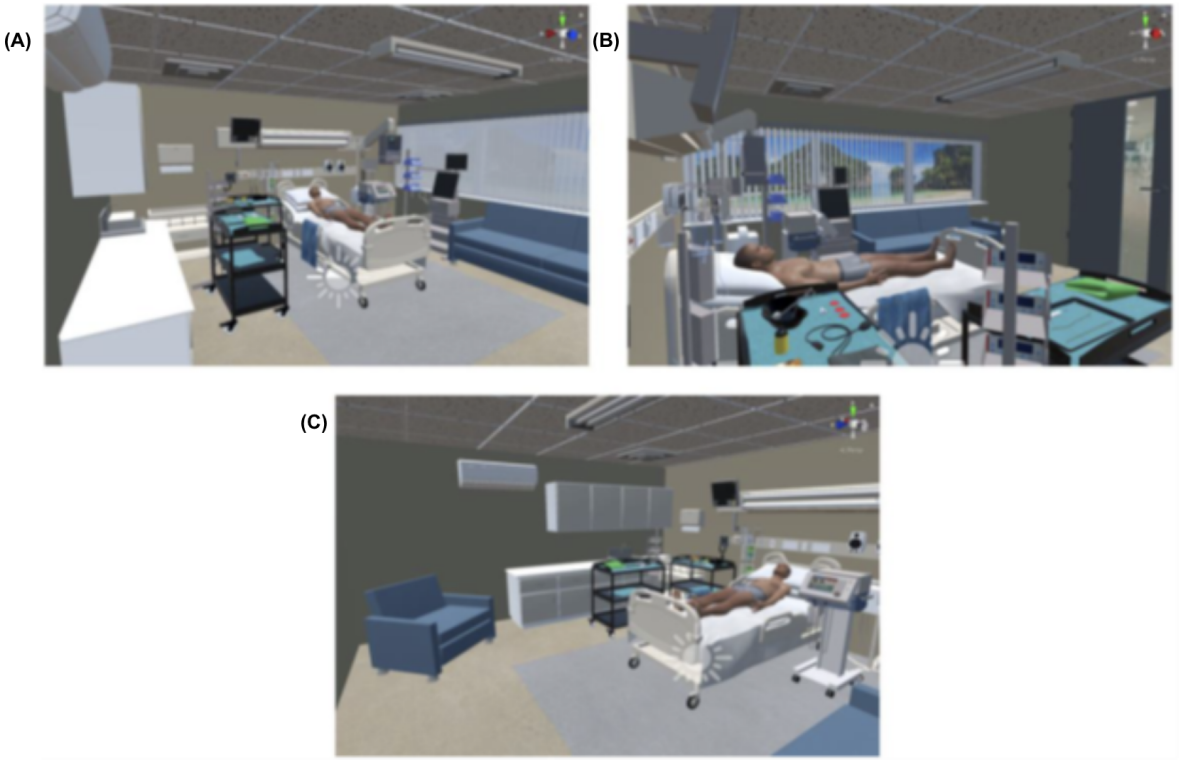
The VR pericardiocentesis simulation was developed in Unity 2021.3.24f1 (Unity Technologies) with the Oculus Integration SDK. The environment included a virtual hospital room, a patient bed, vital signs monitor, and interactive procedural tools.

On top of the bed lies a male patient, and besides him are 2 easily accessible tables with the following instruments: three electrodes, a pressure cuff, a pulse oximeter, an oxygen mask, a box of gloves, a Betadine bottle, forceps with a clean gauze, a surgical cloth, a tray with a syringe, a straight endovascular device, and a bent endovascular device. All these items are interactive and should be used to perform the procedure in the proper order and places.

The vital signs monitor updates when the pressure cuff, the electrodes, the pulse oximeter, and the oxygen masks are placed on the patient, and when the procedure finishes. The data shown on the monitor are accompanied by a beeping sound, recorded from a real device. The complete setup can be seen in Figure 2.

XSL•FO
RenderX

Figure 2. Virtual hospital environment for pericardiocentesis training. The virtual reality simulation presents a high-fidelity virtual hospital room containing a patient bed, vital signs monitor, procedural instruments, and interactive medical equipment necessary for pericardiocentesis. The immersive design was developed using Unity 2021.3.24f1 and aimed to enhance skill acquisition and clinical decision-making among medical students. (A) General view of the simulated room environment. (B) Lateral view of the scenario showing the patient and surrounding procedural equipment. (C) Opposite-side perspective providing a panoramic overview of the full training environment.



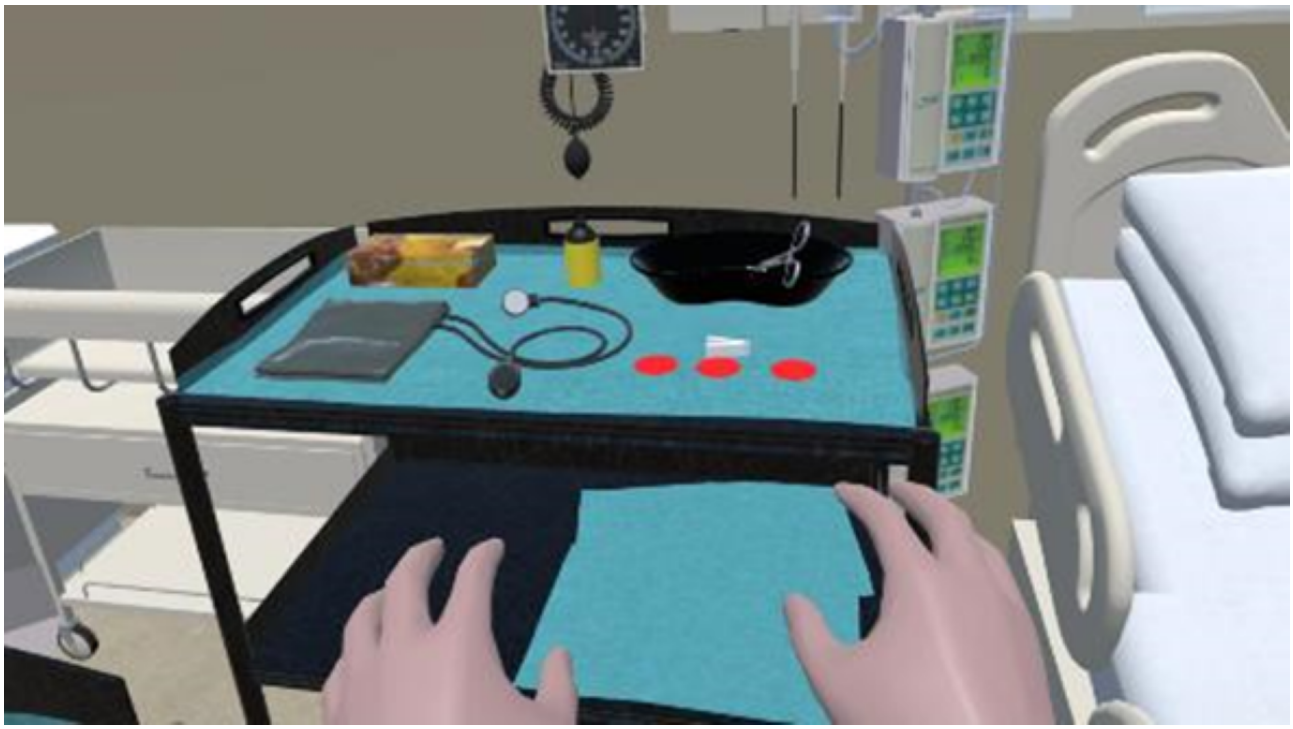
The user does not see their own body but can see 2 virtual hands (Figure 3), which also show the surgical gloves, if appropriate. All the interactions inside the simulation were diegetic: using virtual elements present in the scenario that replicate real items, without any menu, teleportation mechanism, heads-up display, or other immersion-breaking controls (Figure 4). The user can displace themselves only by direct movement, without any teleportation of stick-controlled movement mechanisms.

XSL•FO
RenderX

Figure 3. Virtual hand interface used for procedural interactions. The virtual reality system uses diegetic interaction controls, where users interact with virtual elements directly replicating real-world medical tools. The hands, visible within the simulation, dynamically update to include surgical gloves when appropriate. This design eliminates elements not present in the real world such as teleportation mechanics or external heads-up display, ensuring a realistic and immersive training experience.



Figure 4. User engagement within the virtual reality (VR) simulation. Still image from a recording of a participant performing pericardiocentesis inside the VR environment. The participant is navigating the simulation using handheld Oculus Rift controllers, interacting with virtual instruments to complete procedural steps. The VR session was part of a controlled study evaluating usability, immersion, and physiological stress responses in final-year medical students.



XSL•FO
RenderX

The VR simulation was conducted using the Oculus Rift S headset, which was calibrated before each session to ensure accurate tracking and optimal user experience. The calibration process included adjusting the interpupillary distance, ensuring proper headset fit, and setting the default room-scale boundary to prevent tracking errors. Participants were instructed on how to adjust the headset for comfort, ensuring that the lenses were correctly aligned to minimize motion blur and discomfort. Environmental conditions were controlled to ensure consistency across all sessions. The simulation took place in a dedicated laboratory with dim lighting to reduce glare and reflections and a quiet setting to minimize distractions. Participants were allowed limited standing movement within a 3×3 m play area, which was predefined using the Guardian System of the Oculus Rift S to prevent accidental collisions. The temperature was maintained at a comfortable level to avoid physiological stress unrelated to the simulation. Each session lasted approximately 30 minutes, including headset fitting, familiarization with the controls, and the execution of the pericardiocentesis procedure. To ensure consistency across participants, all interactions were standardized, and the experimenter provided identical procedural instructions before each session.

Procedure

Participants completed the VR pericardiocentesis simulation in three phases:

1. Pre-VR phase: before starting the simulation, participants completed demographic and health behavior questionnaires, and their baseline HRV data were recorded. Participants also reported their prior experience with VR through a binary yes/no response.
2. VR session: throughout the VR simulation, HRV was continuously measured to assess participants' physiological responses to stress. All tasks were time-stamped to facilitate precise analysis of stress responses during different phases of the simulation
3. Post-VR phase: after completing the simulation, questionnaires were administered on the web using Microsoft Forms to ensure ease of completion and data collection. Completion of the questionnaires required approximately 15 minutes. All responses were automatically anonymized upon submission to protect participant confidentiality.

Ethical Considerations

The study was approved by the University Hospital research ethics committee (code: 18.12.1339.GHM), following the Declaration of Helsinki. Participants provided written informed consent after a full briefing on study objectives, risks, and withdrawal rights. No financial or material incentives were provided, and all participant data were anonymized.

Data Analysis

Descriptive Statistics

Descriptive analyses were conducted to summarize participants' responses across all questionnaires, including SUS, PSSUQ, PQ, and SSQ, as well as key HRV parameters. Measures of central tendency (mean), dispersion (SD), and distribution (percentiles) were computed to characterize usability perceptions, simulator sickness symptoms, presence levels, and physiological stress markers. These summary statistics provided the foundation for subsequent inferential analyses and helped identify potential differences based on prior VR experience.

Reliability Analysis

To evaluate the internal consistency of the instruments used in this study, Cronbach alpha was calculated for the SUS, PSSUQ, PQ, and SSQ scales, including their respective subscales. Reliability estimates provide insight into the coherence of participants' responses across items within each scale, indicating how well the instruments measure usability, presence, and simulator sickness. Alpha values above 0.8 were interpreted as strong reliability, while values below 0.7 were considered to indicate potential measurement inconsistency. These analyses ensured the robustness of the self-reported measures used throughout the study.

Statistical Tests

Spearman correlation examined relationships between HRV metrics, usability, and simulator sickness. Mann-Whitney *U* tests compared usability and stress responses between participants with and with no VR experience, while Friedman tests compared overall usability scores across scales. All data were securely stored and anonymized following institutional data protection guidelines. Statistical analyses were performed using IBM SPSS Statistics (version 25.0; IBM Corp), and physiological data were processed using OpenSignals software to ensure accuracy and reliability in HRV analysis.

Results

Usability Scores

The usability of the VR simulation was assessed using the SUS, PSSUQ, PQ, and SSQ. The internal reliability of these scales was evaluated using Cronbach α (Table 1). High reliability ($\alpha>0.8$) was found for SUS ($\alpha=0.810$) and PSSUQ ($\alpha=0.980$), confirming their consistency in assessing usability and user satisfaction. Some subscales, such as Nausea Symptoms (SSQ; $\alpha=0.472$) and Perceived Naturalness (PQ; $\alpha=0.608$), exhibited lower reliability, likely due to individual variability in motion sickness susceptibility and expectations regarding VR realism.

Despite these variations, the overall reliability of the usability scales remains robust, supporting the validity of the findings. Future studies could refine these subscales or integrate biometric indicators (eg, galvanic skin response and pupil dilation) to improve measurement consistency.

XSL•FO
RenderX

Table . Reliability analysis of usability and experience assessment scales[a].

Scale and subscale	Cronbach α values
System Usability Scale	
Overall Usability Score	0.810
Perceived Ease of Use	0.790
Ease of Learning	0.789
Presence Questionnaire	
Overall Presence Score	0.821
Involvement and Control	0.810
Perceived Naturalness of the Environment	0.608
Interface Responsiveness	0.394
Auditory Factors	0.638
Post-Study System Usability Questionnaire	
Overall User Satisfaction	0.980
System Usefulness	0.978
Information Clarity and Accessibility	0.922
Interface Efficiency	0.966
Simulator Sickness Questionnaire	
Overall Simulator Sickness Score	0.782
Nausea Symptoms	0.472
Oculomotor Discomfort	0.800
Disorientation Symptoms	0.342

[a]Cronbach α values above 0.8 indicate good reliability, while values below 0.7 suggest potential inconsistencies within subscales.

Usability and Presence Outcomes

System Usability Scale

Participants rated the VR simulation as highly usable (mean 75.00, SD 6.41; 95% CI 73.42 - 76.58; *P*=.002). This score exceeds the conventional usability benchmark of 68, indicating that users found the system intuitive, efficient, and well suited for medical training. SUS subscales (Perceived Ease of Use and Ease of Learning) suggest that participants found the interface intuitive and required minimal effort to learn the system, a critical factor in medical training scenarios where cognitive load must remain manageable.

Post-Study System Usability Questionnaire

The overall PSSUQ score (mean 2.92, SD 1.83; 95% CI 2.55 - 3.29) indicated moderate-to-high user satisfaction. System Usefulness (mean 2.89, SD 2.05; 95% CI 2.41 - 3.37) and Information Clarity (mean 2.99, SD 1.70; 95% CI 2.56 - 3.42) suggest that the system was perceived as informative and functional, supporting learning engagement.

Presence Questionnaire

Participants reported a high sense of immersion (mean 109.46, SD 9.12). Involvement and Control scored high (mean 55.24, SD 7.46), highlighting realistic and interactive elements. However, Interface Responsiveness (mean 21.40, SD 5.16) was rated lower, indicating areas for improvement in system feedback and response times.

Simulator Sickness Questionnaire

Symptoms of simulator sickness were mild (mean 12.43, SD 15.41; 95% CI: 9.28 - 15.58), primarily affecting Oculomotor Discomfort (mean 2.41, SD 3.12; 95% CI 1.84 - 2.98), and Nausea Symptoms (mean 1.24, SD 1.66; 95% CI 0.94 - 1.54). Participants with prior VR experience reported significantly lower nausea symptoms (*P*=.02; 95% CI 2.14 - 5.32), reinforcing the importance of adaptation to immersive environments. The short VR session duration (10 minutes) likely contributed to the low incidence of sickness, aligning with best practices in VR exposure time management.

VR Experience and Stress Measurement

Correlational analysis examined physiological stress (HRV) and simulator sickness symptoms (Figure 5):

- rMSSD (mean 281.27 ms, SD 98.99; 95% CI 259.45 - 303.09) was negatively correlated with nausea symptoms (*P*=.04), suggesting that greater autonomic relaxation was linked to lower simulator sickness.
- PNN50 (mean 56.85%, SD 19.70; 95% CI 52.23 - 61.47) also showed a negative correlation with nausea and oculomotor discomfort (*P*=.04; Spearman ρ=−0.23).

XSL•FO
RenderX

- LF/HF ratio (mean 0.88, SD 0.47) indicated a balanced autonomic state, but no significant correlations were found between LF/HF and usability or presence scores.
- SD1/SD2 and Poincaré area did not correlate significantly with usability or immersion, suggesting that HRV stress markers were more related to physical discomfort than usability perceptions.

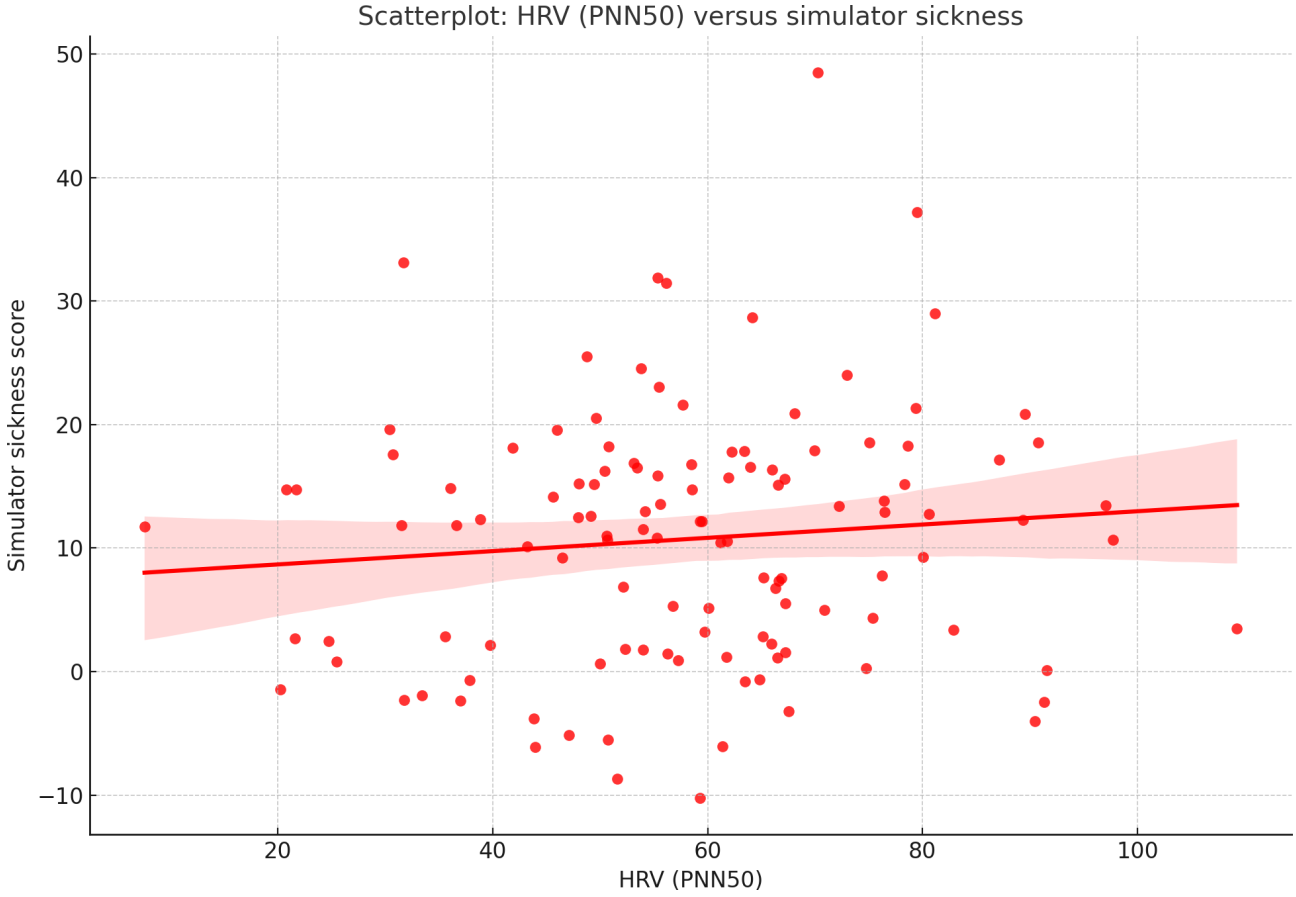
HRV analysis revealed that higher parasympathetic activity was associated with lower simulator sickness symptoms. However, HRV metrics did not significantly correlate with usability or presence scores, suggesting that physiological stress responses were more related to physical discomfort than to system usability. No significant correlations were found between HRV (rMSSD) and usability scores, as shown in Figure 6.

Figure 5. Scatterplot: HRV (PNN50) versus Simulator Sickness. This scatterplot illustrates the correlation between HRV (PNN50) and Simulator Sickness scores, showing a negative trend where higher parasympathetic activity is associated with lower simulator sickness symptoms. HRV: heart rate variability; rMSSD: root-mean-square of successive differences; SUS: System Usability Scale.



XSL•FO
RenderX

Figure 6. Scatterplot: HRV (root-mean-square of successive differences [rMSSD]) versus usability (System Usability Scale score). This scatterplot shows the lack of significant correlation between HRV (rMSSD) and usability scores, supporting the finding that physiological stress does not strongly influence usability perceptions. HRV: heart rate variability; PNN50: proportion of successive heartbeat intervals differing by more than 50 ms.



Influence of Prior VR Experience on Results

Participants with prior VR experience reported lower simulator sickness and higher usability perceptions (Table 2):

- Simulator sickness: novice users reported significantly higher nausea symptoms (*U*=997.0; *P*=.02) and oculomotor discomfort (*U*=1047.0; *P*=.045) than experienced users.
- SUS score: VR-experienced participants rated usability significantly higher (+5.2 points; *P*=.03; 95% CI 3.12 - 7.28), indicating greater ease of use and reduced cognitive load.
- PQ score: no significant differences in overall presence scores were found, but the Naturalness subscale showed borderline significance (*U*=1088.0; *P*=.09), with experienced users reporting slightly higher realism perception.
- HRV metrics: no significant differences in HRV stress markers between VR-experienced and VR-naive participants, suggesting that prior exposure to VR does not substantially alter physiological stress responses during high-risk simulations.

Prior VR experience was linked to lower levels of simulator sickness but did not significantly affect HRV stress indicators, suggesting that the system was user-friendly across different experience levels.

XSL•FO
RenderX

Table . Mann-Whitney *U* test comparing participants with and with no prior virtual reality experience[a].

Outcome measure	No VR[b] experience (n=88), mean rank	VR experience (n=31), mean rank	*U*	*Z*	*P* value	Significance
SSQ[c]: Nausea	64.17	48.16	997.0	–2.364	.02	Significant
SSQ: Oculomotor	63.60	49.77	1047.0	–2.003	.045	Significant
SSQ: Disorientation	61.94	54.50	1193.5	–1.248	.21	Not significant
SUS[d]: Usability	58.88	63.18	1072.0	–1.770	.08	Marginal
SUS: Learnability	58.03	65.58	1191.0	–1.053	.29	Not significant
PQ[e]: Naturalness	56.86	68.90	1088.0	–1.686	.09	Marginal
HRV[f]: rMSSD[g]	61.14	56.76	1263.5	–0.608	.54	Not significant

[a]Statistical comparisons were conducted using the Mann-Whitney *U* test. "Marginal" refers to trends toward significance (.05≤*P*<.10). Higher SSQ scores indicate greater simulator sickness and higher SUS scores reflect better usability perceptions.

[b]VR: virtual reality.

[c]SSQ: Simulator Sickness Questionnaire.

[d]SUS: System Usability Scale.

[e]PQ: Presence Questionnaire.

[f]HRV: heart rate variability.

[g]rMSSD: root-mean-square of successive difference.

Discussion

Principal Findings

The usability assessments revealed that participants rated the VR system as highly usable, with a mean SUS score of 75 (SD 6.41, 95% CI 73.42-76.58; *P*=.002), exceeding the commonly used benchmark of 68. This indicates that the system was well designed and accessible, allowing users to interact with it effectively. Both novice and experienced users provided comparable usability ratings, suggesting that the reliance on direct movement and the absence of nondiegetic interactions eliminated the need for prior VR navigation experience. These findings align with previous research emphasizing the importance of designing VR systems that are intuitive and inclusive, ensuring accessibility across users with varying levels of expertise [15,16]. Moreover, usability-focused approaches tailored to novice users have been recommended, highlighting the importance of gradual adaptation strategies to facilitate immersion and reduce initial cognitive load [17].

Although usability scores were high, simulator sickness remained a challenge, particularly for novice users, who reported significantly greater nausea and oculomotor discomfort than experienced users. These results are consistent with prior studies showing that previous VR exposure reduces discomfort, likely due to increased adaptation to immersive environments [18]. Adaptation mechanisms such as shorter initial exposure sessions or gradual immersion through simplified VR experiences could facilitate acclimatization for first-time users [19]. The 10-minute session duration may have contributed to minimizing simulator sickness symptoms, as shorter exposure times are recommended to optimize user comfort.

Although simulator sickness was more pronounced among novices, this study did not track symptom progression over time, limiting the ability to assess whether participants acclimated during the session. Research suggests that VR-induced discomfort often declines with repeated exposure as users adapt to the sensorimotor feedback of the environment [20]. Motion sickness symptoms typically peak within the first few minutes of VR exposure and gradually subside, particularly with repeated use. However, because symptoms were measured only at the end of the session, it remains unclear whether adaptation occurred during the simulation itself.

The high SUS score of 75 suggests that participants found the VR system not only intuitive but also effective for training. Previous research evaluating medical VR applications has reported lower initial usability ratings, which tend to improve with repeated exposure. In contrast, the high usability scores observed in this study suggest that the system was well optimized for first-time users, reducing potential barriers to interaction. The Perceived Ease of Use subscale further supports this, indicating that while users found the system easy to navigate, some effort was required to learn the procedural steps. The PSSUQ results, reflecting moderate user satisfaction, are consistent with prior studies evaluating VR-based procedural training, reinforcing the system's effectiveness for medical education.

The reliability analysis of usability and experience assessment scales confirmed that SUS and PSSUQ demonstrated strong internal consistency. However, some subscales, such as the Nausea Symptoms subscale of the SSQ and the Perceived Naturalness subscale of the PQ, had lower-than-expected reliability. Several factors may account for this variability. Simulator sickness symptoms can vary widely across individuals, influenced by differences in motion sensitivity, vestibular adaptation, and prior VR exposure. Similarly, expectations regarding VR realism may have affected the Perceived Naturalness subscale, particularly among experienced

XSL•FO
RenderX

users, who may have had higher expectations for environmental fidelity. While these lower reliability values indicate variability in participant responses, they do not necessarily invalidate the subscales.

The PQ scores indicated a strong sense of immersion among participants, with no significant differences between novice and experienced users, except for a borderline significant difference in the Perceived Naturalness subscale. This suggests that while experienced users may perceive VR environments as more natural, both groups experienced a high level of presence [21]. Maintaining a strong sense of presence is essential in medical simulations, as immersive learning enhances procedural training and stress management skills. Prior studies suggest that enhancing interactivity, improving haptic feedback, and refining visual realism could further increase presence scores in VR-based medical training.

The analysis of HRV data revealed no significant correlations between physiological stress and usability or presence scores. This suggests that participants' perceptions of usability were primarily cognitive, influenced by their interaction with the interface rather than underlying physiological stress responses. These findings are consistent with previous literature indicating that usability is predominantly shaped by system design and functionality rather than physiological arousal [22]. However, the absence of correlation between HRV and usability or presence does not preclude interactions between cognitive and physiological factors. Several potential mediators or confounding variables could explain this finding. Individual stress tolerance, task difficulty, cognitive load, and personal coping strategies may have influenced HRV responses in ways not fully captured by usability assessments [23]. For instance, some participants may have experienced heightened physiological stress without perceiving usability issues, while others may have found the interface challenging but did not exhibit strong physiological responses. Prior research indicates that task complexity and prior clinical experience can modulate physiological stress responses in simulation-based training [24].

HRV metrics are also influenced by various physiological and psychological factors beyond stress, including baseline autonomic function, fatigue, and attentional demands. Given that VR simulations inherently require motor adaptation and spatial awareness adjustments, some users may have exhibited physiological responses unrelated to usability perceptions. These findings highlight the importance of multidimensional stress assessment approaches that incorporate both HRV and subjective stress measures (eg, NASA-TLX, perceived workload scales) to better capture cognitive and emotional load during VR training.

Interestingly, HRV parameters did not significantly correlate with presence, suggesting that stress responses do not strongly influence immersion in virtual environments. Instead, presence appears to be more dependent on design quality, realism, and interactivity than physiological arousal [25]. These findings align with prior research indicating that immersion in VR is primarily driven by environmental realism and user engagement rather than stress levels [26]. However, subtle interactions between physiological responses and presence may still exist, potentially influenced by individual factors such as trait anxiety or cognitive flexibility. Future research should explore these relationships further.

Recent studies have demonstrated that VR-based stress management interventions, including adaptive exposure and biofeedback integration, can enhance stress resilience and user engagement [27]. The use of real-time physiological feedback, such as HRV-guided VR adjustments, has been proposed as a strategy to optimize stress regulation in immersive environments. Given that stress responses in medical training simulations are complex and influenced by multiple factors, incorporating similar adaptive VR mechanisms in future iterations of medical VR training tools could improve user experience, reduce discomfort, and enhance long-term skill retention [4].

The results demonstrated significant differences in simulator sickness symptoms between participants with and with no prior VR experience. Novice users reported significantly higher levels of nausea and oculomotor discomfort, reinforcing the idea that familiarity with immersive environments helps mitigate the physical discomfort associated with VR [28]. These findings align with prior studies suggesting that repeated VR exposure can lead to adaptation, gradually reducing the severity of simulator sickness symptoms over time [29]. This adaptation effect highlights the need for tailored protocols for novice users, particularly in medical and emergency training contexts, to minimize discomfort and optimize the learning experience.

Despite differences in simulator sickness, there were no significant differences between novice and experienced users in PSSUQ or PQ scores. This suggests that the system was designed to be accessible and intuitive for users regardless of prior VR experience, reinforcing its inclusiveness. The lack of differences in usability scores further supports the idea that first-time users can effectively engage with VR simulations without a steep learning curve [30].

The borderline significance observed in the Perceived Naturalness subscale of the PQ suggests that experienced users may be more attuned to subtle environmental details, contributing to a heightened perception of realism. For novice users, improving environmental naturalness could enhance immersion. While the current simulation featured low-polygon, cartoon-like graphics, incorporating more photorealistic elements may improve engagement, provided that the "uncanny valley" effect is avoided.

Limitations and Future Directions

This study has several limitations. First, the sample consisted solely of final-year medical students from a single institution, limiting the generalizability of findings to practice clinicians or professionals with different levels of experience. Since clinical expertise influences usability perceptions and physiological stress responses in VR environments, future studies should include more diverse populations across various career stages and medical specialties. Second, usability (SUS and PSSUQ) and simulator sickness (SSQ) were assessed using self-reported measures, which may introduce response biases. External factors such as fatigue or prior VR exposure could

XSL•FO
RenderX

have influenced participants' perceptions. Future research should complement self-reported data with objective behavioral metrics, such as task completion times or interaction errors, to enhance measurement validity. Third, while HRV provided an objective measure of autonomic stress responses, it did not capture cognitive load or psychological stress. The lack of subjective stress assessments (eg, NASA-TLX) and neurocognitive performance measures (eg, reaction times) limits the ability to fully understand the interaction between stress and usability in VR-based procedural training. Fourth, the study's short VR session (10 minutes) restricted the assessment of long-term adaptation to simulator sickness and stress. Research suggests that symptoms and stress responses diminish with repeated exposure, but this study only measured outcomes postsession. Future studies should track symptom progression at multiple time points to assess adaptation over time. Fifth, the absence of real-world consequences for procedural errors may have reduced perceived urgency and stress. In clinical settings, mistakes during pericardiocentesis can have severe outcomes, whereas errors in VR had no direct impact. Introducing performance-based feedback or simulated consequences could enhance realism and improve training effectiveness. In addition, the low-polygon VR environment, while minimizing cognitive load, may have affected immersion, particularly for users with prior VR experience. Higher-fidelity graphics could enhance realism, although careful design is needed to avoid cognitive overload or the "uncanny valley" effect. Furthermore, a limitation of this study is the binary classification of prior VR experience (yes/no), which does not capture potential differences in familiarity with VR applications, such as gaming, medical simulations, or other interactive experiences. A more granular classification—differentiating casual VR users, gaming-oriented users, and those with professional training in VR environments—could provide deeper insights into how prior exposure influences usability and simulator sickness outcomes.

To address these limitations, future research should include broader participant samples, incorporate both objective and subjective stress measures, evaluate longer VR sessions, and refine simulation realism. Exploring different VR platforms and interaction methods could also inform the best practices for improving usability and immersion in medical training simulations.

Practical Implications

The findings of this study provide practical insights into optimizing VR-based medical training. Given the high usability of the system (SUS=75), future improvements should prioritize reducing simulator sickness rather than addressing usability concerns. Implementing adaptive exposure strategies—such as shorter initial sessions and gradual complexity increases—could help novice users acclimate, minimizing nausea and oculomotor discomfort. The negative correlation between HRV parameters and simulator sickness symptoms suggests that HRV could be integrated as a real-time monitoring tool to identify users experiencing discomfort. This would allow proactive adjustments, such as pausing sessions or modifying simulation parameters, to enhance user comfort. To maintain usability and realism, VR training should continue using diegetic interactions that mirror real-world clinical procedures, minimizing artificial UI elements that could disrupt immersion. In addition, incorporating performance-based feedback—such as error tracking or consequence-driven scenarios—could reinforce learning and better simulate real-world clinical pressures. By refining exposure strategies, integrating real-time stress monitoring, and enhancing realism, VR-based simulations can provide a more immersive, effective, and user-friendly training experience for medical learners.

Conclusions

This study suggests that VR-based simulations for high-stress medical procedures, such as pericardiocentesis, offer high usability (SUS=75) and strong immersion, making them effective training tools for users with varying VR experience levels. Physiological stress responses (HRV) correlated more with simulator sickness than with usability or presence, indicating that physical discomfort plays a key role in stress modulation within VR environments.

Prior VR experience was associated with lower simulator sickness symptoms, reinforcing the potential benefits of gradual exposure strategies for first-time users. These findings highlight the complex interaction between usability, stress responses, and prior experience in VR-based medical training.

While these results contribute to understanding VR usability and stress responses, further research is needed to validate these findings in larger and more diverse samples. Future studies should also assess longitudinal effects to determine whether adaptation to simulator sickness occurs over repeated exposures.

Acknowledgments

The authors wish to express their sincere gratitude to all the students who participated in this study. Their involvement and dedication were essential to this research, and their contributions have greatly enhanced the authors' understanding of virtual reality applications in medical training. This work was supported by the Spanish Ministry of Science, Innovation and Universities, under the grant European Regional Development Fund (ERDF) with project reference IDI-20191120.

Data Availability

To support reproducibility, transparency, and further research, we have made the following resources publicly available:

• Biometric data: the anonymized biometric data collected during the study, including heart rate variability (HRV) parameters are available at figshare [31].

XSL•FO
RenderX

• 3D pericardium model: the 3D model of the pericardium designed for this study can be downloaded from figshare [32]. The model was created using Tinkercad and Adobe Fusion 360 (version 15.3.0.1657) and is freely available for educational and research purposes.

• VR simulator repository: the source code of the virtual reality pericardiocentesis simulator, developed in Unity 2018.4.36f1, is available under an open-source license on GitHub [33]. The compiled application can also be downloaded freely.

• Virtual simulation video: a demonstration video showcasing the virtual reality pericardiocentesis simulation is hosted on YouTube [34].

• Study questionnaires: the System Usability Scale (SUS), Post-Study System Usability Questionnaire (PSSUQ), Presence Questionnaire (PQ), and Simulator Sickness Questionnaire (SSQ) used in this study are available as Multimedia Appendices 1-4.

• Supplementary data: a summary table of aggregated HRV metrics and usability scores, including subgroup comparisons, is provided as supplementary material in Multimedia Appendix 5.

These resources are provided to promote reproducibility, facilitate further advancements in medical education, and encourage collaboration among researchers and educators.

Authors' Contributions

AR-L designed the study, collected the data, performed the statistical analysis, and wrote the study as the principal investigator. RGC supervised the development and validation of the virtual reality scenario. LZR developed the virtual scenario. ARN participated in the performance tests and in the design and collection of biometric parameters. AG-P participated in the design and adaptation of the simulated scenarios. PC-F participated in proofreading and supervision of the study design, statistical analysis, and writing of the paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Presence Questionnaire.
[PDF File, 1010 KB - games_v13i1e68515_app1.pdf]

Multimedia Appendix 2
Post-Study System Usability Questionnaire.
[PDF File, 164 KB - games_v13i1e68515_app2.pdf]

Multimedia Appendix 3
System Usability Scale questionnaire used in the study.
[PDF File, 105 KB - games_v13i1e68515_app3.pdf]

Multimedia Appendix 4
Simulator Sickness Questionnaire.
[PDF File, 100 KB - games_v13i1e68515_app4.pdf]

Multimedia Appendix 5
Summary table of aggregated heart rate variability metrics and usability scores, including subgroup comparisons.
[PDF File, 333 KB - games_v13i1e68515_app5.pdf]

References

1. Mergen M, Graf N, Meyerheim M. Reviewing the current state of virtual reality integration in medical education—a scoping review. BMC Med Educ 2024 Jul 23;24(1):788. [doi: 10.1186/s12909-024-05777-5] [Medline: 39044186]
2. Wybraniec MT, Kampka Z, Drabczyk M, et al. Clinical characteristics and risk factors of in-hospital mortality among patients undergoing percutaneous pericardiocentesis. Front Cardiovasc Med 2023;10:1252525. [doi: 10.3389/fcvm.2023.1252525] [Medline: 37781300]
3. Marín-Morales J, Higuera-Trujillo JL, Guixeres J, Llinares C, Alcañiz M, Valenza G. Heart rate variability analysis for the assessment of immersive emotional arousal using virtual reality: comparing real and virtual scenarios. PLoS ONE 2021;16(7):e0254098. [doi: 10.1371/journal.pone.0254098] [Medline: 34197553]

XSL•FO
RenderX

4. Weibel RP, Kerr JI, Naegelin M, et al. Virtual reality-supported biofeedback for stress management: beneficial effects on heart rate variability and user experience. Comput Human Behav 2023 Apr;141:107607. [doi: 10.1016/j.chb.2022.107607]
5. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the System Usability Scale. Int J Hum Comput Interact 2008 Jul 29;24(6):574-594. [doi: 10.1080/10447310802205776]
6. Lewis JR. Psychometric evaluation of the Post-Study System Usability Questionnaire: the PSSUQ. Proc Hum Factors Ergon Soc Annu Meet 1992 Oct;36(16):1259-1260. [doi: 10.1177/154193129203601617]
7. Pears M, Antoniou PE, Schiza E, et al. Virtual reality reusable e-resources for clinical skills training: a mixed-methods evaluation. Pers Ubiquit Comput 2024 Aug;28(3-4):643-654. [doi: 10.1007/s00779-024-01805-9]
8. Witmer BG, Singer MJ. Measuring presence in virtual environments: a Presence Questionnaire. Presence (Camb) 1998 Jun;7(3):225-240. [doi: 10.1162/105474698565686]
9. Johnston M, O'Mahony M, O'Brien N, et al. The feasibility and usability of mixed reality teaching in a hospital setting based on self-reported perceptions of medical students. BMC Med Educ 2024 Jun 27;24(1):701. [doi: 10.1186/s12909-024-05591-z] [Medline: 38937764]
10. Zante B, Schefold JC. Simulation training for emergency skills: effects on ICU fellows' performance and supervision levels. BMC Med Educ 2020 Dec 9;20(1):498. [doi: 10.1186/s12909-020-02419-4] [Medline: 33298042]
11. Luo J, Zhang G, Su Y, et al. Quantitative analysis of heart rate variability parameter and mental stress index. Front Cardiovasc Med 2022;9:930745. [doi: 10.3389/fcvm.2022.930745] [Medline: 35958396]
12. Ewing DJ, Neilson JMM, Travis P. New method for assessing cardiac parasympathetic activity using 24 hour electrocardiograms. Br Heart J 1984 Oct;52(4):396-402. [doi: 10.1136/hrt.52.4.396] [Medline: 6383446]
13. Bennett MM, Tomas CW, Fitzgerald JM. Relationship between heart rate variability and differential patterns of cortisol response to acute stressors in mid-life adults: a data-driven investigation. Stress Health 2024 Jun;40(3):e3327. [doi: 10.1002/smi.3327] [Medline: 37786944]
14. Lee BM, Kang CY, Li L, Rami FZ, Chung YC. Physiological responses to the virtual reality-based trier social stress test in patients with psychosis. Virtual Real 2023 Dec;27(4):3115-3123. [doi: 10.1007/s10055-023-00857-1]
15. Kennedy RS, Stanney KM, Dunlap WP. Duration and exposure to virtual environments: sickness curves during and across sessions. Presence Teleoperators Virtual Environ 2000 Oct;9(5):463-472. [doi: 10.1162/105474600566952]
16. Özgen DS, Afacan Y, Sürer E. Usability of virtual reality for basic design education: a comparative study with paper-based design. Int J Technol Des Educ 2021 Apr;31(2):357-377. [doi: 10.1007/s10798-019-09554-0]
17. Lacoche J, Villain E, Foulonneau A. Evaluating usability and user experience of AR applications in VR simulation. Front Virtual Real 2024;3. [doi: 10.3389/frvir.2022.881318]
18. Fang YM. Exploring usability, emotional responses, flow experience, and technology acceptance in VR: a comparative analysis of freeform creativity and goal-directed training. Appl Sci (Basel) 2024;14(15):6737. [doi: 10.3390/app14156737]
19. Voinescu A, Petrini K, Stanton Fraser D. Presence and simulator sickness predict the usability of a virtual reality attention task. Virtual Real 2023 Mar 24;27:1-17. [doi: 10.1007/s10055-023-00782-3] [Medline: 37360806]
20. Kim YS, Won JH, Jang SW, Ko J. Effects of cybersickness caused by head-mounted display-based virtual reality on physiological responses: cross-sectional study. JMIR Serious Games 2022 Oct 17;10(4):e37938. [doi: 10.2196/37938] [Medline: 36251360]
21. De Paolis LT, De Luca V. The effects of touchless interaction on usability and sense of presence in a virtual environment. Virtual Real 2022 Dec;26(4):1551-1571. [doi: 10.1007/s10055-022-00647-1]
22. Kim YM, Rhiu I. Development of a Virtual Reality System Usability Questionnaire (VRSUQ). Appl Ergon 2024 Sep;119:104319. [doi: 10.1016/j.apergo.2024.104319] [Medline: 38797014]
23. Chun JY, Kim HJ, Hur JW, et al. Prediction of specific anxiety symptoms and virtual reality sickness using in situ autonomic physiological signals during virtual reality treatment in patients with social anxiety disorder: mixed methods study. JMIR Serious Games 2022 Sep 16;10(3):e38284. [doi: 10.2196/38284] [Medline: 36112407]
24. Hsin LJ, Chao YP, Chuang HH, et al. Mild simulator sickness can alter heart rate variability, mental workload, and learning outcomes in a 360° virtual reality application for medical education: a post hoc analysis of a randomized controlled trial. Virtual Real 2022 Sep 14;27:1-17. [doi: 10.1007/s10055-022-00688-6] [Medline: 36118174]
25. Sterna R, Cybulski A, Igras-Cybulska M, Pilarczyk J, Segiet N, Kuniecki M. How behavioral, photographic, and interactional realism influence the sense of co-presence in VR. An investigation with psychophysiological measurement. Int J Hum Comput Interact 2024 Dec 16;40(24):8574-8589. [doi: 10.1080/10447318.2023.2285641]
26. Standen B, Anderson J, Sumich A, Heym N. Effects of system- and media-driven immersive capabilities on presence and affective experience. Virtual Real 2023 Mar;27(1):371-384. [doi: 10.1007/s10055-021-00579-2]
27. Riches S, Taylor L, Jeyarajaguru P, Veling W, Valmaggia L. Virtual reality and immersive technologies to promote workplace wellbeing: a systematic review. J Ment Health 2024 Apr;33(2):253-273. [doi: 10.1080/09638237.2023.2182428] [Medline: 36919828]
28. Ciążyńska J, Janowski M, Maciaszek J. Effects of a modern virtual reality 3D head-mounted display exergame on simulator sickness and immersion under specific conditions in young women and men: experimental study. JMIR Serious Games 2022 Nov 29;10(4):e41234. [doi: 10.2196/41234] [Medline: 36445744]

XSL•FO
RenderX

29. Khaleghi A, Narimani A, Aghaei Z, Khorrami Banaraki A, Hassani-Abharian P. A smartphone-gamified virtual reality exposure therapy augmented with biofeedback for ailurophobia: development and evaluation study. JMIR Serious Games 2024 Mar 6;12:e34535. [doi: 10.2196/34535] [Medline: 38446522]
30. Lim CH, Lee SC. The effects of degrees of freedom and field of view on motion sickness in a virtual reality context. Int J Hum Comput Interact 2024 Oct;40(19):5884-5896. [doi: 10.1080/10447318.2023.2241620]
31. BIOMETRIC-DATASET. Figshare. URL: https://figshare.com/articles/dataset/BIOMETRIC-DATASET/26215184/1 [accessed 2025-04-23]
32. Development of a 3D printed pericardiocentesis mannequin. Figshare. URL: https://figshare.com/articles/online_resource/_b_Development_of_a_3D_Printed_Pericardiocentesis_Mannequin_b_/26219135 [accessed 2025-04-23]
33. Rgarciacarmona/VR-pericardiocentesis. GitHub. URL: https://github.com/rgarciacarmona/VR-Pericardiocentesis [accessed 2025-04-23]
34. Pericardiocentesis-VR. YouTube. URL: https://www.youtube.com/watch?v=75eq2mssiRc [accessed 2025-04-23]

Abbreviations

ANS: autonomic nervous system
HF: high frequency
HRV: heart rate variability
LF: low frequency
PNN20: proportion of successive heartbeat intervals differing by more than 20 ms
PNN50: proportion of successive heartbeat intervals differing by more than 50 ms
PQ: Presence Questionnaire
PSSUQ: Post-Study System Usability Questionnaire
rMSSD: root-mean-square of successive difference
SD1: standard deviation 1
SD2: standard deviation 2
SSQ: Simulator Sickness Questionnaire
SUS: System Usability Scale
VR: virtual reality

Edited by A Coristine; submitted 07.11.24; peer-reviewed by M Chanda, M Agbede, S Mitra; revised version received 19.03.25; accepted 19.03.25; published 07.05.25.

Please cite as:
Rubio-López A, García Carmona R, Zarandieta Román L, Rubio Navas A, González-Pinto A, Cardinal-Fernández P
Measuring Stress and Perceptions for a Virtual Reality–Based Pericardiocentesis Procedure Simulation for Medical Training: Usability Study
JMIR Serious Games 2025;13:e68515
URL: https://games.jmir.org/2025/1/e68515
doi:10.2196/68515

© Alberto Rubio-López, Rodrigo García Carmona, Laura Zarandieta Román, Alejandro Rubio Navas, Angel González-Pinto, Pablo Cardinal-Fernández. Originally published in JMIR Serious Games (https://games.jmir.org), 7.5.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Applying Augmented Reality to Convey Medical Knowledge on Osteoclasts to Users of a Serious Game: Vignette Experiment

Jascha Grübel[1,2,3,4,5], Dr sc ETH; Julia Chatain[3,6], Dr sc ETH; Claudio Schmid[7], Dr med dent; Violaine Fayolle[3,8], MA; Fabio Zünd[3], Dr sc ETH; Reinhard Gruber[9], Dr nat techn; Bernd Stadlinger[7,10], Dr med, Dr med dent

[1]Laboratory for Geo-Information Science and Remote Sensing, Department of Environmental Sciences, Wageningen University, Droevendaalsesteeg 4, Wageningen, The Netherlands

[10]ETH AI Center, Zürich, Switzerland

[2]Center for Sustainable Future Mobility, Department of Mechanical and Process Engineering & Department of Civil, Environmental and Geomatic Engineering, ETH Zurich, Zürich, Switzerland

[3]Game Technology Center, Department of Computer Science, ETH Zurich, Zürich, Switzerland

[4]Chair of Cognitive Science, Department of Humanities, Social and Political Sciences, ETH Zurich, Zürich, Switzerland

[5]Visual Computing Group, Harvard John A. Paulson School Of Engineering And Applied Sciences, Harvard University, Boston, MA, United States

[6]Future Embodied Learning Technologies, Singapore-ETH Centre, Singapore, Singapore

[7]Center for Dental Medicine, Clinic of Cranio-Maxillofacial and Oral Surgery, University of Zurich, Zürich, Switzerland

[8]Disney Research, Zürich, Switzerland

[9]Competence Center of Oral Biology, University Clinic of Dentistry, Medical University of Vienna, Vienna, Austria

Corresponding Author:
Jascha Grübel, Dr sc ETH
Laboratory for Geo-Information Science and Remote Sensing, Department of Environmental Sciences, Wageningen University, Droevendaalsesteeg 4, Wageningen, The Netherlands

Abstract

Background: Visualization technology is enhancing interactive learning by merging digital content with real-world environments, offering immersive experiences through augmented reality (AR) in fields like medical education. AR is being increasingly used in medicine and dental education to improve student learning, particularly in understanding complex concepts such as bone remodeling. Active learning strategies, supported by AR, boost student autonomy, reduce cognitive load, and improve learning outcomes across various disciplines. AR is gaining popularity in higher education as it enhances active learning, reduces cognitive load, and improves cognitive, meta-cognitive, and affective outcomes, particularly in medical and nursing education. The effectiveness of immersive AR in enhancing understanding of complex physiological processes is still unclear, with a lack of rigorous studies on its impact and how to effectively convert academic content into AR.

Objective: We assess the capacity of AR-enhanced content for learning medical knowledge with a state-of-the-art AR game published along with a modern cell atlas of the oral cavity. To assess AR-enhanced content for learning, we formulated hypotheses on the general impact on learning (H1), specific improvements in learning (H2), and long-term retention (H3).

Methods: An AR serious game was developed to represent current knowledge on osteoclasts for classroom use. The game was evaluated in an unblinded face-to-face vignette experiment (39 participants). Learning outcomes on "Osteoclasts" were compared between the AR game (17 participants) and a textbook-only option (20 participants) conveying the same content. Participants were randomly assigned and learning success was measured at three time-points, immediately after the experiment session, 1 week later, and 1 month later, via web-based surveys.

Results: The AR serious game elicited strong interest in the topic (perceived relevance in Attention, Relevance, Confidence, and Satisfaction [ARCS], W=10,417; P<.001) and motivated students by increasing self-efficacy (confidence in ARCS, W=11,882.5; P=.02) and satisfaction (in ARCS, W=4561; P<.001). The learning outcomes were comparable to text-based self-learning (t=2.0103; $P_{\text{Bonferroni}}$=.095). Furthermore, curious students benefited more from interactive learning methods compared with text-only methods and had higher learning success (t=−2.518; P=.02).

Conclusions: Introducing new technology such as AR into teaching requires technological investment, updated curricula, and careful application of learning paradigms. We found support for improved motivation (H1) and some evidence of AR's baseline effectiveness (H2a). While we could not confirm AR's impact on visual tasks overall (H2b), we noted an interesting interaction between curiosity and visual task outcomes (H2c), as well as how game design influences student perception of the material (H2d). Due to attrition, long-term learning outcomes (H3) could not be assessed. AR-based learning may particularly benefit

XSL•FO
RenderX

curious students, who often struggle with text-heavy methods. As students are increasingly accustomed to brief, engaging content, teaching approaches must adapt.

(JMIR Serious Games 2025;13:e64751) doi:10.2196/64751

KEYWORDS

tablet-based augmented reality; osteoclasts; educational game; user study; vignette experiment; augmented reality; communication; medical student; oral cavity; serious game

Introduction

Visualization technology is transforming interactive learning by merging digital content with the real world to enhance educational experiences [1]. Augmented reality (AR) superimposes computer-generated content over real-world environments, enriching sensory perception with 3D images, videos, and audio [2], enabling interaction in a mixed-reality space. AR integrates real environmental stimuli with digital representations [3,4], offering immersive media for fields like medical education (Figure 1). Examples of AR in teaching include augmenting books with digital content [5] and the augmentation of physical objects [6] with contextualized information. Since the 1990s, AR has evolved with advancements in hardware, software, and head-mounted displays (eg, Microsoft HoloLens), expanding its potential as a teaching tool [7]. The improvement of smartphones and tablets has further broadened AR's influence [8], which is reflected in everyday applications like Snapchat filters, Pokémon Go, and IKEA Place [9].

Figure 1. The augmented reality osteoclast serious game being played in a lecture hall.



AR is increasingly used in medicine and dental medicine to enhance student learning, particularly in understanding complex concepts. AR is already used for various applications, from patient communication to student learning and to overcome dental phobia [10,11]. This paper focuses on the potential of AR-enhanced learning material to help students [12] understand the complex concepts of bone remodeling. Knowledge of physiology is essential for teaching health sciences, including medicine and dentistry. Bone physiology is of key importance in surgical fields like orthopedics and oral, and maxillofacial surgery [13]. Many treatments, such as dental implant placement, depend on successful bone remodeling [14]. Currently, bone biology and remodeling are taught in medical school mainly through histology, where tissues are examined and discussed at the microscopic level [15]. Histology courses typically include lectures and microscopy labs, but students often struggle with understanding the spatial relationships of 2D tissue sections [16]. AR-enhanced learning materials could help students to better visualize and understand the spatial features of bone biology.

Active learning strategies, supported by AR technology, enhance student autonomy, reduce cognitive load, and improve learning outcomes across various disciplines. To understand AR-enhanced learning, we considered learning frameworks that emphasize student autonomy and agency [17], such as active learning strategies. Matching content with an appropriate medium is crucial for creating a coherent learning experience [1]. Over the past decades, active learning strategies, such as

XSL•FO
RenderX

flipped classrooms [18], project-based learning [19], and productive failure [20], have been shown to improve understanding, long-term learning, ability to transfer understanding to other situations and contexts [21-24].

The use of AR in higher education is gaining attention because it can enhance active learning, reduce cognitive load, and improve both cognitive and affective outcomes. AR in higher education is being increasingly used [25] because it supports active forms of learning [26] by colocating information and its referent, reducing split-attention effects [27]. AR also fosters collaboration and supports social incentives for learning [28]. One potential pathway in which AR context could improve active learning is by reducing cognitive load through visualization [29,30]. AR game-based learning positively impacts not only cognitive outcomes (eg, performance and learning), but also meta-cognitive outcomes (eg, attitude and participation), as well as affective outcomes (eg enjoyment and motivation) [31]. In medical and nursing education, AR has been shown to enhance knowledge and understanding [24,25], practical skills, and social skills [3]. This is important because, in the medical context, the busy schedules of teachers, practitioners, and students reduce available learning time, making additional improvements particularly beneficial. For instance, gamification in histology teaching can enhance student learning [32], and immersive games can support medical learning, especially in the context [33].

The effectiveness of immersive AR experiences in enhancing student understanding of complex physiological processes remains unclear, with a significant knowledge gap in the literature. The literature is inconclusive regarding what makes an AR-enhanced learning experience valuable and how to effectively convert academic content into AR [34-37]. This knowledge gap is compounded by a lack of rigorous studies, preventing meaningful meta-analysis [10]. Investigating when and how AR improves learning outcomes is essential. We investigate curiosity as a construct that could underpin a learning framework for AR-enhanced content [38]. The conversion of academic content into games or AR-enhanced content is little studied in the literature [39,40]. This conversion requires fundamental simplifications that may impact how students perceive and retain the content. It also determines what students take away from the AR-enhanced learning experience in the long term and requires more research to be evaluated [41].

This study explores the creation of AR-based content for immersive learning and its impact on student performance, focusing on enhancing a medical textbook through the AR Osteoclasts app. The aim of this project was to “augment” a medical cell atlas textbook with a strong focus on photorealistic imaging (Visual Biology in Oral Medicine, Quintessence Publishing [13]) by means of an AR-based App. The textbook itself was inspired by 6 computer-animated scientific videos on cell communication processes in a clinical context and state-of-the-art reviews [42-45]. We selected the process of bone resorption and new bone formation from the book chapter on osteoclasts as a basis for the app AR Osteoclasts was designed to provide an interactive visual experience of bone physiology, specifically bone resorption and formation, enabling users to engage with these processes in an AR environment (Figure 1 and Multimedia Appendix 1). To evaluate the app, we assessed the game on a technical level (Multimedia Appendix 1) and on learning outcomes with an experiment. To assess AR-enhanced content for learning, we formulated hypotheses on the general impact on learning (H1), specific improvements in learning (H2), and long-term retention (H3).

- H1: Learning motivation about osteoclasts would be improved with a serious AR game, compared with learning from text.
- H2a: Learning outcomes about osteoclasts are at least as effective with a serious AR game as learning from text.
- H2b: Learning outcomes about osteoclasts with a serious AR game improve on visual questions compared with learning from text.
- H2c: Learning outcomes about osteoclasts with a serious AR game are improved for curious learners compared with learning from text.
- H2d: Learning outcomes about osteoclasts with a serious AR game are improved for topics represented as visual components of the game.
- H3: Long-term learning outcomes about osteoclasts with a serious AR game are improved compared with learning from text.

Methods

Overview

To assess this AR serious gaming for learning in the context of Osteoclasts, we conducted an experiment comparing AR-based learning with traditional text-based learning. To test our hypotheses, we employed a difference-in-differences design for quantitative analysis of learning outcomes and utilized automated text analysis for qualitative assessment. The design of the AR game follows best practices [46,47] and focuses on converting scientific knowledge into gaming mechanisms involving stakeholders and experts [48]. The game design aims to enhance learning (Multimedia Appendix 1) through embodiment [49,50] in minigames [51]. Additionally, a usability study with current best practices [52-57] was conducted to assess the participants’ ability to effectively engage with the game (Multimedia Appendix 2). Last, we provide insights into participants’ experiences with a video of a single playthrough (Multimedia Appendix 3).

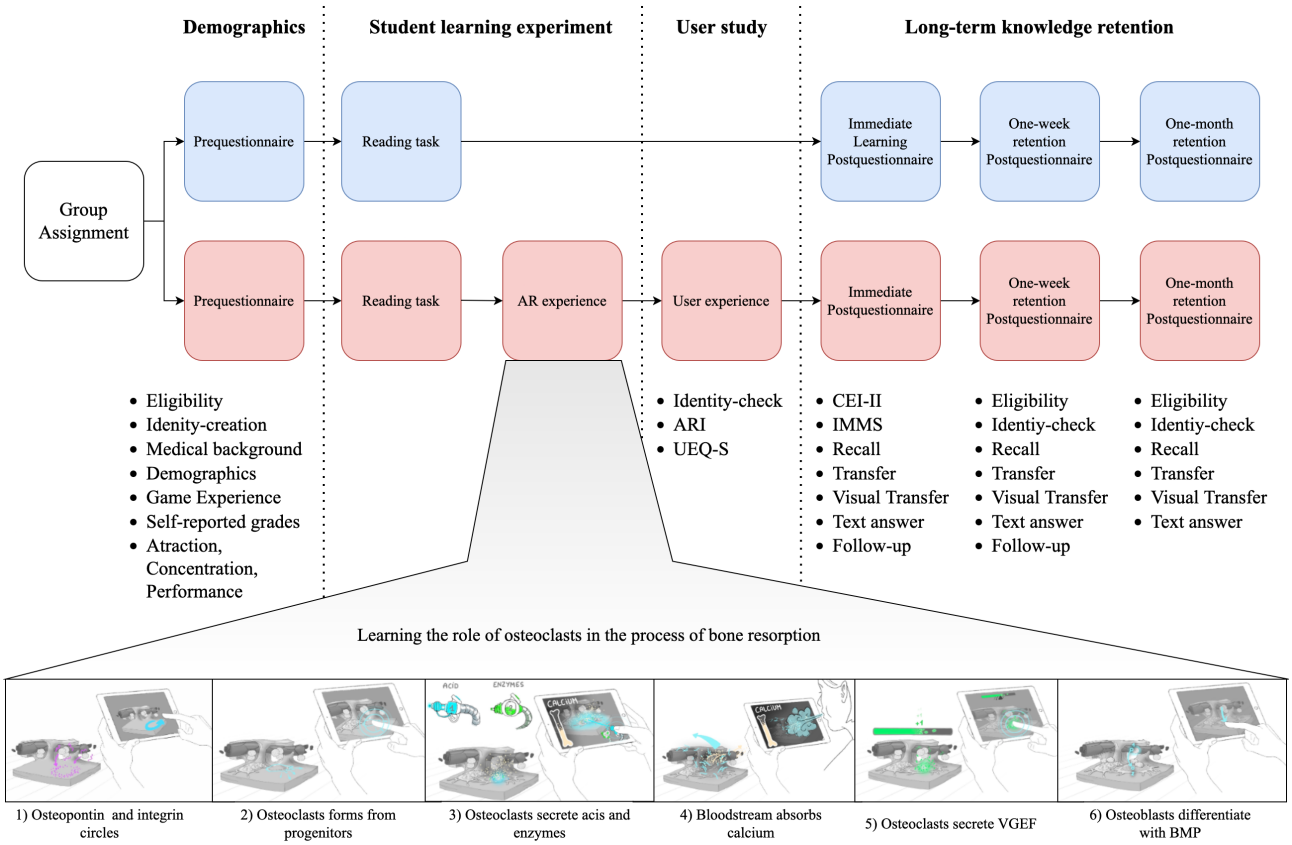
We conducted a power analysis to determine the necessary sample size for our study. We compared the effects of a serious AR game (AR-enhanced treatment) and text reading (text-only treatment) in a vignette experiment. The primary variable of interest was the difference in recall rate (correctly remembering an answer) between the two treatments. We aimed for an effect size of a 10% difference in the recall rate with a SD of 10% in recall rate. Using an α level of 0.05 and power of 0.8, the analysis indicated that a total of 32 participants (at least 16 per group) were required.

Students were randomly assigned to either a text-only group or an AR-enhanced group (Figure 2). Both groups were presented with identical content, though through different learning methods. Both groups read a description of bone remodeling,

XSL•FO
RenderX

while the AR-enhanced group followed up with an interactive AR game designed to provide an additional learning modality (Multimedia Appendix 1). Learning outcomes were assessed immediately after the treatment through a web-based survey, with follow-up assessments conducted at one week and one month to evaluate long-term retention.

Figure 2. Protocol for the student learning experiment. Students were randomly assigned to the text-only group (blue) and the AR experience (red). The AR content (bottom images) included 6 short interactions covering osteoclast differentiation, lacuna formation, and osteoblast differentiation (Multimedia Appendix 1). Demographic information was collected before the experiment, and learning outcomes were measured at 3 points to assess long-term knowledge retention. ARI: augmented reality immersion; BMP: Bone Morphogenetic Protein; CEI-II: Curiosity and Exploratory Inventory-II; IMMS: Instructional Materials Motivation Survey; UEQ-S: User Experience Questionnaire – Short; VEGF: Vascular Endothelial Growth Factor.



Hypothesis H1

First, we examined the relationship between individual learning experiences and motivation, specifically exploring how the game mechanisms influenced students' motivation to learn (H1). To assess this, we used the Attention, Relevance, Confidence, and Satisfaction (ARCS) model of motivation—comprising Attention, Relevance, Confidence, and Satisfaction—which is measured through the Instructional Materials Motivation Survey [58,59]. Attention refers to how effectively the learning materials engage and capture the learner's interest, while Relevance reflects the perceived connection between the materials and the real world. Confidence measures the learners' belief in their potential for success in the discipline when using the materials, and Satisfaction assesses learners' sense of accomplishment during the learning process. The full Instructional Materials Motivation Survey questionnaire [40] is included in Multimedia Appendix 4. To evaluate H1 and understand how participants perceive learning in AR, we analyzed the differences in ARCS model responses using a Wilcoxon signed rank test [58].

Hypothesis H2a, H2b, and H2c

To assess hypothesis H2a, we assessed students' learning outcomes using 3 distinct learning constructs and their composite. The questions are available in Multimedia Appendix 4. First, we included questions that measured recall, which captures the simplest form of knowledge retention. Second, we incorporated questions that assessed the transfer of knowledge, asking students to apply what they had learned in the game to new problems or contexts [60]. Last, we introduced a visual knowledge transfer task, where students were presented with visualizations of processes and asked to transfer knowledge from their learning experience to these new representations. This was particularly important as transferring knowledge to different forms of representation requires both an understanding of the content and its corresponding representation—a difficult task known as the "representation dilemma" [61]. To assess hypothesis H2b, we evaluated whether the game enhanced performance on visual transfer tasks, using visualizations that were not part of the game but depicted the same concepts.

To evaluate hypothesis H2c, our serious game was designed to leverage students' curiosity and exploration affinity. As opposed to linear expository texts, curiosity may be especially impactful for learning outcomes in a gaming context. To account for this,

XSL•FO
RenderX

we employed the Curiosity and Exploratory Inventory-II (CEI-II) [62], specifically focusing on students' general curiosity traits. This approach allowed us to examine how students' curiosity traits impacted their experience, as opposed to how the experience impacted students' curiosity at the time of playing.

To assess H2a, H2b, and H2c, we developed linear regression models for each of the 3 learning constructs: Recall, Transfer, and Visual Transfer. Additionally, we constructed a composite model that combined the performance across all 3 constructs. In each model, the learning construct served as the dependent variable, with the following independent variables:



The Treatment variable distinguishes between AR-enhanced and Text-only learning. The CEI2 variable represents the sum across all CEI-II items. The ARI (augmented reality immersion) variable represents the sum across all ARI items. The Grade variable is the student's self-reported Swiss Grade, ranging from 1 (worst) to 6 (best), with 4 representing a passing grade. The Age variable is self-reported within a 5-year age bracket converted to a continuous scale. The gender variable is either male, female, or other; however, no participants in our sample identified as other. The Experience variable refers to participants' prior knowledge of osteoclasts, rated on a 5-point scale from "no knowledge" to "regular interaction". The Image, Interaction, and Text variables referred to preferred modes of learning as self-reported by the student, the sum of three 5-point scales for each. The coefficient β11 was for the interaction between the treatment group and the CEI-II items sum.

Given the challenges of interpreting interaction models [63], we used marginal effects to better understand how the interaction between treatment and CEI-II impacts learning outcomes. To compute these effects, we used unit-level conditional estimates of the empirical distribution across the full range of the CEI-II scale (from 10 to 50 points), applying heteroskedasticity-robust SEs at a 90% CI [64]. Additionally, we conducted a Simple Slope Analysis [65] to investigate whether curiosity plays a different role in learning outcomes depending on the treatment condition.

Hypothesis H2d

To assess hypothesis H2d, we examined which topics became particularly salient for students under the different treatment conditions. Students were asked 3 text-based questions related to bone remodeling (Table S2 in Multimedia Appendix 5 and survey questions in Multimedia Appendix 4). To encourage more detailed responses, we used encouragement designs [66], prompting students to provide longer and more detailed answers. The text answers provided by students were analyzed with structural topic modeling [67] to explore the relationship between student responses and the treatment conditions. Topic modeling identifies coherent themes within large sets of written documents, while structural topic modeling extends this method by incorporating covariates that include meta-information not present in the documents themselves. In our analysis, we include the treatment condition, the specific question answered, the individual student providing the response, and the time in weeks since the treatment.

Topic modeling is an iterative statistical process that is highly sensitive to initialization [67] and requires fine-tuning to produce reliable results. Therefore, we adhered to established best practices in topic modeling. The number of topics was a priori unknown and needed to be determined. Since the optimal number of topics was unknown a priori, we tested a range of topic counts from 2 to 20. We also examined the impact of including covariates by running models both with and without them (Section S9.1 in Multimedia Appendix 5). Based on these tests, we determined that 10 topics (K=10) were the most suitable for the text data we analyzed. To address potential initialization issues, we ran 100 models with 10 topics (K=10) and selected the best-performing model, balancing exclusivity and coherence [67].

Finally, the topics generated by the model required clear and meaningful labels for interpretation. To ensure high-quality labeling, 3 independent experts assigned labels to the topics, and we calculated the intercoder reliability index, Krippendorff α [68], to assess the consistency and accuracy of the labels. Krippendorff α ranges from –1 to 1, with a score of 1 indicating perfect agreement, 0 representing random labeling, and –1 signifying completely inverse labeling [69]. Generally, an α value greater than 0.80 indicates high reliability, while values above 0.67 are considered acceptable for the purposes of our analysis [68].

Hypothesis H3

To assess hypothesis H3 on knowledge retention, participants completed knowledge assessment questionnaires at 3 time points: immediately following the learning experience, one week later, and one month later. However, due to participant attrition, the hypothesis could not be tested statistically. A multi-level model was initially planned to account for the repeated measurements but was not implemented due to the reduced sample size.

Ethical Considerations

Our participants were selected from dental students in the final year of dental school, attending a course at the end of their studies with appropriate knowledge of the topic and content level. All students were informed in advance about the study taking place, and participation was optional. Students gave informed consent before participating. The data was handled confidentially in line with privacy regulations. Participants were not compensated for their participation. The study was approved by the ethics commission of the Canton of Zürich (Req-2022 - 00290).

Results

Overview

No data points from the 39 student participants were excluded by the principal (23 female, 14 male, 2 NA). However, 2 participants dropped out during the experiment, resulting in attrition. Students were aged 27 years (SD 3.6; measures in 5 y blocks) on average. Most had heard of osteoclasts (36

XSL•FO
RenderX

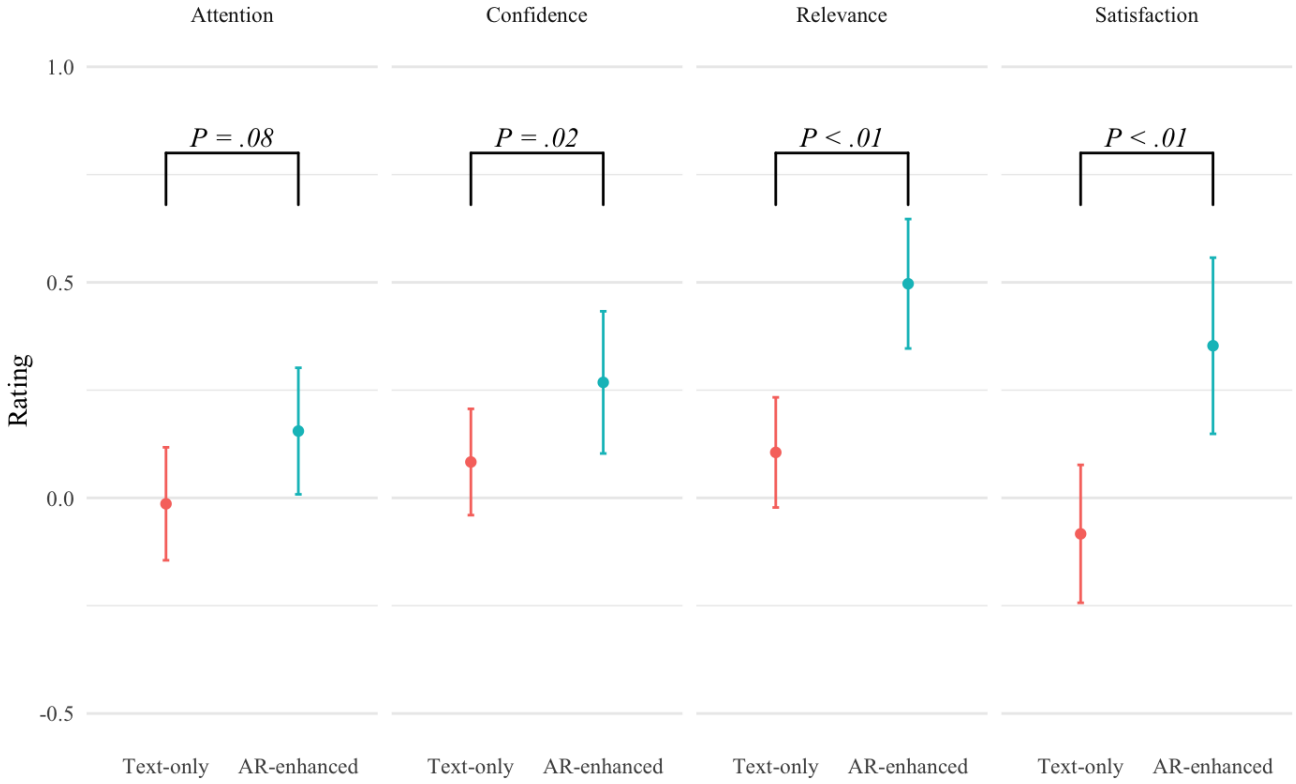
students), and a few had worked with osteoclasts before (2 students). Six students started studying dental medicine with the Master program (0 - 1 y in dental medicine), most students started with dental medicine in their Bachelor program (3 y of experience), and 6 students have extended experience (5 y of experience). Students also self-reported their CEI-II as 32.26 on average (SD 6.74).

For a long-term assessment, we asked the participants to recall the learned material after one week and after one month. Participation in the follow-up survey was voluntary. The first follow-up survey after a week had 6 responses, and the second follow-up survey after a month had 15 responses. Due to limited response rates, statistical analysis of the follow-up data was not feasible according to our power analysis.

Hypothesis H1

We examined the participants' subjective learning experience using the ARCS model (Figure 3). We found significant effects for Confidence (W=11,882.5; *P*=.02), Relevance (W=10,417; *P*<.001), and Satisfaction (W=4561; *P*<.001). Although there was no difference in performance, participants reported higher motivation overall—feeling more confident, perceiving the topic as more relevant, and expressing greater satisfaction with their progress. Interestingly, we did not find significant differences in Attention (W=18,619; *P*=.08), which contrasts with previous findings that interactive learning tools tend to enhance attention [70,71].

Figure 3. Evaluating the Instructional Materials Motivation Survey responses based on the ARCS construct with 95% CIs, we found no significant difference in Attention (W=18,619; *P*=.08) between the learning scenarios. However, participants reported increased Confidence (W=11,882.5; *P*=.02) , greater perceived Relevance (W=10,417; *P*<.001) of the material, and higher Satisfaction (W=4561; *P*<.001) with the learning experience. AR: augmented reality; ARCS: Attention, Relevance, Confidence, and Satisfaction.



Hypothesis H2a

We reported the participant responses as the proportions of correct answers for each question. For recall tasks (Figure S20 in Multimedia Appendix 5), we reported very high correct responses, indicating a ceiling effect. Several questions were answered correctly by 100% of the participants, although there were minor differences between treatment groups, which appeared to be random. For transfer tasks (Figure S21 in Multimedia Appendix 5), we observed that 2 questions appeared simple, with very high correct response rates, similar to recall tasks. The other 3 questions had high correct response rates, ranging from 50% to 70%. For visual transfer tasks (Figure S22 in Multimedia Appendix 5), we reported the lowest correct response rates. A Welch *t* test revealed no significant difference between the AR and text treatments in learning outcomes (t=2.0103; $P_{Bonferroni}$=.095). Although text treatments had nominally higher scores across all transfer tasks, the differences were not statistically significant.

Hypothesis H2b

In the regression analysis (Table 1), we reported no significant results for recall and transfer learning. These models had adjusted R^2 of 0.03 and 0.12, respectively, and their *F* statistics were not significant. However, the models for Visual Transfer and Total cumulative score showed a significant *F* statistic ($F_{11,24,visual\ transfer}$=2.72, *P*=.02; $F_{11,24,total}$=2.81, *P*=.02). Only the visual transfer model revealed significant effects for the AR treatment (t=−2.099; $P_{visual_transfer}$= .046). All other models could not discern a difference between the treatment conditions for the outcome variables. Across models, we also found that

XSL•FO
RenderX

previous experience with osteoclasts was a strong predictor of the result (t_{total}=2.469, P_{total}=.02; $t_{transfer}$=1.865, $P_{transfer}$=.07; $t_{visual\ transfer}$= 1.789, $P_{visual\ transfer}$=.09). For the visual transfer task, we also find a significant effect of the CEI-II ($t_{visual\ transfer}$ =−2.518, $P_{visual\ transfer}$=.02), the interaction effect ($t_{visual_transfer}$=1.913, $P_{visual\ transfer}$=.06) and the preference for text learning ($t_{visual_transfer}$=2.575, $P_{visual_transfer}$=.02). Last, the cumulative score also showed a significant effect for the participants being male (P_{total}=.06) which may be attributed to motivation [72] but cannot be discerned by this study.

XSL•FO
RenderX

Table . Student outcomes for the 3 learning tasks and a cumulative score.

	Dependent variable			
	Total	Recall	Transfer	Visual transfer
AR[a] treatment				
Effect size (SE)	–0.211 (0.172)	0.071 (0.205)	–0.106 (0.358)	–0.615 (0.293)
t value (*df*=11)	−1.261	0.347	−0.296	−2.099
P value	.22	.73	.76	.05
CEI-II[b]				
Effect size (SE)	–0.001 (0.003)	0.005 (0.004)	0.005 (0.007)	–0.014 (0.006)
t value (*df*=11)	−0.401	1.196	0.798	−2.518
P value	.69	.24	.43	.02
ARI[c]				
Effect size (SE)	0.003 (0.002)	0.005 (0.003)	0.007 (0.003)	–0.002 (0.003)
t value (*df*=11)	1.647	1.737	1.842	−0.569
P value	.11	0.095	.07	.57
Swiss grade				
Effect size (SE)	0.044 (0.047)	0.049 (0.056)	0.071 (0.097)	0.008 (0.079)
t value (*df*=11)	0.956	0.979	0.727	0.107
P value	.34	.33	.47	.91
Age (years)				
Effect size (SE)	–0.006 (0.005)	0.004 (0.006)	–0.011 (0.011)	–0.012 (0.009)
t value (*df*=11)	−1.160	0.693	−0.972	−1.338
P value	.25	.49	.34	.19
Male				
Effect size (SE)	0.085 (0.041)	0.084 (0.049)	0.083 (0.085)	0.082 (0.069)
t value (*df*=11)	2.081	1.834	0.978	1.182
P value	.05	.08	.33	.24
Experience				
Effect size (SE)	0.195 (0.079)	0.037 (0.096)	0.307 (0.165)	0.241 (0.134)
t value (*df*=11)	2.469	0.392	1.865	1.789
P value	.02	.69	.07	.09
Image preference				
Effect size (SE)	–0.01 (0.014)	0.002 (0.017)	–0.034 (0.029)	0.002 (0.023)
t value (*df*=11)	−0.708	0.130	−1.177	0.103
P value	.49	.9	.25	.92
Interaction preference				
Effect size (SE)	–0.004 (0.015)	–0.013 (0.018)	0.008 (0.030)	–0.008 (0.025)
t value (*df*=11)	−0.305	−0.740	0.267	−0.346
P value	.76	.46	.79	.73
Text preference				
Effect size (SE)	0.016 (0.012)	–0.019 (0.015)	0.013 (0.026)	0.054 (0.021)
t value (*df*=11)	1.293	−1.329	0.516	2.575
P value	.21	.2	.61	.02

XSL•FO
RenderX

	Dependent variable			
	Total	Recall	Transfer	Visual transfer
Interaction of AR treatment and CEI-II				
Effect size (SE)	0.004 (0.005)	–0.005 (0.006)	–0.001 (0.011)	0.017 (0.009)
t value (*df*=11)	0.772	−0.710	−0.119	1.913
P value	.47	.48	.91	.07
Constant				
Effect size (SE)	0.156 (0.432)	0.537 (0.515)	–0.244 (0.900)	0.175 (0.736)
t value (*df*=11)	0.362	1.043	−0.271	0.238
P value	.72	.31	.79	.81
Observations	36	36	36	36
R^2	0.563	0.354	0.411	0.555
Adjusted R^2	0.362	0.058	0.142	0.351
Residual SE (*df*=25)	0.100	0.117	0.204	0.167
F test (*df*=11, 24)	2.807	1.196	1.526	2.719
P value	.02	.34	.19	.02

[a]AR: augmented reality.

[b]CEI-II: Curiosity and Exploratory Inventory-II.

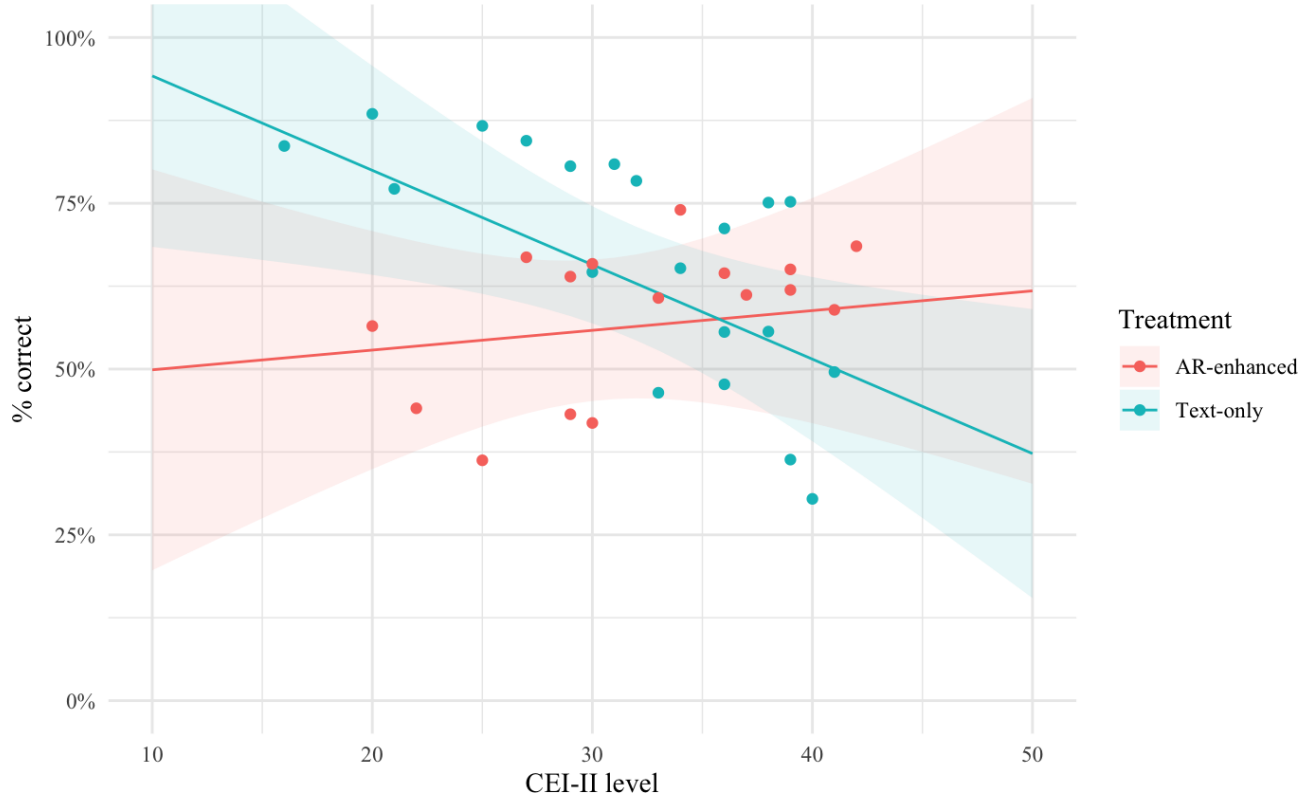
[c]ARI: augmented reality immersion.

Hypothesis H2c

We investigate the marginal effects of the model for visual transfer learning (Figure 4). In the range of the CEI-II index, we detect an inflection point at 36 points. For lower CEI-II scores, text-only learning yields the best results. However, for higher CEI-II scores, AR experiences help participants to score higher. The results are limited by the empirical distribution not having observations for CEI-II under 16 or over 42, making the behavior in the tails of the distribution speculative. We employ a Simple Slope Analysis and detect that the text-only treatment has a significant negative slope for CEI-II (1.4% decrease in correct responses, $t=-2.518$, $P=.02$), whereas the CEI-II has no significant effect under the AR-enhanced treatment ($t=0.347$, $P=.73$).

XSL•FO
RenderX

Figure 4. A simple slope analysis of our model revealed a strong inverse relationship between reading and curiosity in the students (1.4% decrease in correct responses, t_{24}=−2.518, *P*=.02). For students who were curious, this could be offset by experiencing the information in AR. AR: augmented reality; CEI-II: Curiosity and Exploratory Inventory-II.



Hypothesis H2d

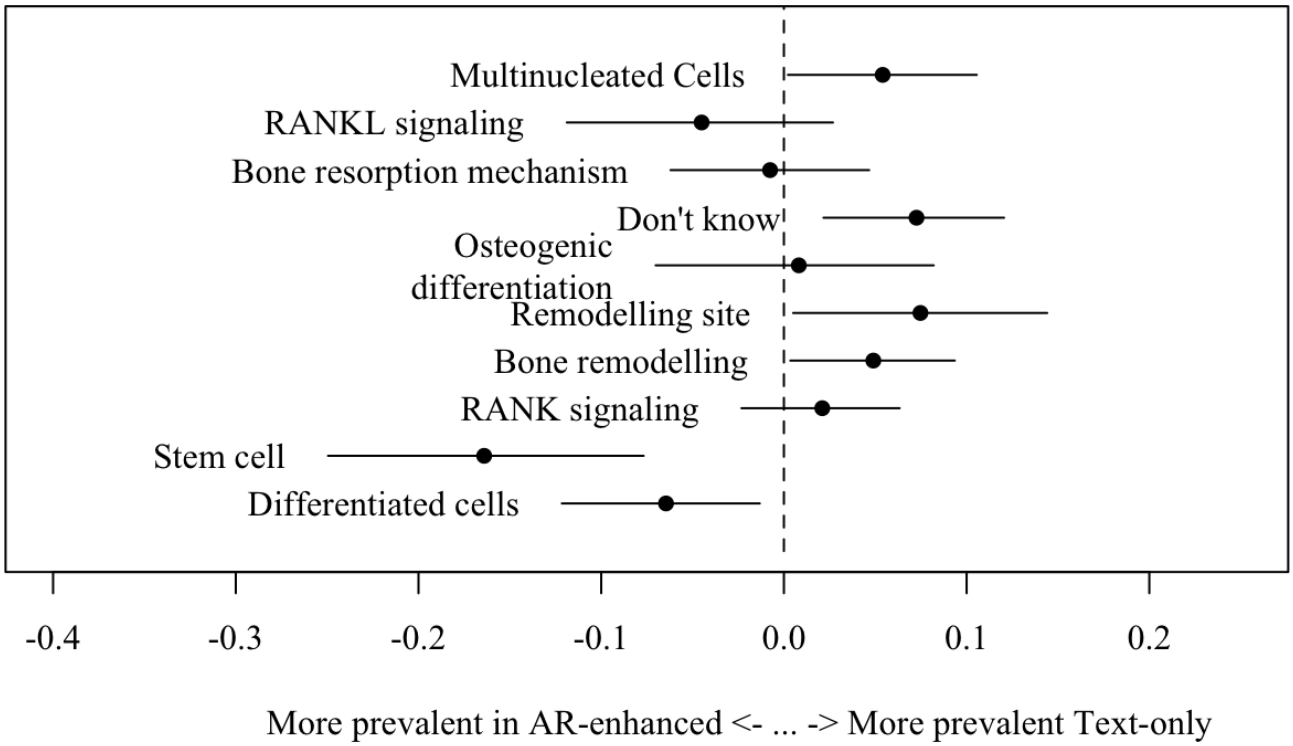
We first analyzed the frequency of word occurrences in participants' responses (Figure S24 in Multimedia Appendix 5). The words identified generally aligned with the topics addressed in each question. When categorizing responses, certain terms are clustered into thematic groups. For the formation of osteoclasts, the most common answers could be grouped into precursors (hemopoetic, progenitor, and stem) and receptors (RANKL [Receptor Activator of NF-κB Ligand], Arginylglycylaspartic acid, and bind). For the function of osteoclasts and osteoblasts, the answers are grouped according to the 2 cell types: osteoclasts (resorb and dissolve), and osteoblasts (form and build). The osteogenic differentiation is separated into cell-related words (bone cell, mesenchyme, osteoblast, osteoclast, and stem) and signaling/receptor-related words (Wingless und Int-1 [Integrated], Bone Morphogenetic Proteins, RANKL, and RANK [Receptor Activator of NF-κB]).

Our topic models were labeled by 3 independent raters, and the intercoder reliability was calculated using Krippendorff α, which reached a value of 0.717. This indicates good intercoder reliability. Although full consensus was not always reached, in cases of disagreement, we relied on the most probable words from the topic contrasts (Section S9.2 in Multimedia Appendix 5) to determine a label for the topics. We looked at the prevalence of topics across all texts, where 100% represents all words written (Figure S25 in Multimedia Appendix 5). The topics of osteogenic differentiation and RANKL signaling were the most frequently discussed, each accounting for approximately 15% or more of the total text. Topics related to differentiated cells, the remodeling site, and stem cells each made up more than 10% of the text. The topics of bone resorption mechanisms, RANK signaling, bone remodeling, and multinucleated cells each contributed more than 5% of the total text. The smallest topic was the "Don't know" category, which captured uncertain or unclear responses from students.

We used the properties of structural topic modeling to assess whether the treatment type influenced the topics students focused on (Figure 5). We found that in the AR-enhanced treatment, students wrote more frequently about the stem cells (t=−3.681, P<.01) and differentiated cells (t=−2.303, P=.02) In contrast, students in the text-only treatment tended to focus more on the remodeling site (t=2.220, P=.03) and the process of bone remodeling (t=2.108, P=.03). People in the text-only treatment were more likely to respond with "Don't know" (t=2.953, P<.01). No significant differences were observed for other topics between the two treatment groups.

XSL•FO
RenderX

Figure 5. Topic prevalence by treatment. Negative values indicate an increased prevalence to appear in the AR-enhanced treatment, while positive values indicate an increased prevalence to appear in the text-only treatment. When the dashed line is crossed, no significant prevalence for either treatment is detected. AR: augmented reality; RANK: Receptor Activator of NF-κB; RANKL: Receptor Activator of NF-κB Ligand.



Topics may also be influenced by the 3 questions we asked (Figure S37 in Multimedia Appendix 5). To determine whether a topic was associated with a particular question, we examined the topic across all 3 paired comparisons between questions. For instance, the topic on multinucleated cells was associated less frequently with the question on osteoclast formation (Q1) but was strongly associated with both the coordinated function of osteoclasts and osteoblast in the canopy (Q2) and osteogenic differentiation (Q3). That indicated that participants were statistically significantly more likely to talk about multinucleated cells in Q2 and Q3.

In the formation question (Q1), participants were less likely to talk about RANK-RANKL signaling and stem cells and multinucleated cells, and more likely to talk about bone remodeling, bone resorption, the remodeling site, differentiated cells, stem cells, or osteogenic differentiation. In the functions question (Q2), participants were less likely to talk about bone remodeling, bone resorption mechanism, the remodeling site, or differentiated cells, and more likely to talk about osteogenic differentiation, RANK-RANKL signaling, and stem cells and multinucleated cells. In the differentiation question (Q3), participants were less likely to talk about bone resorption mechanism, stem cells, RANK signaling, or osteogenic differentiation, and more likely to talk about bone resorption, the remodeling site, differentiated cells, multinucleated cells, and RANKL signaling.

Hypothesis H3

Due to participant attrition, we are unable to report statistical results for the long-term study. However, those participants who did complete the follow-up showed remarkable performance when playing the AR game (Figure S23 in Multimedia Appendix 5). In some cases, participants even showed signs of improved retention over time. However, due to the small sample size, further analysis of these trends was not possible.

Discussion

Principal Findings

In this study, we evaluated the use of an AR medical teaching game in a classroom setting. Before examining the effectiveness of our difference-in-differences design comparing text-based and AR-based learning, we briefly summarize the usability study (Multimedia Appendices 2 and 6), which showed that participants adapted well to the AR experience with minimal difficulty. In our vignette experiment, one group received the AR game with a written text (AR-enhanced), while the second group only received the text (text-only). The goal was to assess whether the AR game enhances student engagement, motivation, and learning outcomes, as well as long-term knowledge retention. Our findings suggest that immersive media positively impacts motivation (H1). The learning experiment showed similar learning outcomes across treatments (H2a), while visual learning was not significantly improved overall (H2b), likely due to ceiling effects from easier questions, though it did improve for curious students (H2c). Topic modeling revealed that AR learners focused more on cells in line with the game design, while text learners focused more on processes (H2d). Due to participant attrition, we could not assess long-term learning (H3). In summary, AR facilitated the learning of key concepts, with curious students benefiting the most.

XSL•FO
RenderX

Principal Results

We found that presenting advanced scientific content through AR significantly improves motivation (H1) by enhancing students' subjective learning experience. AR increased engagement with the process of bone remodeling, helping students develop a deeper appreciation for the topic and perceive it as more important, according to the ARCS model. Students who interacted with the AR scene attributed higher value and relevance to the concepts, forming a stronger personal connection to the content by recognizing its value. Additionally, the playful nature of the AR experience led students to perceive the material as easier to learn, which, despite not improving assessment performance, increases overall satisfaction and may reduce learning pressure.

We also aimed to examine the impact of AR-enhanced content on students' learning outcomes. We started with a conservative hypothesis that students in the AR experience would achieve comparable learning outcomes (H2a). While we anticipated improved performance on visual tasks due to the visual nature of the game (H2b), this expectation was not confirmed. Instead, we found that only highly curious students performed better on visual transfer tasks when learning with AR-enhanced content compared with a text-only approach (H2c). Specifically, a negative relationship between text-only learning and curiosity was observed, which was mitigated by the AR game for visual transfer tasks. Students with higher curiosity performed better in the AR condition than in the text-only condition. We identified a cutoff at CEI-II of 35 beyond which students improved performance on visual transfer tasks following the AR experience. Since dental medicine involves interpreting visual tasks, AR representations may offer a valuable tool to enhance student performance, particularly for curious learners.

We expected that the visual narrative of the AR game would direct students' attention to specific concepts. Indeed, the game's emphasis on cellular mechanisms influenced the text responses of participants and shaped how they learned and retained the material (H2d). Notably, the students in the AR-enhanced treatment group mentioned the cells more compared with the text-only treatments, whose responses focused more on processes and locations. The interactive elements and game mechanics of the AR experience appeared to guide students' language and understanding when answering questions. The choice of how the scientific process was translated into game mechanics played a crucial role in shaping students' learning experiences. For most AR games, the process of translating knowledge into game mechanics is only focused on playability, but our investigation highlights the importance of this aspect, as detailed in the game design and usability study (Multimedia Appendices 1 and 2).

Although our sample size was insufficient for a statistical analysis of long-term knowledge retention (H3), we did observe some significant trends in the text responses (Section S9.4 in Multimedia Appendix 5). Specifically, we found that general topics increased in prevalence, while more specific topics, such as RANK-RANKL signaling, decreased. Most topics maintained their prevalence. This suggests that a stable core of learned concepts persisted over time, though with a slight loss of detail.

Limitations

We identified 4 major limitations of this study that can guide future research on serious gaming in AR for educational purposes. Future investigations should focus on (1) understanding the role of student/player characteristics, (2) effectively translating scientific knowledge into game mechanics, (3) evaluating the effectiveness of long-term assessments, and (4) ensuring that the tests used to measure learning outcomes are appropriate for assessing the game's impact.

First, regarding player characteristics (Multimedia Appendix 2), it is important to recognize that games are played by individuals with diverse backgrounds and gaming skills. Casual gamers [73], who are often less familiar with AR experiences, differ from participants who prefer speedruns or completionist approaches [74] (Figure S41 in Multimedia Appendix 2). Content designed for more advanced players may be less accessible or engaging for casual gamers. To ensure that a serious game effectively imparts knowledge, game design must consider varying play styles. Our observations revealed that most participants were casual gamers, suggesting that performance issues in the average playtime group were more related to user interface and game dynamics than to the learning content itself [75,76].

Second, when translating a scientific process into a game, decisions must be made about what to represent, what to omit, and which aspects of the process to emphasize (visually, interactively, and mechanically) versus those that are only indirectly conveyed [77]. Our findings showed that AR participants focused more on cells, while text-only participants concentrated on location and process. This suggests that AR encouraged students to engage more deeply with the game's core mechanisms, that is, the cells, while somewhat distancing them from other aspects, such as the overall process. Given the strong visual guidance in the AR game, it is crucial that the game not only presents knowledge but also highlights key concepts as part of the immersive experience. To improve AR-enhanced learning, it would be beneficial to follow experimental design frameworks like design, experiment, analyze, and reproduce [78] to ensure that both content delivery and learning outcomes are aligned in a reproducible framework [79].

Third, the small number of participants due to the class size and the participant attrition for the longitudinal study limited our ability to draw conclusions. While some promising trends were observed in the long-term study—such as certain students showing improved knowledge retention after exposure to the AR treatment—it is unclear whether these effects were driven by individual differences or the treatment itself. The interaction between curiosity and the AR treatment is also intriguing but requires further investigation with a larger participant group to provide robust evidence of a positive relationship between curiosity and gamified educational content.

Fourth, another limitation was the selection of tests used to assess learning outcomes. The questions designed for recall, transfer, and visual transfer tasks must not only align with the content but also be sufficiently challenging to differentiate

XSL•FO
RenderX

responses across treatments. In this study, the questions may have been too easy, leading to a ceiling effect that prevented us from detecting significant differences between participants at the expected effect sizes. Most participants scored 75% or higher (at least 8 out of 11 questions) on recall and transfer tasks, with previous knowledge of osteoclasts being the strongest predictor of performance. As a result, the impact of AR on learning outcomes may have been more evident in the visual transfer task, where participants showed more variation.

Comparison With Prior Work

Research on the use of serious gaming for medical education reveals mixed findings regarding the efficacy of AR games, though there is growing evidence of their positive impact on student engagement and learning experiences. We identified 21 studies on serious gaming in medical learning, providing valuable insights that complement our findings. Despite the mixed outcomes in the literature, there is a gradual increase in the adoption of game-based learning approaches in education [80]. Students generally report higher content acceptance and more positive learning experiences when engaging with serious games [81,82]. Furthermore, process learning through serious games, particularly in flipped classroom settings, has been shown to prepare students effectively for tasks [83]. Overall, gamified teaching formats are widely well-received by learners and offer the potential to create immersive educational experiences.

The positive effects of intrinsic benefits are contrasted with the question of why some studies have null-finding and others do not. Null findings are reported regularly in both individual studies [34] and meta-reviews (13 studies [35], 27 studies [36], and 7 studies [37]). These attest to no significant benefits in the use of virtual reality (VR) or AR for learning outcomes. However, they equally suggested that VR and AR could promote intrinsic benefits such as increasing learner immersion and engagement. Our findings replicate these positive effects, while also suggesting a potential moderator that may explain when studies yield null findings versus positive results. Specifically, we identify curiosity to engage the gaming content as a key factor that drives visual understanding of content, and therefore, benefits can only be detected when tasks and assessments capture this criterion.

Studies showing positive outcomes often focus on the enhancement of spatial abilities in medical students, who benefit from the embodied experience offered by VR/AR [84-88]. A study in 2021 compared the teaching efficiency of VR with traditional education [89] and found that VR students had a significantly higher pass rate than those receiving conventional medical training. The VR students also demonstrated a better understanding of various processes and procedures. Another study [90] examined whether AR could improve short-term anatomical knowledge in head and neck anatomy compared with traditional 2D screen-based learning, with the AR group outperforming the traditional learning group. A further study [36] highlighted that VR/AR-based teaching not only improved medical education but also positively influenced student enthusiasm and enjoyment. In the field of occupational medicine, immersive VR serious games help students relate to patients' states and circumstances [33]. In our study, we observed that AR-based learning positively impacted visual transfer tasks, particularly for curious students. This may have been due to eliciting enhanced spatial understanding fostered by the AR experience in the medical context.

Conclusions

AR-enhanced learning holds great promise as an educational tool for the 21st century, particularly in the medical sciences, but also beyond. Immersive serious gaming in AR offers a compelling opportunity to engage students with complex scientific content, which can often be difficult to communicate and may deter new learners. However, the most critical factor that shapes student learning from AR-enhancement is the conversion of scientific knowledge to game mechanics. Additionally, AR has the potential to promote equity in the classroom by offering more inclusive learning opportunities. Our findings reinforce previous research showing that AR fosters a stronger connection with content, resulting in more positive perceptions of learning. While no significant learning benefits were observed for simpler tasks, consistent with existing literature, AR proved especially beneficial for more complex visual transfer tasks. We identified curiosity as a key factor that moderates the effectiveness of AR learning, with curious students experiencing greater learning gains. Ultimately, immersive serious gaming provides students with new perspectives on the content. Our findings indicated that particularly curious students benefited from interactive learning and could improve their learning outcomes compared with a text-only learning environment.

Acknowledgments

The game development was funded by Quintessential Publishing. The 3D models were provided by interActive Systems. JG was supported by an ETH Zurich Doc.Mobility Fellowship. JC was supported by the Future Learning Initiative of ETH Zurich. The students of the 2022 cohort at UZH Dental Medicine are thanked for their participation in the study. We thank everybody who downloaded AR Osteoclasts and played the game; your participation was instrumental in this research. Generative artificial intelligence was used to review the wording of the original manuscript and improve readability (GPT-4o).

Data Availability

All data analyzed during this study are included in this published article and its supplementary information files.

XSL•FO
RenderX

Conflicts of Interest

RG and BS are together with Hendrik Terheyden editors of the book “Visual Biology in Oral Medicine”.

Multimedia Appendix 1
Game design and scientific background for the serious game.
[DOCX File, 6814 KB - games_v13i1e64751_app1.docx]

Multimedia Appendix 2
Usability study of the serious game.
[DOCX File, 1036 KB - games_v13i1e64751_app2.docx]

Multimedia Appendix 3
Outlines of the survey questions fielded in Qualtrics.
[DOCX File, 85 KB - games_v13i1e64751_app3.docx]

Multimedia Appendix 4
Zip file that contains survey data, text response data, and game data used in the analysis.
[ZIP File, 221 KB - games_v13i1e64751_app4.zip]

Multimedia Appendix 5
Advanced statistical analysis of the data not included in the main text.
[DOCX File, 6461 KB - games_v13i1e64751_app5.docx]

Multimedia Appendix 6
Video of playing AR Osteoclasts from beginning to end.
[MP4 File, 191857 KB - games_v13i1e64751_app6.mp4]

References

1. Yu J, Denham AR, Searight E. A systematic review of augmented reality game-based learning in STEM education. Education Tech Research Dev 2022 Aug;70(4):1169-1194. [doi: 10.1007/s11423-022-10122-y]
2. Kwon HB, Park YS, Han JS. Augmented reality in dentistry: a current perspective. Acta Odontol Scand 2018 Oct;76(7):497-503. [doi: 10.1080/00016357.2018.1441437] [Medline: 29465283]
3. Dhar P, Rocks T, Samarasinghe RM, Stephenson G, Smith C. Augmented reality in medical education: students' experiences and learning outcomes. Med Educ Online 2021 Dec;26(1):1953953. [doi: 10.1080/10872981.2021.1953953] [Medline: 34259122]
4. Milgram P, Takemura H, Utsumi A, Kishino F. Augmented reality: a class of displays on the reality-virtuality continuum. : SPIE Presented at: Telemanipulator and Telepresence Technologies; Dec 21, 1995; Boston, MA, United States p. 282-292. [doi: 10.1117/12.197321]
5. Zünd F, Ryffel M, Magnenat S, et al. Augmented creativity: bridging the real and virtual worlds to enhance creative play. : Association for Computing Machinery Presented at: SA '15: SIGGRAPH Asia 2015 Mobile Graphics and Interactive Applications; Nov 2, 2015; Kobe, Japan p. 1-7. [doi: 10.1145/2818427.2818460]
6. Ryffel M, Zünd F, Aksoy Y, et al. AR museum: a mobile augmented reality application for interactive painting recoloring. Presented at: Multi Conference on Computer Science and Information Systems 2017; Aug 23, 2017; Lisbon, Portugal URL: https://studios.disneyresearch.com/wp-content/uploads/2019/03/AR-Museum-A-Mobile-Augmented-Reality-Application-for-Interactive-Painting-Recoloring.pdf [accessed 2025-06-13]
7. Rothbaum BO, Rizzo AS, Difede J. Virtual reality exposure therapy for combat-related posttraumatic stress disorder. Ann N Y Acad Sci 2010 Oct;1208(1):126-132. [doi: 10.1111/j.1749-6632.2010.05691.x] [Medline: 20955334]
8. Molnar A. Content type and perceived multimedia quality in mobile learning. Multimed Tools Appl 2017 Oct;76(20):21613-21627. [doi: 10.1007/s11042-016-4062-2]
9. Dzyuba N, Jandu J, Yates J, Kushnerev E. Virtual and augmented reality in dental education: the good, the bad and the better. Eur J Dent Educ 2022 Nov 6. [doi: 10.1111/eje.12871] [Medline: 36336847]
10. Fahim S, Maqsood A, Das G, et al. Augmented reality and virtual reality in dentistry: highlights from the current research. Appl Sci (Basel) 2022;12(8):3719. [doi: 10.3390/app12083719]
11. Monterubbianesi R, Tosco V, Vitiello F, et al. Augmented, virtual and mixed reality in dentistry: a narrative review on the existing platforms and future challenges. Appl Sci (Basel) 2022 Jan 15;12(2):877. [doi: 10.3390/app12020877]
12. Morris BJ, Croker S, Zimmerman C, Gill D, Romig C. Gaming science: the “Gamification” of scientific thinking. Front Psychol 2013 Sep 9;4(607):607. [doi: 10.3389/fpsyg.2013.00607] [Medline: 24058354]

XSL•FO
RenderX

13. Gruber R, Stadlinger B. Cell-Atlas - Visual Biology in Oral Medicine: Quintessenz Verlag; 2022. URL: https://www.quintessence-publishing.com/deu/de/product/cell-to-cell-communication-cell-atlas-visual-biology-in-oral-medicine [accessed 2025-06-13]
14. Insua A, Monje A, Wang HL, Miron RJ. Basis of bone metabolism around dental implants during osseointegration and peri-implant bone loss. J Biomed Mater Res A 2017 Jul;105(7):2075-2089. [doi: 10.1002/jbm.a.36060] [Medline: 28281321]
15. Eng-Tat A, Liu P, R J, et al. The challenges of histology education. Clin Anat 2023 Mar;36(2):308-314. [doi: 10.1002/ca.23989] [Medline: 36494890]
16. García M, Victory N, Navarro-Sempere A, Segovia Y. Students' views on difficulties in learning histology. Anat Sci Educ 2019 Sep;12(5):541-549. [doi: 10.1002/ase.1838] [Medline: 30378295]
17. McConnell DA, Chapman L, Czajka CD, Jones JP, Ryker KD, Wiggen J. Instructional utility and learning efficacy of common active learning strategies. Journal of Geoscience Education 2017 Nov;65(4):604-625. [doi: 10.5408/17-249.1]
18. Akçayır G, Akçayır M. The flipped classroom: a review of its advantages and challenges. Comput Educ 2018 Nov;126:334-345. [doi: 10.1016/j.compedu.2018.07.021]
19. Kokotsaki D, Menzies V, Wiggins A. Project-based learning: a review of the literature. Improving Schools 2016 Nov;19(3):267-277. [doi: 10.1177/1365480216659733]
20. Kapur M. Productive failure. Cogn Instr 2008 Jul 8;26(3):379-424. [doi: 10.1080/07370000802212669]
21. Taylor DCM, Hamdy H. Adult learning theories: implications for learning and teaching in medical education: AMEE guide no. 83. Med Teach 2013 Nov;35(11):e1561-e1572. [doi: 10.3109/0142159X.2013.828153] [Medline: 24004029]
22. Graffam B. Active learning in medical education: strategies for beginning implementation. Med Teach 2007 Feb;29(1):38-42. [doi: 10.1080/01421590601176398] [Medline: 17538832]
23. Sait MS, Siddiqui Z, Ashraf Y. Advances in medical education and practice: student perceptions of the flipped classroom. Adv Med Educ Pract 2017;8:317-320. [doi: 10.2147/AMEP.S133328] [Medline: 28496377]
24. Kapur M, Bielaczyc K. Designing for productive failure. Journal of the Learning Sciences 2012 Jan;21(1):45-83. [doi: 10.1080/10508406.2011.591717]
25. Avila-Garzon C, Bacca-Acosta J, Kinshuk, Duarte J, Betancourt J. Augmented reality in education: an overview of twenty-five years of research. CONT ED TECHNOLOGY 2021;13(3):ep302. [doi: 10.30935/cedtech/10865]
26. De Micheli AJ, Valentin T, Grillo F, Kapur M, Schuerle S. Mixed reality for an enhanced laboratory course on microfluidics. J Chem Educ 2022 Mar 8;99(3):1272-1279. [doi: 10.1021/acs.jchemed.1c00979]
27. Sweller J, Ayres P, Kalyuga S. Cognitive Load Theory: Springer; 2011. [doi: 10.1007/978-1-4419-8126-4]
28. Hornecker E. The to-and-fro of sense making. ACM Trans Comput-Hum Interact 2016 May 28;23(2):1-48. [doi: 10.1145/2882785]
29. Scaife M, Rogers Y. External cognition: how do graphical representations work? Int J Hum Comput Stud 1996 Aug;45(2):185-213. [doi: 10.1006/ijhc.1996.0048]
30. Grübel J, Gath-Morad M, Aguilar L. Fused twins: a cognitive approach to augmented reality media architecture. Presented at: MAB '20: Proceedings of the 5th Media Architecture Biennale Conference; Oct 22, 2021; Amsterdam and Utrecht Netherlands. [doi: 10.1145/3469410.3469435]
31. Lampropoulos G, Keramopoulos E, Diamantaras K, Evangelidis G. Augmented reality and gamification in education: a systematic literature review of research, applications, and empirical studies. Appl Sci (Basel) 2022 Jul 5;12(13):6809. [doi: 10.3390/app12136809]
32. Felszeghy S, Pasonen-Seppänen S, Koskela A, et al. Using online game-based platforms to improve student performance and engagement in histology teaching. BMC Med Educ 2019 Jul 22;19(1):273. [doi: 10.1186/s12909-019-1701-0] [Medline: 31331319]
33. Hänni RT, Luong T, Chatain J, Mangold F, Dressel H, Holz C. HistoLab VR: a user elicitation study exploring the potential of virtual reality game-based learning for hazard awareness. Presented at: VRST '24: Proceedings of the 30th ACM Symposium on Virtual Reality Software and Technology; Oct 9, 2024; Trier Germany p. 1-11.
34. Moro C, Štromberga Z, Raikos A, Stirling A. The effectiveness of virtual and augmented reality in health sciences and medical anatomy. Anat Sci Educ 2017 Nov;10(6):549-559. [doi: 10.1002/ase.1696] [Medline: 28419750]
35. Baashar Y, Alkawsi G, Ahmad WNW, et al. Effectiveness of using augmented reality for training in the medical professions: meta-analysis. JMIR Serious Games 2022 Jul 5;10(3):e32715. [doi: 10.2196/32715] [Medline: 35787488]
36. Barteit S, Lanfermann L, Bärnighausen T, Neuhann F, Beiersmann C. Augmented, mixed, and virtual reality-based head-mounted devices for medical education: systematic review. JMIR Serious Games 2021 Jul 8;9(3):e29080. [doi: 10.2196/29080] [Medline: 34255668]
37. Lucena-Anton D, Fernandez-Lopez JC, Pacheco-Serrano AI, Garcia-Munoz C, Moral-Munoz JA. Virtual and augmented reality versus traditional methods for teaching physiotherapy: a systematic review. Eur J Investig Health Psychol Educ 2022 Dec 2;12(12):1780-1792. [doi: 10.3390/ejihpe12120125] [Medline: 36547026]
38. Dezza IC, Schulz E, Wu CM. The Drive for Knowledge: The Science of Human Information Seeking: Cambridge University Press; 2022. [doi: 10.1017/9781009026949]
39. Ellington H. Games and Simulations in Science Education: Kogan Page; 1981. URL: https://eric.ed.gov/?id=ED226979 [accessed 2025-06-13]

XSL•FO
RenderX

40. Ekin CC, Polat E, Hopcan S. Drawing the big picture of games in education: a topic modeling-based review of past 55 years. Comput Educ 2023 Mar;194:104700. [doi: 10.1016/j.compedu.2022.104700]
41. Eriksson E, Baykal GE, Torgersson O. The role of learning theory in child-computer interaction - a semi-systematic literature review. : ACM Presented at: IDC '22: Proceedings of the 21st Annual ACM Interaction Design and Children Conference; Jun 27, 2022; Braga Portugal p. 50-68. [doi: 10.1145/3501712.3529728]
42. Gruber R. Osteoimmunology: inflammatory osteolysis and regeneration of the alveolar bone. J Clin Periodontol 2019 Jun;46 Suppl 21:52-69. [doi: 10.1111/jcpe.13056] [Medline: 30623453]
43. Terheyden H, Lang NP, Bierbaum S, Stadlinger B. Osseointegration – communication of cells. Clin Oral Implants Res 2012 Oct;23(10):1127-1135. [doi: 10.1111/j.1600-0501.2011.02327.x] [Medline: 22092345]
44. Gruber R, Stadlinger B, Terheyden H. Cell-to-cell communication in guided bone regeneration: molecular and cellular mechanisms. Clin Oral Implants Res 2017 Sep;28(9):1139-1146. [doi: 10.1111/clr.12929] [Medline: 27550738]
45. Stadlinger B, Jepsen S, Chapple I, Sanz M, Terheyden H. Technology-enhanced learning: a role for video animation. Br Dent J 2021 Jan;230(2):93-96. [doi: 10.1038/s41415-020-2588-1] [Medline: 33483667]
46. Winn B, Heeter C. Resolving conflicts in educational game design through playtesting. Innovate 2006 Dec 1:6.
47. Schade C, Stagge A. Paper prototyping as a method for the evaluation of serious game concepts. Presented at: European Conference on Games Based Learning; Reading; Sep 24-25, 2020; Brighton, UK. [doi: 10.34190/GBL.20.089]
48. Bunt L, Greeff J, Taylor E. Enhancing serious game design: expert-reviewed, stakeholder-centered framework. JMIR Serious Games 2024 May 31;12:e48099. [doi: 10.2196/48099] [Medline: 38820585]
49. Grübel J, Thrash T, Aguilar L, et al. The hitchhiker's guide to fused twins: a review of access to digital twins in situ in smart cities. Remote Sens (Basel) 2022;14(13):3095. [doi: 10.3390/rs14133095]
50. Chatain J, Kapur M, Sumner RW. Three perspectives on embodied learning in virtual reality: opportunities for interaction design. : Association for Computing Machinery Presented at: CHI EA '23: Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems; Apr 19, 2023; Hamburg Germany p. 1-8. [doi: 10.1145/3544549.3585805]
51. Metzmacher H, Rothnie F, Fayolle V, Sumner B, Zünd F. Using augmented reality games to communicate topics in artificial intelligence. Presented at: 54th International Simulation and Gaming Conference (ISAGA 2023) ISAGA; Jul 4-7, 2023; La Rochelle, France p. 282-291 URL: https://www.research-collection.ethz.ch/handle/20.500.11850/634829
52. Chen J. Flow in games (and everything else). Commun ACM 2007 Apr;50(4):31-34. [doi: 10.1145/1232743.1232769]
53. Laugwitz B, Held T, Schrepp M. Construction and evaluation of a user experience questionnaire. Presented at: HCI and Usability for Education and Work, 4th Symposium of the Workgroup Human-Computer Interaction and Usability Engineering of the Austrian Computer Society; Nov 20-21, 2008; Graz, Austria p. 63-76. [doi: 10.1007/978-3-540-89350-9_6]
54. Georgiou Y, Kyza EA. The development and validation of the ARI questionnaire: an instrument for measuring immersion in location-based augmented reality settings. Int J Hum Comput Stud 2017 Feb;98:24-37. [doi: 10.1016/j.ijhcs.2016.09.014]
55. Schrepp M. A comparison of SUS, UMUX-LITE, and UEQ-S. uxpajournal.org. URL: https://uxpajournal.org/wp-content/uploads/sites/7/pdf/JUX_Schrepp_Feb2023_updated.pdf [accessed 2024-05-13]
56. Schrepp M, Hinderks A, Thomaschewski J. Design and evaluation of a short version of the user experience questionnaire (UEQ-S). IJIMAI 2017;4(6):103. [doi: 10.9781/ijimai.2017.09.001]
57. Jacevic M. How the players get their spots: a study of playstyle emergence in digital games. Presented at: 2021 IEEE Conference on Games (CoG); Aug 17-20, 2021; Copenhagen, Denmark. [doi: 10.1109/CoG52621.2021.9619067]
58. Loorbach N, Peters O, Karreman J, Steehouder M. Validation of the instructional materials motivation survey (IMMS) in a self‐directed instructional setting aimed at working with technology. Brit J Educational Tech 2015 Jan;46(1):204-218. [doi: 10.1111/bjet.12138]
59. Keller JM. Development and use of the ARCS model of instructional design. J Instruct Dev 1987(10):2-10. [doi: 10.1007/BF02905780]
60. Perkins DN, Salomon G. Transfer of learning. In: Husén T, Postlethwaite TN, editors. International Encyclopedia of Education, 2nd edition: Pergamon; 1992:425-441.
61. Rau MA. Conditions for the effectiveness of multiple visual representations in enhancing STEM learning. Educ Psychol Rev 2017 Dec;29(4):717-761. [doi: 10.1007/s10648-016-9365-3]
62. Kashdan TB, Gallagher MW, Silvia PJ, et al. The curiosity and exploration inventory-II: development, factor structure, and psychometrics. J Res Pers 2009 Dec 1;43(6):987-998. [doi: 10.1016/j.jrp.2009.04.011] [Medline: 20160913]
63. Hainmueller J, Mummolo J, Xu Y. How much should we trust estimates from multiplicative interaction models? simple tools to improve empirical practice. Polit Anal 2019 Apr;27(2):163-192. [doi: 10.1017/pan.2018.46]
64. Arel-Bundock V. marginaleffects: predictions, comparisons, slopes, marginal means, and hypothesis tests. CRAN. URL: https://cran.r-project.org/web/packages/marginaleffects/index.html [accessed 2023-04-10]
65. Aiken LS, West SG, Reno RR. Multiple Regression: Testing and Interpreting Interactions: SAGE; 1991.
66. Seery K, Barreda A, Hein S, Hiller J. Retention strategies for online students: a systematic literature review. JGER 2021 Jun;5(1):72-84 [FREE Full text] [doi: 10.5038/2577-509X.5.1.1105]
67. Roberts ME, Stewart BM, stm TD. An R package for structural topic models. J Stat Softw 2019 Oct 31;91:1-40. [doi: 10.18637/jss.v091.i02]
68. Krippendorff K. Content Analysis: An Introduction to Its Methodology: SAGE Publications; 2018.

XSL•FO
RenderX

69. Zapf A, Castell S, Morawietz L, Karch A. Measuring inter-rater reliability for nominal data - which coefficients and confidence intervals are appropriate? BMC Med Res Methodol 2016 Aug 5;16(1):93. [doi: 10.1186/s12874-016-0200-9] [Medline: 27495131]
70. Wilujeng I. Increasing the attention, relevance, confidence and satisfaction (ARCS) of students through interactive science learning multimedia. Research in Learning Technology 2021;29. [doi: 10.25304/rlt.v29.2383]
71. Chatain J, Varga R, Fayolle V, Kapur M, Sumner RW. Grounding graph theory in embodied concreteness with virtual reality. : Association for Computing Machinery Presented at: TEI '23: Proceedings of the Seventeenth International Conference on Tangible, Embedded, and Embodied Interaction; Feb 26, 2023; Warsaw Poland p. 1-13. [doi: 10.1145/3569009.3572733]
72. Schinazi VR, Meloni D, Grübel J, et al. Motivation moderates gender differences in navigation performance. Sci Rep 2023 Sep 25;13(1):15995. [doi: 10.1038/s41598-023-43241-4] [Medline: 37749312]
73. Kuittinen J, Kultima A, Niemelä J, Paavilainen J. Casual games discussion. : Association for Computing Machinery Presented at: Future Play '07: Proceedings of the 2007 conference on Future Play; Nov 14, 2007; Toronto, Canada p. 105-112. [doi: 10.1145/1328202.1328221]
74. O'Brien SP. Select the Type of Experience You Would Like to Have": exploring player roles and role affordance in video games. paj: J Initiative Digital Humanit 2015 [FREE Full text]
75. Grübel J, Thrash T, Hölscher C, Schinazi VR. Evaluation of a conceptual framework for predicting navigation performance in virtual reality. PLoS ONE 2017;12(9):e0184682. [doi: 10.1371/journal.pone.0184682] [Medline: 28915266]
76. Colombo G, Grübel J. The spatial performance assessment for cognitive evaluation (SPACE): a novel game for the early detection of cognitive impairment. : Association for Computing Machinery Presented at: CHI EA '23: Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems; Apr 19, 2023; Hamburg Germany p. 1-6. [doi: 10.1145/3544549.3583828]
77. Grübel J. Computer-aided experimentation for human behaviour analysis [Dissertation]. . 2022. [doi: 10.3929/ethz-b-000587537]
78. Grübel J. The design, experiment, analyse, and reproduce principle for experimentation in virtual reality. Front Virtual Real 2023;4. [doi: 10.3389/frvir.2023.1069423]
79. Aguilar L, Gath-Morad M, Grübel J, et al. Experiments as code and its application to VR studies in human-building interaction. Sci Rep 2024 Apr 30;14(1):9883. [doi: 10.1038/s41598-024-60791-3] [Medline: 38688980]
80. Xu M, Luo Y, Zhang Y, Xia R, Qian H, Zou X. Game-based learning in medical education. Front Public Health 2023;11(1113682):1113682. [doi: 10.3389/fpubh.2023.1113682] [Medline: 36935696]
81. Moreno-Ger P, Torrente J, Bustamante J, Fernández-Galaz C, Fernández-Manjón B, Comas-Rengifo MD. Application of a low-cost web-based simulation to improve students' practical skills in medical education. Int J Med Inform 2010 Jun;79(6):459-467. [doi: 10.1016/j.ijmedinf.2010.01.017] [Medline: 20347383]
82. Huber J, Witti M, Schunk M, Fischer MR, Tolks D. The use of the online inverted classroom model for digital teaching with gamification in medical studies. GMS J Med Educ 2021;38(1):Doc3. [doi: 10.3205/zma001399] [Medline: 33659608]
83. Sulong A, Ibrahim AB, Abas A, Bakar AZA. Incorpating gamification in a flipped classroom approach: a review of literature. JPB 2021(14):22-32 [FREE Full text] [doi: 10.37134/bitara.vol14.sp.3.2021]
84. Jung EY, Park DK, Lee YH, Jo HS, Lim YS, Park RW. Evaluation of practical exercises using an intravenous simulator incorporating virtual reality and haptics device technologies. Nurse Educ Today 2012 May;32(4):458-463. [doi: 10.1016/j.nedt.2011.05.012] [Medline: 21664014]
85. Hashimoto DA, Petrusa E, Phitayakorn R, Valle C, Casey B, Gee D. A proficiency-based virtual reality endoscopy curriculum improves performance on the fundamentals of endoscopic surgery examination. Surg Endosc 2018 Mar;32(3):1397-1404. [doi: 10.1007/s00464-017-5821-5] [Medline: 28812161]
86. Yoganathan S, Finch DA, Parkin E, Pollard J. 360° virtual reality video for the acquisition of knot tying skills: a randomised controlled trial. Int J Surg 2018 Jun;54(Pt A):24-27. [doi: 10.1016/j.ijsu.2018.04.002] [Medline: 29649669]
87. Maytin M, Daily TP, Carillo RG. Virtual reality lead extraction as a method for training new physicians: a pilot study. Pacing Clin Electrophysiol 2015 Mar;38(3):319-325. [doi: 10.1111/pace.12546] [Medline: 25494952]
88. Park J, MacRae H, Musselman LJ, et al. Randomized controlled trial of virtual reality simulator training: transfer to live patients. Am J Surg 2007 Aug;194(2):205-211. [doi: 10.1016/j.amjsurg.2006.11.032] [Medline: 17618805]
89. Zhao G, Fan M, Yuan Y, Zhao F, Huang H. The comparison of teaching efficiency between virtual reality and traditional education in medical education: a systematic review and meta-analysis. Ann Transl Med 2021 Feb;9(3):252. [doi: 10.21037/atm-20-2785] [Medline: 33708879]
90. Weeks JK, Pakpoor J, Park BJ, et al. Harnessing augmented reality and CT to teach first-year medical students head and neck anatomy. Acad Radiol 2021 Jun;28(6):871-876. [doi: 10.1016/j.acra.2020.07.008] [Medline: 32828663]

Abbreviations

AR: augmented reality
ARCS: Attention, Relevance, Confidence, and Satisfaction

XSL•FO
RenderX

ARI: augmented reality immersion
CEI-II: Curiosity and Exploratory Inventory-II
RANK: Receptor Activator of NF-κB
RANKL: Receptor Activator of NF-κB Ligand
VR: virtual reality

Edited by A McDaniel; submitted 30.07.24; peer-reviewed by SE Amrahov, X Cheng; revised version received 14.02.25; accepted 07.03.25; published 16.06.25.

Please cite as:
Grübel J, Chatain J, Schmid C, Fayolle V, Zünd F, Gruber R, Stadlinger B
Applying Augmented Reality to Convey Medical Knowledge on Osteoclasts to Users of a Serious Game: Vignette Experiment
JMIR Serious Games 2025;13:e64751
URL: https://games.jmir.org/2025/1/e64751
doi:10.2196/64751

© Jascha Grübel, Julia Chatain, Claudio Schmid, Violaine Fayolle, Fabio Zünd, Reinhard Gruber, Bernd Stadlinger. Originally published in JMIR Serious Games (https://games.jmir.org), 16.6.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Development and Evaluation of a Case-Based Serious Game for Diagnosis and Treatment Planning in Orthodontic Education: Quasi-Experimental Study

Rochaya Chintavalakorn[1], DDS, PhD; Chayuth Chanwanichkulchai[2]; Napat Buranasing[2]; Napatsaporn Parivisutt[2]; Naruchol Patchasri[2]; Kawin Sipiyaruk[1], DDS, PhD

[1]Department of Orthodontics, Faculty of Dentistry, Mahidol University, 6 Yothi Road, Ratchathewi, Bangkok, Thailand

[2]Doctor of Dental Surgery Program, Faculty of Dentistry, Mahidol University, Bangkok, Thailand

Corresponding Author:
Kawin Sipiyaruk, DDS, PhD
Department of Orthodontics, Faculty of Dentistry, Mahidol University, 6 Yothi Road, Ratchathewi, Bangkok, Thailand

Abstract

Background: Orthodontic education requires effective training in diagnosis and treatment planning, but traditional teaching methods may lack engagement and opportunities in offering a safe learning environment. Serious games are gaining momentum in dental education due to their positive educational impact in enhancing learner knowledge and motivation. However, their application in orthodontic diagnosis and treatment planning training remains unexplored.

Objective: The aim of this study was to develop and evaluate a simulation-based serious game for training orthodontic diagnosis and treatment planning in virtual patients (OrthoVirt), examining its impact on student knowledge and satisfaction. This study also explored whether prior gaming experience influenced learning outcomes.

Methods: A quasi-experimental study was conducted with 32 fourth-year dental students, who were requested to complete a preknowledge assessment, 3 simulated patients within OrthoVirt, a postknowledge assessment, and a satisfaction survey. Participants were categorized as gamers (n=16) or nongamers (n=16) based on self-reported weekly gaming time. The primary outcome was knowledge improvement, analyzed using 2-tailed paired *t* tests (Cohen *dz*). Group comparisons were conducted using 2-tailed independent *t* tests (Cohen *d*). User satisfaction was measured using a validated questionnaire based on the technology acceptance model. A stricter significance threshold ($P<.01$) and effect size metrics (Cohen *d* and Cohen *dz*) were used to account for the small sample size, multiple comparisons, and exploratory nature of the study.

Results: Both gamer and nongamer groups showed significant knowledge improvement after using OrthoVirt (mean score increased from 10.75 (SD 2.75) to 14.75 (SD 1.81) out of 20; $P<.001$). The mean scores of the gamer group increased from 10.31 (SD 3.07) to 15.19 (SD 1.83) while those of the nongamer group rose from 11.19 (SD 2.40) to 14.31 (SD 1.74). No statistically significant differences were found between groups in pre- and postknowledge assessments as well as improvement scores ($P>.01$), suggesting that the educational benefit was consistent regardless of gaming background. Participants from both groups rated OrthoVirt positively, particularly for “perceived ease of use.” However, “perceived enjoyment” was rated slightly lower than other aspects, with nongamers scoring it 3.60 (SD 0.81) and gamers 3.45 (SD 0.73) out of 5, indicating a potential area for design enhancement. Overall satisfaction ratings were similar between the 2 groups ($P>.01$).

Conclusions: OrthoVirt demonstrated potential as a supplementary tool for diagnosis and treatment planning in orthodontic education, with statistically significant improvements observed in learner knowledge. While feedback was generally positive, these findings should be interpreted with caution due to the quasi-experimental design and small sample size. Future development should focus on improving user enjoyment and engagement through entertaining design elements. Further research should explore how OrthoVirt can be integrated as a case discussion tool alongside lectures, with the potential to enhance learning not only in orthodontic education but also across other areas of dental training.

(JMIR Serious Games 2025;13:e73956) doi:10.2196/73956

KEYWORDS

dental education; gamification; orthodontics; serious game; simulation

XSL•FO
RenderX

Introduction

In orthodontic practice, diagnosis and treatment planning play a pivotal role in providing patient care and treatment outcomes. This competency is considered an essential component of the expected learning outcomes required for orthodontic practice [1]. Accuracy in diagnosis and treatment planning necessitates a thorough consideration of comprehensive patient information, encompassing various orthodontic needs, such as dental alignment, occlusion, skeletal relationships, and soft tissue considerations [2,3]. To accomplish this, orthodontists rely on the examination and analysis of diagnostic records, including radiographs, photographs, and dental impressions [4,5]. Not only current dental issues but also potential future growth and development are important to ensure optimal results and long-lasting stability. Therefore, effective delivery methods are necessary to equip dental professionals with this competence.

A variety of teaching and learning approaches are important in improving the complex skills required for diagnosis and treatment planning in orthodontics [1]. While traditional lectures provide basic knowledge, case-based learning offers invaluable opportunities for students to apply theoretical concepts to real-world scenarios, improving their critical thinking and decision-making abilities [6-8]. Integrating technology-enhanced learning tools, such as simulations and serious games, further enriches the educational experience, allowing students to explore immersive case scenarios and receive immediate feedback [9-12]. Using a combination of instructional methods can help students develop the full range of skills needed for effective orthodontic diagnosis and treatment planning. However, many current instructional methods tend to teach diagnostic skills separately, without providing a cohesive, interactive setting that mirrors the complexity of real-life orthodontic evaluation. This highlights the need for a simulation-based tool, designed specifically for orthodontic education, that can guide students through the step-by-step reasoning process involved in accurate diagnosis and treatment planning.

Serious games have currently gained momentum in dental education due to their positive educational impact across several aspects [13-16]. By immersing learners in realistic scenarios, serious games facilitate experiential learning, allowing students to apply theoretical knowledge to practical situations while refining clinical decision-making skills [17-19]. Furthermore, the gamified elements inherent in these educational tools can significantly enhance motivation and engagement among students, as they are incentivized to progress through challenges and receive immediate feedback [20,21]. In addition, serious games provide a safe learning environment where students can learn from their mistakes without real-world consequences [13,22,23]. As a result, serious games emerge as a promising interactive learning platform, providing dental learners with immersive experiences to practice orthodontic diagnosis and treatment planning. Despite these advantages, most serious games developed for dental education have focused on foundational sciences, communication training, or procedural skills. Few have been designed to support the clinical reasoning process in orthodontics, which requires multiple sources of diagnostic data, such as patient interviews, cephalometric analysis, and dental models, that are essential for accurate orthodontic diagnosis and treatment planning.

While simulation-based serious games have shown promise in dental education, few have addressed the unique complexity of orthodontic diagnosis and treatment planning, which demands the interpretation of multidimensional clinical records and synthesis of skeletal, dental, and soft tissue information [12,24]. Integrating serious games into this context offers undergraduate dental learners the opportunity to enhance critical thinking skills and apply theoretical knowledge in simulated clinical scenarios [25]. However, the effective adoption of serious games in education may not be uniform across all learners [20,26]. As serious games often incorporate interactive features and mechanics similar to those found in video games, prior gaming experience may influence how students engage with and benefit from these educational tools [27-29]. Learners with higher levels of gaming familiarity may find serious games more engaging, whereas those with less experience might encounter greater cognitive load or usability challenges, potentially impacting their learning outcomes.

Given these considerations, it is important to determine whether the educational benefits of serious games are equally practical to learners with diverse gaming experiences in dental education. Consequently, to serve as a supplementary tool rather than a replacement for conventional orthodontic education, this research aimed to design and evaluate a serious game for training diagnosis and treatment planning in orthodontic practice, comparing its educational impact between gaming and nongaming users. The evidence retrieved from this research may provide valuable insights into the potential benefits and practical considerations of integrating serious games, paving the way for future innovations in teaching and learning in orthodontic education.

Methods

Game Development

Game Concept

A simulation-based serious game for training orthodontic diagnosis and treatment planning in virtual patients (OrthoVirt) was designed and developed using “React,” an open-source JavaScript framework. Within the game, learners were allowed to interact with simulated patients with orthodontic concerns. The development of OrthoVirt was driven by Kolb’s [30] Experiential Learning Theory, which frames learning as a continuous cycle of doing, reflecting, thinking, and applying. The structure of the game reflects the clinical process in orthodontic practice, from gathering patient information to formulating a diagnosis and treatment plan. By working through these stages, learners can gain experience through active learning in a simulated environment. As they receive feedback and revisit their decisions, learners are encouraged to reflect on their thought processes and refine their clinical judgment, helping them improve competencies in diagnosis and treatment planning.

XSL•FO
RenderX

Instructional Design and Learning Tasks

The game login requires a user account generated from a secure back-office system by administrators. Following a log-in process, learners could opt to proceed with the game with 1 of the 3 simulated patients with different orthodontic problems, including deep bite, crossbite, and open bite (learning scenarios). Each scenario began with the patient's chief complaint and involved 5 tasks, which were Task 1: Patient interview, Task 2: Intraoral and extraoral examination, Task 3: Model analysis, Task 4: Radiographic examination, and Task 5: Diagnosis and treatment planning.

The game tasks were designed using a structured and interactive instructional approach to enhance learner competency in orthodontic diagnosis and treatment planning through realistic patient-case scenarios. Learners begin with Task 1, which involves interviewing a simulated patient to understand the chief complaint and gather general information. Completion of this task unlocks Task 2, where students conduct intraoral and extraoral examinations based on standardized visuals of the patient. Once these examinations are completed, students gain access to Tasks 3 and 4, which can be approached in either order. Task 3 presents high-resolution images of dental models, enabling learners to evaluate occlusal relationships, overjet, overbite, and arch form. Task 4 provides cephalometric radiographs accompanied by premeasured angular and linear values. Importantly, while these measurements are provided, learners are expected to interpret them independently, promoting critical thinking and reinforcing the kind of clinical judgment expected in orthodontic practice. As learners progress through each diagnostic step, a dynamic information box compiles key findings, allowing for easy review and synthesis. In the final stage, Task 5, learners draw on all prior assessments to formulate a comprehensive orthodontic diagnosis and propose an appropriate treatment plan. This sequential process mirrors the structure of authentic clinical workflows and encourages deep engagement with each component of the diagnostic pathway.

Game Interface

As presented in Figure 1, the game interface guides students through a structured diagnostic workflow that mirrors clinical practice. The first screen introduces the simulated patient and provides access to key diagnostic stages aligned with the game's intended learning outcomes. In the radiographic examination task, learners apply their knowledge of diagnostic imaging to select appropriate radiographic views, reinforcing concepts taught in preclinical orthodontic courses. The final screen presents a treatment planning scenario, where students must integrate diagnostic findings and theoretical understanding to select suitable appliances. These interactive steps are designed to promote clinical reasoning and strengthen foundational knowledge in orthodontic diagnosis and treatment planning.

Figure 1. Visual representations of the game interface within OrthoVirt.



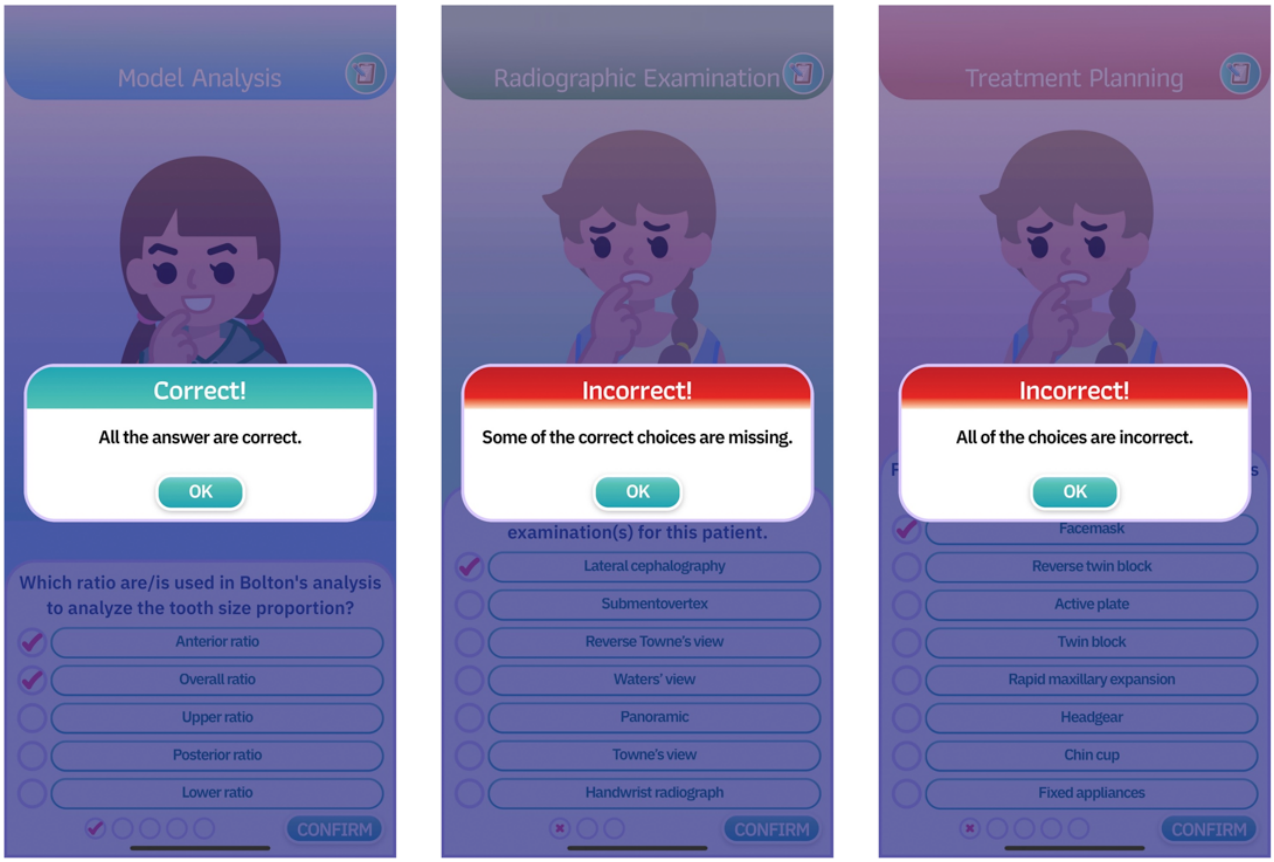
Feedback System

The game incorporated a feedback mechanism that provided learners with immediate responses in diverse formats following their interactions. Notably, the game offered informative feedback explaining the clinical rationale behind each question, regardless of whether the response was correct. In addition,

XSL•FO
RenderX

clues were provided to learners in instances where incorrect responses were selected. Furthermore, the game deducted scores and reflected them through the facial expressions of the simulated patients when learners chose incorrect answers. As shown in Figure 2, the feedback system differentiated between levels of error. For example, it indicated whether only some answers were incorrect or if all choices were wrong, thereby tailoring the guidance to the nature of the mistake. These feedback elements were designed to support learner reflection and maintain engagement throughout the learning process.

Figure 2. Examples of feedback formats provided within the game.



Research Design

This study used a quasi-experimental research design to assess the impact of OrthoVirt in orthodontic education, comparing gaming and nongaming users. The participants voluntarily involved fourth-year dental undergraduates from the Faculty of Dentistry, Mahidol University. Participants initially were assigned to complete a preknowledge assessment. Following a 7-day interval, they engaged with OrthoVirt, completing 3 orthodontic scenarios within the game. Immediately after, participants were required to complete a postknowledge assessment to measure knowledge acquisitions. In addition, participants filled out a survey regarding their experience with OrthoVirt. To maintain data integrity, students were instructed to work independently on all tasks without peer discussion.

Participants and Recruitment

The research participants included fourth-year dental undergraduates at the Faculty of Dentistry, Mahidol University. However, participants who had not passed the course “Diagnosis and Treatment Planning in Orthodontics" during the 2023 academic year were excluded from the research. This was to ensure a consistent baseline of relevant orthodontic knowledge and because the game was designed as a supplementary tool to reinforce content from that course. The participants were categorized into 2 groups: gaming users, who played video games for at least 1 hour per day (7 h/wk), and nongaming users, who played for less than this duration. Participants were recruited for this research using voluntary sampling. This nonprobability sampling technique was considered appropriate for the study, despite potential concerns about nonresponse bias [31]. Participants voluntarily engaged in the learning activity, which led to accurate and reliable research findings with a low possibility of dropout [32].

To determine the appropriate number of participants, a statistical calculation was performed using a method suitable for comparing paired datasets, aligning with the primary objective of assessing knowledge acquisition following OrthoVirt use [33]. The calculation assumed a large effect size (Cohen dz=1.425), based on a previous study [26]. The calculation indicated that to detect a significant difference in scores between the 2 knowledge assessments, with 95% power and a 1% significance threshold, each group required at least 13 participants. Anticipating potential participant attrition, this number was adjusted upward by 20%, resulting in a target of 16 participants per group. Consequently, this study recruited 32 dental students, comprising 16 gaming users and 16 nongaming users. A 1% significance level (P<.01) was adopted to reduce the risk of Type I error, considering the relatively

XSL•FO
RenderX

small sample size, multiple comparisons, and the exploratory nature of the study.

Outcome Measurements

Knowledge Assessments

Participants were required to complete pre- and postknowledge assessments to assess knowledge improvement. The initial development of the knowledge assessment was performed by a research team member with expertise in orthodontics (RC). The assessment was reviewed by 3 experts in orthodontics who were not involved in the study to ensure the content validity and its alignment with the expected learning outcomes of the game. Each assessment included 20 multiple-choice questions. The content reflected a wide range of cognitive skills relevant to orthodontic diagnosis and treatment planning. These included dental and skeletal classification, growth assessment, cephalometric interpretation, space analysis, and treatment selection. Questions were designed to test not only factual knowledge but also clinical reasoning. Specifically, the assessment items spanned multiple levels of Bloom's cognitive domain, ranging from "Remembering and Understanding" to "Applying and Analyzing." Several questions required learners to interpret diagnostic images or synthesize measurements to draw clinical conclusions, thereby promoting higher-order thinking aligned with real-world orthodontic decision-making. Examples of assessment items and their corresponding Bloom's cognitive domains are provided in Multimedia Appendix 1. To reduce potential memory effects, both question order and answer choices were randomized between the 2 assessments.

User Satisfaction

The paper-based questionnaire was constructed to gather user satisfaction based on the technology-acceptance model [34], with questions adapted from existing literature on serious games [25,35,36]. Three experts in health care education were requested to validate the questionnaire using content validity, with items iteratively refined until each achieved an item-objective congruence index above 0.5. Based on the internal consistency reliability assessment, items with low performance were removed until all constructs achieved Cronbach α coefficients above 0.7 (Perceived usefulness: 0.714, Perceived ease of use: 0.727, and Perceived enjoyment: 0.803). The validated questionnaire consisted of 18 five-point Likert scales (Multimedia Appendix 2), covering "perceived usefulness (6 items)," "perceived ease of use (6 items)" and "perceived enjoyment (6 items)," with responses ranging from "Strongly disagree" (1) to "Strongly agree" (5).

Data Analysis

The research data were analyzed using the IBM SPSS Statistics (SPSS version 29, IBM Corp). The distribution of the data was approximately normal, and the assumption of equal variance for the independent *t* test was met, supporting the use of parametric statistical analyses. Descriptive statistics were used to summarize the data. Cognitive improvement was assessed using a paired *t* test, while a comparison of knowledge acquisition between the gaming and nongaming groups was made using an independent *t* test. The threshold for statistical significance was set at $P<.01$. In addition, effect sizes were reported using Cohen *d* for independent *t* tests and Cohen *dz* for paired *t* tests to quantify the magnitude of observed effects.

Ethical Considerations

This research protocol was reviewed and approved by the Institutional Review Board of Faculty of Dentistry/Faculty of Pharmacy, Mahidol University (the certificate of approval number: MU-DT/PY-IRB 2024/012.0103). All methods were performed in accordance with the relevant guidelines and regulations. Informed consent was obtained from all participants prior to their involvement in the study. They were clearly informed about the research purpose, procedures, and their rights, including the right to withdraw at any time. To protect participants' privacy and confidentiality, all collected data were anonymized prior to analysis. Participants did not receive financial compensation or material incentives for their participation. However, they were provided with an opportunity to learn from the OrthoVirt game, which was positioned as an educational activity to support their academic development. No identifiable images of participants are included in the paper or supplementary materials.

Results

Research Participants

A total of 32 dental students participated in the study, comprising 26 females and 6 males. The participants reported a mean gaming time of 9.59 (SD 13.61) hours per week, with a range of 0 - 63 hours. Among the participants, 16 were classified as gaming users, playing an average of 17.63 (SD 15.56; range 7 - 63) hours per week. The remaining 16 were categorized as nongaming users, with a mean gaming time of 1.56 (SD 1.71; range: 0 - 5) hours per week. The gamer group included 4 male and 12 female participants, while the nongamer group included 2 male and 14 female participants. All participants were enrolled in the same academic year (fourth year) and were therefore presumed to be of similar age.

Knowledge Acquisition

According to paired *t* test analysis, both gamer and nongamer groups exhibited statistically significant improvements in their assessment scores ($P<.001$; Cohen $dz=1.65$), as shown in Table 1. The gamer group demonstrated an increase in mean score from 10.31 (SD 3.07) to 15.19 (SD 1.83), with a large effect size (Cohen $dz=1.82$). Similarly, the nongamer group showed improvement, with mean scores rising from 11.19 (SD 2.40) to 14.31 (SD 1.74), yielding a large effect size (Cohen $dz=1.45$). These results indicate substantial performance enhancement in both groups following the intervention.

XSL•FO
RenderX

Table . Analysis of pre- and postknowledge assessment scores.

Groups	Preknowledge assessment[a] Mean (SD)	Postknowledge assessment[a] Mean (SD)	Mean difference (95% CI)	*P* value[b]
Gamer group	10.31 (3.07)	15.19 (1.83)	4.88 (3.23-6.52)	<.001
Nongamer group	11.19 (2.40)	14.31 (1.74)	3.12 (1.53-4.72)	<.001
Total	10.75 (2.75)	14.75 (1.81)	4.00 (2.87-5.12)	<.001

[a]The maximum possible score for both knowledge assessments was 20.

[b]All *P* values are 2-tailed, with significance set at *P*<.01.

Comparison of Knowledge Acquisitions Between Groups

An independent *t* test was used to compare knowledge acquisition between the gamer and nongamer groups (Table 2). No statistically significant differences were observed in preknowledge assessment, postknowledge assessment, or score improvement between the 2 groups (*P*>.01 for all comparisons). The corresponding effect sizes were small to moderate, that is, Cohen *d*=0.32 for preknowledge assessment, Cohen *d*=0.49 for postknowledge assessment, and Cohen *d*=0.57 for score improvement. These results suggest that gaming experience did not have a significant impact on knowledge gains from the intervention.

Table . Comparison of knowledge acquisitions between gamer and nongamer groups.

Assessments[a]	Gamer group Mean (SD)	Nongamer group Mean (SD)	Mean difference (95% CI)	*P* value[b]
Preknowledge assessment	10.31 (3.07)	11.19 (2.40)	0.88 (–2.87 to 1.12)	.38
Postknowledge assessment	15.19 (1.83)	14.31(1.74)	–0.88 (–0.42 to 2.17)	.18
Score improvement	4.88 (3.10)	3.13 (2.99)	–1.75 (–0.45 to 3.95)	.11

[a]The maximum possible score for both knowledge assessments was 20.

[b]All *P* values are 2-tailed, with significance set at *P*<.01.

User Satisfaction

Both gamer and nongamer groups rated OrthoVirt favorably across all assessed aspects (Table 3). An independent *t* test showed no statistically significant differences in satisfaction scores between the 2 groups across all aspects (*P*>.01 for all comparisons). The highest-rated aspect in both groups was “perceived ease of use,” with mean scores of 4.07 (SD 0.54) and 4.06 (SD 0.61) out of 5 for nongamer and gamer groups, respectively (Cohen *d*=0.02). Perceived usefulness was also rated highly, with scores of 3.93 (SD 0.42) for nongamers and 3.98 (SD 0.54) for gamers (Cohen *d*=0.10). In contrast, “perceived enjoyment” received slightly lower ratings compared to other aspects, with nongamers scoring it 3.60 (SD 0.81) and gamers 3.45 (SD 0.73; Cohen *d*=0.10).

Table . Comparison of satisfaction scores between gamer and non-gamer groups.

Aspects[a]	Gamer group Mean (SD)	Nongamer group Mean (SD)	Mean difference (95% CI)	*P* value[b]
Perceived usefulness	3.98 (0.54)	3.93 (0.42)	–0.05 (–0.30 to 0.40)	.77
Perceived ease of use	4.06 (0.61)	4.07 (0.54)	0.01 (–0.42 to 0.41)	.98
Perceived enjoyment	3.45 (0.73)	3.60 (0.81)	0.15 (–0.71 to 0.41)	.59
Overall	3.83 (0.55)	3.86 (0.52)	0.03 (–0.42 to 0.35)	.85

[a]User satisfaction was measured on a five-point Likert scale, ranging from “Strongly disagree” (1) to “Strongly agree” (5).

[b]All *P* values are 2-tailed, with significance set at *P*<.01.

Discussion

Knowledge Gains and Educational Value

The implementation of simulation-based serious games in orthodontic education has shown promising results in cognitive improvement. This research demonstrated significant knowledge improvements after engaging with OrthoVirt, aligning with previous research highlighting the efficacy of serious games as educational tools [20,22,25,37-40]. The observed knowledge improvement, evidenced by increased assessment scores, can be attributed to the game’s dynamic learning environment, which allows students to learn from mistakes without real-world consequences [14,35,41]. While other games have demonstrated knowledge improvements in other areas, OrthoVirt is designed specifically for orthodontic education, incorporating discipline-specific data, such as model analysis and cephalometric interpretation. In contrast to studies comparing

XSL•FO
RenderX

serious games with traditional instruction, this study adopted a quasi-experimental design comparing gamer and nongamer groups to assess the game's applicability across varied learner profiles. Satisfaction results further support the educational value of the game. These results suggest that OrthoVirt represents a valuable addition to traditional orthodontic education methods, offering an engaging and effective platform for knowledge acquisition in orthodontic diagnosis and treatment planning. In addition, its core design may be adaptable for use in other areas of dental education that emphasize diagnostic reasoning and clinical decision-making.

The absence of significant differences in knowledge gain between gamer and nongamer groups suggests that the educational benefits of OrthoVirt are accessible to a diverse student population, regardless of prior gaming experience. This finding is particularly important as it indicates the broad applicability of the tool, consistent with existing evidence [42,43]. However, it is worth considering that gamers might experience lower stress levels compared to nongamers, potentially due to their ability to relieve stress through gaming and their higher tolerance for electronic devices [44,45]. Furthermore, the intrinsic motivation generated by games may inspire students in the gamer group to engage more deeply with the learning process, which could explain why this group tends to show greater improvement after using OrthoVirt [46]. However, these factors were not directly measured in this study; thus, while such explanations are theoretically plausible, they remain speculative. These observations underscore the potential of serious games to enhance knowledge and positively influence the learning experience across diverse student profiles in dental education.

User Experience and Satisfaction

The user experience and satisfaction with OrthoVirt were evaluated, with a particular focus on its ease of use, yielding insights into the game's effectiveness and areas for potential improvement. The satisfaction level regarding ease of use was rated highest by both gamer and nongamer groups, suggesting that users found the interface intuitive and user-friendly, which is crucial for educational software adoption, including serious games [47,48]. This ease of use is a critical factor in serious games, as it reduces cognitive load associated with learning the software itself, allowing users to focus more on the educational content [49]. This reflects the alignments between the interactive functions and pedagogical design dimensions of the evaluation framework of serious games [20]. The volume of information presented in each section could potentially impact the perceived ease of use. This feedback highlights the delicate balance required in designing educational games that are both comprehensive and user-friendly [50,51]. The high ease of use score suggests that OrthoVirt successfully achieved this balance for most users despite these concerns. These findings underscore the importance of user-centered design in educational games, where ease of use plays a vital role in optimizing engagement and, consequently, educational outcomes.

The game functions in OrthoVirt, including interactive activities and feedback mechanisms, aimed at enhancing learner enjoyment and engagement. Interactivity facilitates easier gameplay and provides improved performance feedback, thereby augmenting the overall learning experience for the user [20,25,52]. However, the enjoyment satisfaction ratings were rated notably lower than the ease of use and usefulness. This finding suggests a gap in emotional and motivational engagement, potentially because learners may have compared this serious game to commercial entertainment games, which are primarily designed to maximize immersion and enjoyment. Considering the core difference from entertainment games, serious games are not solely focused on providing enjoyment but are designed to facilitate knowledge acquisition alongside engagement [53,54]. As a result, they may not foster the same level of intrinsic interest in their subject matter [51]. According to the review of Maxim and Arnedo-Moreno's synthesis [55], effective serious game design involves the balance of educational content with motivational elements throughout 4 phases of exploration, design, development, and assessment. Therefore, these findings indicate a need for more integration of motivational and emotional engagement components during the design phase, even as usability and learning objectives were effectively addressed.

With the relatively lower enjoyment scores of OrthoVirt, the prioritization of enhancements that directly target motivational engagement should be recommended. As proposed in the exploration and design phases of the framework [55], understanding learner profiles and collaborative dynamics can support the integration of multiplayer elements or peer consultation tools, promoting social learning and critical thinking. First, incorporating social interaction features, such as competitive features or peer-to-peer interactivity, may promote deeper engagement and potentially reduce learner stress [56,57]. Future iterations could also explore collaborative consultation mechanisms, fostering critical thinking through group problem-solving and brainstorming, which leverages social constructivism principles [58,59]. These enhancements will address the identified enjoyment gap by fostering social and emotional connections. Adaptive features supported by artificial intelligence could personalize the difficulty level and learning path based on user performance, directly addressing user feedback related to engagement and enjoyment [60]. These enhancements could transform OrthoVirt into a more dynamic and engaging learning environment, in which further research could evaluate the impact of these proposed changes on learning engagement in dental education.

Research Limitations

This study had limitations that should be considered. First, voluntary sampling may have introduced selection bias, as students more interested in technology-enhanced learning were more likely to participate. While this approach facilitated recruitment and ensured engagement, as evidenced by zero dropout, it may limit the generalizability of the findings. In addition, the gender imbalance in the sample (predominantly female) may further affect generalizability. Moreover, categorizing participants as gamers or nongamers based on self-reported gaming habits may have led to recall bias and misclassification. In addition, although using identical questions in the pre- and postknowledge assessments ensured consistent difficulty, some memory effect may still have occurred, despite

XSL•FO
RenderX

efforts to mitigate this through shuffling item order. Furthermore, the absence of a control group limits direct comparisons between OrthoVirt and traditional methods. Academic performance data were not collected, which represents an additional limitation of this study. However, since OrthoVirt is designed as a supplementary educational tool rather than a standalone replacement, this limitation does not invalidate the exploratory value of the study. Importantly, as this research used a quasi-experimental design with voluntary participation and a relatively small sample, the ability to draw causal inferences or broadly generalize the findings is limited.

Recommendations for Future Research

Future research should investigate how OrthoVirt can be integrated as a case discussion tool alongside traditional lectures, potentially enhancing the learning experience in orthodontic education. Further studies with larger and more diverse cohorts are required to confirm the findings of this study. In addition, research on the long-term retention of knowledge, as well as the impact of enhanced entertainment and social interaction elements, may inform improvements in learner engagement and broader applicability. Expanding the concept and underlying framework of OrthoVirt to other areas within dental education could also be considered.

Conclusions

This study provides evidence that OrthoVirt can be used as a supplementary tool to support knowledge acquisition in orthodontic diagnosis and treatment planning. Statistically significant improvements in knowledge were observed following use of the game, with similar benefits reported regardless of prior gaming experience. While participants generally provided positive feedback on OrthoVirt, particularly highlighting its usefulness and ease of use, the study revealed opportunities for enhancement, especially in terms of user enjoyment. Future development of OrthoVirt should therefore prioritize improvements in entertainment and social interaction elements to further enhance learner engagement and motivation. These findings may also inform the development and integration of serious games across broader areas of dental education beyond orthodontics.

Acknowledgments

The authors would like to express their sincere gratitude to all participants for their contributions to this research. We also extend our appreciation to the experts who provided insightful suggestions during the validation of the data collection tools. The authors used generative artificial intelligence for the purpose of grammar correction and language refinement and take full responsibility for the final, reviewed content of this manuscript. This research project and the game development were supported by Mahidol University.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request. However, the datasets are not publicly accessible due to privacy considerations and the presence of information that could potentially compromise the confidentiality of research participants. OrthoVirt, the serious game developed for this study, is publicly available for download on the App Store. Access to the game requires a username and password, which can be provided by the corresponding author upon request for academic or non-commercial research purposes. The source code for OrthoVirt is not publicly available but may be shared upon reasonable request, subject to approval by both the corresponding author and the private developer company involved in the game’s development.

Authors' Contributions

RC and KS contributed to the conceptualization, methodology development, validation, project administration, resource management, and manuscript writing (original draft and revisions). RC, CC, NB, NPar, NPat, and KS conducted the investigation, formal analysis, and data curation, and contributed to the discussion.

Conflicts of Interest

The authors declare no conflict of interest. Although OrthoVirt has been designed and developed by the researcher team, its primary purpose is exclusively for academic purposes, with no commercial objectives associated with it.

Multimedia Appendix 1
Examples of assessment items and their corresponding Bloom’s cognitive domains.
[PDF File, 68 KB - games_v13i1e73956_app1.pdf]

Multimedia Appendix 2
Satisfaction questionnaire toward the use of OrthoVirt.
[PDF File, 83 KB - games_v13i1e73956_app2.pdf]

XSL•FO
RenderX

References

1. Ono T, Pangrazio-Kulbersh V, Perillo L, et al. World Federation of Orthodontists guidelines for postgraduate orthodontic education. J World Fed Orthod 2023 Apr;12(2):41-49. [doi: 10.1016/j.ejwf.2023.03.002] [Medline: 36964071]
2. Nanda R, Upadhyay M. Skeletal and dental considerations in orthodontic treatment mechanics: a contemporary view. Eur J Orthod 2013 Oct;35(5):634-643. [doi: 10.1093/ejo/cjs054] [Medline: 24068287]
3. Maaz M, Fida M. Dental, skeletal, and soft-tissue changes in adult orthodontic patients treated with premolar extraction and nonextraction: a cross-sectional study. Am J Orthod Dentofacial Orthop 2022 Sep;162(3):360-366. [doi: 10.1016/j.ajodo.2021.04.026] [Medline: 35570094]
4. Nucera R, Lo Giudice A, Bellocchio M, Spinuzza P, Caprioglio A, Cordasco G. Diagnostic concordance between skeletal cephalometrics, radiograph-based soft-tissue cephalometrics, and photograph-based soft-tissue cephalometrics. Eur J Orthod 2017 Aug 1;39(4):352-357. [doi: 10.1093/ejo/cjw072] [Medline: 27932406]
5. Rischen RJ, Breuning KH, Bronkhorst EM, Kuijpers-Jagtman AM. Records needed for orthodontic diagnosis and treatment planning: a systematic review. PLoS ONE 2013;8(11):e74186. [doi: 10.1371/journal.pone.0074186] [Medline: 24265669]
6. McLean SF. Case-based learning and its application in medical and health-care fields: a review of worldwide literature. J Med Educ Curric Dev 2016;3:JMECD.S20377. [doi: 10.4137/JMECD.S20377] [Medline: 29349306]
7. Alhazmi A, Quadri MFA. Comparing case-based and lecture-based learning strategies for orthodontic case diagnosis: a randomized controlled trial. J Dent Educ 2020 Aug;84(8):857-863. [doi: 10.1002/jdd.12171] [Medline: 32374442]
8. Thistlethwaite JE, Davies D, Ekeocha S, et al. The effectiveness of case-based learning in health professional education. A BEME systematic review: BEME Guide No. 23. Med Teach 2012;34(6):e421-e444. [doi: 10.3109/0142159X.2012.680939] [Medline: 22578051]
9. Ahmed A, Sutton MJD. Gamification, serious games, simulations, and immersive learning environments in knowledge management initiatives. World J Sci Technol Sustain Dev 2017 Apr 6;14(2-3):78-83. [doi: 10.1108/WJSTSD-02-2017-0005]
10. Imlig-Iten N, Petko D. Comparing serious games and educational simulations: effects on enjoyment, deep thinking, interest and cognitive learning gains. Simul Gaming 2018 Aug;49(4):401-422. [doi: 10.1177/1046878118779088]
11. Wong JYH, Ko J, Nam S, et al. Virtual ER, a serious game for interprofessional education to enhance teamwork in medical and nursing undergraduates: development and evaluation study. JMIR Serious Games 2022 Jul 14;10(3):e35269. [doi: 10.2196/35269] [Medline: 35834309]
12. Sipiyaruk K, Kaewsirirat P, Santiwong P. Technology-enhanced simulation-based learning in orthodontic education: a scoping review. Dental Press J Orthod 2023;28(3):e2321354. [doi: 10.1590/2177-6709.28.3.e2321354.oar] [Medline: 37466506]
13. Sipiyaruk K, Gallagher JE, Hatzipanagos S, Reynolds PA. A rapid review of serious games: from healthcare education to dental education. Eur J Dent Educ 2018 Nov;22(4):243-257. [doi: 10.1111/eje.12338] [Medline: 29573165]
14. Sipiyaruk K, Hatzipanagos S, Reynolds PA, Gallagher JE. Serious games and the COVID-19 pandemic in dental education: an integrative review of the literature. Computers 2021;10(4):42. [doi: 10.3390/computers10040042]
15. Zaror C, Mariño R, Atala-Acevedo C. Current state of serious games in dentistry: a scoping review. Games Health J 2021 Apr;10(2):95-108. [doi: 10.1089/g4h.2020.0042] [Medline: 33818135]
16. Pang M, Zhao X, Lu D, et al. Preliminary user evaluation of a new dental technology virtual simulation system: development and validation study. JMIR Serious Games 2022 Sep 12;10(3):e36079. [doi: 10.2196/36079] [Medline: 36094803]
17. Gentry SV, Gauthier A, L'Estrade Ehrstrom B, et al. Serious gaming and gamification education in health professions: systematic review. J Med Internet Res 2019 Mar 28;21(3):e12994. [doi: 10.2196/12994] [Medline: 30920375]
18. Middeke A, Anders S, Raupach T, Schuelper N. Transfer of clinical reasoning trained with a serious game to comparable clinical problems: a prospective randomized study. Simul Healthc 2020 Apr;15(2):75-81. [doi: 10.1097/SIH.0000000000000407] [Medline: 32044851]
19. Neher AN, Wespi R, Rapphold BD, Sauter TC, Kämmer JE, Birrenbach T. Interprofessional team training with virtual reality: acceptance, learning outcome, and feasibility evaluation study. JMIR Serious Games 2024 Nov 4;12:e57117. [doi: 10.2196/57117] [Medline: 39496167]
20. Sipiyaruk K, Hatzipanagos S, Vichayanrat T, Reynolds PA, Gallagher JE. Evaluating a dental public-health game across two learning contexts. Educ Sci 2022;12(8):517. [doi: 10.3390/educsci12080517]
21. Amir LR, Leonardy IC, Dewatmoko SN, et al. Serious game as oral histology learning strategy for undergraduate dental students; crossover randomized controlled trial. BMC Oral Health 2023 Aug 23;23(1):585. [doi: 10.1186/s12903-023-03286-3] [Medline: 37612722]
22. Aksoy E. Comparing the effects on learning outcomes of tablet-based and virtual reality-based serious gaming modules for basic life support training: randomized trial. JMIR Serious Games 2019 May 1;7(2):e13442. [doi: 10.2196/13442] [Medline: 31042153]
23. Song YL, Foo LH, Ong MAM. Exploring gamified learning for inculcating patient safety concepts in dentistry: a mixed-methods study. Eur J Dent Educ 2023 Aug;27(3):662-678. [doi: 10.1111/eje.12853] [Medline: 36065500]
24. Sipiyaruk K, Reynolds PA, Nakornnoi T, Santiwong P, Chintavalakorn R. Game-based learning in orthodontic education: a systematic review. BDJ Open 2024 Jul 4;10(1):56. [doi: 10.1038/s41405-024-00218-3] [Medline: 38965234]

XSL•FO
RenderX

25. Buajeeb W, Chokpipatkun J, Achalanan N, Kriwattanawong N, Sipiyaruk K. The development of an online serious game for oral diagnosis and treatment planning: evaluation of knowledge acquisition and retention. BMC Med Educ 2023 Nov 3;23(1):830. [doi: 10.1186/s12909-023-04789-x] [Medline: 37924052]
26. Buajeeb W, Reynolds PA, Boontub H, Tangmanpuwadol Y, Sipiyaruk K. Comparison of the effectiveness of a serious game and pre-recorded lecture in diagnosis and treatment planning of oral lesions for dental students. Sci Rep 2024 Dec 27;14(1):30641. [doi: 10.1038/s41598-024-83433-0] [Medline: 39730901]
27. Bourgonjon J, Valcke M, Soetaert R, Schellens T. Students' perceptions about the use of video games in the classroom. Comput Educ 2010 May;54(4):1145-1156. [doi: 10.1016/j.compedu.2009.10.022]
28. Behm-Morawitz E, Pennell H, Speno AG. The effects of virtual racial embodiment in a gaming app on reducing prejudice. Commun Monogr 2016 Jul 2;83(3):396-418. [doi: 10.1080/03637751.2015.1128556]
29. Jossan KS, Gauthier A, Jenkinson J. Cultural implications in the acceptability of game-based learning. Comput Educ 2021 Dec;174:104305. [doi: 10.1016/j.compedu.2021.104305]
30. Kolb AY, Kolb DA. Experiential learning theory as a guide for experiential educators in higher education. ELTHE 2017;1(1):38. [doi: 10.46787/elthe.v1i1.3362]
31. Cheung KL, Ten Klooster PM, Smit C, de Vries H, Pieterse ME. The impact of non-response bias due to sampling in public health studies: a comparison of voluntary versus mandatory recruitment in a Dutch national survey on adolescent health. BMC Public Health 2017 Mar 23;17(1):276. [doi: 10.1186/s12889-017-4189-8] [Medline: 28330465]
32. Murairwa S. Voluntary sampling design. Int J Adv Res Manag Soc Sci 2015;4(2):185-200 [FREE Full text]
33. Chow SC, Shao J, Wang H, Lokhnygina Y. Sample Size Calculations in Clinical Research, 3rd edition: CRC Press; 2017.
34. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319. [doi: 10.2307/249008]
35. Teerawongpairoj C, Tantipoj C, Sipiyaruk K. The design and evaluation of gamified online role-play as a telehealth training strategy in dental education: an explanatory sequential mixed-methods study. Sci Rep 2024 Apr 22;14(1):9216. [doi: 10.1038/s41598-024-58425-9] [Medline: 38649451]
36. Peanchitlertkajorn S, Reynolds PA, Chaisawas M, Krungthong Y, Boonkaew B, Sipiyaruk K. The impact of an online gamified virtual tour on cognitive enhancement in dental practice management. Sci Rep 2024 Nov 6;14(1):26975. [doi: 10.1038/s41598-024-75128-3] [Medline: 39505891]
37. Chon SH, Timmermann F, Dratsch T, et al. Serious games in surgical medical education: a virtual emergency department as a tool for teaching clinical reasoning to medical students. JMIR Serious Games 2019 Mar 5;7(1):e13028. [doi: 10.2196/13028] [Medline: 30835239]
38. Wu JH, Du JK, Lee CY. Development and questionnaire-based evaluation of virtual dental clinic: a serious game for training dental students. Med Educ Online 2021 Dec;26(1):1983927. [doi: 10.1080/10872981.2021.1983927] [Medline: 34694980]
39. Germa A, Gosset M, Gaucher C, et al. OSCEGame: A serious game for OSCE training. Eur J Dent Educ 2021 Nov;25(4):657-663. [doi: 10.1111/eje.12643] [Medline: 33314541]
40. Lemos M, Wolfart S, Rittich AB. Assessment and evaluation of a serious game for teaching factual knowledge in dental education. BMC Med Educ 2023 Jul 20;23(1):521. [doi: 10.1186/s12909-023-04498-5] [Medline: 37468897]
41. Damaševičius R, Maskeliūnas R, Blažauskas T. Serious games and gamification in healthcare: a meta-review. Information 2023;14(2):105 [FREE Full text]
42. Orvis KA, Horn DB, Belanich J. The roles of task difficulty and prior videogame experience on performance and motivation in instructional videogames. Comput Human Behav 2008 Sep;24(5):2415-2433. [doi: 10.1016/j.chb.2008.02.016]
43. Romero Rodríguez L, Sánchez‐Alzola A, Salazar A. Serious games in engineering education: assessing novelty effects and the influence of prior gaming experience. Comp Applic In Engineering 2025 May;33(3):e70021 [FREE Full text] [doi: 10.1002/cae.70021]
44. Allahverdipour H, Bazargan M, Farhadinasab A, Moeini B. Correlates of video games playing among adolescents in an Islamic country. BMC Public Health 2010 May 27;10(1):286. [doi: 10.1186/1471-2458-10-286] [Medline: 20507610]
45. Russoniello CV, O'Brien K, Parks JM. The effectiveness of casual video games in improving mood and decreasing stress. J Cyber Ther Rehabil 2009;2(1):53-66 [FREE Full text]
46. Gee JP. What video games have to teach us about learning and literacy. Comput Entertain 2003 Oct;1(1):20. [doi: 10.1145/950566.950595]
47. Malaquias RF, Malaquias FFO, Hwang Y. Understanding technology acceptance features in learning through a serious game. Comput Human Behav 2018 Oct;87:395-402. [doi: 10.1016/j.chb.2018.06.008]
48. Chauhan S, Mittal M, Woźniak M, Gupta S, Pérez de Prado R. A technology acceptance model-based analytics for online mobile games using machine learning techniques. Symmetry (Basel) 2021;13(8):1545. [doi: 10.3390/sym13081545]
49. Wouters P, van Nimwegen C, van Oostendorp H, van der Spek ED. A meta-analysis of the cognitive and motivational effects of serious games. J Educ Psychol 2013;105(2):249-265. [doi: 10.1037/a0031311]
50. Kiili K. Digital game-based learning: towards an experiential gaming model. Internet High Educ 2005 Jan;8(1):13-24. [doi: 10.1016/j.iheduc.2004.12.001]
51. Ravyse WS, Seugnet Blignaut A, Leendertz V, Woolner A. Success factors for serious games to enhance learning: a systematic review. Virtual Real 2017 Mar;21(1):31-58. [doi: 10.1007/s10055-016-0298-4]

XSL•FO
RenderX

52. Putz LM, Hofbauer F, Treiblmaier H. Can gamification help to improve education? Findings from a longitudinal study. Comput Human Behav 2020 Sep;110:106392. [doi: 10.1016/j.chb.2020.106392]
53. Rosenthal S, Ratan RA. Balancing learning and enjoyment in serious games: Kerbal Space Program and the communication mediation model. Comput Educ 2022 Jun;182:104480. [doi: 10.1016/j.compedu.2022.104480]
54. Martey RM, Stromer-Galley J, Shaw A, et al. Balancing play and formal training in the design of serious games. Games Cult 2017 May;12(3):269-291. [doi: 10.1177/1555412016674809]
55. Maxim RI, Arnedo-Moreno J. Identifying key principles and commonalities in digital serious game design frameworks: scoping review. JMIR Serious Games 2025 Mar 5;13:e54075. [doi: 10.2196/54075] [Medline: 40053743]
56. Yu FY. The mediating effects of anonymity and proximity in an online synchronized competitive learning environment. J Educ Comput Res 2003 Sep;29(2):153-167. [doi: 10.2190/59CX-3M7L-KKB4-UYDD]
57. Chen CH, Shih CC, Law V. The effects of competition in digital game-based learning (DGBL): a meta-analysis. Education Tech Research Dev 2020 Aug;68(4):1855-1873. [doi: 10.1007/s11423-020-09794-1]
58. Arayapisit T, Pojmonpiti D, Dansirisomboon K, Jitverananrangsri K, Poosontipong D, Sipiyaruk K. An educational board game for learning orofacial spaces: an experimental study comparing collaborative and competitive approaches. Anat Sci Educ 2023;16(4):666-676. [doi: 10.1002/ase.2257] [Medline: 36695535]
59. Lehtinen A, Kostiainen E, Näykki P. Co-construction of knowledge and socioemotional interaction in pre-service teachers' video-based online collaborative learning. Teach Teach Educ 2023 Oct;133:104299. [doi: 10.1016/j.tate.2023.104299]
60. Tolks D, Schmidt JJ, Kuhn S. The role of AI in serious games and gamification for health: scoping review. JMIR Serious Games 2024 Jan 15;12:e48258. [doi: 10.2196/48258] [Medline: 38224472]

Abbreviations

Edited by A Coristine; submitted 14.03.25; peer-reviewed by S kavuluri, S Nam, V Reis; revised version received 03.07.25; accepted 03.07.25; published 27.08.25.

Please cite as:
Chintavalakorn R, Chanwanichkulchai C, Buranasing N, Parivisutt N, Patchasri N, Sipiyaruk K
Development and Evaluation of a Case-Based Serious Game for Diagnosis and Treatment Planning in Orthodontic Education: Quasi-Experimental Study
JMIR Serious Games 2025;13:e73956
URL: https://games.jmir.org/2025/1/e73956
doi:10.2196/73956

© Rochaya Chintavalakorn, Chayuth Chanwanichkulchai, Napat Buranasing, Napatsaporn Parivisutt, Naruchol Patchasri, Kawin Sipiyaruk. Originally published in JMIR Serious Games (https://games.jmir.org), 27.8.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Use of an Immersive Virtual Reality Application to Educate Medical Students in Patient Handover: Pilot Study

Laura Isabel Hanke[1], Dr med; Patrick Schwoerer[1]; Florentine Huettl[1], Dr med; Lukas Vradelis[1], Dr med; Kai-Uwe Strelow[2]; Christian Boedecker[1], Dr med; Patrick Saalfeld[3], Dr; Vuthea Chheang[3], Dr Ing; Holger Buggenhagen[2], Dr med; Hauke Lang[1], Prof Dr Med; Christian Hansen[3], Prof Dr; Tobias Huber[1], Prof Dr Med

[1]Department for General, Visceral and Tranplant Surgery, University Medical Center of the Johannes Gutenberg-University Mainz, Langenbeckstr. 1, Mainz, Germany

[2]Rudolf Frey Educational Clinic, University Medical Center of the Johannes Gutenberg-University Mainz, Mainz, Germany

[3]Virtual and Augmented Reality Group, Faculty of Computer Science, Otto-von-Guericke University Magdeburg, Magdeburg, Germany

Corresponding Author:
Tobias Huber, Prof Dr Med
Department for General, Visceral and Tranplant Surgery, University Medical Center of the Johannes Gutenberg-University Mainz, Langenbeckstr. 1, Mainz, Germany

Abstract

Background: Patient handover is a daily task for doctors and nurses, and structured handovers have been proven to positively impact patient outcomes. To teach the handover procedure, different communication tools have been applied, such as the ISBAR (introduction and identification, situation, background, assessment and actions, and recommendation) method.

Objective: This study aimed to assess the effectiveness and user engagement of the first-time use of supplementary handover training in virtual reality (VR) for medical students as an addition to an existing curriculum. Furthermore, the VR program was tested for its usability, immersion, visually induced motion sickness (VIMS), and eye strain. Participants were evaluated for their motivation, time spent studying, and experience in VR, as well as their impressions of the use of VR in medical education.

Methods: Handover training using the ISBAR method and patient actors is part of the curriculum in surgery of the eighth semester of human medicine studies in Mainz. Knowledge is tested via an Objective Structured Clinical Examination (OSCE) using patient actors. We developed an immersive VR application using 360° video surroundings with structured patient cases. This application was offered as an optional supplementary training in groups of three with a peer tutor. Parameters evaluated included participants' characteristics, usability, and VIMS. Furthermore, a survey of the entire semester was conducted regarding their experience using VR and their enjoyment of studying. Finally, OSCE scores were collected and compared between the groups.

Results: The study was conducted over two semesters, and 92 of 385 (23.9%) volunteering students were recruited. The median age was 25 (IQR 23-25) years, and the majority were female (n=61, 68.5%). There were few to no issues regarding VIMS and eye strain (median eye strain 1, IQR 1-2; median VIMS 1, IQR 1-2). There was no significant difference in students' motivation (mid rank participant 107.84; mid rank nonparticipant 122.61; P=.11) and the amount studied for the subject (mid rank participant 113.88; mid rank nonparticipants 119.42; P=.54). Students felt significantly more confident in patient handover after the additional training (7-point Likert scale; mean pretraining 3.96, SD 1.39; mean post-training 3.17, SD 1.41; P<.01) and reported significantly more fun studying than their peers who did not participate in the additional training (mean participants 2.8, SD 1.54; mean nonparticipant 3.69, SD 1.73; P<.01). OSCE scores did not differ between the groups (median score 17 in both groups, IQR participants 16-19; IQR nonparticipants 16-18; P=.62).

Conclusions: This study shows that applications in VR, if implemented in a structured curriculum, can be a helpful and safe addition to the teaching of communication skills. VR applications should be considered as a time-flexible, safe, fun, and motivating educational tool as an addition to curricular teaching.

(JMIR Serious Games 2025;13:e73907) doi:10.2196/73907

KEYWORDS

immersive virtual reality; medical education; head-mounted display; handover; gamification; new tools

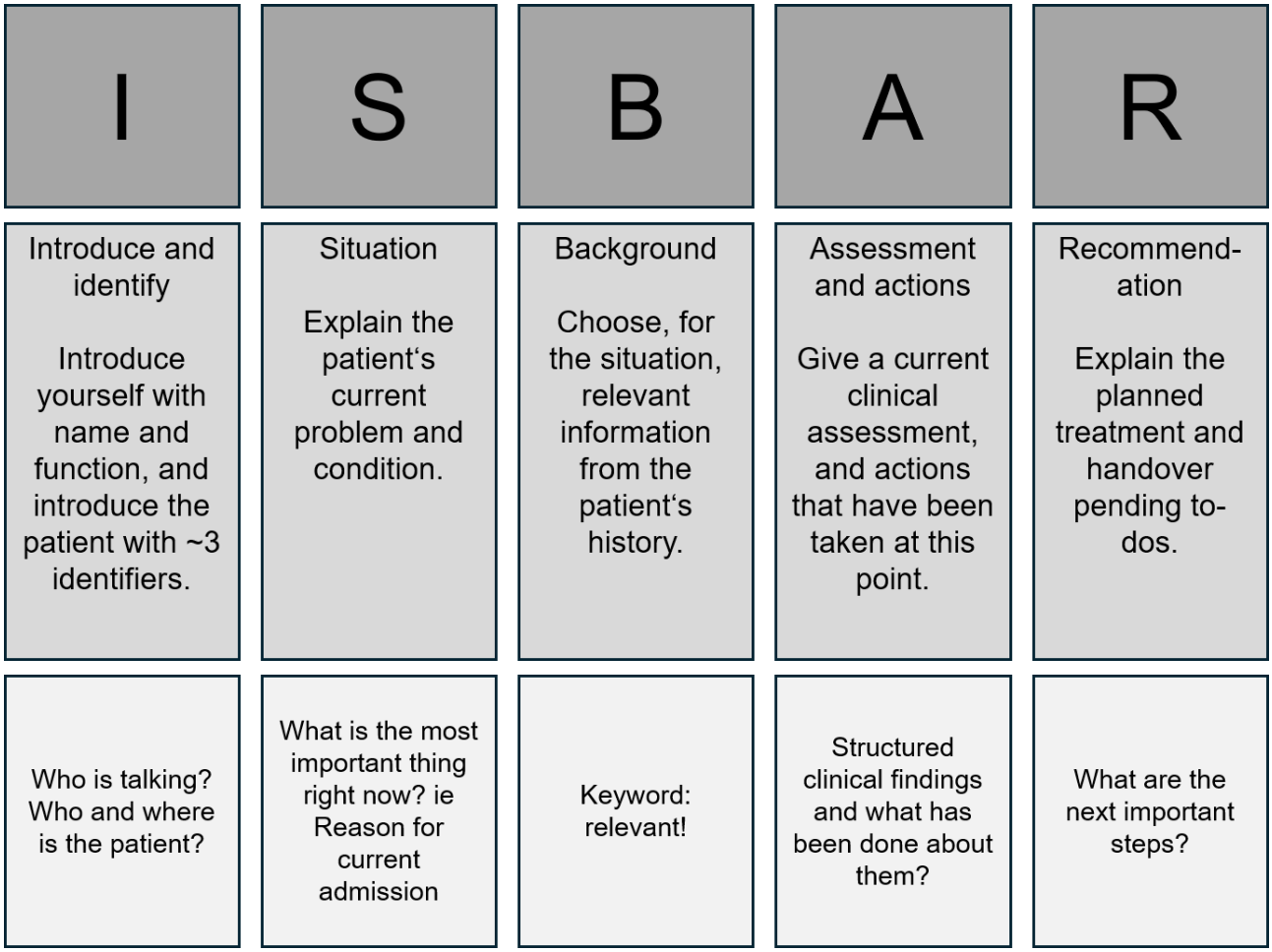
XSL•FO
RenderX

Introduction

Patient handover is a routine and essential component of daily clinical practice in hospitals across the world. Regardless of whether inpatient or outpatient, elective or emergency care, all departments involved in patient care must regularly conduct handovers of patient cases. This involves all medical staff and includes intra- and interprofessional as well as intra- and interdisciplinary communication [1].

It has been shown that patient handover is a key element for adequate patient care, directly influencing patient outcomes and yielding an important aspect of patient safety [2,3]. To facilitate comprehension by the receiving party, the handover should follow a clear structure, use concise language, and focus on essential information. Furthermore, patient handover integrates a variety of communication skills into one task that can be taught and practiced. In response to this need, the ISBAR (introduction and identification, situation, background, assessment and actions, and recommendation) framework was developed, tested, and validated across different medical professions [4-6]. ISBAR provides a structured backbone for every handover situation [7]. The details of the ISBAR concept are presented in Figure 1.

Figure 1. Introduction and identification, situation, background, assessment and actions, and recommendation (ISBAR) method in schematic form. Row 1: abbreviation, row 2: title and description, and row 3: additional cues.



The teaching of communication skills is essential for all medical professionals and has been increasingly incorporated over the last few decades [8]. In Germany, clear learning objectives for communication have been implemented in medical licensing regulations, posing new challenges for educators and academic institutions [9]. Most medical professionals practicing today did not receive formal communication training until years into their careers, if at all, because such training was not emphasized during their undergraduate and postgraduate education [10]. As a result, a crucial first step is to train the trainers, because good clinical expertise does not automatically translate into teaching competence [11].

Moreover, traditional courses, lectures, and written examinations are often ill-suited for communication training. Effective communication training requires a good balance of theoretical knowledge, practical application, self-reflection, and feedback from both instructors and peers. A sound theoretical basis is indispensable to ensure reproducible and consistent results. However, practical skills can only be trained effectively in practical courses where learners actively apply their newly acquired competencies [8].

This experiential learning setting is frequently supported by employing actors as simulated patients, which provides a valuable but costly asset [12].

XSL•FO
RenderX

Self-reflection is key to learning communication skills. This might take dedicated time in or after class to reflect individually or in a small group on why certain situations proved challenging to handle or why some interactions were perceived to be more stressful than others. Feedback from instructors and peers is important not only to correct theoretical and contextual shortcomings but also to raise awareness of how one is perceived by others. Many communication classes incorporate video reviews, allowing students to observe and evaluate their own behavior from an external perspective [13,14].

Applying the principle of constructive alignment, a practical skill such as patient handover cannot be adequately tested using multiple-choice or written examinations, but should instead be tested through an Objective Structured Clinical Examination (OSCE) [15]. This style of examination has been successfully used in the medical field for years [16]. In contrast to a written examination, which hundreds of students can take simultaneously, the OSCE requires students to rotate through different stations and complete each task individually. Consequently, OSCEs are more time-consuming and inherently more resource-intensive.

In recent years, virtual reality (VR) technology using head-mounted displays has become easily accessible to the public, especially in recreational computer science, and has also seen widespread use in different research settings. Among its many applications, VR presents an interesting addition to educational programs, facilitating learning in a safe environment at the learner's own pace and schedule [17]. Furthermore, learning in VR has been reported to be enjoyable for students and capable of fostering autobiographical memory, which may improve knowledge retention and recall [18]. VR applications have been developed for procedural and communication skill training [19].

In recent years, VR has gained additional interest due to its capacity to bridge the gap in situations where direct contact with patients or peers was not feasible, as experienced during the global COVID-19 pandemic [20]. While VR has not yet been able to replace interactions with peers, patients, and teachers, it is steadily evolving into a useful tool to enrich educational programs [17,21,22].

This study aimed to develop and test a supplemental educational tool for training patient handover in VR within a curricular course for medical students.

Methods

Study Setting and Educational Concept Design

At the University Medical Center of Johannes Gutenberg-University Mainz, Germany, communication skills are taught through a longitudinal communication curriculum encompassing various disciplines over multiple semesters. One component of the curriculum is teaching patient handover within the framework of the Surgical Practical Course (SPC), which takes place at the beginning of the 8th semester.

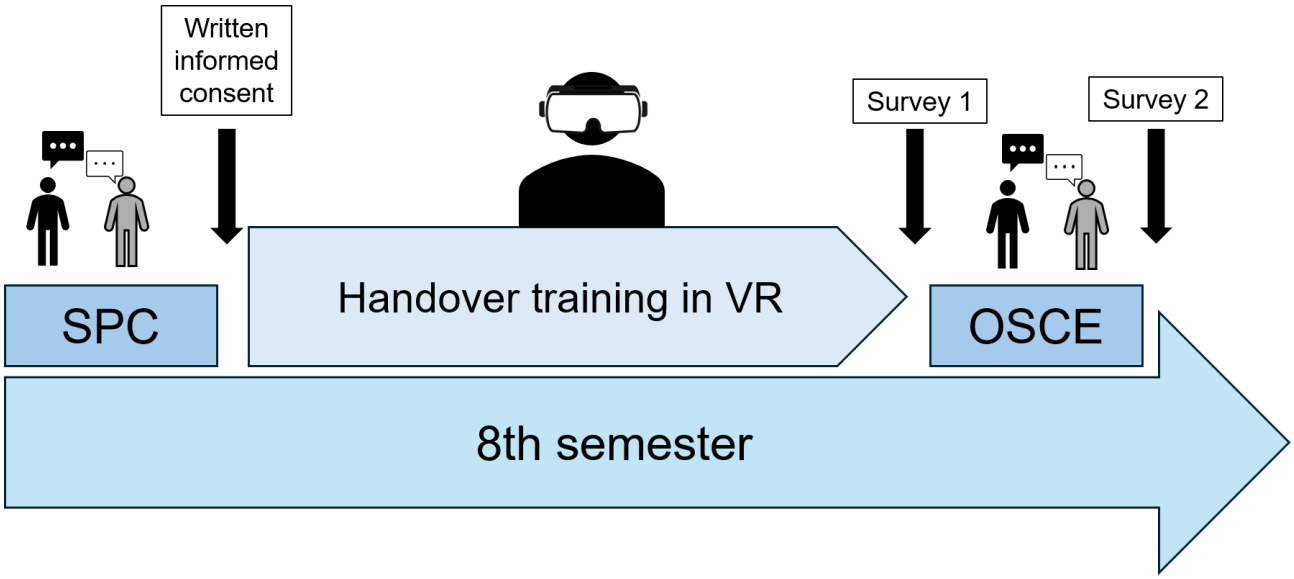
In this unit, students are taught the ISBAR concept through a short theoretical overview followed by practical training with standardized patient actors. They receive feedback from both instructors and peers. The unit lasts 45 minutes and involves small groups of 4-6 students. In addition, videos and texts are made available online for review. At the end of the semester, students undergo an OSCE using patient actors, complying with the concept of constructive alignment [15].

While other components of the SPC, such as surgical knot tying or suturing, can be practiced independently at home or in a skills laboratory, patient handover is more challenging to practice. To address this limitation, we developed a VR training suite consisting of three cases with increasing urgency. The 3 scenarios involve patient actors and SPC instructors as actors and were recorded. They are embedded in a structured curriculum to complement the VR training suite.

During VR training, students progress through each scenario by answering short multiple-choice questions to obtain a patient's history. They also receive a data sheet similar to a patient chart. The training sessions in VR are conducted in groups of three, each led by a peer instructor, thus enabling each student to perform one scenario. The training took place in the newly established VR training center at the Rudolf-Frey-Lernklinik for the semester. Figure 2 provides an overview of the course of the study.

XSL•FO
RenderX

Figure 2. Course of the study over one semester. In the first week of the semester, the Surgical Practical Course (SPC) took place, including the subject of patient handover taught with patient actors. Over the course of the semester, there were multiple time slots for a one-time handover training in VR. In the last week of the semester, the Objective Structured Clinical Examination (OSCE) took place, including examination of the subject of patient handover using patient actors. Black arrows indicate time points of informed consent and survey of participants and nonparticipants in the same semester. OSCE: Objective Structured Clinical Examination; SPC: Surgical Practical Course; VR: virtual reality.



Scenario Design and Video Production

An interdisciplinary team conceptualized scenarios involving the same preoperative patient undergoing Warfarin therapy before inguinal hernia repair, with an increasingly urgent need for medical attention. Detailed scripts for all scenarios are provided in Multimedia Appendix 1.

Scenario 1

The participant arrives in the patient's room and tells the participant that he has continued his daily medication of Warfarin until this morning, despite undergoing inguinal hernia repair tomorrow. Shortly thereafter, the attending physician arrives for the handover.

Scenario 2

The participant arrives in the patient's room, who tells the participant that he has continued his daily medication of Warfarin up until this morning and has noticed black stools in the past few days. He also reports feeling faint. As additional information, the participant is given laboratory results with decreased hemoglobin levels. Shortly thereafter, the attending physician arrives for the handover.

Scenario 3

The nurse calls the participant to assist her because the patient has fainted, and his bed is covered in bloody stool. His blood pressure is low and heart rate elevated. Shortly thereafter, the medical emergency team and the attending physician arrive in the room for a handover.

The participants perform a quick history and then give a structured handover to the arriving doctor using the ISBAR concept. A detailed script of the scenarios is provided in the Multimedia Appendix 1. The videos were recorded in 8k resolution using a 360° camera (Insta360 Pro 2; Insta360) in an actual patient room. The decision to use highly immersive video footage with actors in an actual hospital setting, rather than animated surroundings, was supported by prior research, which showed that participants preferred more realistic scenarios [23]. The shoot was supported by the company Visual-Impressions, which specialized in 360° video shooting (Visual-Impressions GmbH).

Development of the 360° Immersive VR Environment

We used the game engine Unity (version 2019.2.18f1; Unity Technologies) for an immersive VR development environment. The Virtual Reality Toolkit (version 3.3.0; Unity) was used to provide basic user interactions in VR, including locomotion and user interface (UI) interactions. The recorded 360° videos were edited and stitched according to the scenarios using the Shotcut (Meltytech, LLC) video editing tool. The stitched videos resulted in an equirectangular format that can be used for spherical projection mapping in VR. We incorporated the VIVE Media Decoder (HTC Vive Pro), a powerful video decoding plug-in, to achieve high-quality video streaming in Unity. The VR headset used for this study was the HTC Vive Pro (HTC Vive Pro), with a resolution of 1440×1600 pixels per eye (2880×1600 pixels combined), a refresh rate of 90 Hz, and a field of view of 110 degrees. Additional information regarding the scenarios was integrated under close medical and educational supervision and displayed on UI panels in VR. Participants were able to use their controllers with the ray-casting technique to interact with the UI interfaces. We also implemented an interactive questionnaire using the Virtual Reality Questionnaire Toolkit, allowing participants to answer questions directly in VR. If the participant answered the question correctly, the UI panel turned green and the scenario continued. If answered incorrectly, the scenario stopped at the question until it was answered correctly. Unlimited attempts were allowed. Figure 3 provides an impression of the VR environment.

XSL•FO
RenderX

Figure 3. Screenshots of the virtual reality (VR) environment. In (A), a nurse finds the patient, who feels dizzy and unwell. In (B), she checks his vital signs, which are depicted in the box above her head. In (C), the emergency team arrives. In (D), the patient is connected to monitoring. In (E), the attending arrives to receive the patient handover. HF: heart frequency; RR: blood pressure; SO_2: oxygen saturation; Temp: temperature.



Recruiting and Data Collection

Volunteer participants were recruited over two semesters (summer semester 2023 and winter semester 2023/2024). Participants underwent VR training in groups of three, together with a student teacher, and received corrections and feedback after each handover from their peers and the teacher. Students were made aware of the study via email and personal communication during the curricular SPC, and volunteers were recruited after the end of the SPC and throughout the semester. Participants were asked to complete questionnaires at the end of the VR training regarding demographic information, personal motivation, and the Immersion Scale according to Nichols et al [24], as well as rating how confident they were in the subject of patient handover. Furthermore, data regarding visually induced motion sickness (VIMS) and eye strain were collected [25]. At the end of the semester, after completion of the OSCE, participants were questioned again regarding motivation, confidence, and preparedness for the OSCE, as well as their impression of VR as a useful tool for education. Participants were contacted directly via email for the second survey and reminded once. In addition, the entire semester was questioned as a control group. The OSCE results for the handover examination and overall OSCE results were recorded for analysis. For data collection, the online platform LimeSurvey (LimeSurvey GmbH, open source) was used. Questionnaires, in addition to the Immersion Scale by Nichols et al [24], were custom-made for this study and have not been published previously. The custom-made questionnaires are available in Multimedia Appendix 2.

Statistical Analysis

Statistical analyses were performed using SPSS (version 29.0; IBM SPSS Statistics). Participant characteristics were analyzed using descriptive parameters. Questionnaire items were analyzed using the Mann-Whitney *U* test and *t* tests, as appropriate.

Ethical Considerations

All participants were volunteer medical students in the 8th semester of the human medicine program at Johannes Gutenberg-University Mainz, enrolled in the SPC. Written informed consent was obtained at the beginning of the study. Participation was not reported to OSCE examiners or any other teaching personnel. Ethical approval was obtained under the reference code 2023 - 17025–andere Forschung erstvotierend from the ethics committee of the regional medical association of Rhineland-Palatinate (Landesärztekammer Rheinland-Pfalz). Participants did not receive compensation. People visible in pictures (ie, Figure 3) are paid actors or coauthors and have provided written consent for publication.

Results

Overall, 92 participants of 385 (23.9%) students were recruited for the study, of which 3 participants (amounting to one group) dropped out because of exceptionally bad weather conditions, which made them unable to attend training. After obtaining written informed consent, the participants were divided into groups of three and attended VR training. Data analysis was performed on 89 participants who attended training (per-protocol analysis). The median age of the participants was 25 years (IQR 23-28), and the groups consisted of 61 (68.5%) female and 28

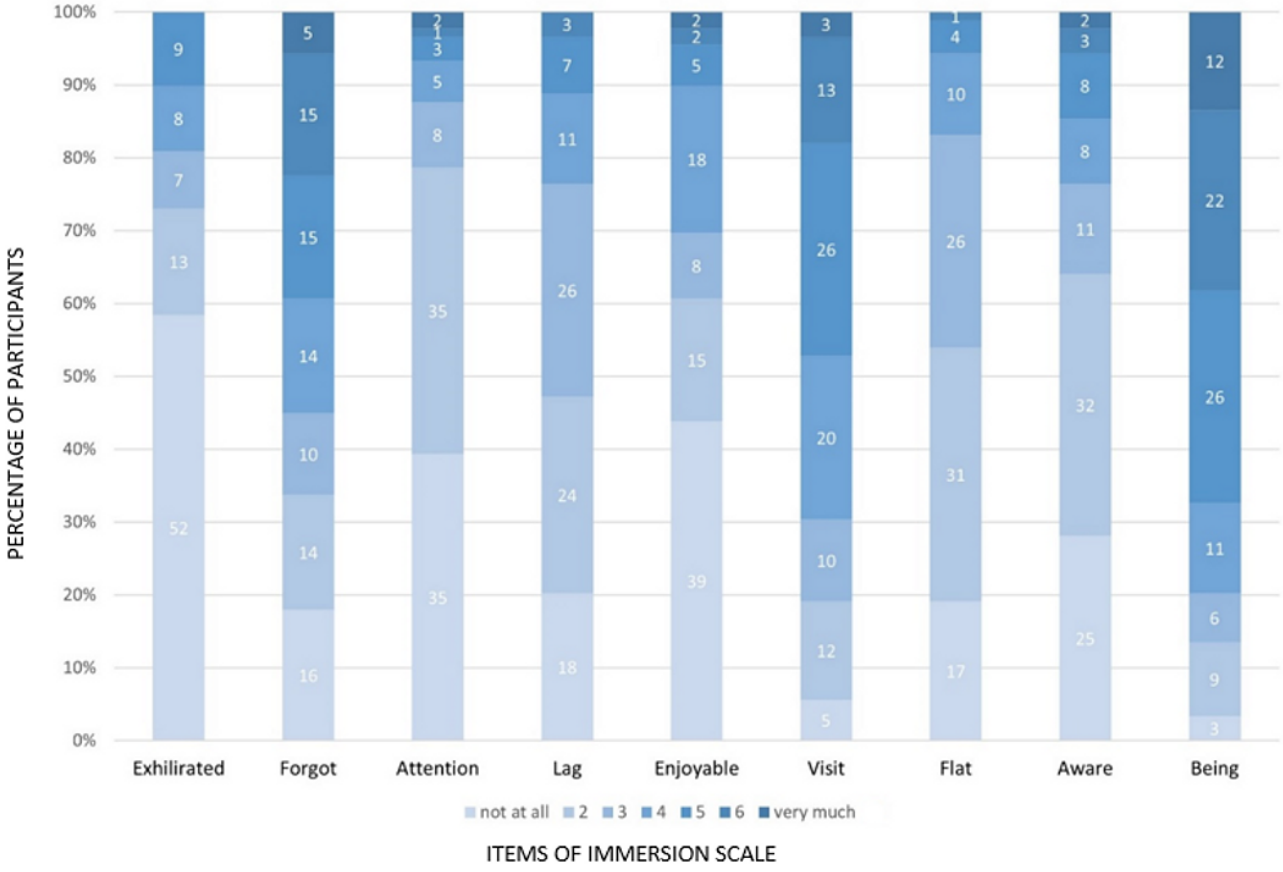
XSL•FO
RenderX

(31.5%) male participants. In total, 40% (n=36) of the participants had prior experience in the medical field (ie, nurses, physical therapists, and paramedics). The second survey, after the OSCE, was completed by 81 of the 89 (91%) participants.

Most participants had minimal to no experience using VR equipment or head-mounted displays (never: n=49, 55.1%; <5 times: n=33, 37.1%). Eye strain and VIMS were reported to be minimal, with one singular outlier who was able to complete the training despite dizziness (median eye strain 1, IQR 1-2, min 1, max 9; median VIMS 1, IQR 1-2, min 1, max 16). The majority of participants stated they felt present in the virtual world (scale of 1 - 7, 1=not at all, 7=very much; n [5 or more]=60, 67.4%). Details of the Immersion Scale Questionnaire results are shown in Figure 4.

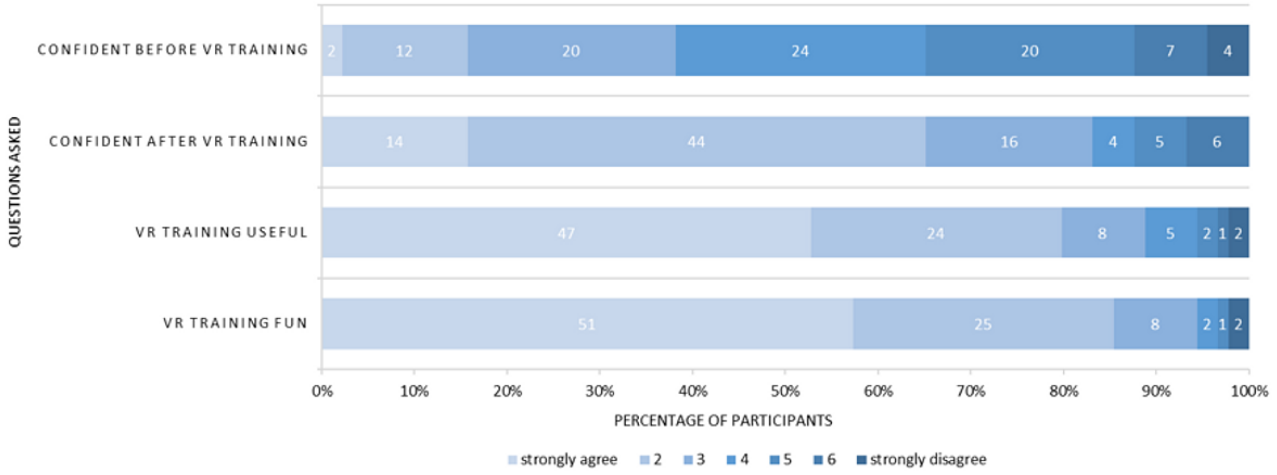
Participants reported feeling significantly more confident in giving a patient handover after the training (median pretraining 4, IQR 3-5; median post-training 3, IQR 2-4; *P*<.001). The majority stated that they enjoyed the handover training in VR and had more fun than in other courses during the training (scale 1 - 7; 1=a lot more fun and enjoyable than other courses; 7=a lot less fun and enjoyable; n (1)=47; 52.8%, n (2)=24; 27%). Further results are shown in Figure 5.

Figure 4. Result of the Immersion Scale as described by Nichols et al, based on a 7-point Likert scale. The questions of the scale are displayed with abbreviated titles; results range from 1 (not at all, at the bottom of the graph in lighter colors) to 7 (very much, at the top of the graph in darker colors). The results show a high grade of presence with minimal lag and minimal exhilaration, indicating a high grade of immersion.



XSL•FO
RenderX

Figure 5. Results of pre- and post-training questionnaires. Short-form questions are displayed on the left, and answers are listed from left to right, starting with "strongly agree" (lighter colors) and ending with "strongly disagree" (darker colors). Rows 1 and 2: participants felt significantly more confident in giving a patient handover after VR training (mean pretraining 3.96, SD 1.39; mean post-training 3.17, SD 1.41; *P*<.001). Rows 3 and 4: after experiencing the training, they deemed virtual reality (VR) programs for medical education to be useful and stated that they had more fun in the VR training than in other courses (7-point Likert scale, 1=strongly agree, 7=strongly disagree; mean useful 1.56, SD 1.34; mean fun 1.71, SD 1.16).

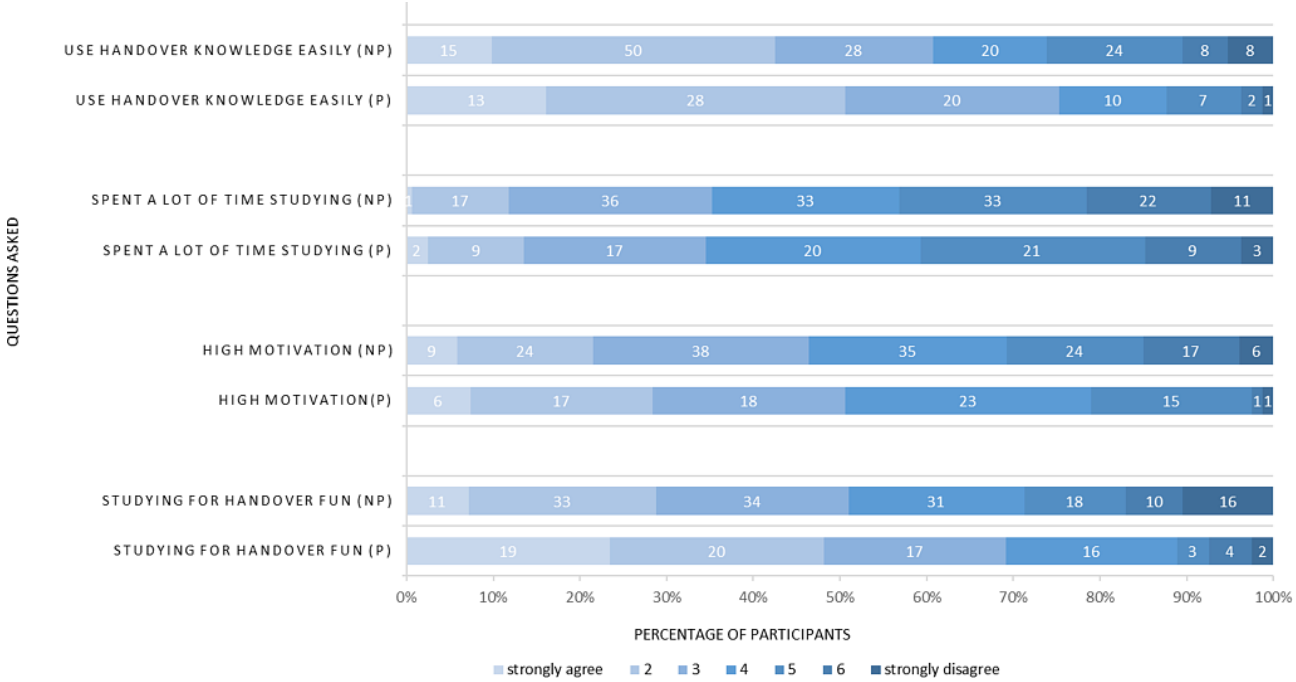


A total of 153 of 296 (51.7%) students from the entire semester who had not participated in the training replied to the survey. When comparing survey results between VR-training participants and nonparticipants, a few significant differences were noted. Students were asked if they felt they were able to recall and use what they had learned about patient handover well in the OSCE (scale 1 - 7; 1=very much, 7=not at all; mean participant 2.75, SD 1.36; mean nonparticipant 3.29, SD 1.65; middle rank participant 104.38; middle rank nonparticipant 124.45; *P*=.03; *r*=0.14). When asked whether they had fun studying for the patient handover, participants stated to have had significantly more fun (scale 1 - 7; 1=very much, 7=not at all). Mean participant 2.8, SD 1.54; mean nonparticipant 3.69, SD 1.73; middle rank participant 94.82; middle rank nonparticipant 129.51; *P*<.001; *r*=0.25. Having experienced the training, significantly more participants deemed VR equipment and head-mounted displays as a useful supplement for medical education (scale 1 - 7; 1=very much, 7=not at all; mean participant 1.56, SD 0.96; mean nonparticipant 3.03, SD 1.62; middle rank participant 54.98; middle rank nonparticipant 110.73; *P*<.001; *r*=0.42). It should be noted that all students were highly motivated (scale 1 - 7; 1=very much, 7=not at all), and no significant difference in motivation was observed between groups (mean participant 3.38, SD 1.31; mean nonparticipant 3.76, SD 1.52; middle rank participant 107.84; middle rank nonparticipant 122.61; *P*=.11; *r*=0.11). Further results of the second survey are shown in Figure 6: (1) rows 1 and 2: participants stated more often than nonparticipants that they were able to recall and apply what they had learned about patient handover in the OSCE (mean participant 2.75, SD 1.36; mean nonparticipant 3.29, SD 1.65; middle rank participant 104.38; middle rank nonparticipant 124.45; *P*=.03; *r*=0.14), (2) rows 3 and 4: participants and nonparticipants spent a similar amount of time studying for patient handover, with no significant difference between groups (mean participants 4.09, SD 1.39; mean nonparticipants 4.24, SD 1.46; middle rank participants 113.88; middle rank nonparticipants 119.42; *P*=.54; *r*=0.04), (3) rows 5 and 6: motivation between participants and nonparticipants was not significantly different (mean participants 3.38, SD 1.31; mean nonparticipants 3.76, SD 1.52; middle rank participants 107.84; middle rank nonparticipants 122.61; *P*=.11; r=0.11), (4) rows 7 and 8: p stated to have significantly more fun than np when studying about patient handover (mean participants 2.8, SD 1.54; mean nonparticipants 3.69, SD 1.73; middle rank participants 94.82; middle rank nonparticipants 129.51; *P*<.001; *r*=0.25).

It should be noted that there was no significant difference in OSCE results, neither in the handover station nor in the overall grade (median grade participants 17, IQR 16-19; median grade nonparticipants 17, IQR 16-18; *P*=.62), with a small effect size (Cohen *d*=0.014).

XSL•FO
RenderX

Figure 6. Comparison of nonparticipants and participant responses in the second survey. Short-form questions are displayed on the left, a 7-point Likert scale, answers are listed from left to right, starting with strongly agree (lighter colors) and ending with strongly disagree (darker colors).



Discussion

Principal Findings

Patient handover is an integral component of clinical practice that directly influences patient outcomes. The ISBAR method provides a clear structure that is easily remembered and reproducible [7]. However, the development of practical skills requires hands-on training, which can be time- and cost-intensive [8].

This study presents a novel VR-based training program for patient handover in a surgical context, which is tailored to our didactic framework and can be easily used with a student tutor. Interest in the study was high, and participant recruitment proceeded rapidly. When asked about their motivation to participate, most students stated that they were curious to explore innovative technologies and learning methods.

Our findings indicate that the VR training program was perceived as an engaging and enjoyable addition to other educational concepts, without drawbacks such as severe VIMS or eye strain. While there was no significant effect on OSCE results, which is consistent with prior study results [26], participants felt more confident after handover training in VR. This is particularly interesting in the context of communication skills, as confidence can play an important role in skill demonstration and application.

Comparison With Prior Work

To the authors' knowledge, there are no similar studies or programs using 360° immersive VR in a curricular setting currently available. In 2023, Andreasen et al [27] compared a VR tool with traditional paper-based learning methods. However, the study used a desktop VR application without head-mounted displays or a fully immersive 360° environment. Furthermore, the evaluation between groups was conducted using a written test. We deliberately decided against this approach to respect the concept of constructive alignment [15]. In the past, our group was able to show that VR simulation using 360° immersive video is preferred by participants; hence, we chose this mode of demonstration over others [28]. Similar to our results, the authors described similar outcomes in the examination, but students using the VR application reported finding this way of learning more enjoyable [27].

Neher et al [29] presented a feasibility study of interprofessional team training for medical and nursing students in VR in 2024. While this study uses immersive VR with head-mounted displays, it is a computer-generated environment in contrast to the program we are presenting. As this was a feasibility study, there was no implementation in curricular teaching or comparison with a control group, as in our study.

Prior studies have evaluated the teaching of structured patient handovers in theoretical form, which mostly includes lectures and observation of handovers in video material with subsequent group discussion [30,31]. Darbyshire et al [32] used structured teaching materials and group discussions of videos, with high satisfaction among students, even though there were no practical exercises involved. Gordon et al [33] and Thaeter et al [34] added role-play scenarios to the theoretical teaching approach to facilitate practice. All three aforementioned studies differ from the presented paper not only in the use of VR but also in the fact that there was no curricular teaching of patient handover, and the handover courses offered were optional and not graded. Thaeter et al [34] compared students who completed the handover course to a control group with no handover teaching at all and found an expectedly high statistical significance, which cannot be reached with a purely supplemental teaching effort, as described in the presented paper.

However, simulation can be achieved not only in VR but also with simulated patients, as is used in the SPC described.

XSL•FO
RenderX

Simulated patients are effectively used in medical education, mainly for training communication skills [35]. Losfeld et al [36] were able to show that handover curricula can successfully be augmented with e-learning in a blended learning concept. To the best of our knowledge, there is currently no handover curriculum comparing peer-to-peer role-play with simulated patients and VR simulation, which would certainly be an interesting topic for future research.

It should be noted that studies on handover naturally include all medical professions and curricula, and study designs and approaches may differ between different target groups.

Limitations

Although our study yielded interesting results, several limitations should be acknowledged. First, studies that rely on voluntary participation, such as ours, tend to attract the interest of highly motivated students, potentially introducing selection bias in the results. These students tend to exhibit greater intrinsic motivation and engage more intensely with teachers and learning materials than others do. To assess this, we specifically asked both participants and nonparticipants about their motivation and time invested in studying. No significant differences were found between the groups in terms of self-reported motivation or time spent preparing for the patient handover. This suggests that the selection bias between groups was minimal; however, as these measures were based on self-assessment, their validity may be limited.

Second, our study compared additional VR training with no additional training. A valuable direction for future studies could be the inclusion of a third group receiving additional training with standardized patient actors. We deliberately chose not to add more theoretical learning materials (ie, written case studies or video footage), since this would have shifted the focus to a comparison between practical and theoretical teaching modalities.

Third, the OSCE itself may present a limiting factor. The OSCE results were overall very good for all stations, with most students receiving very good grades and no examination failures requiring repetition. This raises the possibility that the OSCE lacked sufficient discriminatory power to detect differences in performance because all students did well on the examination. Nonetheless, the strong alignment between the SPC and the OSCE emphasizes the high degree of constructive alignment in the curriculum—an important feature for teaching quality, even if it complicates research evaluation [15].

Future Directions

Although handover training in VR was perceived well by participants, it still required students to attend in-person training sessions at the University Medical Center Mainz, Germany. While the availability of multiple time slots provided greater flexibility compared to training with standardized patients, the program does not yet offer true on-demand accessibility, such as an at-home application. Achieving this level of flexibility in communication training remains challenging, as peer and tutor feedback are critical components of skill development. In the future, applications may be able to assess human communication and provide immediate feedback. In the near term, however, this feedback will most likely be limited to the assessment of correct and complete content and the order of information in line with the ISBAR format. During the handover training sessions, students were enthusiastic about the VR environment and provided valuable feedback for future renditions of the program. A common request was for more time to practice, including access to additional scenarios as well as the opportunity to repeat existing ones.

Future research is currently directed at the implementation of different scenarios; yet, technological advances such as the integration of speech recognition and natural language processing tools to enable automated assessment of handover performance will need further exploration. Implementing such a system would require substantial redevelopment and technical sophistication, but it represents a promising area of development. The evaluation of more advanced features, such as tone, facial expressions, and other nonverbal aspects of communication, would require highly sophisticated technical solutions that are not yet widely available [37,38]. These additions could potentially facilitate communication training unconstrained by time or instructor availability and will certainly be subjects of future research.

However, it should be critically evaluated whether communication training without human interaction is a goal we are trying to achieve. Although the VR program was positively received, students stated that they would not prefer training in VR over bedside teaching. The authors fully support this sentiment: contact with patients and peers should not be replaced, but rather enhanced by technological innovations.

Conclusions

Overall, our study was able to show that VR communication training can be a valuable addition to conventional educational programs, leading to students feeling confident in giving a patient handover. While additional teaching programs in VR can be an interesting and motivating supplementary aid, they are not able to replace current teaching methods, especially bedside teaching. VR applications should be considered as a time-flexible, safe, and fun educational tool to supplement curricular teaching.

Acknowledgments

This publication is part of Patrick Schwoerer's doctoral thesis. We thank Visual-Impressions GmbH from Magdeburg, Germany, for their help during the 360° 8K video shoot. The application for this study was developed during the research project AVATAR, funded by the Federal Ministry of Education and Research, Germany (FKZ: 16SV0857), and the Gutenberg Lehrkolleg. Generative artificial intelligence or large language models were used for language check, but not in any other part of the study or the analyses. Preliminary results were presented at the German National Surgical Congress (April 26, 2024, Leipzig).

XSL•FO
RenderX

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

LH and TH conceptualized the study. LH and PS recruited participants and conducted the study. PS managed the data collection. LH, FH, LV, KUS, CB, HB, and TH conceptualized the script and conducted the video shoot. LH, KUS, HB, and TH conceptualized the educational concept. LH and TH prepared the statistical analysis and the manuscript. PS, VC, and CH created the virtual reality environment and program. HL and TH supervised the study and its implementation.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Scripts for virtual reality scenarios.
[DOCX File, 17 KB - games_v13i1e73907_app1.docx]

Multimedia Appendix 2
Translated questionnaires.
[DOCX File, 19 KB - games_v13i1e73907_app2.docx]

References

1. Burgess A, van Diggele C, Roberts C, Mellis C. Teaching clinical handover with ISBAR. BMC Med Educ 2020 Dec 3;20(Suppl 2):459. [doi: 10.1186/s12909-020-02285-0] [Medline: 33272274]
2. Farina CL, Moreno J, Schneidereith T. Using simulation to improve communication skills. Nurs Clin North Am 2024 Sep;59(3):437-448. [doi: 10.1016/j.cnur.2024.02.007] [Medline: 39059862]
3. Communication during patient hand-over. World Health Organization. 2007. URL: https://tinyurl.com/ctney83a [accessed 2025-08-16]
4. Thompson JE, Collett LW, Langbart MJ, et al. Using the ISBAR handover tool in junior medical officer handover: a study in an Australian tertiary hospital. Postgrad Med J 2011 May;87(1027):340-344. [doi: 10.1136/pgmj.2010.105569] [Medline: 21310805]
5. Alolayan A, Alkaiyat M, Ali Y, Alshami M, Al-Surimi K, Jazieh AR. Improving physician's hand over among oncology staff using standardized communication tool. BMJ Qual Improv Rep 2017;6(1):u211844.w6141. [doi: 10.1136/bmjquality.u211844.w6141] [Medline: 28174657]
6. Chaica V, Marques R, Pontífice-Sousa P. ISBAR: a handover nursing strategy in emergency departments, scoping review. Healthcare (Basel) 2024 Feb 4;12(3):399. [doi: 10.3390/healthcare12030399] [Medline: 38338283]
7. Leonard M, Graham S, Bonacum D. The human factor: the critical importance of effective teamwork and communication in providing safe care. Qual Saf Health Care 2004 Oct;13 Suppl 1(Suppl 1):i85-i90. [doi: 10.1136/qhc.13.suppl_1.i85] [Medline: 15465961]
8. Deveugele M. Communication training: skills and beyond. Patient Educ Couns 2015 Oct;98(10):1287-1291. [doi: 10.1016/j.pec.2015.08.011] [Medline: 26298220]
9. Leitfaden: kommunikative kompetenzen von ärztinnen und ärzten. : Bundesministerium für Gesundheit; 2020 URL: https://www.bundesgesundheitsministerium.de/fileadmin/Dateien/5_Publikationen/Gesundheit/Broschueren/Leitfaden_Kommunikative_Kompetenzen_med._Ausbildung.pdf [accessed 2025-02-14]
10. Venktaramana V, Loh EKY, Wong CJW, et al. A systematic scoping review of communication skills training in medical schools between 2000 and 2020. Med Teach 2022 Sep;44(9):997-1006. [doi: 10.1080/0142159X.2022.2054693] [Medline: 35653622]
11. Sator M, Holler P, Rosenbaum M. National train-the-trainer certificate programme for improving healthcare communication in Austria. Patient Educ Couns 2021 Dec;104(12):2857-2866. [doi: 10.1016/j.pec.2021.07.046] [Medline: 34454798]
12. Isaksson J, Krabbe J, Ramklint M. Medical students' experiences of working with simulated patients in challenging communication training. Adv Simul (Lond) 2022 Oct 10;7(1):32. [doi: 10.1186/s41077-022-00230-3] [Medline: 36217210]
13. Pless A, Hari R, Brem B, Woermamm U, Schnabel KP. Using self and peer video annotations of simulated patient encounters in communication training to facilitate the reflection of communication skills: an implementation study. GMS J Med Educ 2021;38(3):Doc55. [doi: 10.3205/zma001451] [Medline: 33824891]
14. Marita P, Leena L, Tarja K. Nurses' self-reflection via videotaping to improve communication skills in health counseling. Patient Educ Couns 1999 Jan;36(1):3-11. [doi: 10.1016/s0738-3991(98)00069-x] [Medline: 10036555]
15. Biggs J, Tang C, Kennedy G. Teaching for Quality Learning at University: McGraw-Hill Education; 2022.

XSL•FO
RenderX

16. Patrício MF, Julião M, Fareleira F, Carneiro AV. Is the OSCE a feasible tool to assess competencies in undergraduate medical education? Med Teach 2013 Jun;35(6):503-514. [doi: 10.3109/0142159X.2013.774330] [Medline: 23521582]
17. Alvarado ER, Rehder KJ, Turner DA. Doing virtually anything to educate students. Pediatr Crit Care Med 2020 May;21(5):502-503. [doi: 10.1097/PCC.0000000000002251] [Medline: 32358334]
18. Schöne B, Wessels M, Gruber T. Experiences in virtual reality: a window to autobiographical memory. Curr Psychol 2019;106:1-5 [FREE Full text]
19. Kyaw BM, Saxena N, Posadzki P, et al. Virtual reality for health professions education: systematic review and meta-analysis by the digital health education collaboration. J Med Internet Res 2019 Jan 22;21(1):e12959. [doi: 10.2196/12959] [Medline: 30668519]
20. Miltykh I, Kafarov ES, Covantsev S, Dadashev AS, Skarlis AA, Zenin OK. A new dimension in medical education: virtual reality in anatomy during COVID-19 pandemic. Clin Anat 2023 Oct;36(7):1007-1015. [doi: 10.1002/ca.24098] [Medline: 37485993]
21. Tursø-Finnich T, Jensen RO, Jensen LX, Konge L, Thinggaard E. Virtual reality head-mounted displays in medical education: a systematic review. Simul Healthc 2023 Feb 1;18(1):42-50. [doi: 10.1097/SIH.0000000000000636] [Medline: 35136005]
22. Hanke LI, Vradelis L, Boedecker C, et al. Immersive virtual reality for interdisciplinary trauma management - initial evaluation of a training tool prototype. BMC Med Educ 2024 Jul 18;24(1):769. [doi: 10.1186/s12909-024-05764-w] [Medline: 39026193]
23. Huber T, Paschold M, Hansen C, Wunderling T, Lang H, Kneist W. New dimensions in surgical training: immersive virtual reality laparoscopic simulation exhilarates surgical staff. Surg Endosc 2017 Nov;31(11):4472-4477. [doi: 10.1007/s00464-017-5500-6] [Medline: 28378077]
24. Nichols S, Haldane C, Wilson JR. Measurement of presence and its consequences in virtual environments. Int J Hum Comput Stud 2000 Mar;52(3):471-491. [doi: 10.1006/ijhc.1999.0343]
25. Keshavarz B, Hecht H. Validating an efficient method to quantify motion sickness. Hum Factors 2011 Aug;53(4):415-426. [doi: 10.1177/0018720811403736] [Medline: 21901938]
26. Liaw SY, Ooi SW, Rusli KDB, Lau TC, Tam WWS, Chua WL. Nurse-physician communication team training in virtual reality versus live simulations: randomized controlled trial on team communication and teamwork attitudes. J Med Internet Res 2020 Apr 8;22(4):e17279. [doi: 10.2196/17279] [Medline: 32267235]
27. Andreasen EM, Berg H, Steinsbekk A, Høigaard R, Haraldstad K. The effect of using desktop VR to practice preoperative handovers with the ISBAR approach: a randomized controlled trial. BMC Med Educ 2023 Dec 20;23(1):983. [doi: 10.1186/s12909-023-04966-y] [Medline: 38124094]
28. Huber T, Paschold M, Hansen C, Lang H, Kneist W. Artificial versus video-based immersive virtual surroundings: analysis of performance and user's preference. Surg Innov 2018 Jun;25(3):280-285. [doi: 10.1177/1553350618761756] [Medline: 29504470]
29. Neher AN, Wespi R, Rapphold BD, Sauter TC, Kämmer JE, Birrenbach T. Interprofessional team training with virtual reality: acceptance, learning outcome, and feasibility evaluation study. JMIR Serious Games 2024 Nov 4;12:e57117. [doi: 10.2196/57117] [Medline: 39496167]
30. Masterson MF, Gill RS, Turner SR, Shrichand P, Giuliani M. A systematic review of educational resources for teaching patient handover skills to resident physicians and other healthcare professionals. Can Med Educ J 2013;4(1):e96-e110. [Medline: 26451207]
31. Desmedt M, Ulenaers D, Grosemans J, Hellings J, Bergs J. Clinical handover and handoff in healthcare: a systematic review of systematic reviews. Int J Qual Health Care 2021 Feb 20;33(1):mzaa170. [doi: 10.1093/intqhc/mzaa170] [Medline: 33325520]
32. Darbyshire D, Gordon M, Baker P. Teaching handover of care to medical students. Clin Teach 2013 Feb;10(1):32-37. [doi: 10.1111/j.1743-498X.2012.00610.x] [Medline: 23294741]
33. Gordon M, Grafton-Clarke C, Hill E. Teaching handover in undergraduate education: an evidence-based multi-disciplinary approach. MedEdPublish (2016) 2019;8:100. [doi: 10.15694/mep.2019.000100.1] [Medline: 38089334]
34. Thaeter L, Schröder H, Henze L, et al. Handover training for medical students: a controlled educational trial of a pilot curriculum in Germany. BMJ Open 2018 Sep 12;8(9):e021202. [doi: 10.1136/bmjopen-2017-021202] [Medline: 30209154]
35. Shin S, Park JH, Kim JH. Effectiveness of patient simulation in nursing education: meta-analysis. Nurse Educ Today 2015 Jan;35(1):176-182. [doi: 10.1016/j.nedt.2014.09.009] [Medline: 25459172]
36. Losfeld X, Istas L, Schoonvaere Q, Vergnion M, Bergs J. Impact of a blended curriculum on nursing handover quality: a quality improvement project. BMJ Open Qual 2021 Mar;10(1):e001024. [doi: 10.1136/bmjoq-2020-001024] [Medline: 33781991]
37. Riess H, Kraft-Todd G. E.M.P.A.T.H.Y.: a tool to enhance nonverbal communication between clinicians and their patients. Acad Med 2014 Aug;89(8):1108-1112. [doi: 10.1097/ACM.0000000000000287] [Medline: 24826853]
38. Xu J, Yang L, Guo M. Designing and evaluating an emotionally responsive virtual patient simulation. Simul Healthc 2024 Jun 1;19(3):196-203. [doi: 10.1097/SIH.0000000000000730] [Medline: 37651599]

XSL•FO
RenderX

Abbreviations

ISBAR: introduction and identification, situation, background, assessment and actions, and recommendation
OSCE: Objective Structured Clinical Examination
SPC: Surgical Practical Course
UI: user interface
VIMS: visually induced motion sickness
VR: virtual reality

Edited by A Coristine; submitted 13.03.25; peer-reviewed by A Rubio-López, V Wattanasoontorn, X Cheng; revised version received 30.06.25; accepted 30.06.25; published 27.08.25.

Please cite as:
Hanke LI, Schwoerer P, Huettl F, Vradelis L, Strelow KU, Boedecker C, Saalfeld P, Chheang V, Buggenhagen H, Lang H, Hansen C, Huber T
Use of an Immersive Virtual Reality Application to Educate Medical Students in Patient Handover: Pilot Study
JMIR Serious Games 2025;13:e73907
URL: https://games.jmir.org/2025/1/e73907
doi:10.2196/73907

©Laura Isabel Hanke, Patrick Schwoerer, Florentine Huettl, Lukas Vradelis, Kai-Uwe Strelow, Christian Boedecker, Patrick Saalfeld, Vuthea Chheang, Holger Buggenhagen, Hauke Lang, Christian Hansen, Tobias Huber. Originally published in JMIR Serious Games (https://games.jmir.org), 27.8.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

ADRENALINE, a Learning Game to Improve Prescribing Skills in Undergraduate Medical Students: Descriptive Study

Cecile Marie Yelnik[1,2], MD, PhD; Aurélie Daumas[3], MD, PhD; Yanele Poteaux[4], MA; Annie Standaert[5,6], PharmD, PhD; Natacha Grimbert[6], BSc; Raphaël Favory[2,7], MD, PhD; Pierre Ravaux[6], PhD; Marc Lambert[1,2], MD, PhD; Katia Quelennec[4], MA

[1]CHU Lille, Département de Médecine Polyvalentes et de Post-Urgences, Univ. Lille, Hôpital Calmette, Bd du Professeur Jules Leclercq, Lille, France

[2]U1167 -RID-AGE- Facteurs de risque et déterminants moléculaires des maladies liées au vieillissement, Institut Pasteur de Lille, Inserm, Lille, France

[3]Internal Medicine and Therapeutics Department, La Timone, Assistance Publique-Hôpitaux de Marseille (AP-HM), Aix-Marseille Univ, Marseille, France

[4]DAPI, Univ. Lille, Lille, France

[5]Inserm, CHU Lille, U1286 -INFINITE- Institute for Translational Research in Inflammation, Univ. Lille, Lille, France

[6]Département de Pharmacie Officinale, UFR3S - pharmacie, Lille, France

[7]CHU Lille, Département de médecine intensive et de réanimation, Univ. Lille, Lille, France

Corresponding Author:
Cecile Marie Yelnik, MD, PhD
CHU Lille, Département de Médecine Polyvalentes et de Post-Urgences, Univ. Lille, Hôpital Calmette, Bd du Professeur Jules Leclercq, Lille, France

Abstract

Background: Junior doctors often demonstrate insufficient prescribing skills, highlighting the need to enhance undergraduate medical training in this area. Serious games (SGs) have proven effective in teaching knowledge and skills across various medical specialties including surgery and emergency care. To our knowledge, no SG specifically dedicated to prescribing has been developed to date. Our objective was to develop a new educational program, based on a learning game designed to enhance medical students' competencies in safe and effective prescribing.

Objective: This study aimed to describe ADRENALINE, a learning game designed to promote safe and effective prescribing, and to report feedback from sixth-year undergraduate medical students at our medical school after the first year of its implementation in the therapeutics curriculum.

Methods: This study implemented an interactive educational program based on Kolb experiential learning theory to enhance safe and effective prescribing skills among sixth-year medical students. The program followed 3 phases: a preliminary in-person lecture introducing the SG ADRENALINE, autonomous gameplay, and a final debriefing lecture. ADRENALINE, accessible via university platform Moodle (Andrews Lyons) on multiple devices, was developed using MOSAIC software (Katia Quelennec), a software program created to develop evolutive SG based on real-life professional situations, and includes 20 realistic clinical scenarios of varying difficulty, requiring students to make therapeutic decisions and receive immediate feedback. Players advance through levels based on performance, with ongoing support from professors via feedback and a dedicated forum. The program was integrated into the therapeutic curriculum of Lille University, and participation was voluntary. All 598 sixth-year students were invited to access the game via email and to participate in pre- and postintervention surveys assessing usage patterns, satisfaction, and learning outcomes.

Results: Between November 2023 and March 2024, 272 sixth-year students accessed the ADRENALINE program. Of these, 201/272 (73.9%) students completed at least one scenario and obtained scores ranging from 16.5 to 100 out of 100. Pretest survey responses (n=99 answers) indicated that 92/99 (93%) students identified as gamers and believed that SGs could be relevant for their medical education. Posttest survey responses (n=50 answers) reflected a high level of satisfaction among participants. Most students reported that ADRENALINE helps them apply academic knowledge in real-world context, feel more confident with prescribing and managing adverse drug reactions, improve their prescribing skills, and better prepare for the national Objective Structured Clinical Examination.

Conclusions: We developed a learning game focused on medical prescribing, designed to be easily shared with other French-speaking medical schools. Although only 201/598 (33.6%) students engaged with this initial version, 85% (42/50) of the feedback received was positive, indicating strong student interest and supporting the educational value of a game-based approach to enhance prescribing skills among undergraduate medical students.

XSL•FO
RenderX

(JMIR Serious Games 2025;13:e66334) doi:10.2196/66334

KEYWORDS

serious game; game-based learning; therapeutic; medical prescription; medical education; learning games

Introduction

Prescribing is a complex task that requires a broad range of medical knowledge, clinical skills, and professional attitudes, all applied within a dynamic health care system and social context. In a systematic literature review, Brinkman et al [1] addressed the question of which competencies medical students should acquire by graduation to prescribe safely and effectively. They identified over 40 prescribing competencies, despite a lack of a strong consensus among Clinical Pharmacology and Therapeutics (CPT) teachers. Their review also revealed that many students perceive limited learning opportunities related to rational prescribing during their medical training and feel inadequately prepared for their future prescribing responsibilities [1]. Prescribing errors are frequent and can result in serious adverse events. Junior doctors, who are often the primary prescribers in hospital settings, are particularly prone to such errors, with evidence suggesting a higher error rate compared to their senior colleagues [2-6]. Educational interventions and structured training have been shown to reduce prescribing errors [7]. However, a study on CPT education across Europe found that traditional teaching methods (eg, lectures and written examinations) still dominate and are associated with lower levels of prescribing competence than active, problem-based learning methods such as clinical simulations or real-life prescribing practice [8]. In response, there is a growing need to adapt CPT education to better prepare students for rational and safe prescribing.

Digital educational resources, including e-learning programs, podcasts, simulations, serious games (SGs), and virtual or augmented reality, offer several advantages over traditional methods. These include flexibility in time and location, integration of multimedia and interactive content, and opportunities for repeated exposure to key concepts. A systematic review suggested that such tools could be effectively used to enhance prescribing education [9]. The European Association for Clinical Pharmacology and Therapeutics conducted a cross-sectional survey across Europe on the use of digital tools in CPT education [10]. The findings showed that although digital resources are widely adopted, most are limited to knowledge acquisition, as they are predominantly e-learning platforms. SGs, defined as games designed for purposes beyond entertainment, such as education or training, represent a promising pedagogical tool. Educational SGs, often referred to as learning games, have the potential to foster cognitive skills, behavioral change, and practical competencies [11]. Over the past decade, SGs have gained traction in medical education, with studies reporting high levels of student motivation, engagement, and satisfaction [12,13]. For example, a study involving 108 third-year medical students evaluated an SG focused on pulmonary embolism and found that it significantly enhanced student engagement and learning [14]. Although several SGs have been developed for health care students, particularly in fields such as surgery, emergency care, and pharmacy, to our knowledge, no SG to date has been specifically designed to develop undergraduate medical students' competencies in clinical prescribing.

The objective of this study was to design a learning game aimed at enhancing undergraduate medical students' competencies in safe and effective prescribing. This paper provides a detailed description of the ADRENALINE SG and presents user feedback from the first cohort of sixth-year medical students who piloted its initial version.

Methods

General Presentation of the Program

This interactive program was designed by the Professors of Therapeutics team, who have a background as medical doctors, to address our main objective of improving prescribing education. Our program was developed based on Kolb experiential learning theory (1984), which is particularly well suited to the learning of medical prescribing. This process involves complex decision-making and the application of knowledge in varied clinical situations, making experiential learning an ideal framework. The theory outlines 4 sequential phases that structured our program: concrete experience, reflective observation, abstract conceptualization, and active experimentation. The program was implemented in 3 phases: first, an in-person lecture introducing the SG and its learning objectives; second, independent gameplay by the students, and finally, a second in-person lecture dedicated to debriefing, conducted by the 3 professors of therapeutic from our university.

During the game, students engage in concrete experience by actively participating in realistic prescribing scenarios. This immersive approach provides learners with additional opportunities to encounter diverse clinical cases beyond those typically experienced in real-life settings, which can be limited or occur randomly during clinical placements. Importantly, real-time debriefing is provided at each stage of the game, allowing students to immediately reflect on their decisions and receive feedback on their performance. Following gameplay, students further engage in reflective observation by reviewing their overall performance and decisions, facilitated through detailed feedback on correct and incorrect answers. The subsequent phase, abstract conceptualization, takes place during debriefing sessions where theoretical knowledge is connected with practical experience, and any misunderstandings are clarified. Finally, active experimentation occurs during the final debriefing with instructors, allowing students to consolidate their learning and prepare for future clinical practice. By combining experiential learning with guided instruction and real-time feedback, the game enhances students' competence and confidence in medical prescribing, providing a structured yet flexible environment to develop critical skills in patient safety.

XSL•FO
RenderX

All 598 sixth-year undergraduate medical students registered at Lille University were invited to attend the nonmandatory in-person lectures, and all students (including those who did not attend the in-person lecture) received internet-based access to ADRENALINE via email invitation. Participation was based on volunteering and integrated into the therapeutic education of our faculty. At each phase of the program, all students received email invitations, regardless of whether they attended the in-person sessions or accessed the game.

SG Characteristics

ADRENALINE was developed using the MOSAIC platform (Katia Quelennec), which is based on the ScenariChain Opale, ScenariChain Topaze, and JavaScript software. MOSAIC was specifically designed to create adaptive SG rooted in real-life professional situations. ADRENALINE was adapted from e-Caducée, a learning game initially developed by our faculty of pharmacy to support the professional training of undergraduate pharmacy students [15]. The development of ADRENALINE resulted from a collaborative effort between professors of therapeutics from the Universities of Lille and Marseille. The project was approved and supported by the French National College of Therapeutics (CNET). ADRENALINE is accessible via the university's Moodle (Andrews Lyons) platform only, not through open access or commercial apps, and is compatible with multiple devices, including tablets, computers, and smartphones.

ADRENALINE is a 2-dimensional, scenario-based SG designed to simulate common prescribing situations encountered in hospital settings. Players engage in interactive dialogues with game patients and health care professionals by actively selecting responses through a click-based interface. While the game does not allow free navigation within a simulated environment, players progress through clinical scenarios by making diagnostic and therapeutic decisions, ordering tests, and selecting appropriate prescriptions. In the game, the player assumes the role of a sixth-year undergraduate medical student at the fictional University Hospital of Berdeghem (Figure 1).

Figure 1. Home page of the ADRENALINE program. English translation of the French text: "Will you be the top medical student during your rotation? Will you succeed in your medical studies and become a resident, then a senior doctor at Berdeghem Hospital — or will you open your own private practice? Challenge your colleagues and jump into the game!"



The gameplay revolves around realistic clinical situation scenarios where therapeutic decision-making and medical prescribing are central to the clinical problem. These scenarios were designed to reflect a wide range of clinical situations that medical students may encounter during their training, for example, answering questions from a supervising physician during hospital rounds, managing patients in the emergency department, or conducting consultations in an outpatient clinic (Figure 2).

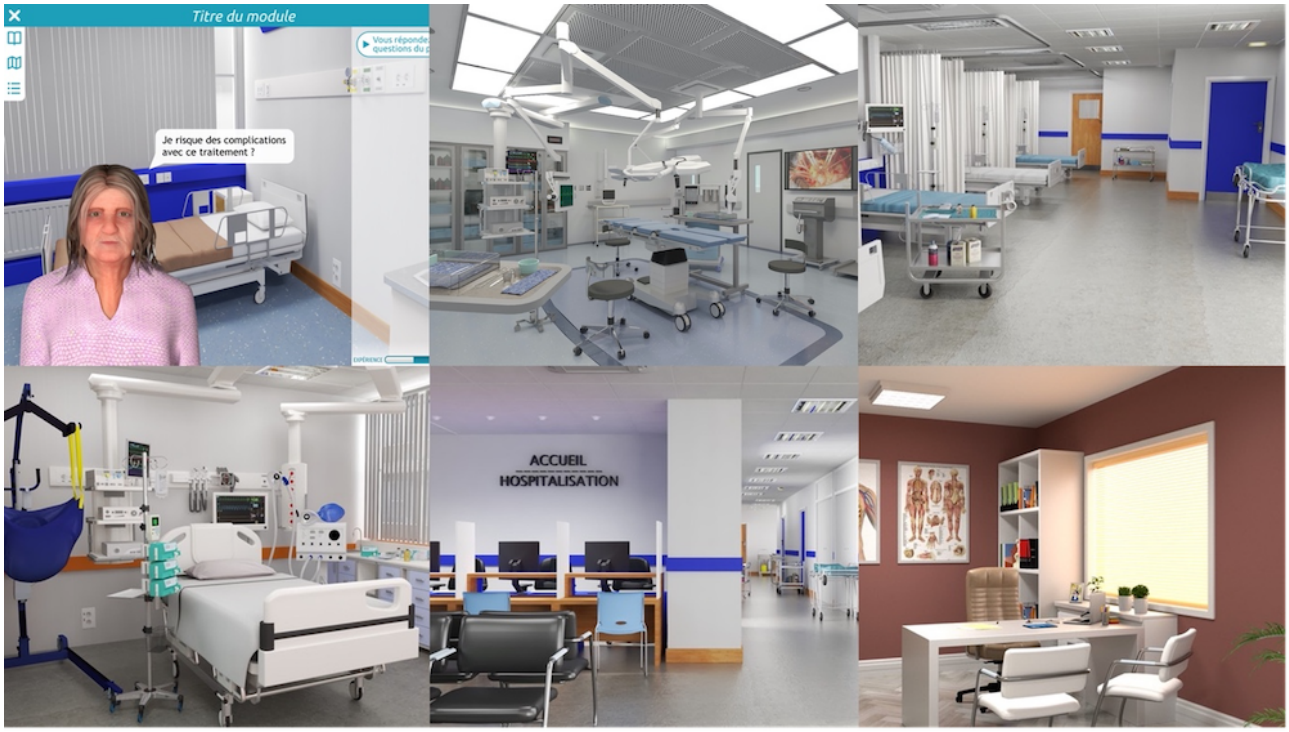
The primary goal of the program is to progressively address all major therapeutic classes included in the official curriculum of the third cycle of medical education. This initial version of the game included 20 scenarios. Students should expect to spend approximately 2 hours to complete this initial version of ADRENALINE. Students respond to various types of questions, including multiple-choice questions, gap-fill exercises, and interactive activities such as identifying errors on prescriptions. The game comprises 3 levels of difficulty. As students progress, they gain increased responsibility in clinical decision-making, managing increasingly complex clinical situations. Their roles evolve from following patient visits to independently handling patient care, then taking emergency shifts, and finally conducting consultations alone directly with patients. A minimum of 60% correct answers is required to successfully complete the scenario, and students must successfully complete at least 80% of the scenarios to unlock the next level. Formative

XSL•FO
RenderX

feedback from teachers, including links to additional resources when relevant, is provided throughout the game following each completed exercise. Players also had opportunities to provide open-ended feedback during the game by communicating with the Professors of Therapeutics via a pedagogical forum on the Moodle platform. They were additionally invited to provide open-ended feedback during the second in-person lecture, which was based on students' questions and the most common errors. It provided professors with the opportunity to clarify medical reasoning in the context of iatrogenic events, explain the rationale behind specific prescriptions, and discuss practical prescribing procedures accompanied by patient counseling advice and real-life follow-up examples.

Prior to this study, students had acquired the necessary clinical knowledge and skills during their therapeutic courses at university.

Figure 2. Examples of ADRENALINE's graphical environment.



Evaluation Measurements

Students were invited to complete an anonymous internet-based pretest survey at the end of the first in-person lecture, as well as a posttest survey at the conclusion of the program, administered during the second in-person lecture. A reminder email containing a QR code linking to the surveys was sent to all sixth-year students, allowing those who did not attend the in-person lecture to complete the posttest survey if applicable. Pretest surveys included questions about students' gaming habits and their intended use of ADRENALINE. The posttest survey was designed to assess their satisfaction after playing the game. Many questions asked students to indicate their level of agreement with statements using a 4-point Likert scale (ranging from "not at all" to "yes, completely") or with yes or no responses. An open-ended question was also included at the end of the posttest survey to capture students' limitations and suggestions for improving ADRENALINE.

Data Management and Analysis

Data were collected anonymously through an internet-based questionnaire. A descriptive analysis of students' responses to the pre- and posttest surveys, their usage patterns of ADRENALINE, and their game performance was conducted using Excel (Microsoft) software.

Ethical Considerations

The study was conducted in accordance with the principles outlined in the Declaration of Helsinki. All data collected were fully anonymized, and all participants provided informed consent electronically for the use of their questionnaire data for research purposes. As this was an observational, noninterventional study using anonymized data, it did not require approval from an ethics review board according to French legislation, specifically in compliance with the Public Health Code [16], which exempts noninterventional studies using anonymous data from mandatory ethics committee review.

Results

Participants

Among the 598 sixth-year students enrolled at Lille University, 272 (45%) logged into the ADRENALINE program at least once and 201 (34%) completed at least one ADRENALINE scenario between November 2023 and March 2024. Among the players, 49/201 (24%) restarted the game at least once. Final score ranged from 16.25 to 100 and the mean score was 71/100 (SD 31.6). Ten (5%) students achieved a perfect score of 100.

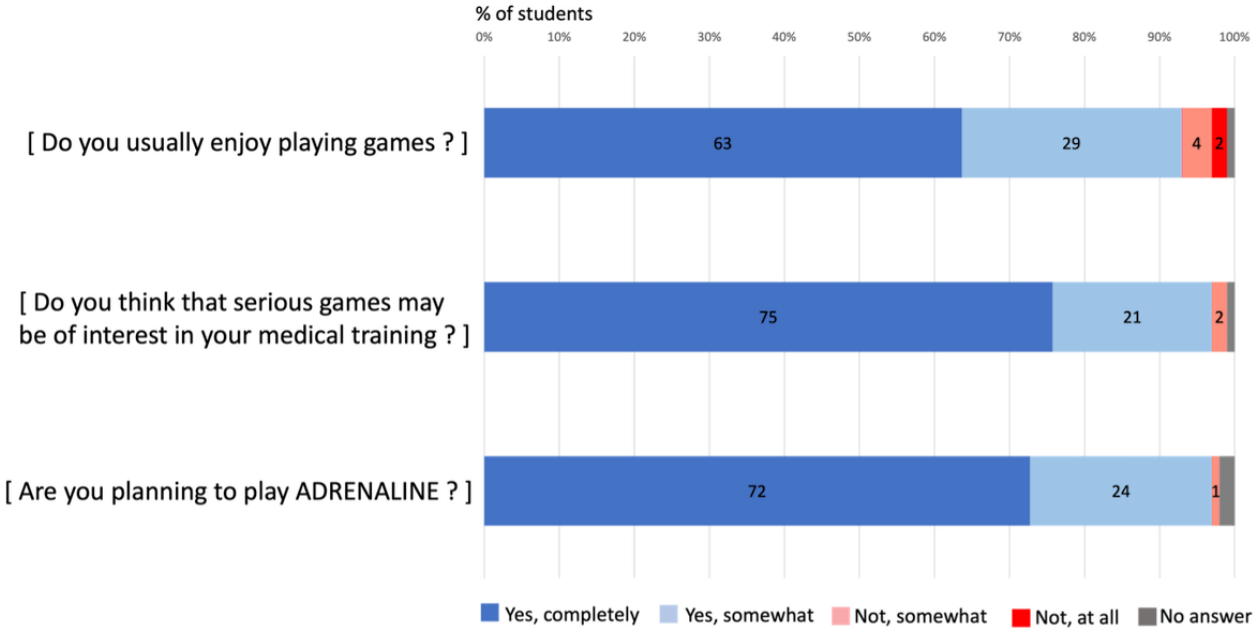
XSL•FO
RenderX

Results of Pretest Surveys

Pretest surveys were collected from 99/598 (17%) sixth-year students. The majority, 92/99 students (93%), identified themselves as gamers and believed that SG could be relevant to their medical training (Figure 3).

Regarding usage pattern, 68/99 (69%) students planned to play alone, 25 (25%) planned to play either alone or in groups, and 5 (5%) planned to play exclusively in groups. Most students intended to use their personal computer to play ADRENALINE; approximately half also planned to use their smartphone, while only 3 (3%) students planned to use a university computer.

Figure 3. Pretest assessment of students' interest in serious games.



Results of Posttest Surveys

Overall, 50/598 (8%) sixth-year students completed the posttest survey. Among them, 7 (14%) did not play to ADRENALINE. Reasons for not playing included lack of time (n=3), insufficient information about the game (n=1), intention to play later in preparation for the national Objective Structured Clinical Examinations (OSCEs) (n=2), and missing data (n=1). Consequently, responses were analyzed for 43 (15.8%) players.

From a technical perspective, students' agreement regarding ease of use and the absence of technical issues was high. Except for 2 students, all played alone, mostly using their own computers. Twelve (24%) students played once, while the others played at least twice. Fifteen students played ADRENALINE more than 4 times.

Approximately 42/50 (85%) players enjoyed playing ADRENALINE and were satisfied with the wording of exercises and their level of difficulty (Figure 4).

The majority of players also believed that ADRENALINE helped improve their prescribing knowledge and skills (Figure 5). For 44/50 (88%) players, ADRENALINE facilitated the application of academic knowledge to real-life clinical settings. Compared to conventional clinical case training, 33/50 (65%) players considered the SG to add significant value for their medical education. Finally, 42/50 (85%) players found ADRENALINE useful for preparing for the national OSCEs.

XSL•FO
RenderX

Figure 4. Posttest assessment of global students' satisfaction.

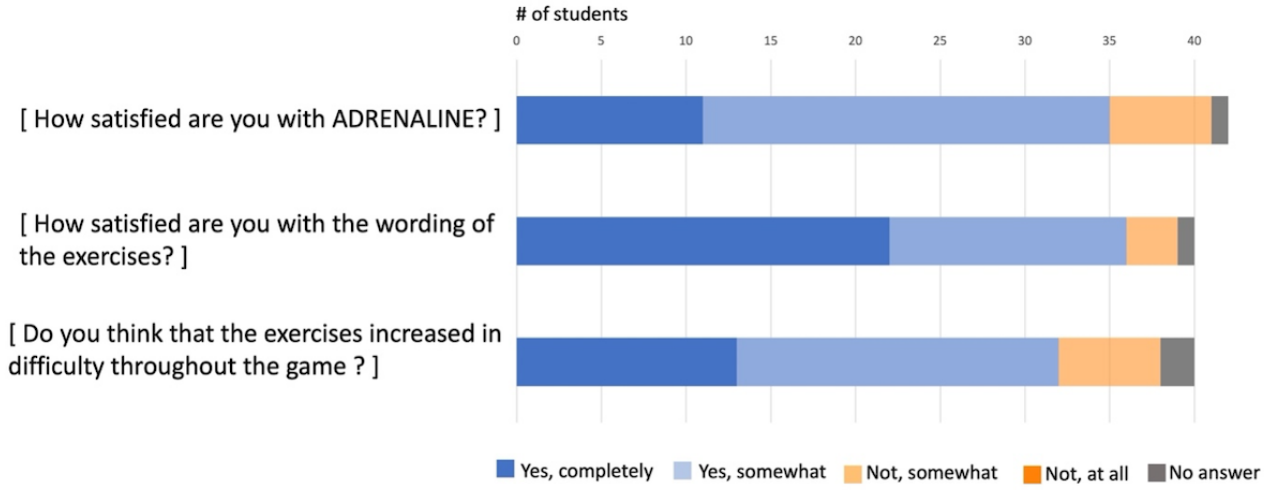
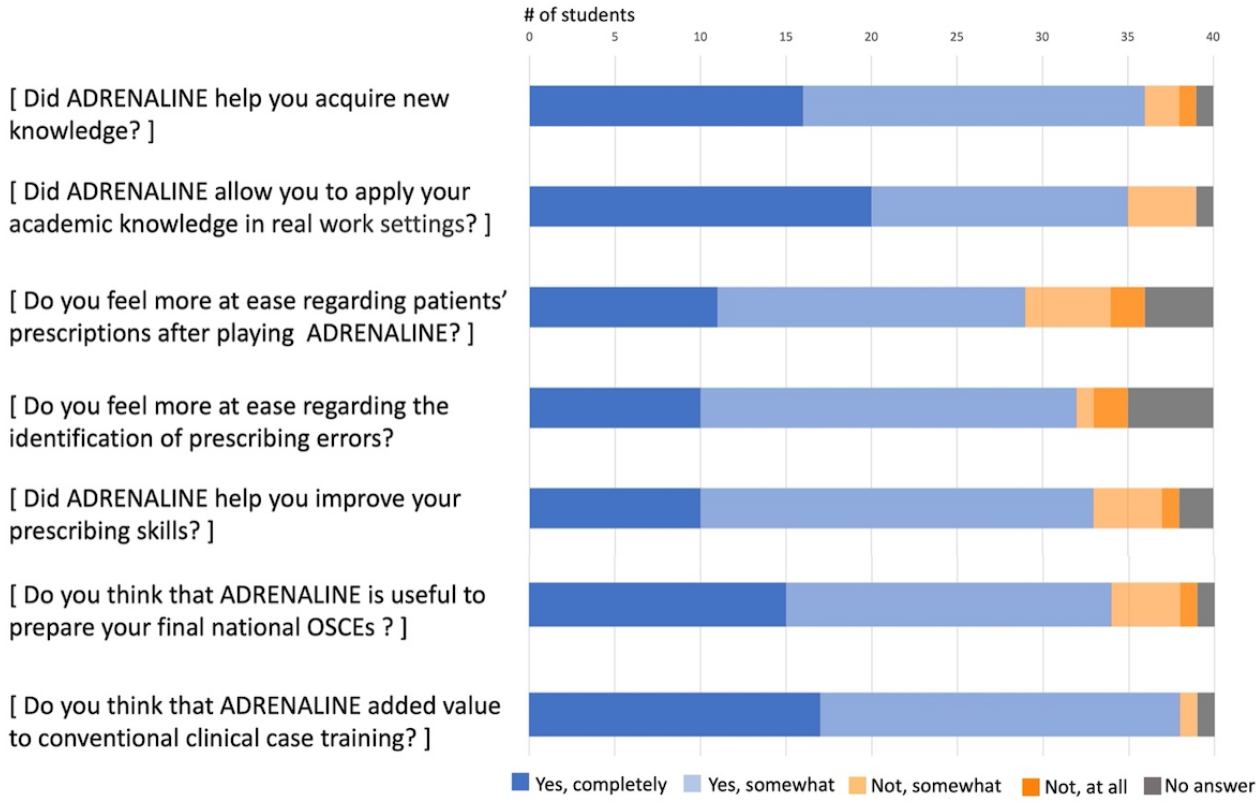


Figure 5. Posttest assessment of students' opinion on ADRENALINE interest in their training.



Discussion

Principal Findings

Our study reported a high level of interest and satisfaction among sixth-year undergraduate medical students with ADRENALINE, an SG designed to enhance prescribing skills prior to their transition to the junior doctor role. Most students found the game enjoyable and perceived it as a valuable tool for improving both their knowledge and prescribing skills, offering added benefit compared to traditional lectures and clinical case–based teaching.

Two major trends are reshaping the traditional landscape of medical education: the decline in classroom attendance and the growing emphasis on competency-based learning that includes not only knowledge, but also skills and attitudes. A consistent global decline in undergraduate medical students' attendance at lectures has been reported worldwide [17,18]. Technological advances in the digital and internet age have likely influenced medical students' expectations for their education. The widespread availability of internet-based, on-demand educational resources may have contributed to a decreased interest in in-person lectures. One study examining the perspectives of medical students and faculty on the benefits of classroom attendance found that most students favored autonomy in choosing learning formats and viewed lecture

XSL•FO
RenderX

recordings as acceptable substitutes for physical attendance [19]. In contrast, only 15% of faculty agreed with this view, instead emphasizing that attendance contributes positively to the development of professionalism. Another study involving first-year osteopathic medical students found no significant correlation between lecture attendance and academic performance, with students showing a preference for lecture capture and podcast as primary modes of content delivery [20]. However, these modalities may be insufficient for developing competencies beyond theoretical knowledge. Prescribing exemplifies this challenge; purely didactic instruction on pharmacology does not necessarily translate into safe and effective prescribing practice. The case-based learning method has been introduced to address this gap by contextualizing medical problems. Evidence from a randomized controlled trial (RCT) involving 80 postgraduate medical students indicated that case-based learning was more effective than traditional lectures in enhancing problem-solving skills [21]. Therefore, educational approaches that combine student autonomy with contextualized case-based methods are particularly needed for complex skills development such as prescribing.

Digital education methods and tools are increasingly being integrated into medical curricula and may help address both the decline in attendance and the need for skill-based learning by fostering autonomy and engagement. Game-based learning, in particular, has seen rapid growth over the past 2 decades [12,22,23]. A substantial majority of students report positive perceptions of SG in terms of engagement and learning [14,20]. The challenge-based structure of SG enables learners to actively participate in decision-making processes and receive immediate feedback, as exemplified by ADRENALINE. Nonetheless, existing meta-analyses have pointed to limitations in the evidence supporting the effectiveness of SG in medical education [24-27]. Methodological shortcomings include a limited number of RCTs, reliance on subjective evaluation metrics (eg, surveys and open-ended questions), and single-center study design. Authors have also noted heterogeneity among the SGs studied and the lack of direct comparisons between similar games in comparable settings. The most recent meta-analysis, which included only RCTs, found no significant differences between SGs and traditional educational methods in terms of knowledge acquisition, knowledge retention, cognitive or procedural skill development, or changes in attitude or behavior [27]. A minor improvement in self-reported confidence was the only consistent benefit associated with SGs. Thus, the added educational value of SGs remains to be firmly established. Our assessment of student satisfaction with ADRENALINE represents an initial step required in its implementation at our institution. The next step will involve evaluating its pedagogical impact by comparing student performance on national OSCEs according to their engagement with the game ADRENALINE. While we recognize that observational study design provides lower-quality evidence than RCTs, we believe conducting an RCT in this context may raise ethical concerns. Restricting access to ADRENALINE for some students could create inequality in preparation for a high-stakes national examination. Furthermore, RCTs may be subject to novelty bias, in which students' enthusiasm for a new educational tool affects outcomes independently of its true pedagogical value.

For educators wishing to develop a similar SG in their own institutions, several key considerations emerged from our experience as developers and were echoed by the Professors of Therapeutics involved in the project. First, the content of clinical cases must be grounded in real-life medical practice to ensure authenticity and relevance. Designing scenarios that accurately reflect the types of prescribing decisions encountered by students and junior doctors was essential. This required close collaboration between content experts and pedagogical designers to strike the right balance between educational clarity and clinical realism. Second, the progression of difficulty within the game was a critical pedagogical choice. Based on our reflections and faculty feedback, we structured the game to move from relatively simple and sometimes exaggerated cases used to introduce basic principles toward more complex, nuanced clinical scenarios. This gradual increase in complexity supported learners' engagement and confidence while fostering deeper integration of knowledge. Third, alignment with learners' training level is essential. We specifically tailored the game to sixth-year medical students, ensuring consistency with the national curriculum and focusing on core pharmacological classes and prescribing competencies. We emphasized the importance of respecting curriculum standards while creating space for clinical reasoning and decision-making. From a development perspective, one of the main challenges was designing cases that were both didactic and clinically realistic, while remaining feasible within the format of a short interactive game. Another key learning was the value of integrating real-time feedback and debriefing after each scenario, which enhanced learner engagement and reflection. The involvement of experienced faculty members also helped ensure that the learning goals remained central throughout the design process. Finally, a major strength of this approach is its adaptability. From a perspective point of view, the game was designed to be transferable across institutions, allowing the platform to be enriched through the inclusion of a wider variety of clinical scenarios contributed by different faculties. This collaborative approach also supports the harmonization of therapeutics education across French medical schools. Furthermore, the national validation of the game's content by the CNET ensures consistency with curricular standards and reinforces its pedagogical legitimacy. We believe these insights and practical considerations may assist educators in developing effective, scalable, and pedagogically robust SGs in medical education.

ADRENALINE has several limitations. First, only 201/598 (30%) sixth-year medical students at our institution participated in the game ADRENALINE during its first year of implementation, and the reasons for nonparticipation remain unknown. To improve uptake, we plan to send reminder emails highlighting the platform's availability and its relevance for national OSCE preparation. The program will also be introduced during the university's induction day. In addition, we aim to encourage student-led sharing of the game through informal networks and social media to increase awareness and engagement. Second, the high cost and time investment required to develop an SG may limit its widespread adoption. In the case

XSL•FO
RenderX

of ADRENALINE, the licensing costs for the MOSAIC software were already covered by our institution, and its modular design allows for adaptation to various professional contexts. This enables potential implementation in other institutions without additional development cost, as the game has already been developed and tested at Marseille University. An agreement must be signed between Lille University and any institution requesting access to ADRENALINE for their students. We only ask instructors in therapeutics to contribute 5 new scenarios to enrich the game. This contrasts with the majority of previously published SGs (90.4%), which were limited to single-institution use. Third, as we tested a pilot version of the game, the number of available scenarios was limited. This restricts the range of prescribing situations represented, which may not fully reflect the diversity of real-world cases encountered by junior doctors. Therefore, the current version of ADRENALINE is not yet sufficient to evaluate its overall impact on students' prescribing abilities. Finally, the types of exercises within the game are constrained by the capabilities of the MOSAIC software, which limits students' ability to provide free-text or open-ended responses, unlike in real-life practice. Future developments, potentially incorporating artificial intelligence and chatbot technologies, could help address this limitation by enabling more naturalistic, flexible interaction.

ADRENALINE also has many strengths. First, the game integrated key factors known to enhance learning: attention, facilitated by the game's visually engaging interface, active cognitive engagement supported by interactive, case-based real-life scenarios and a scoring system, feedback within the game from educators, and consolidation achieved through repetition of scenarios and a progressive increase in complexity. Second, ADRENALINE offers learners the opportunity to engage with a wide variety of realistic scenarios in a hospital setting, promoting contextualized learning. By playing as themselves, students are immersed in authentic clinical situations that closely mirror clinical reality. This experiential approach aligns with Kolb experiential learning theory, enhancing retention and transfer of skills to real practice. Finally, ADRENALINE is compatible with a wide range of digital devices and can be accessed anytime and anywhere, either individually or collaboratively. Its open accessibility allows for large-scale deployment across institutions, supporting wider dissemination and potentially contributing to national standardization of CPT education.

Conclusion

In conclusion, despite a relatively low participation rate, with only 201/598 (33.6%) sixth-year students engaging with the game, a substantial proportion of users reported high levels of interest and satisfaction with the SG designed to improve prescribing skills. ADRENALINE appears to offer added educational value alongside conventional teaching methods in preparing students both for the national OSCE and for their future responsibilities as prescribers. Further research is now needed to assess the educational effectiveness of ADRENALINE, either as a standalone tool or in combination with other learning resources currently available within our institution.

Acknowledgments

We thank all members of GIVRE from the faculty of pharmacy of Lille for their support for the ADRENALINE design. For the financial support, we thank Aix-Marseille University with the grant from Fonds d'Intervention Pédagogique (FIP) 2020 campaign, and the French National College of Therapeutics (CNET).

Data Availability

All data generated or analyzed during this study are included in this published article.

Authors' Contributions

CMY, YP, and KQ contributed to drafting the article, gave substantial contributions to study conception and design, and to the interpretation of data. All co-authors are revising the manuscript critically for important intellectual content and final approval of the version of the article to be published.

Conflicts of Interest

None declared.

References

1. Brinkman DJ, Tichelaar J, Graaf S, Otten RHJ, Richir MC, van Agtmael MA. Do final-year medical students have sufficient prescribing competencies? A systematic literature review. Br J Clin Pharmacol 2018 Apr;84(4):615-635. [doi: 10.1111/bcp.13491] [Medline: 29315721]
2. Dean B, Schachter M, Vincent C, Barber N. Causes of prescribing errors in hospital inpatients: a prospective study. Lancet 2002 Apr 20;359(9315):1373-1378. [doi: 10.1016/S0140-6736(02)08350-2] [Medline: 11978334]
3. Leendertse AJ, Egberts ACG, Stoker LJ, van den Bemt PMLA, HARM Study Group. Frequency of and risk factors for preventable medication-related hospital admissions in the Netherlands. Arch Intern Med 2008 Sep 22;168(17):1890-1896. [doi: 10.1001/archinternmed.2008.3] [Medline: 18809816]

XSL•FO
RenderX

4. Ryan C, Ross S, Davey P, et al. Prevalence and causes of prescribing errors: the PRescribing Outcomes for Trainee Doctors Engaged in Clinical Training (PROTECT) study. PLoS One 2014;9(1):e79802. [doi: 10.1371/journal.pone.0079802] [Medline: 24404122]
5. Ryan C, Ross S, Davey P, et al. Prevalence and causes of prescribing errors: the PRescribing Outcomes for Trainee Doctors Engaged in Clinical Training (PROTECT) study. PLoS ONE 2014;9(1):e79802. [doi: 10.1371/journal.pone.0079802] [Medline: 24404122]
6. Ashcroft DM, Lewis PJ, Tully MP, et al. Prevalence, nature, severity and risk factors for prescribing errors in hospital inpatients: prospective study in 20 UK hospitals. Drug Saf 2015 Sep;38(9):833-843. [doi: 10.1007/s40264-015-0320-x] [Medline: 26115701]
7. Brinkman DJ, Tichelaar J, Schutte T, et al. Essential competencies in prescribing:a first European cross-sectional study among 895 final-year medical students. Clin Pharmacol Ther 2017 Feb;101(2):281-289. [doi: 10.1002/cpt.521] [Medline: 27648725]
8. Brinkman DJ, Tichelaar J, Okorie M, et al. Pharmacology and therapeutics education in the European Union needs harmonization and modernization: a cross-sectional survey among 185 medical schools in 27 countries. Clin Pharmacol Ther 2017 Nov;102(5):815-822. [doi: 10.1002/cpt.682] [Medline: 28295236]
9. Bakkum MJ, Tichelaar J, Wellink A, Richir MC, van Agtmael MA. Digital learning to improve safe and effective prescribing: a systematic review. Clin Pharmacol Ther 2019 Dec;106(6):1236-1245. [doi: 10.1002/cpt.1549] [Medline: 31206612]
10. Bakkum MJ, Tichelaar J, Papaioannidou P, et al. Harmonizing and improving European education in prescribing: an overview of digital educational resources used in clinical pharmacology and therapeutics. Br J Clin Pharmacol 2021 Mar;87(3):1001-1011. [doi: 10.1111/bcp.14453] [Medline: 32638391]
11. George S. Games, simulations, immersive environments, and emerging technologies tatnall a. In: Encyclopedia of Education and Information Technologies: Springer International Publishing; 2019:1-10. [doi: 10.1007/978-3-319-60013-0_36-1]
12. Xu M, Luo Y, Zhang Y, Xia R, Qian H, Zou X. Game-based learning in medical education. Front Public Health 2023;11:1113682. [doi: 10.3389/fpubh.2023.1113682] [Medline: 36935696]
13. Graafland M, Schraagen JM, Schijven MP. Systematic review of serious games for medical education and surgical skills training. Br J Surg 2012 Oct;99(10):1322-1330. [doi: 10.1002/bjs.8819] [Medline: 22961509]
14. Zairi I, Ben Dhiab M, Mzoughi K, Ben Mrad I, Kraiem S. Assessing medical student satisfaction and interest with serious game. Tunis Med 2021;99(11):1030-1035. [Medline: 35288906]
15. Oliver-Quelennec K, Bouchet F, Carron T, Pinçon C. Analyzing the impact of e-Caducée, a serious game in pharmacy on students' professional skills over multiple years. Presented at: CSEDU 2021 - 13th International Conference on Computer Supported Education; Sep 24, 2021 p. 331-338. [doi: 10.5220/0010453903310338]
16. Public Health Code: Articles L1110-1 to L6441-1 [Article in French]. Legifrance. URL: https://www.legifrance.gouv.fr/codes/section_lc/LEGITEXT000006072665/LEGISCTA000006103419/ [accessed 2025-09-01]
17. Gardner G, Feldman M, Santen SA, Mui P, Biskobing D. Determinants and outcomes of in-person lecture attendance in medical school. Med Sci Educ 2022 Aug;32(4):883-890. [doi: 10.1007/s40670-022-01581-2] [Medline: 35821745]
18. Kanter SL. To be there or not to be there: is attendance really the question? Acad Med 2012 Jun;87(6):679. [doi: 10.1097/ACM.0b013e31825a5a2a] [Medline: 22643365]
19. Zazulia AR, Goldhoff P. Faculty and medical student attitudes about preclinical classroom attendance. Teach Learn Med 2014;26(4):327-334. [doi: 10.1080/10401334.2014.945028] [Medline: 25318026]
20. Pettit RK, McCoy L, Kinney M, Schwartz FN. Student perceptions of gamified audience response system interactions in large group lectures and via lecture capture technology. BMC Med Educ 2015 May 22;15:92. [doi: 10.1186/s12909-015-0373-7] [Medline: 25997953]
21. Bi M, Zhao Z, Yang J, Wang Y. Comparison of case-based learning and traditional method in teaching postgraduate students of medical oncology. Med Teach 2019 Oct;41(10):1124-1128. [doi: 10.1080/0142159X.2019.1617414] [Medline: 31215320]
22. Oestreich JH, Guy JW. Game-based learning in pharmacy education. Pharmacy (Basel) 2022 Jan 6;10(1):11. [doi: 10.3390/pharmacy10010011] [Medline: 35076609]
23. Byusa E, Kampire E, Mwesigye AR. Game-based learning approach on students' motivation and understanding of chemistry concepts: a systematic review of literature. Heliyon 2022 May;8(5):e09541. [doi: 10.1016/j.heliyon.2022.e09541] [Medline: 35663764]
24. Gorbanev I, Agudelo-Londoño S, González RA, et al. A systematic review of serious games in medical education: quality of evidence and pedagogical strategy. Med Educ Online 2018 Dec;23(1):1438718. [doi: 10.1080/10872981.2018.1438718] [Medline: 29457760]
25. Sipiyaruk K, Gallagher JE, Hatzipanagos S, Reynolds PA. A rapid review of serious games: from healthcare education to dental education. Eur J Dent Educ 2018 Nov;22(4):243-257. [doi: 10.1111/eje.12338] [Medline: 29573165]
26. Gentry SV, Gauthier A, L'Estrade Ehrstrom B, et al. Serious gaming and gamification education in health professions: systematic review. J Med Internet Res 2019 Mar 28;21(3):e12994. [doi: 10.2196/12994] [Medline: 30920375]
27. Maheu-Cadotte MA, Cossette S, Dubé V, et al. Efficacy of serious games in healthcare professions education: a systematic review and meta-analysis. Simul Healthc 2021 Jun 1;16(3):199-212. [doi: 10.1097/SIH.0000000000000512] [Medline: 33196609]

XSL•FO
RenderX

Abbreviations:

CNET: French National College of Therapeutics
CPT: Clinical Pharmacology and Therapeutics
OSCE: Objective Structured Clinical Examination
RCT: randomized controlled trials
SG: serious game

Edited by A Coristine; submitted 17.09.24; peer-reviewed by E Vallefuoco, L Ng; revised version received 24.07.25; accepted 30.07.25; published 11.09.25.

Please cite as:
Yelnik CM, Daumas A, Poteaux Y, Standaert A, Grimbert N, Favory R, Ravaux P, Lambert M, Quelennec K
ADRENALINE, a Learning Game to Improve Prescribing Skills in Undergraduate Medical Students: Descriptive Study
JMIR Serious Games 2025;13:e66334
URL: https://games.jmir.org/2025/1/e66334
doi:10.2196/66334

© Cecile Marie Yelnik, Aurélie Daumas, Yanele Poteaux, Annie Standaert, Natacha Grimbert, Raphaël Favory, Pierre Ravaux, Marc Lambert, Katia Quelennec. Originally published in JMIR Serious Games (https://games.jmir.org), 11.9.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Efficacy of the Web-Based Gamified Infection Control Training System on Practices for Health Care Workers in Residential Care Homes: Clustered Randomized Controlled Trial

Angela Y M Leung[1,2,3], PhD; Doris Y P Leung[1], PhD; Terence K Lau[1], PhD; Justina Y W Liu[1,3,4], PhD; Teris Cheung[1], PhD; Daphne S K Cheung[1,5,6], PhD; Simon C Lam[7], PhD; Eliza M L Wong[7], PhD; Mimi M Y Tse[8], PhD; Alex Molassiotis[9], PhD

[1]School of Nursing, Hong Kong Polytechnic University, Hong Kong, China (Hong Kong)

[2]WHO Collaborating Centre for Community Health Services, School of Nursing, Hong Kong Polytechnic University, Hong Kong, China (Hong Kong)

[3]Research Institute for Smart Ageing, Hong Kong Polytechnic University, Hong Kong, China (Hong Kong)

[4]Research Centre for Assistive Technology, Hong Kong Polytechnic University, Hong Kong, China (Hong Kong)

[5]School of Nursing and Midwifery, Faculty of Health, Deakin University, Melbourne, Australia

[6]Centre for Quality and Patient Safety Research/Alfred Health Partnership, Institute for Health Transformation, Deakin University, Melbourne, Australia

[7]Tung Wah College, Hong Kong, China (Hong Kong)

[8]Hong Kong Metropolitan University, Hong Kong, China (Hong Kong)

[9]College Arts, Humanities and Education, University of Derby, Derby, United Kingdom

Corresponding Author:
Angela Y M Leung, PhD
School of Nursing
Hong Kong Polytechnic University
5th floor, Core G
Hong Kong
China (Hong Kong)
Phone: 852 27665587
Fax: 852 23649663
Email: angela.ym.leung@polyu.edu.hk

Abstract

Background: Staff working in residential care homes (RCHs) have played a significant role in preventing the spread of infection among residents, visitors, and staff. Providing continuous professional training to the staff is essential. Current infection control training mostly rests on short educational talks or one-to-one reminders in the RCHs. A blended mode of online interactive games and face-to-face consultations was now proposed as a new way to conduct infection control training in the RCHs.

Objective: This study aims to assess the efficacy of the Blended Gaming COVID-19 Training System (BGCTS) on infection control practices and self-reported knowledge, attitude, and practices of standard precautions among health care workers in RCHs.

Methods: A 2-arm, single-blinded, parallel cluster randomized controlled trial was designed, and 30 RCHs were recruited and randomized into an intervention group to receive the BGCTS and a control group to receive usual care on infection control training. Due to the COVID-19 pandemic and infected cases in the homes, 17 RCHs refused or delayed the on-site observations. The BGCTS intervention, developed based on "The COVID-19 Risk Communication Package for Healthcare Facilities" of the World Health Organization, consists of two parts: (1) an eHealth mode of a 120-minute web-based training system covering 8 topics, delivered in short-clip videos and games, and (2) two 30-minute face-to-face interactive sessions for concept clarification. The 2 infection control practices, "use of gloves and personal protective equipment (PPE) and performing respiratory hygiene" and "hand rub," were assessed by on-site unobtrusive observations, and self-reported infection control practices and knowledge and attitude toward infection control were measured via online survey post intervention.

Results: A total of 212 staff from 13 RCHs were involved in the analysis, with 7 RCHs from the intervention group (n=114) and 6 RCHs from the control group (n=98). A significantly greater increase in the proportions of proper use of gloves and PPE and respiratory hygiene performance (β=.195, 95% CI 0.046-0.344; *P*=.02) and properly performed hand rub (β=.068, 95% CI

XSL•FO
RenderX

0.005-0.132; *P*=.04) was observed in the intervention group. The changes in the self-reported outcomes were not statistically significant.

Conclusions: BGCTS improved RCH staff's performance in 2 infection control practices by objective measurement, "gloves and PPE use and performance in respiratory hygiene" and "hand rub." BGCTS was shown to be an effective training, although it was a 2-week intervention. The BGCTS did not perform better than infection control briefing sessions in self-reported infection control knowledge, attitude, and practices. This electronic-based infection control training with 2 intensive interactive sessions has good potential to be adopted as regular training in RCHs.

Trial Registration: Clinicaltrials.gov NCT04783025; http://clinicaltrials.gov/ct2/show/NCT04783025

(JMIR Serious Games 2025;13:e71593) doi:10.2196/71593

KEYWORDS

clustered randomized controlled trial; gamification; hand washing; infection control training; personal protective equipment; residential care homes

Introduction

Residents in residential care homes (RCHs) are at a high risk of being infected by severe acute respiratory diseases (SARD) due to their advanced age, associated comorbidities, and state of dependence [1]. The price paid for an outbreak in RCHs is high. The pace at which SARD is spreading in RCHs is alarming. In a case study in Washington, after a first resident in RCHs was diagnosed with SARD, 101 residents, 50 health care personnel, and 16 visitors were infected, and the cases were proven to be epidemiologically linked [2]. One outbreak of a cluster of SARD in one RCH was enough to fill all of the beds set aside for SARD cases in a whole province in Korea [3]. Based on such evidence and many other reports around the globe, it appears clear that there is an urgent need to develop and implement an effective strategy to prevent the spread of infectious diseases in RCHs [4].

Staff working in RCHs have played a significant role in responding to the outbreak, particularly in preventing the spread of infection among residents, visitors, and staff. RCH staff have the responsibility to protect the residents, themselves, and visitors. Training the staff in RCHs is crucial because they are the gatekeepers of the safety of the residents. Unfortunately, most RCHs are unable to educate and train their employees due to a lack of resources [3]. It is suggested that governments provide an educational system for health care personnel in RCHs [3].

Globally, there is a lack of infection control training for staff in RCHs; even if training is provided, the existing online training programs are generic in nature, not specifically for SARD or COVID-19. Australia makes available e-learning modules to provide information on the principles of infection prevention and control in Australian health care settings [5]. These e-learning modules are suitable for all health care workers, including staff of RCHs. They include information on the principles of infection prevention and control; risk management systems for infectious agents and infectious diseases; basic microbiology and multiresistant organisms; cleaning, disinfection, and sterilization; infectious agent health screening and the immunization of health care workers; the investigation and management of outbreaks; the management of occupational exposure; risk management in renovation, repairs, and redevelopment; basic epidemiology; and surveillance and quality improvement. However, none of the material is specific to SARD or COVID-19. Only links to online resources about COVID-19 are provided, but no training is provided to RCH staff. A similar situation was noted in the United Kingdom and Hong Kong.

Infection control is a core element in preventing the spread of respiratory infections. However, low compliance with infection control practices poses significant challenges in the long-term care setting; in particular, nonmedical professionals were less frequently adherent to these practices than medical professionals [2,6,7]. Before the outbreak of the COVID-19 pandemic, various educational interventions for infection control had been developed. A recent systematic review (SR), based on a narrative synthesis of 11 interventional studies, suggested the preliminary benefits of educational interventions on attitudes and practices related to respiratory infection prevention. Meta-analytic results on 3 studies further showed significant improvements in knowledge about preventing respiratory diseases, although the evidence was of low certainty. The review also suggested that the use of multiple resources that can favor learning and stimulate hearing, vision, and touch in educational interventions may produce better results [8]. Despite these favorable findings, none of the 11 studies included RCH staff, which limits the generalizability to the long-term care facilities setting. Another SR reviewed articles evaluating the effect of infection prevention and control measures in the long-term care facilities setting and concluded that hand hygiene and oral hygiene have a beneficial effect on infection rates in the nonoutbreak setting. In contrast, infection prevention and control bundles seemed to be promising in an outbreak setting, but the finding was based on nonexperimental studies with a high risk of bias [9]. More recently, studies have been conducted to evaluate the impact of educational interventions that were developed in response to the COVID-19 pandemic. For example, an educational training program targeting health care workers in isolation centers could increase knowledge, attitude, and practices of infection prevention and control for COVID-19 [10]. Another study tested the effect of hand hygiene training among health care workers in a respiratory disease hospital and reported significant improvements in compliance with hand hygiene protocol and knowledge [11]. Other studies tested video-based educational interventions on knowledge, attitude, and practices during the

XSL•FO
RenderX

COVID-19 pandemic, with promising results [12,13]. Again, these studies usually had used a pre-post study design without a comparison control group, affecting the internal validity of the study findings. The literature indicates that more studies with high quality are in urgent need for testing educational interventions in infection control, in particular to respond to outbreaks of SARD in the future.

Gamification is widely used as an educational tool in different disciplines. Various reviews have highlighted the wide applications and benefits of adopting gamification as a multidisciplinary educational tool. Gamified interventions have been applied across multiple domains, including science, technology, engineering, mathematics, linguistics, business, and medicine, to improve participants' learning motivation [14]. A recent pre-post interventional study has applied gamified infection control training to health care workers in the hospital setting and reported significantly increased hand-rub consumption in a gastroenterological ward [15]. Furthermore, there are several potential benefits of the use of gamified training in infection control, in particular in the COVID-19 context. First, gamified training for infection control helps learners to understand the relationship between the various steps in infection control. If a learner does not strictly follow the infection control practices, widespread infection will occur in an RCH. If such a situation had taken place in reality, the residents would have suffered real harm. Second, gamified training provides feedback to learners, which helps them to understand the serious consequences of improper infection control practices and provides opportunities for improvement. Third, learners can choose to review the contents in a playful way, resulting in an increase in engagement in training. In summary, the gamified training for COVID-19 infection control provides a safe and flexible learning environment for the staff of RCHs to learn and practice decision-making.

In Hong Kong, infection control training is available largely to health care personnel working in hospitals or clinics and other health care settings. The Hong Kong Training Portal on Infection Control and Infectious Diseases offers programs only to staff from hospitals or clinics and other health care settings [16]. Since all RCHs in Hong Kong are under the jurisdiction of the Social Welfare Department or are privately operated, health care personnel working in RCHs are not eligible to receive this kind of infection control training. Overall, 2 universities are offering year-long infection control certificate programs to health professionals; the contents focus mostly on the daily management of infection control measures in hospitals and clinics [17]. In addition, the high turnover rate and demanding working time commitment in RCHs [18] also call for a flexible, interesting learning module in delivering infection control educational programs. Reviewing the existing infection control training offered around the globe, it is clear that there is a significant gap in the area of providing training in the prevention and control of SARD infections for health care personnel working in RCHs.

It was reported that nearly 5 billion people in the world have a mobile device [19]. Most staff in RCHs own a smartphone and have the technological competence to manage digital information and an electronic device [20]. Hence, smartphones and electronic devices are good platforms for training. In view of the demand for infection control training for staff in RCHs, the widespread use of smartphones, and the significant gaps in knowledge about the efficacy of infection control training on their actual practices, we developed a system called "Blended Gaming COVID-19 Training System (BGCTS)" for the staff of RCHs. Both electronic learning (with games and interactive quizzes) and face-to-face counseling sessions were used to help RCH staff to improve their infection control practices.

This study aims to assess the effect of the BGCTS on infection control practices among all staff in RCHs. The specific objectives of the study include:

1. To assess whether more health care workers in the intervention group performed proper infection control practices (objectively measured in unobtrusive on-site observations) than those in the control group after receiving infection control training.

2. To assess whether there were greater improvements in self-reported infection control knowledge, positive attitudes toward infection control, and infection control practices in health care workers in the intervention group than those in the control group.

Methods

Design and Setting

This is a 2-arm, single-blinded, cluster randomized controlled trial (RCT) conducted at RCHs in Hong Kong. The clusters of this study were RCHs in Hong Kong. Health care providers working in the same RCH formed a cluster to avoid potential contamination across study arms due to the sharing of the intervention content among staff in the same RCH. A total of 2 assessments were conducted at baseline and post intervention, respectively. This study adhered to the reporting guidelines of the Consolidated Standards of Reporting Trials (CONSORT) with an extension to cluster RCTs (Multimedia Appendix 1) [21].

Participants

Inclusion criteria for participants were adult health care personnel who (1) provided direct care to the residents in the RCH, (2) were able to read Chinese, (3) possessed electronic devices (smartphones or tablets) for downloading the BGCTS, and (4) knew how to and were able to swipe the screens of electronic devices with their fingers. Exclusion criteria included (1) temporary staff who were going to cease their employment, be on maternity leave, or on a long vacation during the data collection period; (2) students or trainees (including nursing students, physiotherapy students, and health-related trainees), because training institutions were responsible for their training, and such training might affect their performance; (3) volunteer personnel who came to the RCHs at a specific period (eg, to organize activities for the residents during festivals); (4) clerical and administrative staff, kitchen staff, security staff, and engineering and facilities management staff; and (5) ambulance officers who came to the RCHs to take residents to or from hospitals.

XSL•FO
RenderX

The sample size was determined based on infection control practices. An observational study conducted by the project team in 2019 showed that the mean rate of compliance with hand hygiene measures among health care workers in RCHs was about 10% [22]. Assuming that the hand hygiene compliance rate in the control group would increase to 20% (ie, an increment of 10%) because of the COVID-19 pandemic and that the BGCTS could boost the compliance rate of the intervention group to 40% (ie, an increment of 30%), 138 health care workers would be needed to achieve 80% power at a 5% confidence level in an RCT. For a cluster RCT, assuming that 10 health care providers could be recruited from each RCH, that the intrainstitutional correlation (as measured by the intraclass correlation coefficient) was 0.04, and hence the design factor = 1+(10–1) × 0.04 = 1.36, the total sample size required was 188 (=138×1.36) participants or 20 RCHs. Further, allowing an attrition rate of 30% post intervention, we planned to recruit 30 RCHs to achieve the target sample size.

Randomization and Masking

The clusters were randomly allocated in a 1:1 ratio to one of the two arms: (1) web-based gamified infection control training for 2 weeks and (2) usual care of a 30-minute infection control briefing. A statistician, who was not involved in the participant recruitment, generated a list of random numbers for group assignment using an online randomizer [23] and concealed the sequence using opaque sealed envelopes, which were opened by the research assistant during the group allocation procedure. The outcome assessor was blinded to the group assignment to avoid undue influence during the data collection process.

Study Arms

The Intervention—BGCTS

The health care providers in the intervention group received a web-based gamified infection control training program (BGCTS) on an individual basis. The aims of the BGCTS are to promote the principles of infection control and reinforce the actions needed to stop the spread of COVID-19 within RCHs and how to provide care for those residents suspected of having COVID-19. The BGCTS has eight topics, namely (1) preparing for COVID-19 at your health care facility, (2) managing patients with suspected or confirmed COVID-19 at your health care facilities, (3) protecting yourself at work from COVID-19, (4) personal protective equipment (PPE) according to health care activities, (5) communicating with patients with suspected or confirmed COVID-19, (6) information about COVID-19, (7) coping with stress, and (8) my 5 moments for hand hygiene. These contents were developed based on the evidence-based contents from the World Health Organization (WHO) guideline "The COVID-19 Risk Communication Package for Healthcare Facilities" [24] and the infection control guidelines for care facilities developed by Hong Kong Government [16,25], with modifications based on feedback from stakeholders (Multimedia Appendix 2).

Elements of game design (ie, short video clips and interactive video games in the form of quizzes) in nongame environments (RCHs) within nongame contexts (infection control training) were adopted in the program development. In one of the games, the participants were asked to choose the essential PPE for their work at different sites in the RCHs. For example, in the triage scenario, where the staff received relatives or visitors who intended to enter the RCHs, staff should put on a medical mask only. A score as a reward was given for giving the correct answer to the question. Scores and a progress bar were shown to the staff so as to motivate their continuous participation in the training.

The participants were asked to complete the 8 topics in 2 weeks via the web-based system, which can be operated on laptops or mobile phones. They were encouraged to learn each topic independently in 15 minutes, resulting in a total of 120 minutes to complete all 8 topics in 2 weeks. In addition, in each week, a 30-minute face-to-face or Zoom (Zoom Communications, Inc) interactive session after they played the games was provided to allow the participants to clarify the infection control concepts with the infection control officer, the nurse who was trained for infection control, of the RCH. The final score achieved by the participant was shown on the screen and was also recorded in the scoring system so that the infection control officer can check which questions the participant could not answer correctly. Each participant was given an individual account with a unique username and password to access the platform [26]. The password could be reset by clicking the button "forgot the password." Then, a new password would be sent to the participant's email. Access to the platform was unlimited in the designated 2-week period. If the participant stopped playing the game at a particular point, the system had the record. When the participant played the game again, the system would continue the game and let the participant play the game at the point where the game was stopped. Online reminder was not made; however, the nurse managers in the RCHs verbally reminded the participant to complete the training within the designated period. No subscription fee to the BGCTS was charged in this study; however, a nominal subscription fee will be applied in the future to maintain the operation of the system.

Usual Care

Participants in the control group received the infection control briefing session by the infection control officer, who was either a registered nurse or a health worker appointed to handle matters related to infection control in the RCH. The appointment of an infection control officer is a licensing requirement of RCH in Hong Kong [27]. However, the briefing session was usually irregular, unstructured, and determined by infection control officers, depending on when the infection control officers receive the information about infection control practice from the Centre for Health Protection of the Hong Kong Government [25]. In addition, the format of the briefing session was mostly face-to-face in the form of a verbal announcement or short talk that lasted for less than 30 minutes, determined by the infection control officer. It is also common that the infection control information was delivered in the form of posters or a written document and circulated to all staff in the RCH, partly due to the high turnover rate in RCHs [18].

XSL•FO
RenderX

Measures

Overview

The study outcomes were measured at 2 time points, baseline and post intervention, following group allocation. The primary outcome of the study was infection control practices performed by RCH staff, which is an objective measure made by on-site unobtrusive observations by trained research nurses (offline). Secondary outcomes included self-reported infection control practice and knowledge and attitudes toward infection control [28], which were included in an online self-reported survey completed by all the participants at baseline and upon completion of the training. Participants were encouraged to complete the survey within 7 days after completion of the intervention. Data at the backend of the system were retrieved to monitor the use of the system (how many times the staff used each chapter and the correctness of the answers to each question).

Infection Control Practices Performed by RCH Staff

On-site observation of the infection control practices performed by the RCH staff was the primary outcome of this study. Research nurses were trained to conduct unobtrusive on-site observations in the RCHs when the participants were conducting direct care to the residents (older adults). Unobtrusive means the research nurse would not interfere with the participants' performance during observation even though the practices were not in line with the infection control guidelines. To minimize bias, each RCH was observed for at least 200 moments at baseline and another 200 moments post intervention. More than one observation session was conducted by the trained research nurse in each RCH in order to obtain at least 200 moments.

An e-rub system was developed by a Swedish company based on the WHO guideline [29], and it was adopted to support the recordings of the observations in the RCHs. The following five infection control practices were observed: (1) hand wash, (2) hand rub, (3) use of gloves, (4) use of PPE, and (5) time spent on hand wash. Haze factors such as wearing rings, bracelets, watches, and long sleeves were also logged in the system. Each observation moment was judged by the research nurse as to whether it was performed properly, improperly, missing, or not applicable. Considering the actual practice in RCHs, "hand hygiene" was defined as "proper hand wash or hand rub." Proper hand wash referred to hand washing that lasted for 20 seconds or more. Hand rub refers to the use of 70% alcohol for rubbing both hands. We created 2 binary outcome variables based on the on-site observations, namely "use of gloves and PPE and performing respiratory hygiene" and "hand rub," with 2 options of properly performed versus improperly performed. For the first variable, "properly performed" refers to the situation that the participants had properly used the gloves, performed respiratory hygiene (ie, covered mouth and nose with the upper part of the sleeve while coughing or sneezing) [30], and used the PPE, while for the second variable, "hand rub," properly performed refers to the situation that the participants did the hand rub properly. The variable, hand rub, is of particular importance, as the staff could not change their gloves and wash hands all the time during the pandemic. Observation moments varied according to the complexity of the care given to the residents; therefore, the total number of moments in each observation varied.

Training was provided to the research nurses to ensure consistent data collection. They were qualified registered nurses with rich experience in patient care and infection control training. All of them were introduced to the e-rub system and were invited to rate the procedures on 4 training videos covering oral feeding, Ryle's tube feeding, turning handling, and manual handling. Debriefing was given to the research nurses right after the training, ensuring they understood how to rate the observation moments and what we meant by unobtrusive observations. Before they went for on-site assessments, a briefing meeting was held by the project coordinator with the research nurse and to introduce the staff profile of the RCHs. In addition, biweekly regular meetings were conducted with the research nurses and the research team coordinator to ensure the quality of data collected in the RCHs. The research nurses recorded observations at any time in different shifts (morning shift or afternoon shift) on different days (Monday to Saturday), as this arrangement captured a variety of staff behavior at different times. Observations were made at any time without notifying the RCH staff, and they were conducted anywhere in the RCHs.

Self-Reported Infection Control Practices

Self-reported infection control practices were measured with the 10-item self-administered questionnaire developed by O'Neil and colleagues [28], asking the participants to indicate whether they and their coworkers often performed the infection control practices, which included hand hygiene before entering and leaving the residents' rooms, whether they educated the RCH residents about hand and respiratory hygiene, and whether the staff stayed at home if they had respiratory infection symptoms. A 5-point Likert scale with 5 = always, 4 = usually, 3 = sometimes, 2 = rarely, and 1 = never perform was given for selection. A total score was created by summing the scores of all 10 items, with higher scores indicating more frequent performance of the infection control practices. The questionnaire has been used in a large-scale study on health care workers in China [31].

Knowledge and Attitudes Toward Infection Control

Knowledge and attitudes toward infection control was measured using the 24-item self-administered instrument developed by O'Neil et al [28], with 12 items about knowledge and 12 items on attitudes in the following four areas: (1) clinical burden and the transmission of acute respiratory infection (7 items), (2) PPE and hand hygiene (6 items), (3) facility infection prevention and control policies for acute respiratory infection (6 items), and (4) influenza vaccination (5 items). Since the last 5 items are not relevant to COVID-19, we did not include these items in the survey. Respondents responded to the 19 statements using a 5-point Likert scale (from strongly disagree to strongly agree). For each knowledge item, respondents were classified as (1) "having knowledge" if their answers were either "strongly agree" or "agree" to the positive statements and "strongly disagree" or "disagree" to the negative statements, or (2) "lack of knowledge" if otherwise. For each attitude item, respondents were classified as (1) "positive attitudes" if their answers were either "strongly agree" or "agree" to the positive statements and

XSL•FO
RenderX

"strongly disagree" or "disagree" to the negative statements, or (2) "negative attitudes," if otherwise. This instrument has also been included in the large-scale study on health care workers in China [31].

Total Number of BGCTS Visits

Based on the data retrieved from the backend of the system, we counted the total number of times the health care workers viewed the chapters of the BGCTS.

Demographics of the participants, including age, sex, educational level, profession, previous training in infection control, and months of working experience in the current RCH, and information of the RCH, including financial type, number of staff, and number of residents, were also collected in the baseline survey (Multimedia Appendix 3). Only age and months of working experience were treated as continuous variables in the analysis.

Data Analyses

To compare sample characteristics at baseline between the intervention and control groups, for categorical variables at the cluster level, Fisher's exact tests were used, while for variables at the individual level, chi-square tests for categorical variables and independent 2-sided *t* tests for continuous variables were used. Cluster-level analysis was used to examine the between-group difference in the changes in the proportion of infection control practices by on-site observations, with RCHs as clusters. The analysis included a 2-stage approach consisting first of estimating the proportions of the 2 infection control practices (proper use and hand rub) by cluster, and then a weighted cluster-level linear regression was fitted on these proportions. The weights in the regression were set to the number of staff in the RCH and the number of residents in the RCH as a covariate. Demographic variables that have statistical significance between groups at baseline were also included in the regression. The financial type of RCH was not included in the model because of collinearity in the regression model. The distribution of the total number of BGCTS visits among the intervention participants was summarized using descriptive statistics.

To estimate the intervention effect on the self-reported secondary outcomes, we followed Van Breukelen's [32] recommendation for cluster RCT with pretest and posttest using linear mixed effects models. At the individual level, analyses were performed on the change scores within person (ie, postintervention score minus baseline score) for repeated measures, with a fixed effect for group, a random intercept for cluster, and control for 6 individual-level covariates (sex, age, educational level, baseline staff position, whether they had an infection control program before, and working experience in the current RCH) and 3 cluster-level covariates (number of staff, number of residents, and financial type of RCH) as fixed effects. The intervention effect was reflected by the coefficient of the group. We also conducted sensitivity analysis for self-reported secondary outcomes using multiple imputation to impute missing values using chained equations in regression models that included outcomes for each cluster. This group-based multiple imputation technique performs quite well for the estimates of the fixed effects, which are the primary interest of the study, although it also artificially inflates the true variation between groups [33,34]. A total of 500 imputed datasets were generated. The pooled estimates from the linear mixed-effects model analyses and the corresponding 95% CI were calculated based on the 500 imputed datasets. The analyses were performed using SPSS software (version 30.0; IBM Corp).

Ethical Considerations

Ethical approval was obtained from the Hong Kong Polytechnic University Institutional Review Board (reference number HSEARS20200903007). Offline informed consent from participants was obtained after a thorough explanation, which was given via Zoom meeting (Zoom Communications, Inc). The intervention participants knew which intervention was the intervention of interest, as they were granted access to the BGCTS, while the control participants also knew they were not given the intervention of interest (BGCTS). Since this intervention lasted for a short period of time (2 weeks only), the control participants did not consider they were in a disadvantage. Privacy consideration was reassured by providing a unique username and password to each participant, and the data were stored in a secured server with no one (except the study team) able to access the data. The nurse manager was given aggregated data of the performance of their staff (individual performance was not disclosed to the nurse managers).

The training system (BGCTS) was demonstrated to the staff via Zoom meetings, as many RCHs closed the door during the COVID-19 pandemic. The study team explained the features of the BGCTS, the planned duration of the training, and the access frequency of the system during the designated period and demonstrated the access method. All this information was included in the written information sheet as well, which was distributed to the staff when they signed the written consent forms.

To protect the privacy of the RCH staff, the staff in the RCH were deidentified and anonymized, and all direct nursing care actions were examined in confidentiality for patients' conditions. Since the research nurses stayed in the RCH for 2-3 weeks, this design minimized the Hawthorne effect due to observations (ie, RCH staff perform differently when being observed) because the observer's presence gradually becomes insignificant to the RCH staff [22]. In an RCH unit, more than one health care personnel provided care to the residents. Observations were made on a random basis. There was no interference with the care procedure during the observations. Similar to this previous study [22], no observations would be made if the residents or the relatives of the residents refused to allow the observations around.

Results

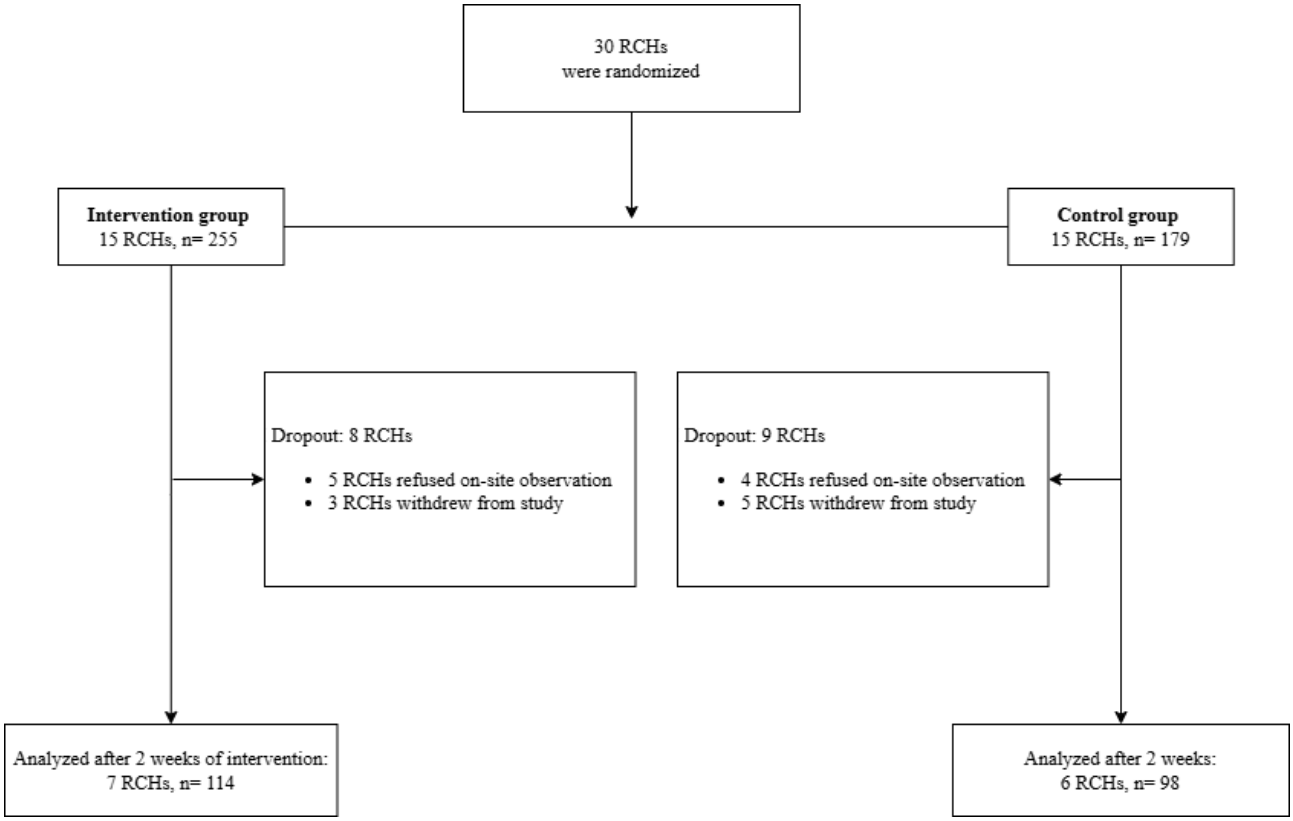
Overview

A total of 30 RCH units were recruited and randomized into either the intervention group (15 RCHs and 255 health care workers) or the control group (15 RCHs and 179 health care workers). Since COVID-19 infection happened during the study

XSL•FO
RenderX

period, 9 RCHs did not undergo observations, 2 RCHs delayed the postintervention observations (postintervention observations were conducted 2 months after intervention), 2 RCHs had only 1 observation session with only 20-30 moments in postintervention observations, and 4 RCHs withdrew from the study. Figure 1 shows the CONSORT diagram of recruitment and data collection. A total of 212 participants were recruited from the 13 RCHs, with 114 in the intervention group and 98 in the control group, and included in the analysis. All residents in the 13 RCHs allowed the observations around.

Figure 1. The Consolidated Standards of Reporting Trials diagram of recruitment and data collection. RCH: residential care home.



The participating RCHs were similar in terms of number of residents, number of staff, and financial type, with no statistical significance between the 2 groups. At the individual level, the majority of the participants were female, were in their middle age on average, had secondary school education, and were nonprofessional health care workers. Many of them had previous training in infection control and had been working in the current position for years (Table 1). There was no significant difference between the intervention and control groups in all variables at the individual level, except that the proportion of participants having previous training in infection control was significantly larger in the control group (χ^2_1=13.8; *P*<.001). Post intervention, a total of 142 (67%) participants had completed the survey after training, with 63 (55.3%) in the intervention group and 79 (80.6%) in the control group, and the difference was statistically significant (χ^2_1=15.3; *P*<.001). The completers were similar to those who were lost to follow-up, except that a significantly larger proportion of completers had previous training in infection control (180/142, 76.1% vs 43/70, 61.4%; χ^2_1=4.9; *P*=.04). No adverse events were reported during the study period.

XSL•FO
RenderX

Table 1. Demographics of the participants in the survey.

Characteristics	Intervention group, n (%)	Control group, n (%)	Statistical test values	*P* value
Cluster level (RCH[a]; n=7 intervention group and n=6 control group)				
Number of staff			__[b]	.82
<99	4 (57.1)	4 (56.7)		
100-199	2 (28.6)	2 (33.3)		
≥200	1 (14.3)	0 (0)		
Number of residents			__[b]	.21
<50	4 (57.1)	1 (16.7)		
50-100	0 (0)	2 (33.3)		
>100	3 (42.9)	3 (50)		
Financial type			__[b]	.59
Government subsidized	2 (28.6)	3 (50)		
Private organization	5 (71.4)	3 (50)		
Individual level (n=114 intervention group and n=98 control group)				
Sex			0.006 (1)[c]	.94
Male	12 (10.5)	10 (10.2)		
Female	102 (89.5)	88 (89.8)		
Education			2.4 (2)[c]	.31
Secondary school	70 (61.4)	63 (64.3)		
Postsecondary school	15 (13.2)	8 (8.2)		
University or above	14 (12.3)	7 (7.1)		
Profession[d]			0.5 (1)[c]	.57
Nurses or therapists	19 (16.7)	13 (13.3)		
Nonprofessional workers	95 (83.3)	85 (86.7)		
Previous training in infection control			13.8 (1)[c]	<.001
Yes	69 (60.5)	82 (83.7)		
No	45 (39.5)	16 (16.3)		
Age (years), mean (SD)	47.06 (12)	48.89 (11.78)	0.059 (190)[e]	.81
Working experience in the current RCH (months), mean (SD)	64.20 (61.72)	82.65 (83.13)	–1.754 (162.8)[e]	.08

[a]RCH: residential care home.

[b]Fisher exact test (statistical values are not applicable).

[c]Chi-square (*df*).

[d]Nonprofessional workers include care-related support workers, health workers, personal care workers, physiotherapy assistants, and occupational therapy assistants who do not have medical-related professional licenses.

[e]*t* test (*df*).

Infection Control Practices (Primary Outcome)

A total of 5168 moments were observed; 2666 moments (1273 at baseline and 1393 post intervention) were observed in the intervention group, while 2502 moments (1237 at baseline and 1265 post intervention) were recorded in the control group. Proper use of gloves and PPE and performed respiratory hygiene was not common in the RCHs, with 3.1% (37/1197) in the intervention group and 8.3% (97/1168) in the control group at baseline. The proportions of moments showed an increase to 9.3% (124/1339) in the intervention group after the BGCTS training, while a decrease to 5.4% (67/1246) in the control group after the usual care training. On the other hand, the proportion of moments with hand rub performed properly at baseline was high in both groups, with 89.5% (1139/1273) in the intervention group and 96% (1187/1237) in the control group. After training, the proportion increased to 94.9% (1322/1393) in the intervention group, while it remained similar in the control

XSL•FO
RenderX

group (Table 2). In addition to the number of residents in the RCH, the proportion of participants having previous training in infection control in the RCH was also included in the weighted cluster-level regression. Results of the weighted cluster-level regressions also supported that the BGCTS had a beneficial effect on both objective primary outcomes when compared to the usual care of the infection control program (Table 2). The BGCTS group had increased significantly more in the proportion of proper use of gloves and PPE and performed respiratory hygiene than the control group (β=.195, 95% CI 0.046-0.344; t_1=2.962; P=.02) and significantly more in the proportion of hand rub in the intervention group than that in the control group (β=.068, 95% CI 0.005-0.132; t_1=2.445; P=.04). Among the intervention participants, the mean of the total number of BGCTS visits was 7.68 (SD 6.92; median 8, range 0-52). A total of 23 (20.18%) participants did not visit any of the 8 BGCTS chapters, and 18 (78.26%) of them had previous training in infection control.

Table 2. Results from analyses of the primary outcomes, use of gloves and personal protective equipment (PPE) and perform respiratory hygiene and hand rub between the intervention and control groups.

Outcomes	Intervention group, n (%)		Control group, n (%)		Weighted cluster-level linear regression[a]		
	Baseline	Post intervention	Baseline	Post intervention	β (95% CI)	*t* test (*df*)	*P* value
Use of gloves and PPE and perform respiratory hygiene					.195 (0.046-0.344)	2.962 (1)	.02
Improperly performed	1160 (96.9)	1215 (90.7)	1071 (91.7)	1179 (94.6)			
Properly performed	37 (3.1)	124 (9.3)	97 (8.3)	67 (5.4)			
Hand rub					.068 (0.005-0.132)	2.445 (1)	.04
Improperly performed	134 (10.5)	71 (5.1)	50 (4)	36 (2.8)			
Properly performed	1139 (89.5)	1322 (94.9)	1187 (96)	1229 (97.2)			

[a]The number of residents in the residential care home and the proportion of participants having previous training in infection control in the residential care home were included as covariates.

Self-Reported Infection Control Practices and Knowledge and Attitudes Toward Infection Control (Secondary Outcome)

Results on comparing the changes in the secondary outcomes in the 2 groups are provided in Tables 3 and 4, on their mean scores at baseline and post intervention and mixed-effects model results, respectively. Table 3 showed that the mean score in self-reported infection control practices increased in the intervention group from 40.23 (SD 5.82) to 42.81 (SD 4.22) while reduced from 42.21 (SD 4.33) to 41.66 (SD 3.88) in the control group after training. Slight increments in mean scores in the 3 outcomes in knowledge and attitudes toward infection control were observed in both groups after training. However, the mixed effects model results revealed that there were no statistically significant differences in the mean score changes between the 2 groups in all 4 secondary outcomes in both the complete data and multiple imputation analyses (Table 4). For other controlling factors, only number of staff was statistically and significantly associated with knowledge about clinical burden and the transmission of acute respiratory infection in the multiple imputation analysis (Table 4).

Table 3. Mean and SD of self-reported infection control practices and knowledge and attitudes toward infection control by group and time.

Outcomes	Intervention group, mean (SD)		Control group, mean (SD)	
	Baseline	Post intervention	Baseline	Post intervention
Self-reported infection control practices	40.23 (5.82)	42.81 (4.22)	42.21 (4.33)	41.66 (3.88)
Knowledge about clinical burden and the transmission of acute respiratory infection	27.90 (4.74)	29.05 (4.42)	28.71 (4.56)	29.63 (4.15)
Knowledge about personal protective equipment and hand hygiene	23.17 (4.16)	24.98 (3.48)	24.62 (2.97)	25.49 (2.81)
Knowledge about facility infection prevention and control policies for acute respiratory infection	23.26 (3.45)	23.94 (2.98)	23.83 (3.33)	24.52 (3.43)

XSL•FO
RenderX

Table 4. Results of linear mixed effects models on change scores in secondary outcomes (self-reported infection control practices and knowledge and attitudes toward infection control).

Outcomes	Complete data			Multiple imputation		
	β (95% CI)	*t* test (*df*)	*P* value	β (95% CI)	*t* test (*df*)	*P* value
Self-reported infection control practices[a]						
Group	.13 (–8.44 to 8.71)	0.043 (1)	.97	–.92 (–3.63 to 1.80)	–0.663 (1)	.51
Sex	–.65 (–3.87 to 2.57)	–0.403 (1)	.69	–1.66 (–4.67 to 1.35)	–01.080 (1)	.28
Education	–.61 (–2.28 to 1.06)	–0.729 (1)	.47	–.49 (–1.98 to 0.99)	–0.651 (1)	.52
Age	.04 (–0.05 to 0.14)	0.904 (1)	.37	.04 (–0.04 to 0.12)	0.916 (1)	.36
Working experience	.00 (–0.01 to 0.02)	0.683 (1)	.50	.00 (–0.01 to 0.02)	0.663 (1)	.51
Profession	1.87 (–1.00 to 4.74)	1.292 (1)	.20	1.86 (–1.09 to 4.82)	1.235 (1)	.22
Previous infection control training	.38 (–2.16 to 2.92)	0.299 (1)	.77	.49 (–1.91 to 2.88)	0.399 (1)	.69
Number of staff	.04 (–0.03 to 0.11)	1.141 (1)	.26	.03 (–0.03 to 0.08)	1.053 (1)	.29
Number of residents	.01 (–0.05 to 0.08)	0.502 (1)	.63	.00 (–0.03 to 0.04)	0.138 (1)	.89
Financial type	1.51 (–1.07 to 4.09)	1.234 (1)	.23	1.06 (–0.49 to 2.62)	1.338 (1)	.18
Knowledge about clinical burden and the transmission of acute respiratory infection[a]						
Group	.03 (–4.42 to 4.49)	0.021 (1)	.98	.23 (–1.83 to 2.28)	0.216 (1)	.73
Sex	–.30 (–3.28 to 2.68)	–0.201 (1)	.84	–.88 (–3.47 to 1.71)	–0.665 (1)	.51
Education	–.30 (–1.85 to 1.25)	–0.380 (1)	.71	–.65 (–1.95 to 0.64)	–0.989 (1)	.32
Age	.01 (–0.07 to 0.10)	0.329 (1)	.74	–.01 (–0.09 to 0.07)	–0.235 (1)	.81
Working experience	.00 (–0.01 to 0.01)	0.302 (1)	.76	.00 (–0.01 to 0.01)	0.110 (1)	.91
Profession	1.05 (–1.62 to 3.71)	0.780 (1)	.44	1.78 (–0.78 to 4.33)	1.363 (1)	.17
Previous infection control training	–.04 (–2.35 to 2.27)	–0.033 (1)	.97	.61 (–1.49 to 2.72)	0.573 (1)	.57
Number of staff	.05 (–0.01 to 0.10)	1.839 (1)	.10	.05 (0.01 to 0.09)	2.287 (1)	.02
Number of residents	–.01 (–0.05 to 0.03)	–0.712 (1)	.50	–.01 (–0.05 to 0.01)	–1.020 (1)	.31
Financial type	.55 (–1.37 to 2.47)	0.715 (1)	.50	.83 (–0.33 to 2.00)	1.399 (1)	.16
Knowledge about personal protective equipment and hand hygiene[a]						
Group	.25 (–4.73 to 5.23)	0.197 (1)	.88	.32 (–2.36 to 2.98)	0.232 (1)	.82
Sex	.10 (–2.31 to 2.51)	0.086 (1)	.93	–1.10 (–3.13 to 0.92)	–1.067 (1)	.29
Education	–1.20 (–2.45 to 0.06)	–1.894 (1)	.06	–.94 (–1.98 to 0.09)	–1.783 (1)	.08
Age	–.04 (–0.11 to 0.04)	–1.009 (1)	.32	–.02 (–0.08 to 0.04)	–0.565 (1)	.57
Working experience	.00 (–0.01 to 0.01)	.502 (1)	.62	.00 (–0.01 to 0.01)	0.492 (1)	.82
Profession	1.05 (–1.10 to 3.20)	0.968 (1)	.34	.85 (–1.13 to 2.82)	0.840 (1)	.40
Previous infection control training	1.21 (–0.67 to 3.09)	1.279 (1)	.20	1.11 (–0.62 to 2.85)	1.257 (1)	.21
Number of staff	.03 (–0.02 to 0.08)	1.493 (1)	.18	.03 (–0.01 to 0.07)	1.515 (1)	.13
Number of residents	–.01 (–0.50 to 0.02)	–1.052 (1)	.36	–.01 (–0.04 to 0.02)	–0.694 (1)	.49
Financial type	–.70 (–2.51 to 1.12)	–1.083 (1)	.34	–.78 (–2.09 to 0.52)	–1.175 (1)	.24
Knowledge about facility infection prevention and control policies for acute respiratory infection[a]						
Group	.84 (–3.53 to 5.20)	0.550 (1)	.61	.08 (–1.86 to 2.02)	0.079 (1)	.94
Sex	–1.67 (–3.64 to 0.31)	–1.677 (1)	.10	–1.43 (–3.51 to 0.65)	–1.346 (1)	.18
Education	–.52 (–1.54 to 0.51)	–1.001 (1)	.32	–.23 (–1.33 to 0.87)	–0.408 (1)	.88
Age	.01 (–0.05 to 0.07)	0.316 (1)	.75	.01 (–0.06 to 0.08)	0.194 (1)	.85

XSL•FO
RenderX

Outcomes	Complete data			Multiple imputation		
	β (95% CI)	*t* test (*df*)	*P* value	β (95% CI)	*t* test (*df*)	*P* value
Working experience	.00 (–0.01 to 0.01)	0.539 (1)	.59	.00 (–0.01 to 0.01)	0.287 (1)	.77
Profession	–.14 (–1.90 to 1.62)	–0.160 (1)	.87	–.19 (–2.26 to 1.89)	–0.177 (1)	.86
Previous infection control training	–.73 (–2.29 to 0.82)	–0.937 (1)	.35	–.83 (–2.56 to 0.90)	–0.937 (1)	.35
Number of staff	.03 (–0.01 to 0.07)	1.524 (1)	.14	.02 (–0.02 to 0.06)	1.151 (1)	.25
Number of residents	–.00 (–0.04 to 0.03)	–0.308 (1)	.77	–.01 (–0.04 to 0.01)	0.913 (1)	.36
Financial type	–.24 (–1.71 to 1.23)	–0.356 (1)	.73	.22 (–0.82 to 1.27)	0.287 (1)	.77

[a]Six individual-level covariates (sex, age, educational level, baseline staff position, whether had infection control program before, and working experience in the current RCH) and 3 cluster-level covariates (number of staff, number of residents, and financial type of RCH) were included as fixed effects in the linear mixed-effects model.

Discussion

Overview

On the basis of the cluster-level analysis, the trial showed benefits of the 2-week blended infection control training BGCTS over usual care of an infection control training session for infection control practices of hand rub and the use of gloves and PPE and proper performance in respiratory hygiene among health care workers in RCHs. The study findings have extended the literature on the effectiveness of educational interventions on infection control practices by providing evidence from more than 5500 on-site observations, rather than data from self-reported questionnaires, as well as evidence in long-term care facilities with an experimental design during an outbreak setting [8,9]. After attending the BGCTS training, more RCH staff in the intervention group performed properly in the use of gloves and PPE and respiratory hygiene and rubbed their hands properly than those in the control group. These infection control practices are crucial to protect the staff from being infected with COVID-19. One of the stories in the BGCTS allowed the staff to choose appropriate PPE for different scenarios in RCHs (please refer to Multimedia Appendix 2 for the key learning points in the BGCTS). This is helpful to the staff and makes them understand the relevant guidelines in glove and PPE use. Wearing insufficient PPE cannot protect the staff, while wearing inappropriate PPE may waste the resources in the RCHs, leading to insufficient resources when they are needed. Educating the RCH staff on how to choose the most appropriate PPE and gloves for particular procedures is crucial in infection control training and making good use of the resources in the RCH facilities [35].

These study findings have provided preliminary evidence that gamification was a good strategy for the infection control practices among the staff in RCHs. This observation echoes the potential benefits of blended learning and active learning in delivering evidence-based health care educational interventions recommended by a recent SR including 61 RCTs [36] and is consistent with the findings from a previous study in which the correct concepts and information regarding the choice and handling of PPE can be delivered to the staff (game players) in interactive manners [37]. In this study, the majority of the participants were nonprofessional staff in the RCHs whose medical knowledge was expected to be lower than that of professional staff. This result implies the wide applicability of the BGCTS in public health training, in particular, to nonprofessional staff who are always the key frontline staff in long-term care settings. The key feature in the BGCTS of the voice-over function in the mother tongue language (Cantonese) in the games has ensured the staff access the information without language barrier or reading demand, which is critical to nonprofessional staff in strengthening their ability to apply knowledge and analytical skills in infection control practices.

Hand hygiene is the key in infection control practices. According to the WHO's recommendation [29], proper hand wash or hand rub is considered proper hand hygiene. Since RCHs are busy clinical sites in which staff always struggle with time when they provide direct care to the residents, many staff perform hand rubs rather than hand washes. Training about the conditions when hand rub could replace hand wash would be needed, especially in the SARD or COVID-19 condition, where hand wash is nearly impossible when the staff is fully gowned. The percentage of hand rubs at baseline was very high in the control group (1187/1237, 96%), which has limited its ability to demonstrate significant gains in the control participants and potentially underestimated the impact of usual care and thereby overestimated the impact of BGCTS. We noted the percentage of hand rub at baseline in both groups was very high. Such a high hand rub percentage may be due to the occurrence of the COVID-19 pandemic in the study period [38] because the percentage was reported as low as 15% in an observational study in the RCHs in Hong Kong in 2019 [2]. Upon the outbreak of the COVID-19 pandemic, many RCH staff had been educated to pay attention to hand hygiene before we started the BGCTS intervention. Nevertheless, the results suggested that the BGCTS training has good efficacy to promote the staff's performance in proper hand hygiene, in particular, hand rub.

Although the measures of these 2 outcomes were made objectively, our results might have been confounded by the implementation of the study because the halt or delay of all research activities in some of the participating RCHs due to the COVID-19 pandemic in the study period finally led to the dropout, delay in observations, or insufficient observations in 17 RCHs. The large reduction in the number of clusters from 30 to 13 RCHs has resulted in an underpowered analysis.

XSL•FO
RenderX

In the survey, the intervention group had reported an increase while the control group reported a decrease in the scores of self-reported infection control practices, whereas both groups showed increases in the scores in the 3 knowledge and attitude toward infection control scales, although the between-group differences were not statistically significant. These findings were in line with the literature that self-reported knowledge, attitude, and practices increased after receiving educational training [10-13]. The nonsignificant results could be attributed to several possible reasons in the design and the implementation of the study. First, the use of active control may have produced improvements in knowledge, attitude, and practices in infection control in the control participants. During the COVID-19 pandemic, staff in RCHs were at high risk of infection, and hence they would treat these briefing sessions more seriously and pay more attention to infection control, which may result in self-searching of infection control knowledge. Second, the significant between-group attrition rates post intervention might have introduced attrition biases. Third, the reduced sample size had lowered the study power in detecting between-group differences. Another possible reason was the measurement tools themselves. The measurement tools for the self-reported infection control practices and the 3 knowledge and attitude toward infection control scales have been used in infection control practice studies [31]; however, their psychometric properties have not yet been reported [28], and the use of self-reported questionnaires for outcome assessment also may lead to response bias by providing socially desirable responses from the participants. Thus, these findings might be subject to measurement error. We recommend that the psychometric properties of these scales be fully examined before their use in future studies. Meanwhile, these nonsignificant findings also probably reflected that the contents of the BGCTS training have not well addressed all the issues mentioned in these scales; for example, the clinical burden of acute respiratory infection and the policies in relation to facility infection prevention and control. Since infection control practices cover a broad spectrum of knowledge, showing a broad array of pathogens and different modes of transmission across various settings, the intensity of gamification in the BGCTS may not be sufficient to cover all these. Alternatively, these nonsignificant results might also be attributed to the low engagement in the BGCTS in some of the intervention participants. Increasing the contents or developing more videos and games could be considered to foster learning. Further studies are warranted to test this claim.

Limitations

This study has some limitations. The main study limitation was that the trial was carried out when Hong Kong experienced the fifth wave of the COVID-19 pandemic, and the recruitment was severely affected, leading to less satisfactory power in the analysis. Many RCHs refused the on-site observations or turned down the postintervention observations due to infected cases in the homes. This did not only induce attrition bias but also reduced the sample size for analysis, leading to the underpower of the study as well as the reduced generalizability of the study findings. Second, the occurrence of the pandemic may have altered the performance of the RCH staff in infection control, leading to a ceiling effect of hand rub. The findings about hand rub should further be investigated in postpandemic or nonpandemic periods when the RCH staff become less stressed so that their usual practices can be captured in observations. Third, the attrition bias might be further compounded by the significantly higher attrition rate post intervention in the intervention group as well as the fact that significantly more completers had previous training in infection control, leading to an unpredictable direction of the attrition bias. Fourth, the study findings on the 4 secondary outcomes might also be subject to measurement errors because they were assessed by the use of self-reported questionnaires, including measurement tools with unclear psychometric properties. Lastly, there may be observer bias, Hawthorne effects, or interrater reliability among research nurses in collecting objective on-site measurements, although we have attempted to minimize these biases by providing training in data collection to the nurses, taking the observations on a random basis, and making sure there is no interference with the care procedure during the observations.

Recommendations

This study has implications for practices and future research. First, this is the first-of-its-kind electronic platform for infection control training. With the current evidence from this study, the system should be further tested with studies to be conducted in the postpandemic condition. Nevertheless, this BGCTS training still can be considered as a tool for either training for new staff or reinforcement training for existing RCH staff. Since this system provides data about staff's engagement in the infection control training, the management team of the RCHs can use these data to plan for reinforcement training workshops for the staff. For future research, this study is a good foundation for understanding the efficacy of e-learning in RCH staff development. The study design of this study is robust with the collection of objective and subjective data and thus can be used as a reference for future research in other clinical settings.

Implementing good infection control measures and practices in RCHs protects residents and staff in the RCH from being infected. It was very unfortunate that we had a very high mortality rate for older people in the fifth wave of COVID-19 in early 2022, with 87% of the cumulative deaths being people aged 70 years or above [4,39]. RCHs were the most traumatized places, where many residents were infected and deceased due to COVID-19. This tragedy demonstrates the weakness of traditional infection control training in RCHs, in which lectures or educational talks are usually given to the staff. The assumption has been made that the RCH staff understand the infection control guidelines after the talks and carry out the infection control measures as told. It is uncommon to check their interpretation and keep track of their practices after training. This study provides some preliminary evidence that the BGCTS is a potentially good model of infection control training that could be applied in the RCHs. The BGCTS provides a record of the training, and this could identify who needs more training. Building up the RCH staff's competence in infection control practice prevents the spread of infectious diseases among the residents and infectious outbreaks in the RCHs.

XSL•FO
RenderX

Conclusions

In conclusion, the BGCTS training was shown to better improve actual performance in infection control practices regarding "the use of gloves and PPE and respiratory hygiene" and "hand rub" when compared to the usual care training, usually in terms of lectures or poster self-learning. However, the BGCTS did not perform better than infection control briefing sessions in self-reported infection control practices as well as knowledge and attitude toward infection control in health care workers in RCHs.

Acknowledgments

We are very grateful for the participants of this study (management team and staff in the participating residential care homes) who are willing to cooperate with us in this trial. Special thanks go to Ms Rachel Szeto, Chief Officer (Elderly Service); Ms Joyce Fu of the Hong Kong Council of Social Services; Dr MF Leung, President of the Hong Kong Association of Gerontology, who supported the recruitment of residential care homes for this study; and Mr Alex Choy (PolyU Industrial Centre) and his team, who provided technical support in camera fixation and setup. The Blended Gaming COVID-19 Training System development will not be successful without the genuine contributions from Ms Agnes Wong (graphic designer) and Mr Rex Cheung (system developer). We also acknowledge all the support provided by Dr Stephanie Szeto, Mr Tim Leong, and all research personnel and nurses in data collection, data input, and data cleaning. This training system was patented as "An Internet-based Infectious Disease Prevention and Control Training System and Method" in the China Patent Registry (CN202310214233.0) and the Hong Kong Short Term Patent Registry (30086664) on March 23, 2023. Patent Cooperation Treaty (PCT) patent application number PCT/CN2024/076227 was filed on February 6, 2024.

Funding

The project is funded by the Health and Medical Research Fund Commissioned Research on the Novel Coronavirus Disease (COVID-19), Food and Health Bureau, the Hong Kong SAR government (reference number COVID190218). The funding party played no role in the design of the study and did not have any role in the collection and interpretation of the data, in the writing of the manuscript, or in the decision to submit results.

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request. The Blended Gaming COVID-19 Training System is now available to be subscribed to by residential care homes or nursing schools according to the number of users in the institutions. The fee covers unlimited access to the system within the period, 2 administrator accounts for monitoring the staff and students' use, and the summary note of the performances of the users. The fee will be waived for collaborative research.

Authors' Contributions

AYML is the principal investigator who formulated the research question and conceptualized the study design and intervention design, as well as contributed to manuscript writing. DYPL is a coinvestigator who contributed to data analysis and manuscript writing. TKL contributed to study implementation and manuscript writing. JYWL, TC, DSKC, SCL, EMLW, and MMYT are coinvestigators who contributed to the intervention design and coordinated participant recruitment. AM is a coinvestigator who contributed to the research design and interpretation of findings. All authors interpreted the findings and reviewed the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

There is no conflict of interest from the funding body, as it has not been involved in the process of evaluation and did not interfere with the evaluation results. A few members of the study team (AYML, TKL, and AM) are the inventors of the training system. Measures, such as blinding of data collectors and independent data analysts, were used to minimize the bias in the evaluation.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 4435 KB - games_v13i1e71593_app1.pdf]

Multimedia Appendix 2
The 8 stories in the Blended Gaming COVID-19 Training System.
[DOCX File , 5103 KB - games_v13i1e71593_app2.docx]

Multimedia Appendix 3

XSL•FO
RenderX

The demographic questionnaire used in the study.
[PDF File (Adobe PDF File), 91 KB - games_v13i1e71593_app3.pdf]

References

1. Cesari M, Proietti M. Editorial: geriatric medicine in Italy in the time of COVID-19. J Nutr Health Aging 2020;24(5):459-460 [FREE Full text] [doi: 10.1007/s12603-020-1354-z] [Medline: 32346679]
2. McMichael TM, Currie DW, Clark S, Pogosjans S, Kay M, Schwartz NG, Public Health–SeattleKing County, EvergreenHealth,CDC COVID-19 Investigation Team. Epidemiology of COVID-19 in a long-term care facility in King County, Washington. N Engl J Med 2020;382(21):2005-2011 [FREE Full text] [doi: 10.1056/NEJMoa2005412] [Medline: 32220208]
3. Kim T. Improving preparedness for and response to coronavirus disease 19 (COVID-19) in long-term care hospitals in Korea. Infect Chemother 2020;52(2):133-141 [FREE Full text] [doi: 10.3947/ic.2020.52.2.133] [Medline: 32406211]
4. Rolland Y, Benetos A, Villars H, Braun H, Blain H. Editorial: a COVID-19 support platform for long term care facilities. J Nutr Health Aging 2020;24(5):461-462 [FREE Full text] [doi: 10.1007/s12603-020-1364-x] [Medline: 32346680]
5. Hand hygiene and infection prevention and control eLearning modules. Australian Commission on Safety and Quality in Health Care. 2024. URL: https://www.safetyandquality.gov.au/our-work/infection-prevention-and-control/hand-hygiene-and-infection-prevention-and-control-elearning-modules [accessed 2025-10-30]
6. Houben F, den Heijer CD, Dukers-Muijrers NH, Nava JB, Theunissen M, van Eck B, et al. Self-reported compliance with infection prevention and control of healthcare workers in Dutch residential care facilities for people with intellectual and developmental disabilities during the COVID-19 pandemic: a cross-sectional study. Disabil Health J 2023;17:101542. [doi: 10.1016/j.dhjo.2023.101542] [Medline: 39492010]
7. Babore G, Eyesu Y, Mengistu D, Foga S, Heliso A, Ashine T. Adherence to infection prevention practice standard protocol and associated factors among healthcare workers. Glob J Qual Saf Healthc 2024;7(2):50-58 [FREE Full text] [doi: 10.36401/JQSH-23-14] [Medline: 38725880]
8. Linhares FMP, Abreu WJCD, Melo PDOC, Mendes RCMG, Silva TAD, Gusmão TLA, et al. Effectiveness of educational interventions in knowledge, attitude, and practice for preventing respiratory infections: a systematic review and meta-analysis. Rev Bras Enferm 2022;75(4):e20210522 [FREE Full text] [doi: 10.1590/0034-7167-2021-0522] [Medline: 35352787]
9. Bloch N, Männer J, Gardiol C, Kohler P, Kuhn J, Münzer T, et al. Effective infection prevention and control measures in long-term care facilities in non-outbreak and outbreak settings: a systematic literature review. Antimicrob Resist Infect Control 2023;12(1):113 [FREE Full text] [doi: 10.1186/s13756-023-01318-9] [Medline: 37853477]
10. Nassar AAH, Al Serouri AA, Al-Shahethi AH, Almoayed KA. Effectiveness of training on health care workers' knowledge, attitude and practice regarding COVID-19 infection prevention and control, Yemen, 2021. BMC Health Serv Res 2024;24(1):1411 [FREE Full text] [doi: 10.1186/s12913-024-11927-8] [Medline: 39548437]
11. Chakma SK, Hossen S, Rakib TM, Hoque S, Islam R, Biswas T, et al. Effectiveness of a hand hygiene training intervention in improving knowledge and compliance rate among healthcare workers in a respiratory disease hospital. Heliyon 2024;10(5):e27286 [FREE Full text] [doi: 10.1016/j.heliyon.2024.e27286] [Medline: 38486731]
12. Gupta S, Goswami B, Madhu SV. Effectiveness of video-based educational intervention on knowledge, attitude, and practice (KAP) of COVID-19 health care workers: lesson for future pandemic preparedness. ANAMS 2023;59(04):233-236 [FREE Full text] [doi: 10.1055/s-0043-1772218]
13. Singh V, Supehia S, Gupta P, Narula H, Sharma M, Devi K, et al. Effectiveness of video modules in infection control trainings during COVID-19 pandemic: a quasi-experimental study in tertiary care institute. J Educ Health Promot 2021;10(1):183 [FREE Full text] [doi: 10.4103/jehp.jehp_1009_20] [Medline: 34250117]
14. Zeybek N, Saygı E. Gamification in education: why, where, when, and how?—a systematic review. Games and Culture 2023;19(2):237-264. [doi: 10.1177/15554120231158625]
15. Reinoso Schiller N, Motaharina G, König A, Benze G, Eichkorn J, Weber-Krüger M, et al. More than rewards: insights into a hospital infection prevention and control gamification strategy. J Hosp Infect 2025;160:74-80 [FREE Full text] [doi: 10.1016/j.jhin.2025.02.019] [Medline: 40096892]
16. Hong Kong training portal on infection control and infectious diseases. Hospital Authority. 2020. URL: https://icidportal.ha.org.hk/Trainings [accessed 2025-10-30]
17. Post-registration courses. School of Nursing, The University of Hong Kong. 2020. URL: https://nursing.hku.hk/education/certificate-in-infection-control [accessed 2020-06-07]
18. Wong GC, Ng T, Li T. Infection control in residential care homes for the elderly in Hong Kong (2005-2014). Hong Kong Med J 2019;25(2):113-119 [FREE Full text] [doi: 10.12809/hkmj187328] [Medline: 30967517]
19. How many people have smartphones worldwide. BankMyCell. 2025. URL: https://www.bankmycell.com/blog/how-many-phones-are-in-the-world [accessed 2025-09-11]
20. Leung AY, Xu XY, Chau PH, Yu YTE, Cheung MK, Wong CK, et al. A mobile app for identifying individuals with undiagnosed diabetes and prediabetes and for promoting behavior change: 2-year prospective study. JMIR Mhealth Uhealth 2018;6(5):e10662 [FREE Full text] [doi: 10.2196/10662] [Medline: 29793901]

XSL•FO
RenderX

21. Campbell MK, Piaggio G, Elbourne DR, Altman DG, CONSORT Group. CONSORT 2010 statement: extension to cluster randomised trials. BMJ 2012;345:e5661. [doi: 10.1136/bmj.e5661] [Medline: 22951546]
22. Au JKL, Suen LKP, Lam SC. Observational study of compliance with infection control practices among healthcare workers in subsidized and private residential care homes. BMC Infect Dis 2021;21(1):75 [FREE Full text] [doi: 10.1186/s12879-021-05767-8] [Medline: 33446137]
23. Random.org. URL: https://www.random.org/lists/ [accessed 2025-11-03]
24. World Health Organization. The COVID-19 risk communication package for healthcare facilities. Western Pacific Region Office. URL: https://iris.wpro.who.int/handle/10665.1/14482 [accessed 2020-06-06]
25. Guidelines for residential care homes for the elderly (RCHEs) and residential care homes for persons with disabilities (RCHDs) for the prevention of coronavirus disease (COVID-19). Center for Health Protection. 2024. URL: https://www.swd.gov.hk/storage/asset/section/447/en/E6_new_eng.pdf [accessed 2024-06-06]
26. RCH gaming. The Hong Kong Polytechnic University. URL: https://www.polyu.edu.hk/sn/rchgaming/signin [accessed 2025-11-03]
27. Code of practice for residential care homes (elderly persons). Social Welfare Department. 2024. URL: https://www.swd.gov.hk/en/pubsvc/lr/rche/rche_orcopref/rche_cop/ [accessed 2024-06-06]
28. O'Neil CA, Kim L, Prill MM, Stone ND, Garg S, Talbot HK, et al. Preventing respiratory viral transmission in long-term care: knowledge, attitudes, and practices of healthcare personnel. Infect Control Hosp Epidemiol 2017;38(12):1449-1456. [doi: 10.1017/ice.2017.232] [Medline: 29173225]
29. WHO guidelines on hand hygiene in health care. World Health Organization. 2025. URL: https://iris.who.int/bitstream/handle/10665/44102/9789241597906_eng.pdf?sequence=1 [accessed 2025-10-30]
30. Biswas D, Ahmed M, Roguski K, Ghosh PK, Parveen S, Nizame FA, et al. Effectiveness of a behavior change intervention with hand sanitizer use and respiratory hygiene in reducing laboratory-confirmed influenza among schoolchildren in Bangladesh: a cluster randomized controlled trial. Am J Trop Med Hyg 2019;101(6):1446-1455 [FREE Full text] [doi: 10.4269/ajtmh.19-0376] [Medline: 31701861]
31. Hong Q, Xu Y. Evaluating hand hygiene knowledge, attitudes, and practices among healthcare workers in post-pandemic H1N1 influenza control: a cross-sectional study from China. Front Public Health 2024;12:1432445 [FREE Full text] [doi: 10.3389/fpubh.2024.1432445] [Medline: 39399703]
32. Van Breukelen GJP. Cluster randomized trials with a pretest and posttest: equivalence of three-, two- and one-level analyses, and sample size calculation. Multivariate Behav Res 2024;59(2):206-228 [FREE Full text] [doi: 10.1080/00273171.2023.2240779] [Medline: 37590444]
33. Andridge RR. Quantifying the impact of fixed effects modeling of clusters in multiple imputation for cluster randomized trials. Biom J 2011;53(1):57-74. [doi: 10.1002/bimj.201000140] [Medline: 21259309]
34. Graham JW. Missing Data: Analysis and Design. New York: Springer; 2012.
35. Tsang CC, Holroyd-Leduc JM, Ewa V, Conly JM, Leslie MM, Leal JR. Barriers and facilitators to the use of personal protective equipment in long-term care: a scoping review. J Am Med Dir Assoc 2023;24(1):82-89.e2. [doi: 10.1016/j.jamda.2022.11.012] [Medline: 36473522]
36. Hill J, Gratton N, Kulkarni A, Hamer O, Harrison J, Harris C, et al. The effectiveness of evidence-based healthcare educational interventions on healthcare professionals' knowledge, skills, attitudes, professional practice and healthcare outcomes: systematic review and meta-analysis. J Eval Clin Pract 2024;30(6):909-935. [doi: 10.1111/jep.14001] [Medline: 38817022]
37. Suppan M, Gartner B, Golay E, Stuby L, White M, Cottet P, et al. Teaching adequate prehospital use of personal protective equipment during the COVID-19 pandemic: development of a gamified e-learning module. JMIR Serious Games 2020;8(2):e20173 [FREE Full text] [doi: 10.2196/20173] [Medline: 32516115]
38. Zhang M, Zhou M, Tang F, Wang Y, Nie H, Zhang L, et al. Knowledge, attitude, and practice regarding COVID-19 among healthcare workers in Henan, China. J Hosp Infect 2020;105(2):183-187 [FREE Full text] [doi: 10.1016/j.jhin.2020.04.012] [Medline: 32278701]
39. Strategies to prevent the spread of COVID-19 in long-term care facilities (LTCF). Centers for Disease Control and Prevention. URL: https://stacks.cdc.gov/view/cdc/85498 [accessed 2024-06-06]

Abbreviations

BGCTS: Blended Gaming COVID-19 Training System
CONSORT: Consolidated Standards of Reporting Trials
PPE: personal protective equipment
RCH: residential care home
RCT: randomized controlled trial
SARD: severe acute respiratory diseases
SR: systematic review
WHO: World Health Organization

XSL•FO
RenderX

Edited by A Coristine; submitted 25.01.25; peer-reviewed by R Mccloskey, J Sarvestan; comments to author 16.04.25; revised version received 23.10.25; accepted 24.10.25; published 27.11.25.

Please cite as:
Leung AYM, Leung DYP, Lau TK, Liu JYW, Cheung T, Cheung DSK, Lam SC, Wong EML, Tse MMY, Molassiotis A
Efficacy of the Web-Based Gamified Infection Control Training System on Practices for Health Care Workers in Residential Care Homes: Clustered Randomized Controlled Trial
JMIR Serious Games 2025;13:e71593
URL: https://games.jmir.org/2025/1/e71593
doi:10.2196/71593
PMID:

©Angela Y M Leung, Doris Y P Leung, Terence K Lau, Justina Y W Liu, Teris Cheung, Daphne S K Cheung, Simon C Lam, Eliza M L Wong, Mimi M Y Tse, Alex Molassiotis. Originally published in JMIR Serious Games (https://games.jmir.org), 27.11.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Effectiveness of a Gamified Mobile App in Enhancing Treatment Adherence for Children With Amblyopia: Explorative Study

Bo Liu*, PhD; Yisheng Fan*, MSc; Meng Xu, MSc; Fangyuan Chang, PhD; Yue Shi, MA; Zhao Liu, PhD

College of Design, Shanghai Jiaotong University, 800 Dongchuan Road, Minhang District, Shanghai, China

*these authors contributed equally

Corresponding Author:
Zhao Liu, PhD
College of Design, Shanghai Jiaotong University, 800 Dongchuan Road, Minhang District, Shanghai, China

Abstract

Background: Amblyopia is the leading cause of visual impairment in children worldwide. The predominant clinical treatment, occlusion therapy, is marred by poor adherence, often attributed to the physical discomfort and social stigma associated with eye patching. Adjunct digital visual trainings have not consistently sustained patient engagement due to their repetitive nature, thereby compromising their efficacy.

Objective: This study aimed to evaluate the effectiveness of a gamified mobile app designed to increase treatment adherence among children with amblyopia by making the therapeutic process more engaging and accessible within home settings.

Methods: An exploratory study was conducted, commencing with qualitative interviews and questionnaires to explore the barriers to traditional treatment adherence. This formative research informed the development of a gamified mobile app, which was shaped by cognitive appraisal theory to address identified emotional and psychological needs, potentially impacting adherence. The subsequent quantitative phase utilized a randomized controlled trial involving 34 children with amblyopia who were aged 7 - 10 years and recruited from a local primary school. These participants were randomly assigned to either the intervention group, which used a novel gamified mobile app developed by our team, or the control group, which utilized another commercially available mobile app. Both groups engaged with their respective apps in a home environment. The 8-item Morisky Medication Adherence Scale was adapted to measure treatment adherence.

Results: Over the 4-week trial, 34 children aged 7 - 10 years with amblyopia were enrolled and randomized into 2 groups: intervention (n=18) and control (n=16). Children in both the intervention and control groups engaged daily for 20 minutes at home, using mobile apps designed for visual rehabilitation. The intervention group (n=18) achieved a significantly higher mean adherence rate (mean 6.56, SD 1.06) on the Morisky Medication Adherence Scale compared to the control group (n=16; mean 5.01, SD 1.22; $P<.001$). Thematic analysis of the design process revealed that integrating cognitive appraisal theory effectively enhanced emotional engagement and adherence.

Conclusions: The integration of cognitive appraisal theory into the design of a gamified mobile app for amblyopia treatment has shown to significantly improve adherence among children.

Trial Registration: ClinicalTrials.gov NCT06372548; https://clinicaltrials.gov/study/NCT06372548

(JMIR Serious Games 2025;13:e60309) doi:10.2196/60309

KEYWORDS

amblyopia; child health care innovation; emotional cognition; gamified therapy; mobile health apps; pediatric eye care; randomized controlled trial; serious game; treatment adherence; visual rehabilitation

Introduction

Background

Amblyopia, the most prevalent cause of visual impairment in children, affects a significant number of young individuals worldwide [1-4]. It leads to deficits in visual acuity, spatial perception, and visual processing, making early intervention essential [5-11]. Currently, the primary treatment involves occlusion therapy—patching the healthy eye to stimulate the amblyopic one [1,5,12,13]. However, adherence to this method is low, often under 50%, due to psychological stress, discomfort, and inadequate parental support, resulting in suboptimal treatment outcomes [14-20]. To overcome these challenges, digital therapy has emerged as a promising supplement to traditional methods [1,21,22]. Visual training tasks delivered through digital platforms, such as contrast detection or letter recognition games, can stimulate the amblyopic eye while engaging children in interactive, structured activities [23-27]. Studies suggest that incorporating gamification into such tasks may enhance participation and adherence [7,27-33]. Despite this promise, many digital therapies remain ineffective due to

XSL•FO
RenderX

repetitive content, limited personalization, and poor user engagement [1,9,21,22,27,28,32-34]. Reliance on bulky equipment or clinical settings often adds logistical burdens for families, whereas mobile-based solutions offer portability, accessibility, and alignment with children's behavioral tendencies, making them more suitable for home-based rehabilitation [8,21,35,36].

Within this landscape, evidence indicates that family-centered care models, particularly those integrating interactive game designs, can significantly enhance adherence and therapeutic outcomes among children with amblyopia [7,21,22]. Gamified approaches that incorporate emotional support and social feedback effectively address resistance to eye patching and repetitive training, thereby improving adherence [37]. Features such as immediate feedback and reward systems within gamified interventions help sustain interest and reinforce habitual therapeutic behaviors over time [38,39]. Integrating cartoon-based education and visual feedback into game design has been shown to markedly boost short-term adherence in children undergoing amblyopia care [37,40,41]. Beyond short-term engagement, gamification can foster long-term rehabilitative habits through immersive and layered situational experiences that scaffold ongoing participation [42]. An increasing body of research is now investigating digital interventions to improve adherence and therapeutic outcomes in pediatric amblyopia, and the current market already offers mobile health apps that combine visual stimulation with simple games to capture children's attention [43].

Rationale

Yet, a critical gap persists: the emotional and psychological needs of children with amblyopia are often overlooked, and while advanced interventions like virtual reality and binocular therapies claim to improve engagement, few solutions are designed with explicit attention to children's cognitive and emotional experiences, which may compromise long-term adherence and therapeutic effectiveness [24,36]. Many existing mobile apps also lack diversity in interactivity and feedback mechanisms, limiting their ability to sustain engagement over time despite providing visual stimulation that can improve adherence [1,34,44]. Research shows that when content richness or dynamic feedback is insufficient, engagement levels diminish, underscoring the need for gamified therapeutic solutions that integrate greater interactivity and real-time feedback to better support home-based rehabilitation [1,34,44]. Moreover, evidence on the long-term vision improvement benefits of digital interventions remains limited, and concerns persist about potential negative effects such as visual fatigue and myopia from prolonged use of small-screen devices, highlighting the need for further research to define their role and impact in amblyopia rehabilitation [45].

To address these limitations, we conducted an exploratory study using qualitative interviews and questionnaires to identify key adherence barriers, and we developed a mobile-based gamified intervention informed by cognitive appraisal theory to target the emotional and behavioral needs of users with pediatric conditions. The intervention was integrated into standard visual training routines and evaluated through a randomized controlled trial (RCT) to test whether a theory-informed, emotionally intelligent design could improve adherence and engagement in home settings.

Theoretical Framework: Cognitive Appraisal Theory

In both psychology and marketing, emotional states are recognized as critical drivers of effective individual actions. Among the various theories exploring the relationship between emotion and cognition, Lazarus's cognitive appraisal theory has garnered significant attention. This theory posits that emotions arise from an individual's subjective evaluation of an event, meaning emotions are the result of appraisals, which subsequently influence behavior [46].

In specific applications, cognitive appraisal theory has been utilized to explain the emotional states elicited by users' appraisals of products, services, environments, or events [47,48]. It has been widely adopted in fields such as tourism and consumer behavior to explore the comprehensive interplay of cognition, emotion, and behavior triggered by interactions with particular events or physical environments [49]. Notably, research related to entertainment games has demonstrated that players' cognitive appraisals of video games can elicit both positive and negative emotional states [50]. Kim et al further argued that compared to other products or services, video games are often perceived as leisure activities from an emotional appraisal perspective [51]. This suggests that users prioritize enjoyment and pleasure [50].

In the context of our study, cognitive appraisal theory offers a valuable framework for understanding and integrating users' emotional responses into the design of gamified interventions. By leveraging this theory, we can systematically guide the emotional responses of children toward specific products or services, aligning with their psychological and emotional needs. Integrating cognitive appraisal theory into our intervention design allows us to address adherence challenges by enhancing emotional engagement and fostering positive emotional states, which are essential for the success of gamified treatments for amblyopia.

Table 1 summarizes the five core dimensions—relevance, valence, likelihood, agency, and coping potential—and their definitional criteria as established in the appraisal literature [46,48,52-54]. Definitions emphasize how appraisals of goal significance, expected outcomes, perceived control, and coping capacity shape emotions that, in turn, drive behavior [46,48,52-54]. Prior work shows that emotional support improves adherence to amblyopia therapy in children, with enhanced relevance and positive valence fostering engagement and role-play strengthening agency and self-efficacy [37]. A multidimensional, appraisal-guided approach has been reported to increase user engagement and positive participation, with particularly strong effects among child users in serious game contexts [39].

XSL•FO
RenderX

Table . Criteria for dividing dimensions of cognitive evaluation theory.

Dimensions	Criteria
Relevance	The anticipation of achieving a more valuable goal or outcome can result in a heightened emotional response. This is closely linked to the experience of interest [52,53], which is a typical reaction to situations that are subjectively significant and elicit high levels of attentional activity. This, in turn, increases the likelihood of responding effectively to events and evokes strong positive emotions, such as enjoyment [54].
Valence	Individuals are more likely to experience positive emotions, such as happiness or gratitude, when they perceive an experience as related to a positive goal and moving them closer to that desired goal [54].
Likelihood	The likelihood dimension assesses the probability of a situation leading to a specific outcome, identifying associated emotions such as hope, fear, or anxiety [46,54].
Agency	Consider the influence of the user, others, or the environment/objects on the outcome of the situation, as well as factors related to controllability, in the evaluation of the agent [48].
Coping potential	Coping potential is the perceived ability to manage or change a situation and is categorized into 3 main groups: benign-positive, neutral, and stressful [46].

Study Aim

Guided by cognitive appraisal theory and the growing evidence for gamified digital interventions, this study developed and evaluated a mobile serious game to address emotional and psychological drivers of adherence in pediatric amblyopia care. The aim was to support sustained, home-based rehabilitation by personalizing mechanics, narratives, and feedback to children's appraisal processes, thereby improving treatment adherence in everyday settings. We tested this approach in an RCT with children aged 7 - 10 years completing daily home training, aligning with best practices for interventional evaluation in digital health. The primary outcome was treatment adherence at 4 weeks measured with an adapted 8-item Morisky Medication Adherence Scale (MMAS-8), and secondary outcomes included user experience dimensions (eg, attractiveness, novelty, usability) and qualitative engagement mapped to appraisal dimensions. We hypothesized that an appraisal-informed, child-centric design would help overcome conventional barriers to occlusion and repetitive training and enhance engagement in pediatric vision care.

Methods

Study Design

We used a 3-stage mixed methods design comprising a qualitative needs assessment (Stage 1), an expert-informed design (Stage 2), and a parallel-group RCT (Stage 3) to develop and evaluate a gamified mobile intervention for pediatric amblyopia rehabilitation. The qualitative components were conducted and reported in line with COREQ (Consolidated Criteria for Reporting Qualitative Studies) and employed inductive thematic analysis with dual coding in NVivo 14 (developed by QSR International) to enhance rigor and credibility. The interventional RCT was designed and reported according to CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) to ensure transparency and reproducibility in digital health trials. Insights from Stages 1 - 2 identified key adherence barriers and specified design levers grounded in cognitive appraisal theory to inform the gamified app's mechanics, narrative, and feedback. The primary outcome for Stage 3 was treatment adherence at 4 weeks measured with an adapted MMAS-8, and secondary outcomes included User Experience Questionnaire (UEQ) scores and posttrial interviews mapping children's emotional engagement to appraisal dimensions. The trial was registered at ClinicalTrials.gov (NCT06372548) and created to cover the entire study, including foundational qualitative and design phases.

Ethical Considerations

This study was approved by the Institutional Review Board of the Eye & ENT Hospital of Fudan University (H20230365I). Written parental consent was obtained for all minors, and child assent was acquired using a plain-language cartoon booklet and an assent form completed (signature or fingerprint) prior to any study procedures. Participation was discontinued immediately if a child indicated distress or refusal, in accordance with established child-safeguarding protocols. All data were deidentified and stored on encrypted hospital servers with dual backup, with a retention policy of 5 years before secure destruction. Participants received a stationery set and ¥20 (approximately US $2.80) upon completion as compensation for time and effort.

The ClinicalTrials.gov registration (NCT06372548) was submitted following completion of the initial qualitative phase but prior to the recruitment and enrollment of participants into the randomized controlled intervention. Importantly, the registration occurred before any RCT participant was recruited or randomized, ensuring that all interventional procedures and outcomes were prospectively registered.

XSL•FO
RenderX

Stage 1: Qualitative Needs Assessment

Participants and Recruitment

We used purposive, maximum-variation sampling to capture a diverse range of experiences and perspectives. The inclusion criteria included (1) parent or legal guardian of a child aged 6‐10 years diagnosed with amblyopia, (2) pediatric ophthalmologist with ≥3 years of clinical experience, or (3) child meeting diagnostic criteria who could assent with parental help. The exclusion criteria included inability to communicate in Mandarin or refusal to be recorded. Parents were approached during clinic visits; clinicians were invited by email; children were recruited via clinician referral. In total, 8 ophthalmologists, 22 parents (8 families), and 6 children participated. The sample size was guided by information power, and theoretical saturation was reached by the sixth parent interview, as no new codes or themes emerged, indicating sufficient coverage of core concepts.

Interview Guide

Semistructured interview guides were pilot-tested with 2 parents and 1 clinician, then refined for clarity and relevance. The interview topics covered barriers to occlusion therapy, child motivation, game preferences, and parental support. The interview outlines for parents and children are available in Multimedia Appendices 1 and 2.

Data Collection

All interviews were conducted by 2 trained researchers (a female design graduate student and a male MSc research assistant in ophthalmology) in private clinic rooms (for parents or clinicians) or a quiet play area (for children). Each interview lasted 40‐60 minutes, was audio-recorded, transcribed verbatim, and securely backed up on encrypted drives. Member-checking was used for credibility: summary transcripts were sent to 4 participants to confirm the accuracy and resonance of identified themes.

Analysis and Credibility

Transcripts and questionnaires were analyzed using inductive thematic analysis in NVivo 14: familiarization with data, initial coding (open coding), theme development (axial and selective coding), theme review, theme definition/naming, and reporting [55]. At least 2 coders independently performed each coding round; disagreements were discussed with a third coder until consensus was reached, resulting in high intercoder reliability (κ=0.82). At least 2 anonymized, supportive quotations per theme are provided in the *Results* section for confirmability.

Reflexivity and trustworthiness were prioritized throughout analysis. Researchers maintained a reflexive diary to document positionality, preconceptions, and their influence on data interpretation. Confirmability was established through audit trails and member checking. Dependability was ensured via transparent documentation of all analytic steps and coder discussions. Credibility was strengthened by participant feedback; thick description; and triangulation between parent, child, and clinician interviews. Transferability is enhanced by detailed reporting of sample characteristics and recruitment context, enabling readers to assess relevance to other settings.

Stage 2: Expert-Informed Design Process

Expert Panel and Sampling

An expert panel was assembled using purposive sampling across hospital, academic, and design networks to capture diverse knowledge relevant to pediatric amblyopia treatment and gamified intervention development. A total of 6 experts were recruited, including 3 pediatric ophthalmologists (each with at least 5 years of clinical experience) and 3 game designers (each with a minimum of 5 years of experience in serious games or health-related gamification). The inclusion criteria comprised verified domain expertise and willingness to engage in 2 rounds of structured consultation.

Round 1 Interviews and Design Mapping

In Round 1, each expert participated in a 45‐ to 60-minute Zoom interview focused on operationalizing the 5 dimensions of cognitive appraisal theory (relevance, valence, likelihood, agency, coping potential) within the concrete mechanics, narrative structures, feedback modalities, and visual styles suitable for pediatric rehabilitation. Interview questions, refined following pilot feedback, are available in Multimedia Appendix 3. All interviews were audio-recorded, transcribed verbatim, and thematically coded by 2 independent coders in NVivo 14 following an inductive-deductive hybrid approach; intercoder reliability for theme extraction was κ=0.79. The team used reflexive diaries to minimize bias, ensuring that both clinical and design perspectives were adequately represented. Coded insights from all experts were synthesized into a design-decision matrix that mapped cognitive appraisal dimensions to actionable game features. This matrix guided ideation of low-fidelity paper sketches, followed by a high-fidelity Figma digital prototype (Prototype 1) that reflected prioritized features and narrative components identified in Round 1.

Round 2 Consensus and Prototype Refinement

In Round 2, experts were provided asynchronous access to Prototype 1 and asked to rate each proposed feature on a 7-point Likert scale for relevance and feasibility. Consensus was defined a priori as at least 80% agreement per domain and item. All items met or exceeded this threshold after minor revision based on expert feedback, resulting in a finalized intervention prototype (Prototype 2) used in the subsequent RCT. Reflexivity was maintained during consensus scoring, and disagreements were reconciled through written clarification and reevaluation to ensure the transferability of core design principles to clinical practice and child end-user contexts.

Stage 3: Randomized Controlled Trial

Participants and Recruitment

Children aged 7-10 years with refractive, strabismic, or mixed amblyopia were recruited via school outreach and clinic referrals. The inclusion criteria were a confirmed diagnosis, ability to follow the training with parental support, and access to a mobile device at home. The exclusion criteria included severe amblyopia (best corrected visual acuity≤0.2), other ocular or neurological comorbidities, and prior exposure to digital amblyopia therapy. Based on pilot data (expected mean difference 1.5 on MMAS-8, SD 1.2; Cohen *d*=1.25), 15

XSL•FO
RenderX

participants per arm were required to achieve 90% power at α=.05 (2-tailed). We therefore enrolled 40 participants to mitigate risk of loss to follow-up and randomized after consent or assent.

Randomization, Concealment, and Masking

An independent statistician generated a computer-based permuted-block sequence (block size=4). Allocation was concealed using sequentially numbered, opaque, sealed envelopes. Outcome assessors and the statistician were blinded to group assignment.

Procedures

After written parental consent and child assent, participants completed a baseline enrollment visit during which we verified eligibility and collected demographic variables on a standardized case report form: age (in years), sex (male/female), and school grade (3rd-5th), as well as device availability for home use; demographic sources were parent report at enrollment using a structured checklist. Participants were then randomized to the intervention (our gamified app, "Find You! Cure My Animal Friends") or the control (a widely used commercial visual-training app) and received a standardized orientation, including a brief instructional video for parents and a printed quick-start guide covering daily use, patching alongside play as clinically indicated, and troubleshooting. Families were instructed to complete home-based training for 20 minutes per day over 4 consecutive weeks; no additional clinical care was altered for either group during the study period. To support consistency, parents were asked to supervise session start/stop and encourage completion at roughly the same time each day.

Attrition and Analysis Set

During data collection, 6 participants (3 per arm) withdrew (reasons reported to the study team: lack of reliable home internet, scheduling conflicts, or unrelated illness). Data from these participants were not used in the analysis. The final analytic sample was N=34 (intervention n=18; control n=16).

Postintervention Assessment

The primary outcome was treatment adherence, assessed with an adapted MMAS-8 (licensed from Morisky LLC; see Multimedia Appendix 4 for item wording and scoring) [56,57]. Within the adaptation, medication-specific terms were reworded to "vision training" and "eye patch" to ensure contextual relevance for pediatric amblyopia. Comprehensibility and content validity were confirmed via cognitive interviews with 5 parent-child dyads prior to trial launch; pilot administration with n=20 yielded Cronbach α=.83, supporting internal consistency. The MMAS-8 was administered at the end of week 4 with parental assistance, and total scores were classified into low, medium, or high adherence according to developer guidance.

Secondary outcomes included the UEQ (26 items across 6 dimensions, see Multimedia Appendix 5) and postintervention semistructured interviews with children and parents (20 - 40 min, conducted via Zoom, see Multimedia Appendix 6). These interviews explored emotional and behavioral engagement with the intervention, thematically mapped to cognitive appraisal theory dimensions (relevance, valence, likelihood, agency, and coping potential).

All interviews were audio-recorded, transcribed verbatim, and analyzed via dual-coder thematic analysis in accordance with COREQ standards. Saturation was reached when no new themes or codes emerged during review, and at least 2 anonymized supporting quotations per theme are reported for confirmability. Intercoder reliability was assessed, and transcript review was supplemented by member-checking for credibility. Reflexivity was maintained by both coders using analytic memos to document positionality and potential biases.

Data Analysis

Quantitative Analysis

Analyses were conducted in SPSS 26.0 (IBM Corp). Normality was assessed with the Shapiro-Wilk normality test. Between-group comparisons used 2-tailed independent-samples *t* tests for normally distributed variables and Mann-Whitney *U* tests otherwise. Two-tailed α was set at .05. We report test statistics, degrees of freedom, 95% CIs, and effect sizes (Cohen *d*) where applicable. Analyses followed a complete-case approach; the trial statistician was blinded to group assignment and had no role in data collection or intervention delivery.

Qualitative Analysis

Transcripts were analyzed using inductive thematic analysis in NVivo [55]. Two coders independently coded all transcripts; discrepancies were resolved via discussion with a third reviewer. Data saturation was deemed reached when no new codes/themes emerged. Trustworthiness was addressed through reflexive memos (researcher positionality); an audit trail; member-checking of summaries; and triangulation across parents, children, and clinicians. At least 2 anonymized quotations per theme are reported; intercoder reliability is provided.

Results

Result 1—Qualitative Needs Assessment: Barriers to Adherence and Design Motivators

Overview of Stakeholder Interviews

Table 2 shows a quantitative summary of stakeholder interview themes and subthemes from Stage 1 qualitative needs assessment in a 3-stage mixed methods study developing a gamified mobile intervention for pediatric amblyopia rehabilitation. Two overarching themes—Desires and Preferences—were derived through inductive thematic analysis with dual coding in NVivo 14 under COREQ-consistent procedures, and each subtheme is reported with participant proportion and an example code. Within Desires, adherence-relevant subthemes included sustained engagement (75%), enhancing the appeal of the eye patch (50%), rewards and immediate feedback (65%), and emotion regulation concerning feelings about being "in treatment," "being ill," and "wearing an eye patch" (95%). Within Preferences, children favored colorful collectibles (60%) and hands-on activities such as puzzles and drawing (45%). Interviews used purposive, maximum-variation sampling and included 8 pediatric ophthalmologists, 22 parents (8 families),

XSL•FO
RenderX

and 6 children, conducted in clinic and child-friendly settings at the Eye & ENT Hospital of Fudan University.

The stakeholder interviews revealed a strong preference among parents and children for games that sustain attention over time. Key design motivators included continuous storylines; meaningful role-play; and clear, simple instructions—preferred over repetitive, single-player mini-games. These findings align with the high endorsement of narrative elements (75%) as reported in Table 2. Quantitative summary of theme frequencies, coded proportions, and representative quotations are presented to illustrate key patterns and their implications for therapy design.

Table . Quantitative results of stakeholder interviews.

Theme and subtheme	Participant proportion (%)	Code example
Desires		
The game needs to sustain attention	75	"Children need a continuous storyline"
Enhance appeal of the eye patch	50	"The eye patch causes discomfort"
Rewards mechanism	65	"Positive feedback motivates children"
Emotion regulation	95	"Reducing stigma, fear, and frustration" around "treatment/illness/eye-patch"
Preferences		
Preference for colorful items	60	"Children enjoy collecting colorful items"
Enjoyment of handicrafts	45	"Children like puzzles, drawing, etc"

Enhancing the Appeal of the Eye Patch

Families described both physical discomfort and peer-related embarrassment as major barriers to consistent patching (50%). Suggestions included attractive, customizable patterns and integrating patch-related actions with meaningful in-game events so that wearing the patch feels purposeful rather than punitive.

Rewards and Immediate Feedback

Parents reported that immediate, varied rewards are more effective than vague, delayed promises in sustaining daily participation (65%). Several children also mentioned striving for a high score as intrinsically motivating, while parents cautioned that repeating the same reward can lead to boredom.

Emotion Regulation

Interviews recurrently pointed to reducing stigma, fear, and frustration associated with treatment routines and visible patching (95% of parents). Typical comments included "If the patch could be part of the story—like a special pass to enter the game world—my child wouldn't feel 'sick'; she'd feel proud to wear it." Parents also highlighted the need for gentle hints and affirming messages when children get stuck, and for empathetic narratives that normalize help-seeking and provide language to address teasing at school.

Mapped to cognitive appraisal theory, these requests correspond to valence (positive affect and feedback), agency (controllability through role-play), coping potential (scaffolding and hints), relevance (linking patch use to core game goals), and likelihood (clear goals and immediate feedback that reduce uncertainty).

Preferences

Under preferences, children spontaneously discussed liking bright, saturated colors (eg, pink, orange, blue), collectible items (stickers, stationery), and hands-on activities (puzzles, building, sketching). Many favored comic-style narratives and "becoming a character," further supporting story-driven, role-based mechanics.

Design Recommendations

Synthesizing these findings, we propose 4 actionable recommendations for pediatric amblyopia training: (1) deliver quick, varied positive feedback to sustain engagement; (2) personalize content to children's preferences and abilities; (3) embed rich storyworlds (animal companions, collectible artifacts) to deepen immersion; and (4) incorporate emotion-regulation affordances—empathetic cues, reframing of patching, and progressive challenges—to mitigate negative feelings surrounding treatment and illness and thereby strengthen adherence.

Result 2—Qualitative Findings: Design Mapping and Prototype

Overview of Expert Interviews

Table 3 shows quantitative support from expert interviews in Stage 2, mapping the 5 cognitive appraisal dimensions to occlusion-therapy optimization and rehabilitation game-design principles for a gamified pediatric amblyopia intervention. This table summarizes expert-endorsed design mappings across Relevance, Valence, Likelihood, Agency, and Coping potential, reporting for each dimension the corresponding occlusion-therapy focus, game-design guideline, proportion of expert agreement, and an illustrative solution to support adherence. The expert process comprised 2 interview rounds with 6 experts (3 pediatric ophthalmologists and 3 game designers; ≥5 years' experience), Zoom-based sessions (45-60 min), thematic coding with intercoder agreement κ=0.79, and an a priori consensus threshold of ≥80% on relevance and feasibility, informing Prototype 2 used in the RCT. Collectively, these principles specify how occlusion-therapy behaviors and game mechanics should be integrated to enhance adherence in pediatric amblyopia care.

XSL•FO
RenderX

Table . Quantitative support from expert interviews.

Appraisal dimension	Occlusion-therapy optimization	Rehabilitation training game design	Proportion (%)	Example of design solution
Relevance	Assign special meaning to the eye patch	Incorporate educational elements and understandable therapeutic goals	85	"Make friends with critters through treatment"
Valence	Link game progress with eye patch usage	Provide positive feedback and guidance	70	"Combines forms of visual training with therapeutic behaviors in play"
Likelihood	Showing the direct game effects of eye patch wear	Clear goal setting and feedback	65	"Unlock new maps after each level, add logos (corresponding eye sticker patterns) to the maps"
Agency	Emphasize the child's role in the game through eye patches	Emphasize the leading role of children and role-playing	55	"Add interaction with animal companions"
Coping potential	Teach the correct use of eye patches	Emphasize progress rather than success	60	"Three achievement levels for each map"

Occlusion-Therapy Optimization

Experts recommended treating the eye patch as a meaningful artifact in the game world (eg, a narrative "pass" or "gift" from animal companions) to reframe patching as purposeful rather than punitive (Relevance). Progress in the game should depend visibly on correct patch use (Valence/Likelihood), and a short tutorial animation should demonstrate proper wear to reduce misuse and frustration (Coping potential).

Rehabilitation Game Mechanics (by Appraisal Dimension)

Rehabilitation game mechanics were mapped to the following appraisal dimensions:

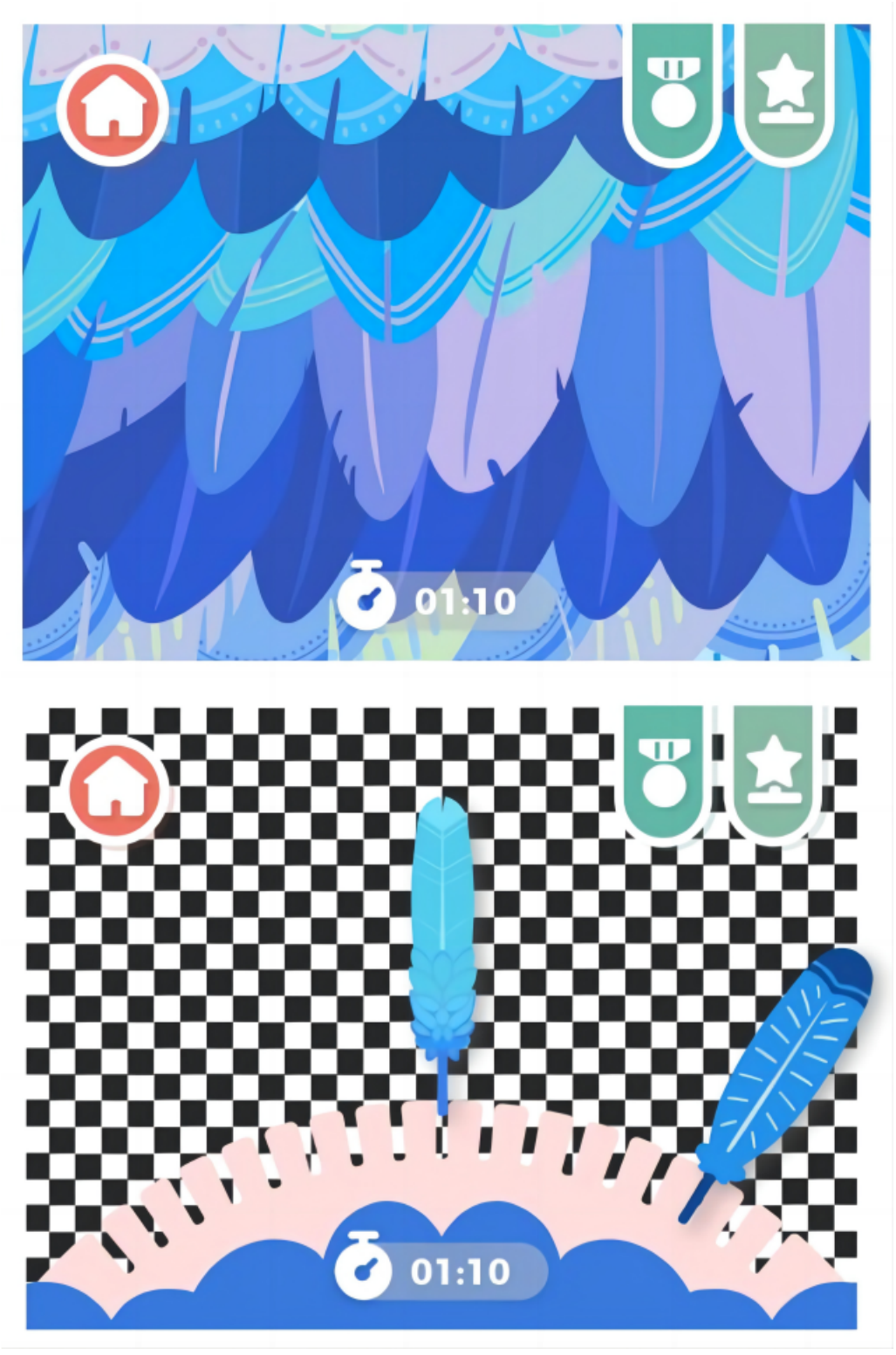
- Relevance (85%): assign special meaning to the patch and incorporate clear therapeutic goals within the storyline.
- Valence (70%): deliver positive, specific feedback and guidance tied to therapeutic behaviors.
- Likelihood (65%): use explicit level goals and immediate feedback; successful sessions can unlock new maps and display patch-pattern "logos" on the map.
- Agency (55%): make the child the protagonist (eg, intern doctor), enable role-play, and add interaction with animal companions.
- Coping potential (60%): emphasize progress over perfection with 3 achievement tiers per map and adjustable difficulty to prevent overload.

Resulting Game Blueprint

Synthesizing these inputs, we adopted interactive storytelling as the primary format. Players act as an intern doctor who completes daily treatment tasks for animal patients (visual-training tasks embedded). The game includes 3 difficulty levels selectable by the child, collectible illustrations and achievements as incentives, and a child-friendly cartoon style. These principles were consolidated into a design-decision matrix and implemented in the intervention app used in Stage 3 (see Figures 1,2; Table 4).

XSL•FO
RenderX

Figure 1. Visual stimulation training and fine eye training.



XSL•FO
RenderX

Figure 2. Main game interface and level interface.



XSL•FO
RenderX

Table . Participant characteristics.

	Experimental group, n (%)	Control group, n (%)
Age (years)		
10	6 (33.33)	6 (37.5)
9	7 (38.88)	6 (37.5)
8	4 (22.22)	3 (18.75)
7	1 (5.55)	1 (6.25)
Sex		
Male	13 (72.22)	11 (68.75)
Female	5 (27.77)	5 (31.25)
Total	18 (100)	16 (100)

Gamified Treatment App

The gamified mobile app "Find You! Cure My Animal Friends" developed for this study was designed to enhance treatment adherence in children with amblyopia by integrating cognitive appraisal theory. It provides a series of visually stimulating activities that are specifically crafted to improve visual acuity and encourage regular use, as shown in Figure 1. The main game interface and level interface are shown in Figure 2.

Result 3—RCT Outcomes

User Statistics

Table 4 shows baseline characteristics of participants by randomized arm in the Stage 3 parallel-group RCT evaluating a gamified mobile app versus a commercial visual-training app for treatment adherence in pediatric amblyopia. Children aged 7-10 years were recruited via school outreach and clinic referrals in Shanghai, China, yielding an analytic sample of N=34 (intervention n=18; control n=16) after attrition. Variables include age distribution (7-10 y) and sex, with no statistically significant between-group differences at baseline.

The intervention and control groups yielded W=0.946 (P=.37) and W=0.963 (P=.71), respectively, indicating no evidence against normality. Normality assessment informed the use of independent-samples t tests alongside nonparametric tests in SPSS 26.0 with 2-tailed α=.05, as prespecified in the statistical analysis plan.

Adherence

After conducting an independent samples t test, a significant difference was found in the adherence scores on the MMAS-8 scale between the experimental group that used the amblyopia training game (t_{32}=3.896; P<.001). The experimental group (N=18) had a mean adherence score of 6.56 (SD 1.06, SE 0.25, 95% CI 6.03-7.08), while the control group (N=16) had a mean score of 5.01 (SD 1.22, SE 0.31, 95% CI 4.36-5.66). The results indicated that the experimental group had significantly higher adherence scores than the control group.

Significant behavioral differences were found when comparing patient adherence to an amblyopia training game (experimental group) with other games on the market (control group). It was revealed that 61.11% (11/18) of patients in the experimental group showed high adherence, whereas only 18.75% (3/16) in the control group achieved the same level of adherence(P=.012). Additionally, the experimental group demonstrated a significantly lower percentage of patients with low adherence: 38.89% (7/18) in the experimental group versus 81.25% (13/16) in the control group(P=.012).

Table 5 shows the comparison of adherence issues between the intervention group and the control group. We provided the percentage of subjects who scored 0 for Q1-Q7, along with the mean score for Q8. The results revealed higher percentages in the control group for Q2, Q5, and Q7, which align with the significant differences observed in our statistical analyses. Significant differences emerged in some questions, notably "Have you missed any doses in the past two weeks?" (Q2: 11.11% vs 56.25%, t_{32}=3.03; P<.001), indicating a higher compliance in the experimental group. Similarly, for "Did you find it challenging to adhere to the treatment program?" (Q7: 11.11% vs 50.00%, t_{32}=2.59; P=.02), the experimental group showed significantly better adherence. No significant differences were observed between the experimental and control groups for questions about forgetting to take medication (Q1: 22.22% vs 31.25%, t_{32}=0.58; P=.57), adjusting medication without informing the doctor (Q3: 11.11% vs 12.50%, t_{32}=0.12; P=.90), and whether they took their medication the previous day (Q5: 11.11% vs 50.00%, t_{32}=2.59; P=.02).

XSL•FO
RenderX

Table . Comparison of adherence issues between intervention and control groups.[a]

Questions	Experimental group (% or mean [SD])	Control group (% or mean [SD])	*t* test (*df*) or *z* statistic	*P* value
1. Do you ever forget to take your medication?	22.22%	31.25%	0.58(32)	.57
2. Have you missed any doses in the past two weeks?	11.11%	56.25%	3.03(32)	*<.001*
3. Do you adjust your medication without consulting your doctor when your symptoms worsen or new symptoms arise during treatment?	11.11%	12.50%	0.12(32)	.90
4. Do you sometimes forget to bring your medication with you when traveling or away from home for extended periods?	22.22%	25.00%	0.18(32)	.85
5. Did you take your medication yesterday?	11.11%	50.00%	2.59(32)	*.02*
6. Have you ever stopped taking your medication before completing the prescribed course, even if your symptoms improved or disappeared?	22.22%	25.00%	0.18(32)	.85
7. Did you find it challenging to adhere to the treatment program?	11.11%	50.00%	2.59(32)	*.02*
8. How frequently do you forget to take your medication?	0.67[b] (0.27)	0.51[b] (0.34)	−1.28[c]	.20

[a]Italic values indicate $P<.05$, showing statistical significance.

[b]Mean values.

[c]z-statistics.

User Experience Questionnaire

Table 6 shows the UEQ scores by randomized group at the 4-week assessment in the Stage 3 parallel-group RCT evaluating a gamified mobile app versus a commercial visual-training app for pediatric amblyopia rehabilitation (children aged 7-10 y, Shanghai, China, March-December 2024). The table reports means, standard deviations, sample sizes, and 95% CI for 6 UEQ dimensions (Attractiveness, Understandability, Efficiency, Reliability, Facilitation, and Novelty) in the intervention (n=18) and control (n=16) arms as a prespecified secondary outcome. The intervention outperformed the control on Attractiveness (2.14 vs 0.39), Understandability (1.88 vs 0.19), Facilitation (1.92 vs 1.14), and Novelty (2.13 vs 1.09), while the control showed a slight advantage on Efficiency (1.20 vs 1.08); Reliability was comparable (1.35 vs 1.30)

XSL•FO
RenderX

Table . User Experience Questionnaire scores.

Dimensions	Experimental group					Control group				
	Mean	SD	N	Confidence	95% CI	Mean	SD	N	Confidence	95% CI
Attractiveness	2.14	0.59	18	0.27	1.87 to 2.41	0.39	0.41	16	0.20	0.18 to 0.59
Understandability	1.88	0.78	18	0.36	1.51 to 2.24	0.19	0.57	16	0.28	–0.09 to 0.47
Efficiency	1.08	0.49	18	0.22	0.86 to 1.31	1.20	0.26	16	0.13	1.07 to 1.33
Reliability	1.35	0.62	18	0.29	1.06 to 1.63	1.30	0.23	16	0.11	1.19 to 1.41
Facilitation	1.92	0.45	18	0.21	1.71 to 2.13	1.14	0.26	16	0.13	1.01 to 1.27
Novelty	2.13	0.38	18	0.17	1.95 to 2.30	1.09	0.32	16	0.16	1.03 to 1.35

Personal Interviews

Table 7 shows postintervention individual interview results by randomized group mapping children's experiences to cognitive appraisal dimensions in a 3-stage mixed methods study of a gamified mobile intervention for pediatric amblyopia rehabilitation (Stage 3 RCT, Shanghai, China, children aged 7-10 y). Interviews with children and parents were conducted via Zoom (20-40 min) after the 4-week intervention and thematically analyzed with dual-coder procedures consistent with COREQ, with codes organized under 3 themes: rehabilitation training experience, attitudes toward rehabilitation, and attitudes toward wearing eye patches. Codes are presented for the intervention and control groups and annotated with the corresponding appraisal dimension (relevance, valence, likelihood, agency, and coping potential) to show directional changes in perceived engagement and coping during treatment.

Table . Results of individual interviews.[a]

Theme	Experimental group (codes and dimensions)	Control group (codes and dimensions)
Rehabilitation training treatment experience	• Children like the background of the story in the game (relevance↑) • Children enjoy playing the role of a doctor (relevance↑, agency↑) • Children enjoy interacting with animal friends in the game (relevance↑, agency↑) • Children like the illustrations in the game to collect/unlock the list of friends (relevance↑, valence, likelihood↑)	• Children find the mini-games a little boring (relevance↓) • Children find the game graphics unattractive (valence↓) • Children like getting high scores in games (relevance↑, likelihood↑)
Attitude toward rehabilitation training	• Children like to play the game repeatedly to achieve high levels of achievement (relevance↑, coping potential↑, likelihood↑) • Children perceive the game as difficult, but are more likely to want to complete the challenge (coping potential↑, likelihood↑) • Children were still motivated to play the game at the end of the 30-day experiment (relevance↑, agency↑)	• Children report resistance the first time they use the game (relevance↓, valence↓) • Children say they are reluctant to repeat the training after experiencing all the mini-games (relevance↓) • Children play the same game 2- - 3 times to get a high score (relevance↑, likelihood↑, coping potential↑)
Attitude toward wearing eye patches	• Children are generally motivated to wear eye patches (relevance↑, agency↑) • Children like eye patches with special patterns and keep them carefully (relevance↑, valence↑)	• Children are often resistant to the eye patch and require parental intervention (valence↓, coping potential↓)

[a]"↑" indicates an increase or improvement, while "↓" signifies a decrease or reduction.

The study found that children in the experimental group had a positive reaction to the gamified rehabilitation training. This positive response may be attributed to the game design being consistent with cognitive appraisal theory. Children particularly enjoyed storytelling, role-playing, interacting with in-game animal characters, and collecting icons. These elements increased their perceptions of goal relevance and congruence, as well as their assessments of the likelihood of successfully

XSL•FO
RenderX

completing the game. Furthermore, enabling the children to actively engage in the game and interact with the surroundings enhanced their sense of control and ability to cope, which could account for the experimental group's greater participation in the game and adherence to rehabilitation training. The experiences and attitudes of the children in the control group differed significantly, as they mostly expressed disinterest in the game content or found the images unattractive. However, even in the control group, children were willing to play the game repeatedly to achieve a high score, indicating that perceptions of likelihood and coping potential can still be improved with appropriate incentives. Furthermore, the experimental group's children exhibited a higher level of positive acceptance toward the specially patterned eye patches, which provides additional evidence that increasing goal relevance and consistency can enhance treatment adherence. Conversely, the control group's children were less receptive to eye patches and needed parental intervention.

Discussion

Principal Results and Contextualization

This study explored a gamified mobile app developed with cognitive appraisal theory to improve treatment adherence among children with amblyopia. Compared with a commercially available alternative, our app significantly improved self-reported adherence and user engagement. These findings resonate with prior research showing that gamified interventions tailored to pediatric needs can increase adherence, particularly when emotional engagement and role-play are incorporated into therapeutic design [58]. Our findings reinforce this, demonstrating that when visual therapy is embedded within a child-centric narrative, such as acting as a doctor for animal patients, treatment becomes more intrinsically motivating and socially meaningful.

Prior digital therapies for amblyopia often emphasized perceptual gains while overlooking emotional dimensions, limiting their long-term efficacy [32,44]. In contrast, our results suggest that integrating emotional resonance through personalized narratives enhances not only short-term adherence but also sustained behavioral engagement. Thematic interview results indicate children developed emotional attachment to in-game characters, aligning with self-determination theory, which emphasizes autonomy, competence, and relatedness as key drivers of motivation [59]. These findings bridge cognitive appraisal theory with game-based learning literature, supporting an integrative framework for digital therapeutic design.

The app's ability to improve perceived coping potential—a critical appraisal dimension—suggests that such designs may empower children by increasing their perceived control and self-efficacy. This finding aligns with pediatric psychology literature, where increased agency and coping predict better adherence in chronic treatment contexts [60]. Similar conclusions have been drawn in pediatric diabetes and asthma care, where gamified self-management tools improved perceived autonomy and long-term outcomes [61,62]. Our results extend these insights to amblyopia care, suggesting cross-condition applicability of emotionally intelligent gamification.

In addition to cognitive appraisal theory, the positive reception of the product aligns with established game engagement models, such as the Player Experience of Need Satisfaction model, which has been shown to predict retention and satisfaction in health games [63]. The high UEQ scores in attractiveness and novelty observed in our trial mirror predictors of engagement identified in broader mobile health research [64]. Together, these results underscore the value of designing from a multidimensional framework that spans affective science, pediatric behavioral health, and human-computer interaction.

Implications for Practice and Design

These findings have several implications for both clinical practice and digital health design. First, clinicians should consider recommending therapeutic applications that support emotional regulation and engagement, not just visual training efficacy. Second, designers should consider embedding rehabilitation goals within game narratives that emphasize positive emotions, role-play, and intrinsic reward mechanisms. Our approach could be adapted to other chronic pediatric conditions requiring repetitive behavior change, such as speech therapy or motor training.

Additionally, co-design with children and caregivers is essential to ensure developmental appropriateness and motivational alignment. Our use of animal avatars, customizable achievements, and therapeutic storytelling was informed by early-stage qualitative interviews, reinforcing the value of participatory design in digital health tool development [65].

Future Research Directions

Longer follow-up trials are needed to determine whether appraisal-informed gamified therapy produces sustained improvements in visual function (eg, visual acuity and binocular outcomes) beyond short-term adherence gains, given the current paucity of long-term evidence for digital interventions in amblyopia. Future studies should incorporate objective adherence metrics (eg, standardized usage logs in both arms) and conduct deeper psychometric evaluation of the adapted MMAS-8, including test-retest reliability, Rasch calibration, and convergent validity against behavioral data, to address limitations of self-reports and strengthen measurement robustness. Multisite and larger-sample RCTs with cross-cultural recruitment are warranted to assess generalizability of emotionally adaptive gamification across diverse pediatric populations and health care settings. Designing and testing explicit parent- or family-centered coengagement components may further improve adherence, building on evidence that family involvement and supportive feedback can enhance therapeutic participation in pediatric amblyopia. Mechanistic studies should isolate how specific appraisal dimensions (relevance, valence, likelihood, agency, coping potential) drive adherence and engagement to optimize narrative, feedback, and role-play elements in serious games for vision rehabilitation. Safety evaluations should also monitor potential adverse effects related to screen exposure (eg, visual fatigue or myopia risk) to inform risk-benefit profiles for home use of small-screen devices in pediatric care. Finally, integration of occlusion-therapy behaviors with in-game mechanics should be quantitatively linked to clinical endpoints (eg, patch-wear patterns and acuity

XSL•FO
RenderX

change) to clarify therapeutic synergy in real-world home settings.

Limitations

This study has several limitations. First, the sample size was modest and drawn from a limited setting, which may affect generalizability. Second, the evaluation focused on short-term adherence (4 weeks); the sustainability of effects and links to long-term visual outcomes require follow-ups. Third, while the adapted MMAS-8 offered a brief and accepted measure for pediatric behavioral adherence, its origin in medication contexts introduces potential semantic transfer; although cognitive interviewing and a pilot supported face validity and internal consistency (α=.83), further psychometric work (eg, Rasch calibration, test-retest, and convergent validity against usage logs and visual-acuity change) is warranted. Fourth, adherence included self-reports, which are susceptible to recall and social-desirability bias. The comparable logs were unavailable for the commercial control app, limiting between-arm objective comparisons. Finally, the game primarily targeted child users; although parents supported use, the intervention did not explicitly structure parental coengagement, which is known to influence adherence and could be incorporated in future iterations.

Conclusion

This 3-stage mixed methods study developed and evaluated a cognitive appraisal–informed mobile serious game to support home-based rehabilitation for pediatric amblyopia, with a prespecified primary focus on treatment adherence and secondary assessments of user experience and qualitative engagement. In a 4-week parallel-group RCT, the intervention significantly improved adherence on the adapted MMAS-8 when compared with a commercial control app (mean 6.56 vs 5.01; t_{35}=3.896; P<.001), with a higher proportion of high-adherence participants and fewer participants with low adherence in the intervention arm. Secondary outcomes showed superior user experience for the intervention in terms of attractiveness, understandability, facilitation, and novelty, which is consistent with stronger engagement profiles for pediatric digital health tools. Posttrial interviews mapped children's experiences to appraisal dimensions, indicating higher relevance, valence, likelihood, agency, and coping potential in the intervention group alongside more positive attitudes toward patch use, aligning with the study's theory-driven design. Collectively, these findings demonstrate that an appraisal-guided, child-centric gamified intervention can enhance short-term adherence and user experience in home amblyopia care, directly meeting the stated study aim of supporting sustained home-based rehabilitation through emotionally intelligent digital therapy.

Acknowledgments

We sincerely thank Dr Dian Zhu for her valuable guidance in methodological design and writing throughout this project. We have used ChatGPT for parts of the manuscript's grammar and spelling checks. The authors have obtained permission to use the 8-item Morisky Medication Adherence Scale (Copyright MMAS 2006 [66])

Data Availability

Data are available from the authors upon reasonable request.

Authors' Contributions

Conceptualization: BL, YF
Data curation: MX, YF
Formal analysis: MX, YF
Investigation: MX
Methodology: FC, YF
Project administration: BL, ZL
Resources: BL
Software: MX, YS
Supervision: BL, FC, ZL
Validation: BL, MX
Visualization: YF
Writing – original draft: YF
Writing – review & editing: BL, FC, YF, ZL
BL and YF are co–first authors.

Conflicts of Interest

None declared.

Multimedia Appendix 1

XSL•FO
RenderX

Interview outline for parents (preintervention).
[DOCX File, 16 KB - games_v13i1e60309_app1.docx]

Multimedia Appendix 2
Children's interest and preference questionnaire.
[DOCX File, 15 KB - games_v13i1e60309_app2.docx]

Multimedia Appendix 3
Interview outline for medical experts.
[DOCX File, 15 KB - games_v13i1e60309_app3.docx]

Multimedia Appendix 4
Adapted 8-item Morisky Medication Adherence Scale (MMAS-8).
[DOCX File, 15 KB - games_v13i1e60309_app4.docx]

Multimedia Appendix 5
User Experience Questionnaire (UEQ).
[DOCX File, 16 KB - games_v13i1e60309_app5.docx]

Multimedia Appendix 6
Postintervention interview outline for parents.
[DOCX File, 15 KB - games_v13i1e60309_app6.docx]

Checklist 1
COREQ 32-item checklist.
[PDF File, 334 KB - games_v13i1e60309_app7.pdf]

Checklist 2
CONSORT 2010 checklist of information to include when reporting a randomized trial.
[PDF File, 202 KB - games_v13i1e60309_app8.pdf]

References

1. Levi DM. Rethinking amblyopia 2020. Vision Res 2020 Nov;176:118-129. [doi: 10.1016/j.visres.2020.07.014] [Medline: 32866759]
2. Fu Z, Hong H, Su Z, Lou B, Pan CW, Liu H. Global prevalence of amblyopia and disease burden projections through 2040: a systematic review and meta-analysis. Br J Ophthalmol 2020 Aug;104(8):1164-1170. [doi: 10.1136/bjophthalmol-2019-314759] [Medline: 31704700]
3. Hashemi H, Pakzad R MSc, Yekta A, et al. Global and regional estimates of prevalence of amblyopia: A systematic review and meta-analysis. Strabismus 2018 Dec;26(4):168-183. [doi: 10.1080/09273972.2018.1500618] [Medline: 30059649]
4. Ciuffreda KJ, Levi DM, Selenow A. Amblyopia: Basic and Clinical Aspects: Butterworth-Heinemann; 1991.
5. Tailor V, Bossi M, Greenwood JA, Dahlmann-Noor A. Childhood amblyopia: current management and new trends. Br Med Bull 2016 Sep;119(1):75-86. [doi: 10.1093/bmb/ldw030] [Medline: 27543498]
6. Birch EE. Amblyopia and binocular vision. Prog Retin Eye Res 2013 Mar;33:67-84. [doi: 10.1016/j.preteyeres.2012.11.001] [Medline: 23201436]
7. Gambacorta C, Nahum M, Vedamurthy I, et al. An action video game for the treatment of amblyopia in children: a feasibility study. Vision Res 2018 Jul;148:1-14. [doi: 10.1016/j.visres.2018.04.005] [Medline: 29709618]
8. Syntosi A, Felizzi F, Bouchet C. A social media listening study to understand the unmet needs and quality of life in adult and pediatric amblyopia patients. Ophthalmol Ther 2022 Dec;11(6):2183-2196. [doi: 10.1007/s40123-022-00571-2] [Medline: 36175822]
9. Holmes JM, Levi DM. Treatment of amblyopia as a function of age. Vis Neurosci 2018 Jan;35:E015. [doi: 10.1017/S0952523817000220] [Medline: 29905125]
10. Holmes JM, Lazar EL, Melia BM, et al. Effect of age on response to amblyopia treatment in children. Arch Ophthalmol 2011 Nov;129(11):1451-1457. [doi: 10.1001/archophthalmol.2011.179] [Medline: 21746970]
11. Fronius M, Cirina L, Ackermann H, Kohnen T, Diehl CM. Efficiency of electronically monitored amblyopia treatment between 5 and 16 years of age: new insight into declining susceptibility of the visual system. Vision Res 2014 Oct;103:11-19. [doi: 10.1016/j.visres.2014.07.018] [Medline: 25130409]

XSL•FO
RenderX

12. Wong AMF. New concepts concerning the neural mechanisms of amblyopia and their clinical implications. Can J Ophthalmol 2012 Oct;47(5):399-409. [doi: 10.1016/j.jcjo.2012.05.002] [Medline: 23036539]
13. Loudon SE, Simonsz HJ. The history of the treatment of amblyopia. Strabismus 2005 Jun;13(2):93-106. [doi: 10.1080/09273970590949818] [Medline: 16020365]
14. Wallace MP, Stewart CE, Moseley MJ, et al. Compliance with occlusion therapy for childhood amblyopia. Invest Ophthalmol Vis Sci 2013 Sep 17;54(9):6158-6166. [doi: 10.1167/iovs.13-11861] [Medline: 23882695]
15. Loudon SE, Polling JR, Simonsz HJ. Electronically measured compliance with occlusion therapy for amblyopia is related to visual acuity increase. Graefe's Arch Clin Exp Ophthalmol 2003 Mar;241(3):176-180. [doi: 10.1007/s00417-002-0570-z]
16. Loudon SE, Passchier J, Chaker L, et al. Psychological causes of non-compliance with electronically monitored occlusion therapy for amblyopia. Br J Ophthalmol 2009 Nov;93(11):1499-1503. [doi: 10.1136/bjo.2008.149815] [Medline: 19661070]
17. Searle A, Vedhara K, Norman P, Frost A, Harrad R. Compliance with eye patching in children and its psychosocial effects: a qualitative application of protection motivation theory. Psychology, Health & Medicine 2000 Feb;5(1):43-54. [doi: 10.1080/135485000105990]
18. Webber AL, Wood JM, Gole GA, Brown B. Effect of amblyopia on self-esteem in children. Optom Vis Sci 2008 Nov;85(11):1074-1081. [doi: 10.1097/OPX.0b013e31818b9911] [Medline: 18981922]
19. Carlton J, Kaltenthaler E. Health-related quality of life measures (HRQoL) in patients with amblyopia and strabismus: a systematic review. Br J Ophthalmol 2011 Mar;95(3):325-330. [doi: 10.1136/bjo.2009.178889] [Medline: 20693563]
20. Carlton J, Kaltenthaler E. Amblyopia and quality of life: a systematic review. Eye (Lond) 2011 Apr;25(4):403-413. [doi: 10.1038/eye.2011.4] [Medline: 21274010]
21. Leal Vega L, Piñero DP, Hernández Rodríguez CJ, et al. Study protocol for a randomized controlled trial of the NEIVATECH virtual reality system to improve visual function in children with anisometropic amblyopia. BMC Ophthalmol 2022 Jun 7;22(1):253. [doi: 10.1186/s12886-022-02466-z] [Medline: 35672688]
22. Molina-Martín A, Leal-Vega L, de Fez D, Martínez-Plaza E, Coco-Martín MB, Piñero DP. Amblyopia treatment through immersive virtual reality: a preliminary experience in anisometropic children. Vision (Basel) 2023 May 19;7(2):42. [doi: 10.3390/vision7020042] [Medline: 37218960]
23. Astle AT, Webb BS, McGraw PV. Can perceptual learning be used to treat amblyopia beyond the critical period of visual development? Ophthalmic Physiol Opt 2011 Nov;31(6):564-573. [doi: 10.1111/j.1475-1313.2011.00873.x] [Medline: 21981034]
24. Martín S, Portela JA, Ding J, Ibarrondo O, Levi DM. Evaluation of a virtual reality implementation of a binocular imbalance test. PLoS One 2020;15(8):e0238047. [doi: 10.1371/journal.pone.0238047] [Medline: 32822405]
25. Hess RF, Thompson B. Amblyopia and the binocular approach to its therapy. Vision Res 2015 Sep;114:4-16. [doi: 10.1016/j.visres.2015.02.009] [Medline: 25906685]
26. Herbison N, Cobb S, Gregson R, et al. Interactive binocular treatment (I-BiT) for amblyopia: results of a pilot study of 3D shutter glasses system. Eye (Lond) 2013 Sep;27(9):1077-1083. [doi: 10.1038/eye.2013.113] [Medline: 23807383]
27. Kelly KR, Jost RM, Dao L, Beauchamp CL, Leffler JN, Birch EE. Binocular iPad game vs patching for treatment of amblyopia in children: a randomized clinical trial. JAMA Ophthalmol 2016 Dec 1;134(12):1402-1408. [doi: 10.1001/jamaophthalmol.2016.4224] [Medline: 27832248]
28. Li L, Xue H, Lai T, Xue Y, Luo G. Comparison of compliance among patients with pediatric amblyopia undergoing virtual reality-based and traditional patching method training. Front Public Health 2022;10:1037412. [doi: 10.3389/fpubh.2022.1037412]
29. Polat U, Ma-Naim T, Belkin M, Sagi D. Improving vision in adult amblyopia by perceptual learning. Proc Natl Acad Sci USA 2004 Apr 27;101(17):6692-6697. [doi: 10.1073/pnas.0401200101] [Medline: 15096608]
30. Zhang JY, Cong LJ, Klein SA, Levi DM, Yu C. Perceptual learning improves adult amblyopic vision through rule-based cognitive compensation. Invest Ophthalmol Vis Sci 2014 Apr 1;55(4):2020-2030. [doi: 10.1167/iovs.13-13739] [Medline: 24550359]
31. Dosher B, Lu ZL. Visual perceptual learning and models. Annu Rev Vis Sci 2017 Sep 15;3:343-363. [doi: 10.1146/annurev-vision-102016-061249] [Medline: 28723311]
32. Noah S, Bayliss J, Vedamurthy I, Nahum M, Levi D, Bavelier D. Comparing dichoptic action video game play to patching in adults with amblyopia. J Vis 2014 Aug 22;14(10):691-691. [doi: 10.1167/14.10.691]
33. Manh VM, Holmes JM, Lazar EL, et al. A randomized trial of a binocular iPad game versus part-time patching in children aged 13 to 16 years with amblyopia. Am J Ophthalmol 2018 Feb;186:104-115. [doi: 10.1016/j.ajo.2017.11.017] [Medline: 29196184]
34. Sen S, Singh P, Saxena R. Management of amblyopia in pediatric patients: current insights. Eye (Lond) 2022 Jan;36(1):44-56. [doi: 10.1038/s41433-021-01669-w] [Medline: 34234293]
35. Herbison N, MacKeith D, Vivian A, et al. Randomised controlled trial of video clips and interactive games to improve vision in children with amblyopia using the I-BiT system. Br J Ophthalmol 2016 Nov;100(11):1511-1516. [doi: 10.1136/bjophthalmol-2015-307798] [Medline: 26951772]

XSL•FO
RenderX

36. Chan HS, Tang YM, Do CW, Ho Yin Wong H, Chan LY, To S. Design and assessment of amblyopia, strabismus, and myopia treatment and vision training using virtual reality. Digit Health 2023;9:20552076231176638. [doi: 10.1177/20552076231176638] [Medline: 37312939]
37. Tjiam AM, Holtslag G, Vukovic E, et al. An educational cartoon accelerates amblyopia therapy and improves compliance, especially among children of immigrants. Ophthalmology 2012 Nov;119(11):2393-2401. [doi: 10.1016/j.ophtha.2012.05.035] [Medline: 22920669]
38. Vacca RA, Augello A, Gallo L, et al. Serious Games in the new era of digital-health interventions: a narrative review of their therapeutic applications to manage neurobehavior in neurodevelopmental disorders. Neurosci Biobehav Rev 2023 Jun;149:105156. [doi: 10.1016/j.neubiorev.2023.105156]
39. Bocci F, Ferrari A, Sarini M. Putting the gaming experience at the center of the therapy—the Video Game Therapy® approach. Healthcare (Basel) 2023;11(12):1767. [doi: 10.3390/healthcare11121767]
40. Loudon SE, Polling JR, Simonsz HJ. A preliminary report about the relation between visual acuity increase and compliance in patching therapy for amblyopia. Strabismus 2002 Jan;10(2):79-82. [doi: 10.1076/stra.10.2.79.8143]
41. Newsham D. A randomised controlled trial of written information: the effect on parental non-concordance with occlusion therapy. Br J Ophthalmol 2002 Jul;86(7):787-791. [doi: 10.1136/bjo.86.7.787] [Medline: 12084751]
42. Azizi-Soleiman F, Heidari-Beni M, Hemati Z, Kelishadi R. Designing and developing an educational-therapeutic game for improving healthy lifestyle in children and adolescents. International Journal of Computer Games Technology 2023 May 18;2023:1-7. [doi: 10.1155/2023/9897496]
43. Bress AP, Bellows BK, King JB, et al. Cost-effectiveness of intensive versus standard blood-pressure control. N Engl J Med 2017 Aug 24;377(8):745-755. [doi: 10.1056/NEJMsa1616035] [Medline: 28834469]
44. Pediatric Eye Disease Investigator Group, Holmes JM, Manny RE, et al. A randomized trial of binocular dig rush game treatment for amblyopia in children aged 7 to 12 years. Ophthalmology 2019 Mar;126(3):456-466. [doi: 10.1016/j.ophtha.2018.10.032] [Medline: 30352226]
45. Pavel IA, Bogdanici CM, Donica VC, et al. Computer vision syndrome: an ophthalmic pathology of the modern era. Med Bogota Colomb 2023;59(2):412. [doi: 10.3390/medicina59020412]
46. Lazarus RS. Emotion and Adaptation: Oxford University Press; 1991. [doi: 10.1093/oso/9780195069945.001.0001]
47. Yih J, Uusberg A, Taxer JL, Gross JJ. Better together: a unified perspective on appraisal and emotion regulation. Cogn Emot 2019 Feb;33(1):41-47. [doi: 10.1080/02699931.2018.1504749] [Medline: 30058449]
48. Watson L, Spence MT. Causes and consequences of emotions on consumer behaviour. Eur J Mark 2007 Jun 5;41(5/6):487-511. [doi: 10.1108/03090560710737570]
49. Ismagilova E, Dwivedi YK, Slade E. Perceived helpfulness of eWOM: emotions, fairness and rationality. Retail Consum Serv 2020 Mar;53:101748. [doi: 10.1016/j.jretconser.2019.02.002]
50. Kim M. Does playing a video game really result in improvements in psychological well-being in the era of COVID-19? J Retail Consum Serv 2021 Jul;61:102577. [doi: 10.1016/j.jretconser.2021.102577]
51. Kim J, Kim M. Spectator e-sport and well-being through live streaming services. Technol Soc 2020 Nov;63:101401. [doi: 10.1016/j.techsoc.2020.101401]
52. Ellsworth PC, Smith CA. Shades of joy: patterns of appraisal differentiating pleasant emotions. Cogn Emot 1988 Oct;2(4):301-331. [doi: 10.1080/02699938808412702]
53. Perugini M, Bagozzi RP. The role of desires and anticipated emotions in goal-directed behaviours: broadening and deepening the theory of planned behaviour. Br J Soc Psychol 2001 Mar;40(Pt 1):79-98. [doi: 10.1348/014466601164704] [Medline: 11329835]
54. Ma J, Gao J, Scott N, Ding P. Customer delight from theme park experiences. Ann Tour Res 2013 Jul;42:359-381. [doi: 10.1016/j.annals.2013.02.018]
55. Fereday J, Muir-Cochrane E. Demonstrating rigor using thematic analysis: a hybrid approach of inductive and deductive coding and theme development. Int J Qual Methods 2006 Mar;5(1):80-92. [doi: 10.1177/160940690600500107]
56. Wang J, Bian RW, Mo YZ. Validation of the Chinese version of the eight-item Morisky medication adherence scale in patients with type 2 diabetes mellitus. Journal of Clinical Gerontology and Geriatrics 2013 Dec;4(4):119-122. [doi: 10.1016/j.jcgg.2013.06.002]
57. Berlowitz DR, Foy CG, Kazis LE, et al. Effect of intensive blood-pressure treatment on patient-reported outcomes. N Engl J Med 2017 Aug 24;377(8):733-744. [doi: 10.1056/NEJMoa1611179] [Medline: 28834483]
58. Mayer I, Bekebrede G, Harteveld C, et al. The research and evaluation of serious games: toward a comprehensive methodology. Brit J Educational Tech 2014 May;45(3):502-527. [doi: 10.1111/bjet.12067]
59. Ryan RM, Deci EL. Self-Determination Theory: Basic Psychological Needs in Motivation, Development, and Wellness: The Guilford Press; 2017.
60. Modi AC, Pai AL, Hommel KA, et al. Pediatric self-management: a framework for research, practice, and policy. Pediatrics 2012 Feb;129(2):e473-e485. [doi: 10.1542/peds.2011-1635] [Medline: 22218838]
61. DeShazo J, Harris L, Pratt W. Effective intervention or child's play? A review of video games for diabetes education. Diabetes Technol Ther 2010 Oct;12(10):815-822. [doi: 10.1089/dia.2010.0030]

XSL•FO
RenderX

62. Fleming TM, Bavin L, Stasiak K, et al. Serious games and gamification for mental health: current status and promising directions. Front Psychiatry 2017 Jan 10;7:215. [doi: 10.3389/fpsyt.2016.00215] [Medline: 28119636]
63. Gee JP. Glued to games: how video games draw us in and hold us spellbound. Am J Play 2012;4(4):510.
64. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Behav Med Pract Policy Res 2017 Jun;7(2):254-267. [doi: 10.1007/s13142-016-0453-1]
65. Druin A. The role of children in the design of new technology. Behav Inf Technol 2002 Jan;21(1):1-25. [doi: 10.1080/01449290110108659]
66. Improving patient adherence with the Morisky Scale. Adherence. URL: https://adherence.cc [accessed 2025-09-12]

Abbreviations

CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
COREQ: Consolidated Criteria for Reporting Qualitative Studies
MMAS-8: 8-item Morisky Medication Adherence Scale
RCT: randomized controlled trial
UEQ: User Experience Questionnaire

Edited by A Coristine; submitted 07.05.24; peer-reviewed by GB Balu, J Zhang; revised version received 19.08.25; accepted 19.08.25; published 28.10.25.

Please cite as:
Liu B, Fan Y, Xu M, Chang F, Shi Y, Liu Z
Effectiveness of a Gamified Mobile App in Enhancing Treatment Adherence for Children With Amblyopia: Explorative Study
JMIR Serious Games 2025;13:e60309
URL: https://games.jmir.org/2025/1/e60309
doi:10.2196/60309

© Bo Liu, Yisheng Fan, Meng Xu, Fangyuan Chang, Yue Shi, Zhao Liu. Originally published in JMIR Serious Games (https://games.jmir.org), 28.10.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Development and Evaluation of a Monocular Camera–Based Mobile Exergame for at-Home Intervention in Individuals at High Risk of Type 2 Diabetes: Randomized Controlled Trial

Jianan Zhao[1], PhD; Dian Zhu[2], PhD; Yanan Wang[1], PhD; Yaqin Xia[1], PhD; Zeshi Zhu[2], MD; Jihong Yu[1], PhD

[1]College of Fashion and Design, Donghua University, East Yanan Road 1882, Shanghai, China

[2]School of Design, Shanghai Jiao Tong University, Shanghai, China

Corresponding Author:
Jihong Yu, PhD
College of Fashion and Design, Donghua University, East Yanan Road 1882, Shanghai, China

Abstract

Background: Exergames have emerged as effective interventions for promoting physical activity and preventing type 2 diabetes (T2D). Kinect-based exergames have demonstrated improvements in exercise adherence and health outcomes, but their high cost and reliance on specialized hardware hinder widespread home-based adoption. Recent advances in computer vision now enable monocular camera–based systems, offering a potentially cost-effective and scalable alternative for promoting physical activity at home.

Objective: This study aimed to evaluate the feasibility and user experience of monocular camera–based exergames as a home-based intervention for individuals at risk for T2D.

Methods: Forty-five community-dwelling individuals at high risk for T2D (mean age 47.12, SD 6.92 years) were recruited and randomized into three groups (n=15 each): (1) control group (traditional offline exercise), (2) Kinect group (Kinect-based exergame), and (3) monocular group (monocular camera–based exergame). Participants engaged in a 10-minute intervention once per week for 7 weeks. Data were collected at 3 time points: baseline (exercise performance: heart rate and perceived fatigue), postintervention (exercise performance and user experience, including game experience and intrinsic motivation), and follow-up (user engagement and qualitative feedback). One-way ANOVA was used for data analysis.

Results: Exercise performance was comparable across all groups, with no significant differences in heart rate (P=.76) or fatigue levels (P=.25). However, participants in the monocular group reported significantly lower fatigue than those in the control group (P=.04). Intrinsic motivation was significantly higher in both the Kinect (mean 35.13, SD 3.20) and monocular (mean 34.00, SD 4.41) groups than in the control group (mean 26.06, SD 1.87; P<.001), with no significant difference between the 2 exergame groups (P=.44). While most user experience measures showed no significant differences, the monocular group reported a higher perceived challenge (mean 3.45, SD 0.51) than the Kinect group (mean 2.96, SD 0.39; P=.09). Additionally, the monocular group exhibited higher engagement, as evidenced by more frequent use, fewer challenges, and a greater intention to continue using the system.

Conclusions: Monocular camera–based exergame is a feasible and effective solution for promoting physical activity in individuals at risk for T2D. It offers motivational and experiential benefits similar to Kinect-based systems but requires less costly and more accessible equipment. These findings suggest that monocular systems have strong potential as scalable tools for home-based chronic disease prevention.

Trial Registration: ClinicalTrials.gov NCT06950528; https://clinicaltrials.gov/study/NCT06950528

(JMIR Serious Games 2025;13:e75823) doi:10.2196/75823

KEYWORDS

monocular camera; exergames; type 2 diabetes; gamification; disease prevention

Introduction

Type 2 diabetes (T2D) is a chronic, progressive condition that poses a major global health challenge, with the number of affected individuals projected to reach 500 million by 2030 [1,2]. The mean age of T2D diagnosis is 45 years [3]. Individuals older than 35 years, those with obesity, having a family history of diabetes, not physically active, and having prediabetes are considered to have high risk of T2D [4]. Among high-risk populations, lifestyle interventions, particularly those that encourage physical activity, are crucial for delaying or preventing T2D onset [5]. Despite the proven benefits of exercise, barriers such as lack of motivation, limited access to

XSL•FO
RenderX

fitness resources, and insufficient guidance persist [6,7]. To address these challenges, technology-driven solutions such as digital exergames have emerged as promising tools to promote physical activity in an engaging and accessible way [8].

Exergames combine exercise with interactive gaming and offer real-time feedback, personalized experiences, and gamified rewards, all of which enhance physical activity [9]. Motion-sensing technology has been central to the development of exergames, with Microsoft's Kinect being a landmark innovation [10]. Released in 2010, Kinect revolutionized motion-sensing games by enabling real-time body tracking and virtual avatar control through natural movements [11]. Kinect-based exergames have been extensively studied for their impact on physical performance, therapy adherence, and clinical outcomes [12]. However, due to reliance on specialized hardware, high costs, ineffective marketing strategies, and limited home usability, the Kinect v2 was discontinued in 2023.

In recent years, advancements in deep learning have enabled a new generation of motion-sensing technologies, including monocular human pose tracking, which relies solely on monocular cameras [13]. This approach allows for real-time mapping of users' movements in virtual environments without requiring expensive depth cameras or wearable devices [14].

The monocular camera–based exergame emerges as a low-cost, noncontact technology and offers a practical solution for home-based exergame training [15]. Its potential for enhancing accessibility and usability makes it an attractive area for exploration. However, while the potential of monocular camera–based pose tracking has been identified, its use in designing and evaluating exergames tailored to patients with T2D remains rarely explored, and the quality assessment of the exergame is rarely discussed as well [16,17]. This creates a need for accessible, scalable solutions that not only use cutting-edge technologies but also align with the practical needs and preferences of the target audience.

To address these gaps, this study introduces an exergame specifically designed for individuals at high risk of diabetes, using monocular camera technology to ensure affordability and ease of use. The game is informed by evidence-based physical activities to promote metabolic health and incorporates features to enhance user engagement and long-term adherence. This study aims to evaluate the effectiveness and user experience of monocular camera–based exergame to understand its potential to replace Kinect in providing home-based exercise intervention. The primary outcome of this study is to demonstrate that there is no significant difference in physical activity levels between the experimental and control groups, as evidenced by the lack of variation in heart rate changes before and after the intervention. The secondary outcome focuses on evaluating the impact of the monocular camera–based group on user experience, providing insights into the acceptability and usability of the intervention from the participants' perspective.

Methods

Study Design

This 3-arm randomized controlled trial used a 1:1:1 allocation ratio to evaluate the effectiveness of an exergame designed using monocular camera technology for individuals at high risk of developing T2D. A total of 45 participants meeting specific inclusion criteria were recruited and randomly assigned to 1 of 3 groups: a control group performing traditional video-guided exercises, a Kinect-based exergame group, and a monocular camera–based exergame group. The sample size was calculated based on a power analysis with an expected medium size at 0.48, where previous research reported medium effect sizes (0.40 - 0.50) in evaluating usability and adherence in Kinect-based exergames for individuals with prediabetes [18], aiming for 80% power and a significance level of .05 to detect meaningful differences between the groups. The intervention was conducted in a controlled setting, ensuring consistency across groups while collecting physiological and experiential data.

Recruitment

A total of 45 individuals at high risk for T2D were recruited through offline outreach facilitated by community managers affiliated with local residential service centers in Minhang District, Shanghai, China. They supported recruitment by identifying eligible participants through existing health records and outreach channels. The inclusion criteria were as follows: (1) 35 years of age and older; (2) a score of 25 or higher on the Chinese Diabetes Risk Score; (3) community-dwelling individuals, not residing in assisted living or long-term care facilities; (4) physically capable of engaging in light to moderate exercise, as determined by self-report and physician clearance; (5) normal cognitive function that enables the participant to complete the experiment independently or with minimal assistance; and (6) written informed consent provided by participants or their families. Exclusion criteria included individuals with (1) diagnosis of T1D or T2D, (2) current participation in another exercise intervention study, (3) severe cognitive impairment with Mini-Mental State Examination score of <24, and (4) major mobility limitations such as severe osteoarthritis and recent orthopedic surgery.

Experimental Setting and Equipment

Experiment Group (Monocular Camera–Based Exergame Design)

Overview

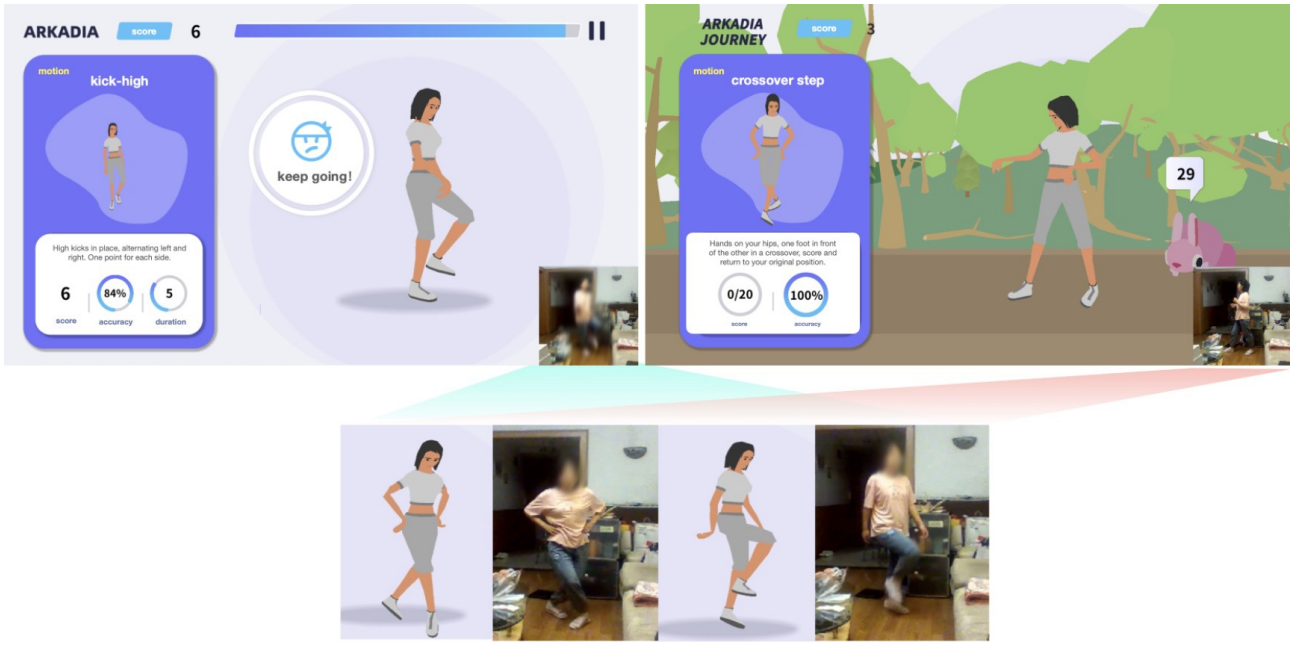
This intervention is an exercise game tailored for individuals at high risk of diabetes. It incorporates 6 movements adapted from the "Diabetes Health Exercises," which include toe stretches, high knees, side arm raises, punching exercises, arm stretches with hip extensions, and elbow-to-chest expansions. These movements were selected for their simplicity, defined training objectives, and distinct characteristics that facilitate recognition and replication. The exercises aim to enhance peripheral blood circulation and target key muscle groups such as the triceps, biceps, gluteal muscles, and lower limbs. Each

XSL•FO
RenderX

session is structured to last 10 minutes, comprising 5 minutes of aerobic exercise and 5 minutes of resistance training, thereby providing a balanced regimen that aligns with the physiological needs of the target population.

The game uses a virtual avatar to replicate users' movements, thereby fostering an immersive and engaging experience (Figure 1). A monocular camera captures users' movements in real time, which are analyzed through pose estimation algorithms and subsequently mapped onto the virtual avatar. Users interact with the game by following on-screen visual demonstrations, presented as either static images or animations, to perform the prescribed exercises. Movement accuracy is evaluated by the system, with scores awarded based on performance. To enhance user motivation and adherence, the game incorporates a reward system, where points earned through accurate execution can be redeemed for in-game rewards, such as unlocking background music, avatar customization options, and new virtual environments.

Figure 1. Interface design of the exergame. The exergame interface includes real-time avatar mirroring, performance feedback indicators, instructional visuals, and a reward progress tracker.



Technical Framework of the Monocular Camera–Based Exergame

The intervention is designed with a robust technical foundation, using Figma for interface design and Unity 2019.4 for game development. The motion capture system leverages the open-source ThreeDPoseTracker framework, based on VNect technology, which processes low-resolution red, green, and blue video streams using Unity's Barracuda system and a pretrained ONNX model. This framework accurately detects 10 critical body landmarks—wrists, elbows, ankles, knees, waist, and head—estimating their 3D spatial positions in real time.

The real-time pose correction process compares the user's detected body posture with a predefined regulated posture within the exergame. The correction formula involves angular calculations between vectors formed by detected body landmarks, as shown in the image. Specifically, each detected angle (θ_i) is compared with the corresponding regulated angle (θ'_i), and the overall difference (D_2) is calculated as $D_2=\sum(\theta i-\theta'_i)$. To quantify the degree of correction required, a recognition threshold parameter (D_{st}) and a baseline recognition rate parameter (S_{st}) are used. The final evaluation metric (S) is determined by the piecewise function:

$$S=\{Dst-D\times100-SstDst+Sst,0\leq D<Dst0,D>Dst\}$$

This ensures that minor deviations from the regulated posture are dynamically adjusted, while significant deviations are flagged as errors. Temporal filtering techniques are applied to improve the smoothness and precision of pose estimations, which are then mapped to a skeletal model in Unity for seamless motion replication.

By integrating real-time motion capture and precise pose correction, the system provides immediate feedback, ensuring accurate and adaptive user movements during gameplay. This enhances the usability and efficacy of the intervention, making it a scientifically grounded and technologically advanced solution for promoting physical activity, particularly for diabetes prevention.

Kinect Group

Kinect group features the same software functionality and design as the application based on bullet-screen cameras, but the hardware is built using a Kinect sensor. The gamified platform is primarily developed and presented using Unity 3D. The Kinect 3D motion-sensing camera incorporates real-time dynamic capture and image recognition capabilities, offering new possibilities for interactive approaches to motion therapy.

Control Group

The control group followed a set of 6 instructor-led exercise videos, which included crossover steps, high knees, lateral

XSL•FO
RenderX

raises, punching movements, downward leg punches from a standing position, and elbow-to-chest expansions. Each session in the control group was designed to reflect the same aerobic and resistance exercise level as the Kinect group and the experimental group, ensuring comparability across groups in terms of physical exertion and activity content.

Procedure

Participant recruitment was conducted from March to April 2024 in the Hongqiao Community. Individuals meeting the inclusion criteria were provided with comprehensive information about the study and signed written informed consent. Baseline demographic data were collected prior to random assignment using block randomization by researcher into one of three groups: (1) control group (traditional exercise control), (2) Kinect group (Kinect-based exergame), or (3) monocular group (monocular camera–based exergame).

Three days prior to the intervention, all participants received instructional videos demonstrating the 6 targeted movements to be performed during the exercise sessions. Participants were instructed to review and practice these movements to ensure familiarity with the protocol.

On the day prior to the intervention, trained researchers and caregivers delivered and assisted in setting up the required equipment—either a Kinect V2 sensor or an Aoni C33 monocular camera (manufactured by Shenzhen Aoni Electronic Industry Co, Ltd, resolution: 1920×1080 pixels, frame rate: 30 fps, field of view: 90°)—in participants' homes. Setup was scheduled in the afternoon, approximately 1 - 2 hours post meal, to ensure participants' comfort and safety during physical activity.

On the first intervention day, baseline data, including age, gender, T2D risk, resting heart rate, and perceived fatigue, were measured prior to the session. Each intervention session lasted 10 minutes and consisted of 2 components: 5 minutes of aerobic exercise followed by 5 minutes of resistance training. Participants in the control group followed a prerecorded instructional video demonstrating the 6 aerobic movements, followed by guided resistance exercises. The monocular group used a custom-developed exergame based on monocular camera technology. Real-time visual instructions were displayed on a television screen, guiding participants through the same movement protocol. The Kinect group used an exergame designed with Kinect-based skeletal tracking, offering a similar visual and instructional format as the monocular system.

During the aerobic component, participants performed 6 distinct movements, each lasting 50 seconds. For the resistance component, participants selected their preferred resistance level (ranging from 0.25 to 1.5 kg), and resistance bands were applied to the arms and legs under caregiver supervision. Caregivers and researchers refrained from intervening during the session unless directly requested by participants to minimize external influence.

Immediately following the first intervention, participants' postexercise heart rate, perceived fatigue (via Borg Rating of Perceived Exertion Scale), intrinsic motivation (via Intrinsic Motivation Inventory [IMI]), and user experience (including perceived enjoyment, challenge, and usability) were assessed using standardized instruments.

Participants were then instructed to continue the 10-minute intervention daily for 7 consecutive days independently without additional notification from the researchers. They were asked to document their participation by recording short videos as proof of session completion. At the end of the 1-week period, researchers returned to collect the devices and retrieve adherence data, including the number of completed sessions and any user feedback.

Outcome Measure

Exercise Performance

To evaluate the amount of exercise during the experiment, heart rate and perceived fatigue were assessed among participants at the baseline and immediately after the intervention. The postintervention heart rate was collected during the first 10 seconds postexercise to estimate physiological responses. Considering the general physical condition of individuals at high risk of diabetes, the study used 50% - 80% of the maximum heart rate as the target exercise intensity. The maximum heart rate was calculated using the formula: 208 − (Age × 0.7). Additionally, perceived fatigue, a widely accepted parameter in exercise assessments for diabetes-related fields, was used to supplement the evaluation of physical activity. The Borg Rating of Perceived Exertion Scale [19], ranging from 6 to 20, was used to assess subjective fatigue and compare perceived exercise intensity across groups.

Perceived User Experience

At the first session of intervention, participants in the Kinect and monocular groups completed the Game Experience Questionnaire (GEQ) to assess their user experience [20]. The GEQ was used to evaluate and compare the impact of different technologies on immersion and the overall experience of exergames. Furthermore, the Interest/Enjoyment Subscale of the IMI was administered to participants in the control group, Kinect group, and monocular group [21]. This subscale evaluated and compared intrinsic motivation and enjoyment associated with physical activity across the 3 groups. At the end of the 1-week experiment period, user engagement was further quantified by tracking the frequency of intervention use over the 1-week period, based on participants' video-recorded usage logs. To complement the quantitative data, a brief qualitative assessment was conducted at the end of the intervention week. Participants responded to open-ended questions such as: "Did you experience any issues while using the exergame?" and "Would you be interested in continuing to use the exergame in the future?"

Data Analysis

Data collected from the physiological measurements (resting and exercise heart rates) and the questionnaires were analyzed to compare the differences between groups and evaluate the effectiveness of the interventions. ANOVA analysis was used to understand the differences of the data collected from 3 experiment groups [22], and 2-tailed *t* test (for normally distributed result) or nonparametric test (for non–normally

XSL•FO
RenderX

distributed result) was used to understand the significance of differences between 2 groups. For pairwise comparisons between 2 groups, independent-samples 2-tailed *t* tests were applied for normally distributed data. For data not meeting normality assumptions, nonparametric tests were used: the Kruskal-Wallis test for comparisons among 3 groups and the Mann-Whitney *U* test for pairwise group comparisons. Statistical analyses were conducted using SPSS (version 26.0; IBM Corp), with a significance level set at $P<.05$. Cohen *f* value is used to understand the effect size, and the critical points for small, medium, and large effect sizes are 0.1, 0.40, and 0.80, respectively.

Ethical Considerations

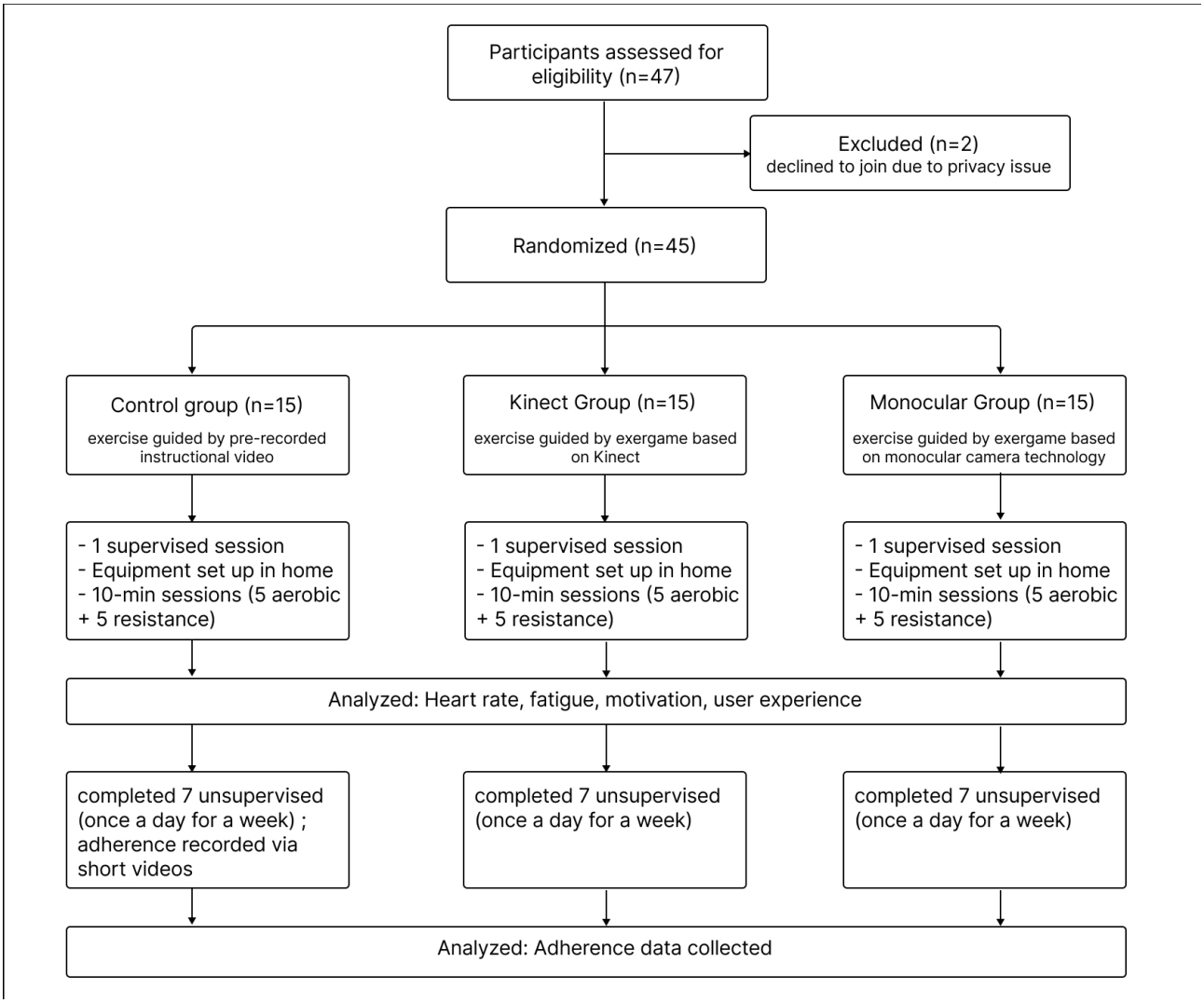
This study was approved by the Shanghai Jiao Tong University's institutional review board (H20230160I), and all procedures were conducted in accordance with the ethical standards outlined in the Declaration of Helsinki. Due to the initial pilot-phase nature of the project, the trial was retrospectively registered at ClinicalTrials.gov (NCT06950528), prior to data analysis. All participants provided written informed consent prior to participation. The study ensured strict confidentiality of participant data, and no personally identifiable information was collected or disclosed. Participants received a US $7 shopping voucher after completing the study. No identifiable images or recordings were used in the manuscript or supplementary materials.

Results

Overview

This study recruited 45 participants identified as being at high risk for T2D. The CONSORT (Consolidated Standards of Reporting Trials) flowchart is shown in Figure 2. Participants aged between 40 and 54 years, with a mean age of 47.12 (SD 6.92) years. Participants were classified as high-risk based on the Chinese Diabetes Risk Score, with an average mean score of 33.38 (SD 5.04). As shown in Table 1, an ANOVA analysis of baseline characteristics among the 3 participant groups revealed no statistically significant differences. No negative incidents were reported.

Figure 2. The CONSORT (Consolidated Standards of Reporting Trials) flowchart of the experiment.



XSL•FO
RenderX

Table . Baseline data of the recruited participants.

Characteristics	Control group (n=15)	Kinect group (n=15)	Monocular group (n=15)	*P* value
Gender (female), n (%)	8 (53.33%)	9 (60.0%)	8 (53.33%)	__[a]
Age (year), mean (SD)	47.27 (6.59)	47.93 (6.85)	46.17 (7.33)	.80
T2D[b] risk, mean (SD)	33.07 (5.02)	33.60 (4.75)	33.47 (5.36)	.84
Heart rate, mean (SD)	79.46 (10.23)	80.06 (10.01)	78.02 (8.66)	.48
Perceived fatigue, mean (SD)	9.27 (1.53)	9.41 (1.12)	9.28 (1.06)	.81

[a]Not available.

[b]T2D: type 2 diabetes.

Outcomes in Exercise Performance

The amount of exercise was assessed through heart rate and perceived fatigue. As shown in Table 2, the average heart rate for the control group, Kinect group, and monocular group was 86.34 (SD 7.48), 90.81 (SD 7.71), and 89.29 (SD 6.52), respectively. The between-group comparison revealed no significant difference in heart rate (P=.76; Cohen f=0.117), indicating a small effect size.

Exercise performance was evaluated using 2 indicators: heart rate as an objective physiological measure and perceived fatigue as a subjective self-report metric. As shown in Table 2, the average heart rates for the control, Kinect, and monocular groups were 86.34 (SD 7.48) bpm, 90.81 (SD 7.71) bpm, and 89.29 (SD 6.52) bpm, respectively. A one-way ANOVA revealed no statistically significant differences among the groups ($F_{2,42}$=0.28; P=.76), with a small effect size (Cohen f=0.117). These results suggest that the type of system used did not substantially impact cardiovascular exertion during the exercise session.

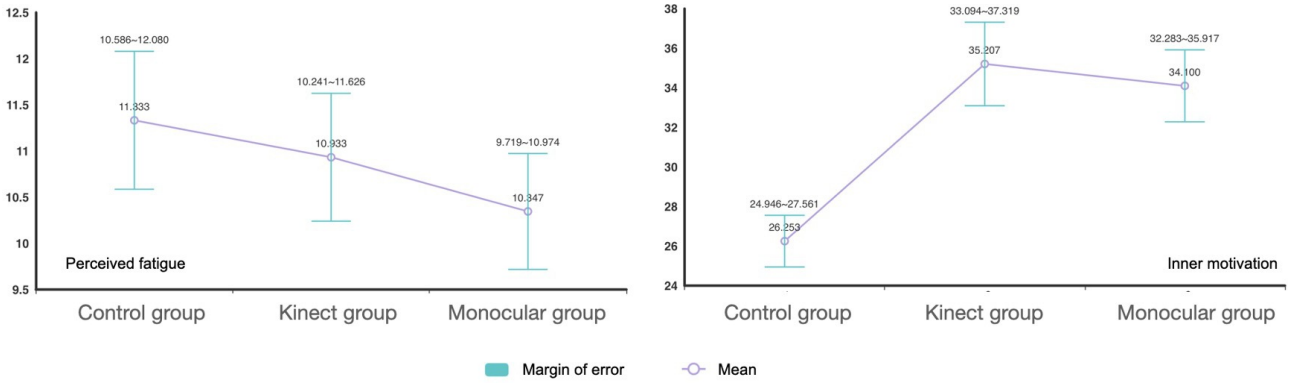
In terms of perceived fatigue, all participants reported increased fatigue after the intervention (P<.05). As shown in Figure 3, the control group reported the highest average fatigue level (mean 11.60, SD 1.12). However, between-group comparisons did not reach statistical significance (P=.25; Cohen f=0.265), indicating a moderate effect size. Post hoc comparisons revealed that the monocular group experienced significantly lower fatigue than the control group (P=.04), while the Kinect group's difference was not significant (P=.15). These findings suggest a potential advantage of the monocular system in reducing perceived exertion during gameplay.

Table . Data collected before and after the intervention in 3 groups of participants.

Variable	Control group	Kinect group	Monocular group	*F* test (*df*)	*P* value	Cohen *f*
Heart rate				0.28 (2,42)	.76	0.117
Data collected at baseline, mean (SD)	79.46 (10.23)	80.06 (10.01)	78.02 (8.66)			
Data collected after experiment, mean (SD)	86.34 (7.48)	90.81 (7.71)	89.29 (6.52)			
P value	.038	.004	.003			
Perceived fatigue				1.442 (2,42)	.25	0.265
Data collected at baseline, mean (SD)	9.27 (1.53)	9.41 (1.12)	9.28 (1.06)			
Data collected after experiment, mean (SD)	11.60 (1.12)	10.60 (1.45)	10.33 (1.49)			
P value	<.001	.007	.026			
Inner motivation, mean (SD)	26.06 (1.87)	35.13 (3.20)	34.00 (4.41)	29.701 (2,42)	<.001	1.204

XSL•FO
RenderX

Figure 3. Comparison of perceived fatigue and intrinsic motivation across 3 participant groups.



Outcomes in User Experience

User experience following the first intervention session was assessed through self-reported measures of intrinsic motivation and game experience. Intrinsic motivation scores, as measured by the Interest/Enjoyment subscale of the IMI, were as follows: control group (mean 26.06, SD 1.87), Kinect group (mean 35.13, SD 3.20), and monocular group (mean 34.00, SD 4.41). A one-way ANOVA revealed a significant between-group difference in intrinsic motivation (P<.001), with a large effect size (Cohen f=1.204). Post hoc comparisons indicated that both the Kinect-based and monocular camera–based exergames significantly enhanced intrinsic motivation compared with those of the control group. However, there was no statistically significant difference between the Kinect and monocular groups (P=.44), suggesting comparable motivational effects across the 2 exergaming modalities.

Game experience, as measured by the GEQ and shown in Table 3, revealed significant differences in perceived challenge. Participants in the monocular group reported a higher challenge score (mean 3.45, SD 0.39) than those in the Kinect group (mean 2.96, SD 0.51), with the difference reaching statistical significance (t_{28}=−2.71; P=.01). One possible explanation is the Kinect system's superior motion-tracking capabilities, which may have facilitated a stronger sense of virtual body ownership and smoother gameplay, thus reducing the perceived difficulty.

Table . Results of the Game Experience Questionnaire for the Kinect and monocular groups.

Dimension	User feedbacks, mean (SD)		*t* test (*df*)	*P* value
	Kinect group (n=15)	Monocular group (n=15)		
Sensory and imaginative immersion	3.40 (0.38)	3.51 (0.79)	−2.034 (28)	.15
Flow	3.40 (0.54)	3.58 (0.63)	−0.811 (28)	.43
Competence	2.90 (0.60)	3.16 (0.71)	−1.049 (28)	.30
Tension or annoyance	0.10 (0.38)	0.08 (0.28)	0.124 (28)	.90
Challenge	2.96 (0.51)	3.45 (0.39)	−2.71 (28)	.01
Positive affect	3.53 (0.44)	3.79 (0.39)	−1.579 (28)	.13
Negative affect	0.33 (0.67)	0.41 (0.70)	−0.688 (28)	.50

For all other dimensions of game experience—including sensory and imaginative immersion, flow, perceived competence, tension or annoyance, and emotional responses (positive and negative)—no statistically significant differences were observed between the 2 exergame groups (P>.05). While not statistically significant, participants in the Kinect group reported slightly higher tension levels, potentially due to unfamiliarity with the hardware. Conversely, the monocular group reported marginally higher negative affect, which may reflect minor limitations in motion-tracking responsiveness or accuracy.

User engagement over the 1-week intervention period was measured by the number of completed exercise sessions. On average, participants in the control group completed 2.33 sessions (SD 1.29), the Kinect group completed 2.93 sessions (SD 1.79), and the monocular group completed 3.60 sessions (SD 1.59). Although the monocular group exhibited the highest level of usage, between-group comparisons did not reach statistical significance (P=.49). Similarly, no significant difference was observed between the Kinect and monocular groups (P=.29).

A thematic analysis of the qualitative user feedback was conducted to contextualize participants' experiences across groups. Feedback was first coded and then categorized into three major themes: usability, motivation, and future intention to use. (1) *Usability*: Most participants across all groups found the systems easy to operate. However, specific issues were reported. Four participants in the Kinect group mentioned challenges related to hardware setup (eg, spatial calibration and sensor alignment), while 1 participant in the monocular group cited difficulties with interface responsiveness. (2) *Motivation*: Lack of engagement was a recurring theme in the control group, with 6 participants indicating low motivation due to the static

XSL•FO
RenderX

and repetitive nature of the game. (3) *Future intention to use*: Despite these challenges, participants in the exergame groups showed strong willingness to continue. In the Kinect group, 66.67% (10/15) expressed interest in future use, while 80% (12/15) of the monocular group reported a clear intention to continue engaging with the exergames.

Discussion

Principal Results

In this study, we evaluated the impact of exercise interventions using 3 modalities—traditional offline exergames, Kinect-based exergames, and monocular camera–based exergames—on exercise performance and user experience in a high-risk T2D population. The findings revealed several noteworthy insights into the potential of monocular camera–based exergames as an affordable and convenient alternative to Kinect-based systems in developing exergame for high-risk T2D population.

From an exercise performance perspective, all 3 modalities—traditional, Kinect-based, and monocular camera–based exergames—elicited comparable elevations in participants' heart rates, indicating similar levels of cardiovascular engagement. This finding aligns with prior research demonstrating that exergames can effectively promote physical activity across diverse technological platforms [23]. While the heart rate differences did not reach statistical significance, the effect size suggests a small practical effect that may still be meaningful in public health applications, especially for preventive care in at-risk populations.

Notably, participants reported lower perceived fatigue when engaging with digital exergames, particularly the monocular camera–based system, and although these differences did not reach statistical significance, the moderate effect size and significant pairwise difference between the monocular and control groups suggest a clinically relevant trend. This trend suggests that digital interventions may offer a more comfortable exercise experience while maintaining effectiveness, corroborating previous studies that have highlighted the potential of exergames to enhance exercise enjoyment and adherence [24]. The observed lower perceived fatigue in the monocular camera–based exergame group may be attributed to the system's ability to provide real-time feedback and adjust to users' movements, thereby enhancing engagement and reducing the subjective experience of exertion. Such features are consistent with findings from studies emphasizing the importance of interactive and adaptive elements in exergame design to optimize user experience and outcomes [25].

User Experience and Engagement

In terms of user experience, digital exergames were found to foster higher levels of inner motivation than traditional offline exergames, highlighting their potential to engage users more effectively [26,27]. The difference in inner motivation showed a large effect size, indicating the robust motivational potential of digital exercise systems for behavior change. Although there was no significant difference in inner motivation between Kinect-based and monocular camera–based exergames, these 2 systems exhibited distinct strengths in the user experience. Monocular camera–based exergames excelled in providing sensory and imaginative immersion, flow, competence, and positive affect, offering an engaging and enjoyable exercise environment [28]. In contrast, Kinect-based exergames performed better in the dimensions of tension, negative affect, and challenge, with the challenge dimension showing a statistically significant advantage, suggesting Kinect's ability to reduce negative emotional responses during gameplay [29]. The higher challenge scores in the monocular group may reflect user frustration stemming from occasional inaccuracies in motion tracking or delayed feedback, which can reduce the sense of control and increase cognitive effort.

To enhance the long-term appeal and sustain user engagement, current research indicating incorporating adaptive difficulty levels, personalization features, and reinforcement mechanisms is crucial [30]. Adaptive difficulty can dynamically adjust the game's challenge to match the user's skill level, preventing boredom and frustration. Reinforcement mechanisms, including real-time feedback and rewards, can further encourage consistent participation [31]. Future design may enhance these design elements to improve adherence and motivation in exergame interventions.

Nevertheless, participants using the monocular camera system reported fewer setup difficulties and a higher willingness to continue using the system postintervention. Compared with Kinect, the monocular system's low hardware burden, environmental adaptability, and plug-and-play configuration make it highly promising for home-based and scalable interventions. These findings highlight the practical potential of monocular camera–based exergames as an accessible solution for promoting physical activity in resource-limited settings, where traditional fitness infrastructure or expensive hardware such as Kinect may not be feasible.

Limitations

Despite encouraging results, this study has several limitations. First, the small sample size and short follow-up duration limit generalizability. Future studies should include larger, more demographically diverse populations. Second, the controlled indoor setup does not fully reflect real-world conditions; testing in-home or community settings is needed to assess ecological validity. Third, the study did not examine long-term adherence or health outcomes (eg, HbA_{1c}, weight, and insulin sensitivity), which are critical for evaluating the sustained impact of digital interventions. Fourth, although Kinect-based exergames scored higher in the challenge dimension, participant feedback suggested that the monocular system's challenge scores may have resulted from tracking errors or interface latency, rather than intentional game difficulty—indicating a need to distinguish between technical and design-related difficulty. Fifth, the study did not conduct subgroup analyses based on gender, which could influence user experience and outcomes due to differences in physical ability, digital familiarity, or motivational preferences. Sixth, the study was originally planned to be conducted at a different location; due to logistical challenges, it was relocated to another community site, which may affect comparability with the original protocol.

XSL•FO
RenderX

Future research should evaluate long-term health outcomes such as HbA_{1c}, insulin sensitivity, and weight control associated with these interventions; conduct trials in real-world environments such as community centers or homes to assess ecological validity; investigate personalized game adaptation strategies that adjust challenge and feedback based on individual skill level and fatigue patterns; and develop cross-platform, low-barrier versions of monocular camera–based systems for scalable deployment in resource-limited settings.

Conclusions

The findings suggest that monocular camera–based exergame offers a viable alternative to Kinect-based exergame for at-home use by T2D high-risk populations. It provides similar exercise outcomes and user experiences while offering the advantages of simplicity and greater adaptability. However, further improvements are needed to enhance user experience, particularly in balancing the challenge dimension.

Acknowledgments

The authors extend their sincere gratitude to all participants who took part in this study. This research is supported by the Shanghai Pujiang Program (project number 22PJC002).

Data Availability

The datasets are available from the corresponding author upon reasonable request.

Conflicts of Interest

None declared.

Editorial Notice

This randomized study was retrospectively registered on April 10, 2025, explained by authors as due to lack of prior awareness of the prospective registration requirement. Institutional review board approval was granted on January 12, 2024, the date of first patient enrollment was March 1, 2024, and the last patient data were collected on April 30, 2024. Authors uploaded the originally approved institutional review board protocol, which is largely similar to the manuscript submitted, save for the following discrepancies: the study was conducted in a different location than originally planned (Minhang District, Shanghai instead of Keqiao District, Shaoxing), and the full intervention schedule (1-week follow-up) was not detailed in the original protocol. Outcome measurement was the same. As a result, and because the risk of bias appears low, the editor granted an exception from ICMJE rules mandating prospective registration of randomized trials. However, readers are advised to carefully assess the validity of any potential explicit or implicit claims related to primary outcomes or effectiveness due to the retrospective registration.

Checklist 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File, 367 KB - games_v13i1e75823_app1.pdf]

References

1. Ahmad E, Lim S, Lamptey R, Webb DR, Davies MJ. Type 2 diabetes. Lancet 2022 Nov;400(10365):1803-1820. [doi: 10.1016/S0140-6736(22)01655-5]
2. Basu S, Yudkin JS, Kehlenbrink S, et al. Estimation of global insulin use for type 2 diabetes, 2018–30: a microsimulation analysis. Lancet Diabetes Endocrinol 2019 Jan;7(1):25-33. [doi: 10.1016/S2213-8587(18)30303-6]
3. Carrillo-Larco RM, Guzman-Vilca WC, Xu X, Bernabe-Ortiz A. Mean age and body mass index at type 2 diabetes diagnosis: pooled analysis of 56 health surveys across income groups and world regions. Diabet Med 2024 Feb;41(2):e15174. [doi: 10.1111/dme.15174] [Medline: 37422703]
4. Risk factors for type 2 diabetes. National Institutes of Health. 2022. URL: https://www.niddk.nih.gov/health-information/diabetes/overview/risk-factors-type-2-diabetes [accessed 2025-07-31]
5. Magkos F, Hjorth MF, Astrup A. Diet and exercise in the prevention and treatment of type 2 diabetes mellitus. Nat Rev Endocrinol 2020 Oct;16(10):545-555. [doi: 10.1038/s41574-020-0381-5] [Medline: 32690918]
6. Dnes N, Coley B, Frisby K, et al. “A little bit of a guidance and a little bit of group support”: a qualitative study of preferences, barriers, and facilitators to participating in community-based exercise opportunities among adults living with chronic pain. Disabil Rehabil 2021 Nov;43(23):3347-3356. [doi: 10.1080/09638288.2020.1742801] [Medline: 32223460]
7. Farholm A, Sørensen M. Motivation for physical activity and exercise in severe mental illness: a systematic review of intervention studies. Int J Ment Health Nurs 2016 Jun;25(3):194-205. [doi: 10.1111/inm.12214] [Medline: 26916699]
8. Choi SD, Guo L, Kang D, Xiong S. Exergame technology and interactive interventions for elderly fall prevention: a systematic literature review. Appl Ergon 2017 Nov;65(65):570-581. [doi: 10.1016/j.apergo.2016.10.013] [Medline: 27825723]

XSL•FO
RenderX

9. Schättin A, Pickles J, Flagmeier D, et al. Development of a novel home-based exergame with on-body feedback: usability study. JMIR Serious Games 2022 Dec 6;10(4):e38703. [doi: 10.2196/38703] [Medline: 36472900]
10. Huang K, Zhao Y, He R, et al. Exergame-based exercise training for depressive symptoms in adults: a systematic review and meta-analysis. Psychol Sport Exerc 2022 Nov;63:102266. [doi: 10.1016/j.psychsport.2022.102266]
11. Hu MC, Chen CW, Cheng WH, Chang CH, Lai JH, Wu JL. Real-time human movement retrieval and assessment with Kinect sensor. IEEE Trans Cybern 2015 Apr;45(4):742-753. [doi: 10.1109/TCYB.2014.2335540] [Medline: 25069133]
12. Da Gama A, Fallavollita P, Teichrieb V, Navab N. Motor rehabilitation using Kinect: a systematic review. Games Health J 2015 Apr;4(2):123-135. [doi: 10.1089/g4h.2014.0047] [Medline: 26181806]
13. Fan Z, Zhu Y, He Y, Sun Q, Liu H, He J. Deep learning on monocular object pose detection and tracking: a comprehensive overview. ACM Comput Surv 2023 Apr 30;55(4):1-40. [doi: 10.1145/3524496]
14. Delibasoglu I, Kosesoy I, Kotan M, Selamet F. Motion detection in moving camera videos using background modeling and FlowNet. J Vis Commun Image Representation 2022 Oct;88(88):103616. [doi: 10.1016/j.jvcir.2022.103616]
15. Chung YY, Annaswamy TM, Prabhakaran B. Performance and user experience studies of HILLES: home-based immersive lower limb exergame system. Presented at: Proceedings of the 14th Conference on ACM Multimedia System; Jun 7-10, 2023. [doi: 10.1145/3587819.3590985]
16. Assad O, Hermann R, Lilla D, Mellies B, Meyer R, Shevach L. Motion-based games for Parkinson's disease patients. In: Entertainment Computing—ICEC 2011: Heidelberg; 2011. URL: https://link.springer.com/chapter/10.1007/978-3-642-24500-8_6 [accessed 2025-07-31]
17. Dill S, Müller PN, Caserman P, Göbel S, Hoog Antink C, Tregel T. Sensing in exergames for efficacy and motion quality: scoping review of recent publications. JMIR Serious Games 2024 Nov 5;12(1):e52153. [doi: 10.2196/52153] [Medline: 39499916]
18. Li J, Theng YL, Foo S. Effect of exergames on depression and anxiety in older adults: a systematic review and meta-analysis. J Med Internet Res 2020;23(6):e16210. [doi: 10.2196/16210]
19. Chen MJ, Fan X, Moe ST. Criterion-related validity of the Borg ratings of perceived exertion scale in healthy individuals: a meta-analysis. J Sports Sci 2002 Nov;20(11):873-899. [doi: 10.1080/026404102320761787] [Medline: 12430990]
20. Johnson D, Gardner MJ, Perry R. Validation of two game experience scales: the Player Experience of Need Satisfaction (PENS) and Game Experience Questionnaire (GEQ). Int J Hum Comput Stud 2018 Oct;118(118):38-46. [doi: 10.1016/j.ijhcs.2018.05.003]
21. Ostrow KS, Heffernan NT. Testing the validity and reliability of intrinsic motivation inventory subscales within assistments. Presented at: 19th International Conference, Artificial Intelligence in Education 2018; Jun 27-30, 2018 URL: https://link.springer.com/chapter/10.1007/978-3-319-93843-1_28 [accessed 2025-09-02]
22. Armstrong RA, Eperjesi F, Gilmartin B. The application of analysis of variance (ANOVA) to different experimental designs in optometry. Ophthalmic Physiol Opt 2002 May;22(3):248-256. [doi: 10.1046/j.1475-1313.2002.00020.x] [Medline: 12090640]
23. Wang C, Lee C, Shin H. Digital therapeutics from bench to bedside. NPJ Digit Med 2023 Mar 10;6(1):38. [doi: 10.1038/s41746-023-00777-z] [Medline: 36899073]
24. Montoya MF, Muñoz J, Henao OA. Fatigue-aware videogame using biocybernetic adaptation: a pilot study for upper-limb rehabilitation with sEMG. Virtual Real 2021:1-14. [doi: 10.1007/s10055-021-00561-y]
25. Li A, Qiang W, Li J, Geng Y, Qiang Y, Zhao J. Retracted: evaluating the clinical efficacy of an exergame-based training program for enhancing physical and cognitive functions in older adults with mild cognitive impairment and dementia residing in rural long-term care facilities: randomized controlled trial. J Med Internet Res 2025 Feb 19;27:e69109. [doi: 10.2196/69109] [Medline: 39969990]
26. Chen X, Wu L, Feng H, et al. Comparison of exergames versus conventional exercises on the health benefits of older adults: systematic review with meta-analysis of randomized controlled trials. JMIR Serious Games 2023 Jun 22;11(1):e42374. [doi: 10.2196/42374] [Medline: 37347534]
27. Chen Y, Zhang Y, Guo Z, Bao D, Zhou J. Comparison between the effects of exergame intervention and traditional physical training on improving balance and fall prevention in healthy older adults: a systematic review and meta-analysis. J Neuroeng Rehabil 2021 Nov 24;18(1):164. [doi: 10.1186/s12984-021-00917-0] [Medline: 34819097]
28. Salti S, Schreer O, Di Stefano L. Real-time 3D arm pose estimation from monocular video for enhanced HCI. Presented at: Proceedings of the 1st ACM Workshop on Vision Networks for Behavior Analysis; Oct 31, 2008. [doi: 10.1145/1461893.1461895]
29. Simonsen D, Popovic MB, Spaich EG, Andersen OK. Design and test of a Microsoft Kinect-based system for delivering adaptive visual feedback to stroke patients during training of upper limb movement. Med Biol Eng Comput 2017 Nov;55(11):1927-1935. [doi: 10.1007/s11517-017-1640-z]
30. Chen J, Yang T, He Q, et al. The impact of gamified interventions on the management of chronic obstructive pulmonary disease: systematic literature review. JMIR Serious Games 2025 May 30;13(1):e69510. [doi: 10.2196/69510] [Medline: 40446290]

XSL•FO
RenderX

31. Chu CH, Biss RK, Cooper L, Quan AML, Matulis H. Exergaming platform for older adults residing in long-term care homes: user-centered design, development, and usability study. JMIR Serious Games 2021 Mar 9;9(1):e22370. [doi: 10.2196/22370] [Medline: 33687337]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
GEQ: Game Experience Questionnaire
IMI: Intrinsic Motivation Inventory
T2D: type 2 diabetes

Edited by A Coristine; submitted 08.05.25; peer-reviewed by C Liu, Y Qian; revised version received 27.06.25; accepted 27.06.25; published 12.12.25.

Please cite as:
Zhao J, Zhu D, Wang Y, Xia Y, Zhu Z, Yu J
Development and Evaluation of a Monocular Camera–Based Mobile Exergame for at-Home Intervention in Individuals at High Risk of Type 2 Diabetes: Randomized Controlled Trial
JMIR Serious Games 2025;13:e75823
URL: https://games.jmir.org/2025/1/e75823
doi:10.2196/75823

© Jianan Zhao, Dian Zhu, Yanan Wang, Yaqin Xia, Zeshi Zhu, Jihong Yu. Originally published in JMIR Serious Games (https://games.jmir.org), 12.12.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Becoming a Rapid Shooter in a Game Using Embodied Electrical Muscle Stimulation: Development and Usability Study

Jihwan Kim[1*], MS; Mingyu Kang[1*], BS; Jejoong Kim[2], PhD; Kwanguk (Kenny) Kim[1], PhD

[1]Department of Computer Science, Hanyang University, 222 Wangsimni-ro, Seongdong-gu, Seoul, Republic of Korea

[2]Department of Psychology, Duksung Women's University, Seoul, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Kwanguk (Kenny) Kim, PhD
Department of Computer Science, Hanyang University, 222 Wangsimni-ro, Seongdong-gu, Seoul, Republic of Korea

Abstract

Background: Electrical muscle stimulation (EMS) systems enhance human capabilities, such as reaction time, by inducing preemptive muscle contractions. One of the key challenges for EMS applications is preserving the user's sense of agency, and it is defined as a subjective experience of initiating and controlling one's actions. Prior research highlights the importance of the "sweet spot"—a balance between sense of agency and preemptive gain—for effective EMS use. However, most prior studies have focused on simplistic tasks, leaving a gap in understanding how the sweet spot functions in complex and ecological game scenarios. Moreover, the potential benefits of a personalized approach were not explored.

Objective: This study examines the effects of embodied EMS on performance and sense of agency in a serious-games–based pistol-shooting scenario. Additionally, we investigated the effects of personalization in the same game scenario.

Methods: Two studies were conducted. Study 1 identified the optimal EMS timing ("sweet spot") to improve response time while preserving agency. A total of 13 participants completed a visual stimulus-response task after EMS calibration. Participants performed 150 right-button clicks on a target using an EMS-equipped mouse, with EMS timings ranging from −200 to +100 ms. An agency questionnaire followed each trial. Logistic regression was used to identify the sweet spot. Study 2 applied the findings of Study 1 to a pistol-shooting game. A total of 10 new participants were recruited. Before gameplay, individual sweet spots were measured for the individually embodied EMS condition. During the game, participants navigated to a target location, distinguished enemies from hostages, and shot enemies. After a practice session, participants completed the game under 4 counterbalanced conditions: averagely embodied EMS (used the average sweet spot value from Study 1), individually embodied EMS (used each participant's own sweet spot value), immediate EMS (stimulation at a target onset timing), and no EMS.

Results: Study 1 identified a sweet spot that significantly improved reaction time while preserving agency. Logistic analysis showed an average sweet spot of 46.8 ms and individual variability across participants. Study 2 confirmed these findings in a pistol-shooting game. Both averagely and individually embodied EMS significantly reduced reaction times compared to no EMS (P=.03 and P=.001, respectively), without compromising agency. Individually embodied EMS yielded additional benefits. In the slower than average group, the sense of agency in the individually embodied EMS condition showed an additional benefit that was not observed in the averagely embodied EMS condition (P=.003 and P=.095, respectively).

Conclusions: The findings indicate that embodied EMS enhances performance in complex game scenarios while maintaining users' sense of agency, with individualization yielding additional benefits. These results extend prior evidence to more complex game contexts and provide valuable implications for the personalized design of EMS systems in sports training, rehabilitation, and serious gaming.

(JMIR Serious Games 2025;13:e69330) doi:10.2196/69330

KEYWORDS

electrical muscle stimulation; embodied interaction; human augmentation; sense of agency; serious game

Introduction

Background

Technologies for augmenting human capabilities have advanced significantly with the development of electrical muscle stimulation (EMS). EMS enhances motor performance and reaction speed by inducing muscle contractions through externally applied electrical impulses [1-5]. A key challenge in EMS application is preserving the user's sense of agency—the subjective perception of control over one's actions—while achieving performance gains [6,7]. Prior studies have shown that carefully timed EMS can improve reaction performance without compromising agency, identifying a precise stimulation

XSL•FO
RenderX

window known as the "sweet spot" [5,6]. However, most existing studies have used simplistic motor tasks, limiting the understanding of EMS in more realistic contexts [8,9]. Furthermore, although embodied EMS can yield a sweet spot, its personalization has not been explored, despite agency being a subjective experience [10-13]. To address these gaps, we investigate the effects of embodied EMS in a pistol-shooting game requiring rapid decision-making and response execution. We also assess the benefits of individualized EMS configurations.

The remainder of this paper is organized as follows: The *Methods* section presents details on the 2 experiments conducted. The *Results* section presents the results of both experiments. The final section discusses the results and offers future research directions.

Previous Studies on EMS Applications

EMS generates electrical impulses that induce muscle contractions [14], typically delivered through electrodes placed on the skin near target muscles [10,15]. Compared to motor-based devices, such as exoskeletons, EMS is significantly more wearable [2], making it suitable for force feedback in augmented and virtual reality applications [16]. When combined with haptic instruction systems requiring rapid and precise movements [3,15-18], EMS has proven effective.

EMS is generally used to enhance physical capabilities. For example, the "Wired Muscle" system preemptively stimulates muscles to improve reflexes in physical tasks [19], although "Stimulated Percussions" guides hand movements in rhythmic tasks, enhancing timing and coordination in musical performance [20]. EMS also improves speed and accuracy in physical actions. The electrical head actuation system enables EMS-based control of head orientation, extending applications beyond the limbs to support gaze-guided selection and immersive head movement [15]. "Affordance++" guides limb movements in tasks such as shaking a spray can, reducing effort and training requirements [21], whereas "Pose-IO" directs wrist movements for accurate gesture-based communication without visual feedback [22]. A back-of-hand actuation system improves finger dexterity by enabling refined, independent control of finger joints [3].

Some systems use EMS to enhance immersion. "Paired-EMS" stimulates antagonistic muscle pairs to provide realistic, stable force feedback in high-intensity VR scenarios, improving user comfort and immersion [17]. "BioSync" synchronizes EMS across users, allowing shared kinesthetic experiences in real time [23]. "ErgoPulse" uses real-time biomechanical simulations to estimate muscle forces and joint angles, guiding lower-limb stimulation during virtual interactions such as walking or jumping [24]. EMS for walls and heavy objects simulates resistance during interaction with massive virtual structures, increasing realism in virtual environments [16,25]. A key advantage of EMS is its ability to reduce reaction time, making it particularly effective for applications requiring rapid responses [1,25]. The details of previous studies are summarized in Table 1.

XSL•FO
RenderX

Table . Previous studies using electrical muscle stimulation (EMS).

Author; year	Purpose	Target	Participants	Performance metrics
Kruijff et al [25], 2006	To create pseudo-haptic feedback using EMS	Biceps or brachioradialis	7 (user evaluation)	Haptic realism, perceived resistance
Lopes et al [22], 2015b	To communicate object-specific gestures using EMS	Flexor digitorum, flexor carpi radialis, extensor digitorum, flexor digitorum profundus, biceps brachii	12 (user evaluation)	Understanding the meaning of induced motion, helpfulness, visual affordance
Lopes et al [21], 2015a	To improve limb position awareness through EMS-based proprioception	Extensor digitorum, flexor digitorum superficialis	10 (system validation), 12 (user evaluation)	Angular posture accuracy, subjective experiences
Ebisu et al [20], 2016	To guide rhythmic movement for music training using EMS	Extensor carpi radialis, gastrocnemius, brachioradial muscle	12 (user evaluation)	Sound timing, rhythm accuracy
Nishida et al [19], 2017	To enable faster kinesthetic synchronization	Extensor digitorum	N/A[a]	Reaction time, subjective reaction fluency
Nishida and Suzuki [23], 2017	To create shared motion experience between users via synchronized EMS	Extensor digitorum	5 (design validation), 5 (formative study 1), 6 (formative study 2), 1 (user experience)	Synchronization quality, subjective response
Lopes et al [16], 2017	To simulate physical resistance in virtual reality using EMS on arm/shoulder	Extensor digitorum, infraspinatus, extensor carpi ulnaris, teres major/minor	13 (design validation), 6 (user evaluation)	Preference, realism, consistency, impermeability
Kasahara et al [6], 2019	To accelerate reaction time using sense of agency preserving EMS	Flexor digitorum profundus	12 (formative study),12 (user evaluation)	Sense of agency, reaction time
Kasahara et al [5], 2021	To evaluate motor adaptation by EMS	Flexor digitorum profundus	18 (formative study),17 (user evaluation)	Sense of agency, reaction time
Takahashi et al [3], 2021	To improve finger dexterity using EMS on the back hand	Extensor muscles	9 (user evaluation)	Movement ratio between finger joints, calibration time
Tanaka et al [15], 2022	To control head orientation via EMS for hands-free virtual reality interaction	Splenius capitis, splenius cervicis, sternocleidomastoid	7 (system validation), 8 (user evaluation)	Head movement accuracy, contribution of head actuation, enjoyment
Hwang et al [24], 2024	To deliver biomechanical EMS for lower-limb virtual reality tasks	Quadriceps, tibialis anterior, hamstrings, gastrocnemius	9 (system validation), 12 (user evaluation)	Haptic force accuracy, sense of presence
Cheng et al [17], 2024	To enhance force feedback by EMS of antagonistic muscle pairs	Biceps, triceps	8 (formative study), 32 (user evaluation)	Preferences for realism, harmony, entertainment, comfort, overall

[a]N/A: not applicable.

Balancing Performance Enhancement and Sense of Agency

Previous research suggests that enhancing performance while preserving the sense of agency is critical to effective EMS application. Performance enhancement is typically measured by preemptive gain over a user's natural response [6]. The sense of agency refers to the neural mechanisms underlying an individual's awareness of initiating, executing, and controlling voluntary actions [7,26,27]. One of the main challenges of EMS use is the potential disruption of agency [1,6], causing users to perceive their movements as externally controlled [5]. Kasahara et al [6] identified a stimulation "sweet spot" that enables users to feel the initiated movement voluntarily, achieving performance gains and moderate preservation of agency. These findings warrant further exploration in applied contexts.

Overstimulation can induce fatigue or impair voluntary motor control, whereas suboptimal stimulation may yield limited benefits [28]. Striking the right balance is critical, as neural adaptations influence how the brain and muscles respond to external stimuli [14]. When delivered at the sweet spot, EMS can promote voluntary contractions without overriding sensory-motor systems, supporting efficient neural adaptation over time [29].

This balance is essential in practical applications—such as entertainment, training, or rehabilitation—where significant preemptive gain must be achieved without cognitive dissonance. Attaining the sweet spot facilitates immediate performance gains and supports long-term improvements in strength, motor control, and functional capacity [14,30]. However, to our

XSL•FO
RenderX

knowledge, no studies have applied this concept in realistic game-based scenarios.

Rationale for Personalization Approaches

Research indicates that personalized EMS settings—adjusted for pulse width, stimulation timing, and electrode placement—produce more efficient and reliable outcomes [10,11,31]. Kono et al [9] emphasized the value of multichannel EMS systems for precise control of distinct muscle groups. Gerovasili et al [12] demonstrated that EMS efficacy depends on muscle mass and neuromuscular condition, whereas Kemmler et al [13] reported considerable variability in muscle adaptation and strength. These findings underscore the importance of personalization, given the substantial influence of individual physiological and neural differences. Variations in baseline motor responses, muscle composition, and nervous system plasticity render standardized EMS approaches inadequate.

Personalized EMS minimizes the risk of fatigue and overstimulation, enabling longer, more effective sessions while maintaining stimulation within the user's optimal tolerance range [10]. This is essential for practical applications, where prolonged EMS use must remain effective without causing discomfort, injury, or reduced motivation. Tailored stimulation also supports long-term neural adaptation [5,14], as users are more likely to achieve sustained improvements in motor control when EMS is aligned with their individual characteristics.

Methods

Study 1: Determination of Embodied EMS

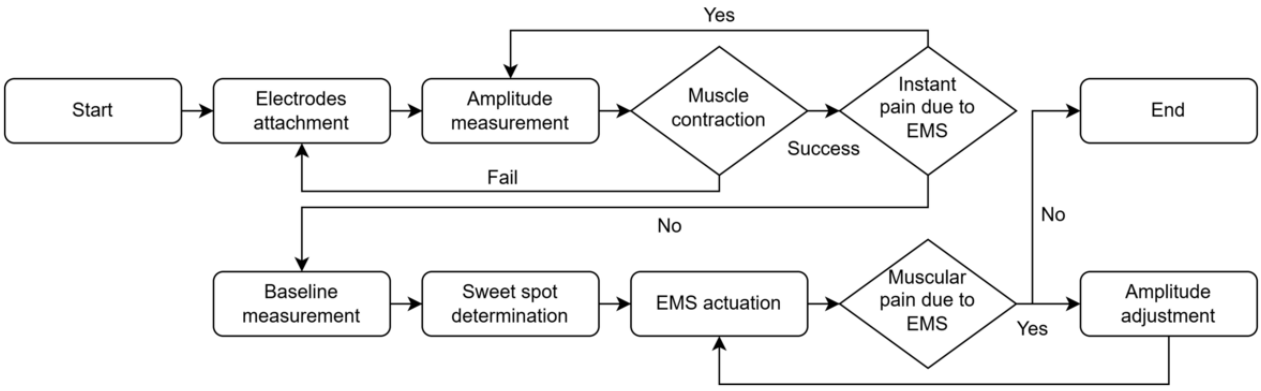
Overview

To understand the effect of embodied EMS, determining the EMS timing corresponding to the sweet spot and comparing the effect of embodied EMS across various conditions are necessary. Therefore, in Study 1, we determined the average of embodied EMS, and this average was used for Study 2.

Proposed Embodied EMS System Flowchart

The embodied EMS system comprises calibration, baseline measurement, sweet spot determination, and user feedback. Initially, electrodes are attached to the target position. Calibration involves setting the individualized EMS amplitude, a parameter crucial for inducing muscle contraction. This personalization is necessary because of interuser variability in appropriate stimulation levels [12,13,25]. If the user does not perceive stimulation, electrodes are repositioned. Conversely, if stimulation is perceived, amplitude is increased until muscle movement is induced without pain, at which point calibration concludes. The sweet spot determination stage identifies the optimal balance between preemptive gain and sense of agency. An elevated agency level may result in minimal preemptive gain, whereas excessively rapid preemptive gain can significantly diminish the sense of agency. Thus, the objective is to establish a moderate agency level. Following sweet spot determination, practice sessions commence with an actual task, enabling detailed adjustments based on user feedback. Participants complete a limited number of trials to confirm consistent muscle induction. Should muscle contraction be intermittent or pain reported, electrode placement is adjusted, and the process reverts to the calibration stage. Figure 1 shows the complete embodied EMS system flowchart.

Figure 1. Flowchart of proposed embodied electrical muscle stimulation (EMS) system. The stages include calibration, sweet spot determination, EMS actuation, and user feedback.



Ethical Considerations

All experimental protocols were approved by the Hanyang University Institutional Review Board (case number: HYUIRB-202304‐022). All participants received a comprehensive overview of the procedure and provided written consent, and their data were deidentified. In appreciation of their involvement, all participants received compensation of $20 per hour. Individual users were not identified in any images of this study.

Participants

A total of 13 participants (5 women and 8 men) were recruited for this study. The participants had a mean age of 25.76 years (SD=3.65). All the participants were right-handed. Recruitment

XSL•FO
RenderX

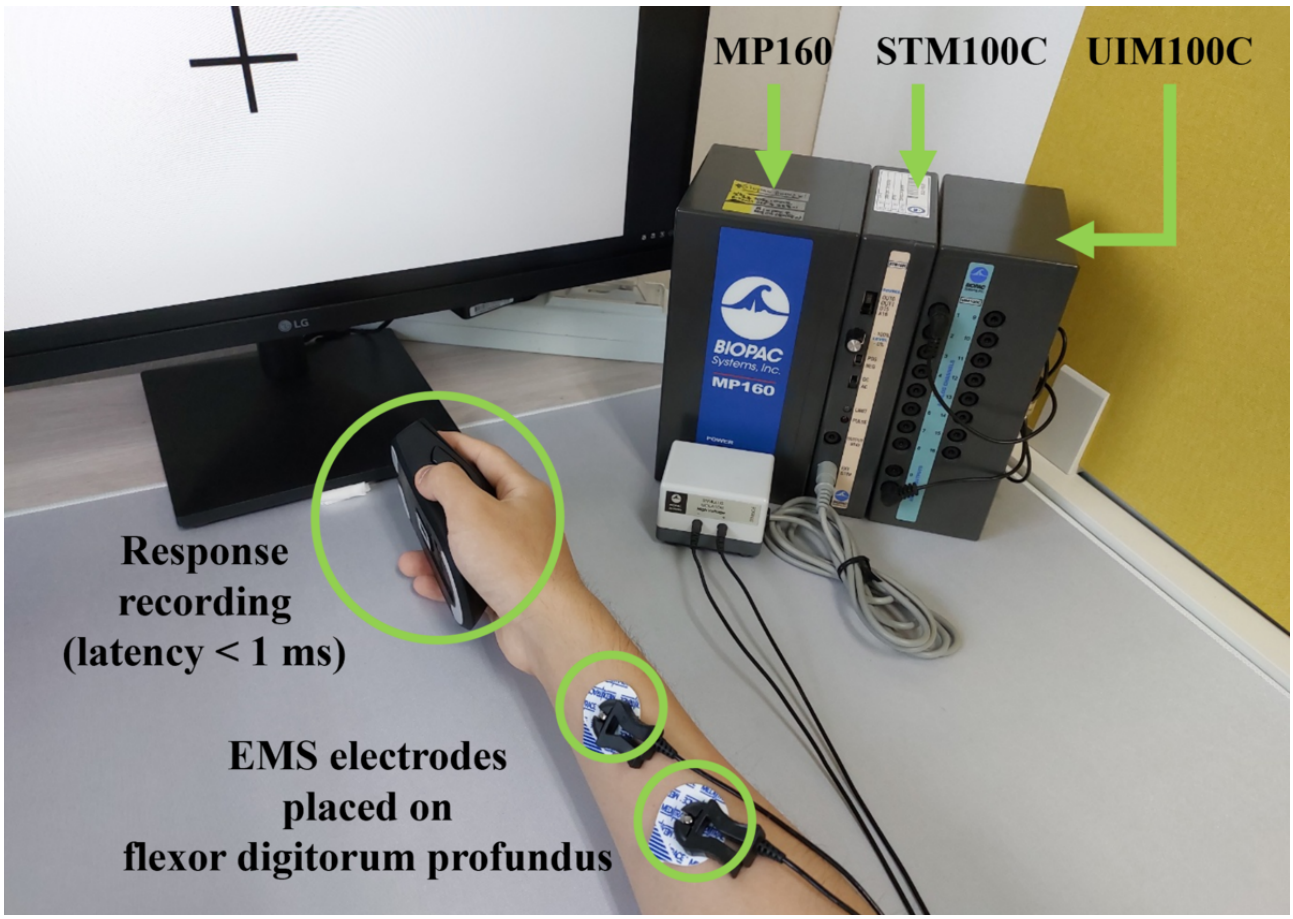
was conducted via bulletin boards on campus and the university website.

Hardware and Software

Figure 2 shows the experimental setup. For EMS control, we utilized the MP160, STM100C, and UIM100C systems (Biopac Systems Inc., USA). AcqKnowledge (Biopac Inc., USA), software specifically designed for Biopac modules, facilitated EMS control and communication between the modules and the experimental program. A pair of electrodes was attached to the flexor digitorum profundus [6] to guide muscle contraction. The flexor digitorum profundus is a compartmental, multitendinous muscle [5,6,32] located in the forearm. It serves as the sole flexor of the distal interphalangeal joint, working in conjunction with other extrinsic (flexor digitorum superficialis and extensor digitorum) and intrinsic hand muscles to control various hand and finger movements, such as grasping and object manipulation. Despite being a single muscle, it comprises 4 compartments capable of relatively autonomous activation at low force levels to flex each fingertip [33].

The experimental program was developed using Unity 2022.3.5f1 (Unity Technologies, USA) and executed on a PC running Windows 10 Home edition (64-bit OS, Microsoft, USA). The system was powered by a Core i5-9600KF processor (3.70 GHz) (Intel Corporation, USA), 32 GB RAM (Samsung, South Korea), and a GeForce GTX 1660 Ti graphics card (NVIDIA, USA).

Figure 2. Experimental settings for Study 1. Electrical muscle stimulation (EMS) was controlled by a stimulation system (MP160, STM100C, and UIM100C). A pair of electrodes was placed on the participants' forearms. Participants respond to the visual stimulus using a wireless gaming mouse.



Participant responses (right-button clicks) to the cues were captured using a DeathAdder V2 Pro wireless gaming mouse (Razer Inc., USA) with a latency of 1 ms. Grip tape was applied to prevent hand slipping, and a supportive structure was used to maintain proper hand posture.

The response time of the system was from the moment of visual stimulus presentation command to the participant's muscle response, and it approximately took 45 ms. It included 5 sequential steps. In step 1, the Unity activates a trigger for the AcqKnowledge software. In step 2, AcqKnowledge activates the EMS trigger. In step 3, the EMS is delivered to the participant's forearm. In step 4, the EMS induces muscle contraction and leads to mouse clicks. In step 5, the Unity detects the mouse click and sends the termination command.

Calibration of EMS

We calibrated the EMS parameters [6,10,11], using a single, symmetric biphasic square wave with a 4-ms pulse width. This pulse width was determined through internal testing to be optimal for inducing precise finger movements. We opted against a fixed amplitude due to its high dependency on individual pain tolerance [6,10], instead of employing an individualized calibration approach. The amplitude, measured in volts, was iteratively calibrated for each participant through a 3-step process, with a 3-minute rest period between each step.

XSL•FO
RenderX

First, the adaptation phase prevented participants from startling. In total, 2 electrodes were attached to the participant's forearm, and a single EMS pulse (4-ms pulse width) was applied. The amplitude gradually increased from 0, ensuring no stimulation caused pain or discomfort. The maximum amplitude at which participants reported no pain or discomfort after several repetitions was recorded as the critical value. The same process was then repeated with continuous EMS, incorporating a 3-ms interpulse interval and a 3-second stimulation duration, also starting from 0 amplitude.

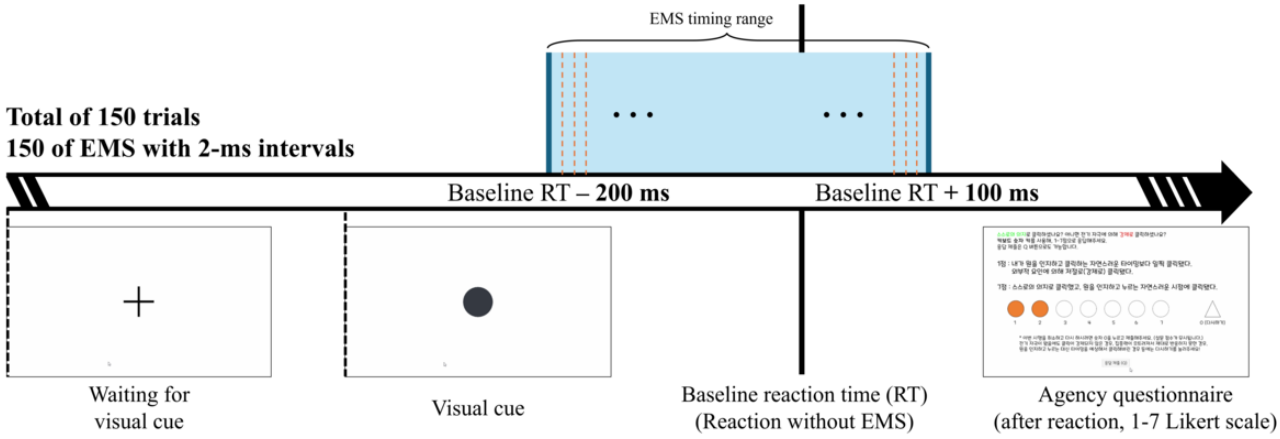
Second, electrode positions were adjusted to ensure the EMS reliably actuated or produced a greater response in the middle finger than in any other part of the hand. A single EMS pulse at the critical amplitude, determined in the adaptation phase, was used. Participants were instructed to adjust their arm and palm angles to optimize middle finger activation.

Finally, participants held the mouse in the optimal position established in the previous step. A single EMS pulse at the critical amplitude was applied to ensure immediate, involuntary actuation of the middle finger, resulting in a right-button click. The amplitude was fine-tuned as needed based on participant response and the degree of middle finger activation.

Sweet Spot Determination: Visual Stimulus-Response Task

The task was a standard reaction-time measurement test (Figure 3). Participants were instructed to click the right button of the mouse upon recognizing the visual cue represented by a circle. Following each reaction, the participants were presented with a Likert scale questionnaire (ranging from 1 to 7) to assess their perceived sense of agency. The scale ranged from 1, meaning "I did not press the right button to click," to 7, meaning "I pressed the right button to click." Participants were asked to rely on their intuition when responding to the questionnaire.

Figure 3. Visual stimulus-response task for sweet spot determination. A baseline of reaction time, which is response ability without electrical muscle stimulation (EMS), was measured first. The timing range of EMS was set by the individual baseline. Subsequently, the participants engaged in 150 trials of the task with the sense of agency questionnaire.



The EMS actuation timing ranged from −200 to +100 ms based on the participant's reaction time without the EMS (baseline). For example, if a participant had a baseline reaction time of 250 ms, the EMS timing window would range from 50 to 350 ms after the visual cue. We assigned 150 points with 2-ms intervals into 5 blocks, ensuring no redundancy across the blocks. To minimize the learning effect, the order of 30 points within each block was randomized.

Procedures

Before engaging in the main task, the participants completed a practice phase of up to 10 trials. Subsequently, the participants were asked to perform 30 trials without the EMS to establish their baseline reaction times. The participants completed the visual stimulus-response task. The task was to perform 150 right-button clicks on the target using an experimental mouse with timing ranging from −200 to +100 ms. The EMS was administered at a moment within the designated time window in every trial. Each trial was conducted using an agency questionnaire. The participants were debriefed after completing the trials.

Statistical Analysis

The relationship between preemptive gain and sense of agency (sweet spots of each participant) was calculated using logistic regression. Linear regression between EMS offset time and reaction time was conducted using the IBM SPSS 27.0 (SPSS Inc., USA) software package. Statistical significance was set at $P<.05$.

Before the analysis, we normalized the data by subtracting each participant's baseline reaction time from each reaction time (with EMS actuation). This allows us the time gained by means of preemption (preemptive gain). We then normalized the agency axis from 0 to 1 and the horizontal axis to depict −400 to 400 ms. Logistic regression computes the relationship between the perceived agency and the preemptive gain per participant.

Study 2: Evaluation of the Effect of Embodied EMS in a Practical Game for Physical Training

Overview

Based on the results of Study 1, we investigated the effects of EMS in a practical game scenario. We employed a

XSL•FO
RenderX

pistol-shooting game that required fast reactions and cognitive judgment (ie, whether to act or not). Study 2 examined the effects of embodied EMS and additional effects of individual approaches. The task was conducted under 4 distinct conditions. The averagely embodied EMS condition used the average sweet spot from the results of Study 1, whereas the individually embodied EMS condition used the sweet spot from individual measurements using a method suggested in Study 1. In the first baseline condition (immediate EMS), the EMS was delivered immediately upon the appearance of the target. The second baseline condition (no EMS) enabled us to record participants' natural reaction times. By comparing performance across these conditions, we assess how different EMS settings influence outcomes.

Ethical Considerations

We employed the same ethical considerations as Study 1.

Participants

A total of 10 participants (3 women and 7 men) who did not participate in Study 1 were included. The participants had a mean age of 24.44 years (SD=3.71), and all were right-handed.

Hardware and Software

We employed the same experimental setting as Study 1.

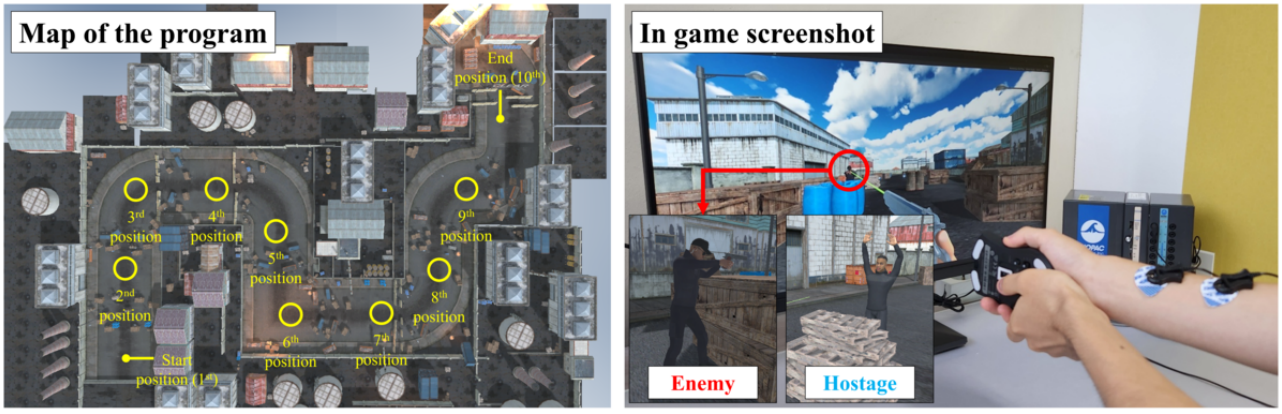
Sweet Spot Determination: Visual Stimulus-Response Task

Before beginning the pistol-shooting training, participants underwent a measurement described in Study 1 to identify individual sweet spots for the individually embodied EMS condition.

Evaluation Task: Pistol-Shooting Task

We used a pistol-shooting program [34] similar to Virtua Cop 2 (SEGA Games, Japan) (Figure 4). In the game, participants were tasked with reaching a designated location, distinguishing enemy targets from hostages, and firing at the enemies. A target board displaying "start" appeared in the virtual space before the game began. The game began when the participant hit the target. Upon reaching the shooting position, 10 agents—comprising 4-6 hostages and 6-4 enemies—appeared (correspondingly) with random jittering time (from 500 to 1500 ms). Participants had to eliminate all enemies at the given location to proceed to the next position. The targets at each position were arranged in a 10×1 (horizontal×vertical) spherical coordinate system, with a horizontal range of −55° to +55°, and each target was spaced at a 10° radial distance. In total, 10 shooting positions were considered; 50 enemies and 50 hostages were presented, accompanied by corresponding sound effects.

Figure 4. Pistol-shooting game with electrical muscle stimulation (EMS). Participants engaged in 4 conditions: averagely embodied EMS (EMS timing was fixed with an average sweet spot from Study 1), individually embodied EMS (EMS timing with an individualized sweet spot), immediate EMS (EMS timing presented when the enemy agent appears), and no EMS (participants respond without EMS).



Procedures

After participants were briefed, personalized values for the individually embodied EMS conditions were measured as in Study 1. In the second visit, participants engaged in a practice session of the pistol-shooting program. The pistol-shooting task was performed at least 3 days after the first visit to minimize the learning effect. Participants did this task under 4 conditions: averagely embodied EMS, individually embodied EMS, immediate EMS, and no EMS, presented in counterbalanced order. During the task, participants completed a sense of agency questionnaire. After the experiment, participants were debriefed.

Statistical Analysis

The relationship between preemptive gain and sense of agency (individual "sweet spots") was calculated using logistic regression, as in Study 1. Owing to variability in subjective feelings and physical conditions, participants were categorized into 2 subgroups for further analysis based on their measured sweet spots: those with sweet spots faster than average (FTA group) and those slower than average (STA group). We then investigated whether the effects of the 4 EMS conditions differed between these 2 groups.

All statistical analyses were performed using SPSS 27.0 (SPSS Inc., USA). Normality, skewness, kurtosis, and Shapiro-Wilk tests were conducted. Sphericity was assessed using the Mauchly test; if violated, the Greenhouse-Geisser correction was applied. A mixed ANOVA investigated the differences between the FTA and STA groups across the 4 EMS conditions: averagely embodied EMS, individually embodied EMS, immediate EMS, and no EMS. Statistical significance was set at $P<.05$. Post hoc analyses for group differences and comparisons across the 4 EMS conditions both utilized paired t tests. Data from 9

XSL•FO
RenderX

participants were used in the analysis, as 1 female participant's data were excluded because of a calibration failure.

Results

Results of Study 1

In total, 1950 trials were collected from all participants, with 2 data points (reaction time and sense of agency) recorded per trial. The average baseline was 238.3 ms (SD=29.2 ms), which aligns with findings in psychophysics research that report an average reaction time of 250 ms in response to visual stimuli [6,35]. The baseline reaction time for the fastest and slowest participants was 212.8 and 304.8 ms, respectively. Consequently, the EMS offset timing ranged from 12.8 to 402.8 ms after the onset of the visual stimulus. Table 2 presents the measurement results for each participant.

Table . Measurement results for Study 1[a].

Participant	Sex	Stimulus timing (ms)	Baseline (ms)	Sweet spot (ms)
1	Male	31.0 - 329.0	231.0	68.2
2	Male	13.6 - 311.6	213.6	20.1
3	Male	12.8 - 310.8	212.8	17.4
4	Male	27.4 - 325.4	227.4	38.8
5	Male	85.9 - 383.9	285.9	28.1
6	Female	13.1 - 311.1	213.1	−4.0
7	Female	19.1 - 317.1	219.1	121.7
8	Male	58.8 - 356.8	258.8	28.1
9	Male	41.1 - 339.1	241.1	57.5
10	Female	16.1 - 314.1	216.1	28.1
11	Male	27.4 - 325.4	227.4	84.3
12	Female	104.8 - 402.8	304.8	20.1
13	Female	46.4 - 344.4	246.4	44.1

[a]All participants first measured baseline response time to visual stimuli in the absence of EMS. Based on the individually measured baseline, stimulus timings ranging from −200 to +100 ms EMS were determined. Afterward, the sweet spot was calculated.

EMS actuation reduced reaction time. Figure 5 shows the relationship between the EMS offset and total reaction time. This relationship was linear between 12.8 ms (the earliest reaction time influenced by EMS) and 238.3 ms (the average reaction time), indicating that EMS increased the reaction times.

Through logistic analysis, the average sweet spot for all participants was 46.8 ms. This general sweet spot was used as the averagely embodied EMS condition in Study 2. We also found the individual differences of those sweet spots. The highest level of preemptive gain among participants was 121.7 ms, whereas the lowest preemptive gain was −4.0 ms. Figure 6 shows the logistic regression analysis for each participant, indicating the relationship between agency and preemptive gain.

XSL•FO
RenderX

Figure 5. Relationship between electrical muscle stimulation (EMS) offset time and reaction time. The linear relationship between EMS timing and reaction time in the area within the gray line suggests that the EMS directly induced the reaction. The values in the area beyond the gray line are dispersed because the EMS was applied after the natural reaction speed.

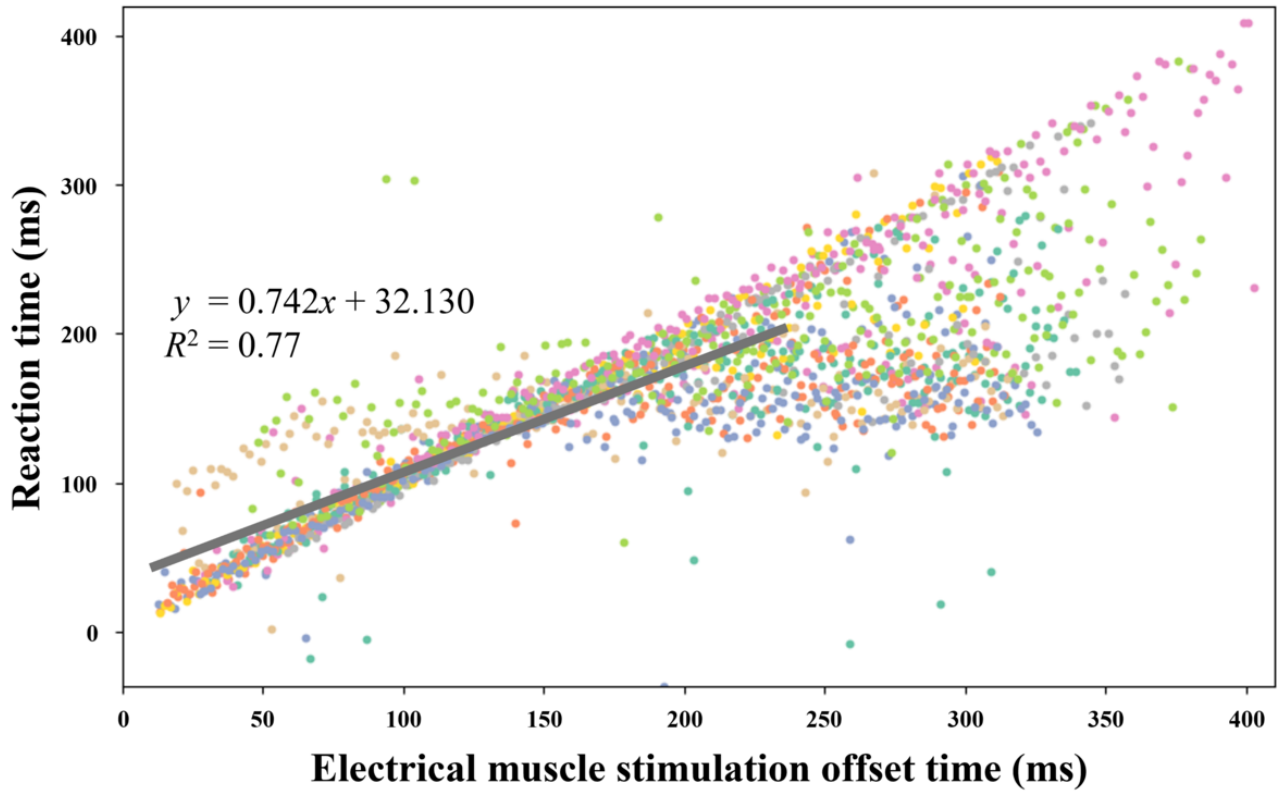
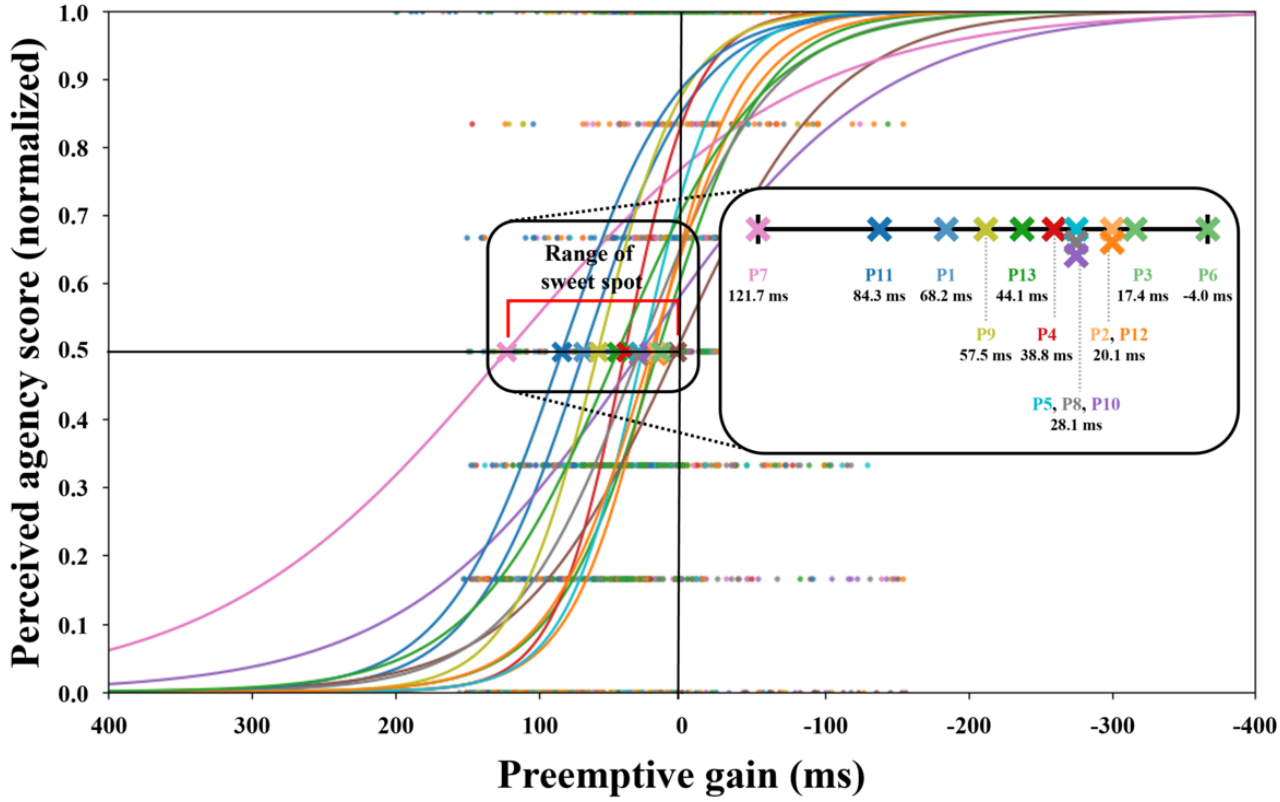


Figure 6. Result of logistic regression analysis between preemptive gain and sense of agency in Study 1. The average sweet spot for the 13 participants was 46.8 ms.



XSL•FO
RenderX

Results of Study 2

Results of Individual Sweet Spot Measurement

We collected 1350 trials from participants, with 2 data points (reaction time and sense of agency response) recorded per trial. The average baseline of Study 2 was 236.8 ms (SD=21.0 ms), which is consistent with the findings from Study 1 (average=238.2 ms) and previous research [6,35]. The baseline reaction time for the fastest and slowest participants was 206.9 and 262.7 ms, respectively. Table 3 presents the measurement results for each participant.

Using logistic analysis, we calculated the sweet spots for each participant in Study 2. The average sweet spot for all participants was 54.8 ms. The levels of preemptive gain for the fastest and slowest participants were 76.3 and 14.7 ms, respectively. Figure 7 shows the logistic regression analysis for each participant, indicating the relationship between agency and preemptive gain.

Table . Measurement results of Study 2[a].

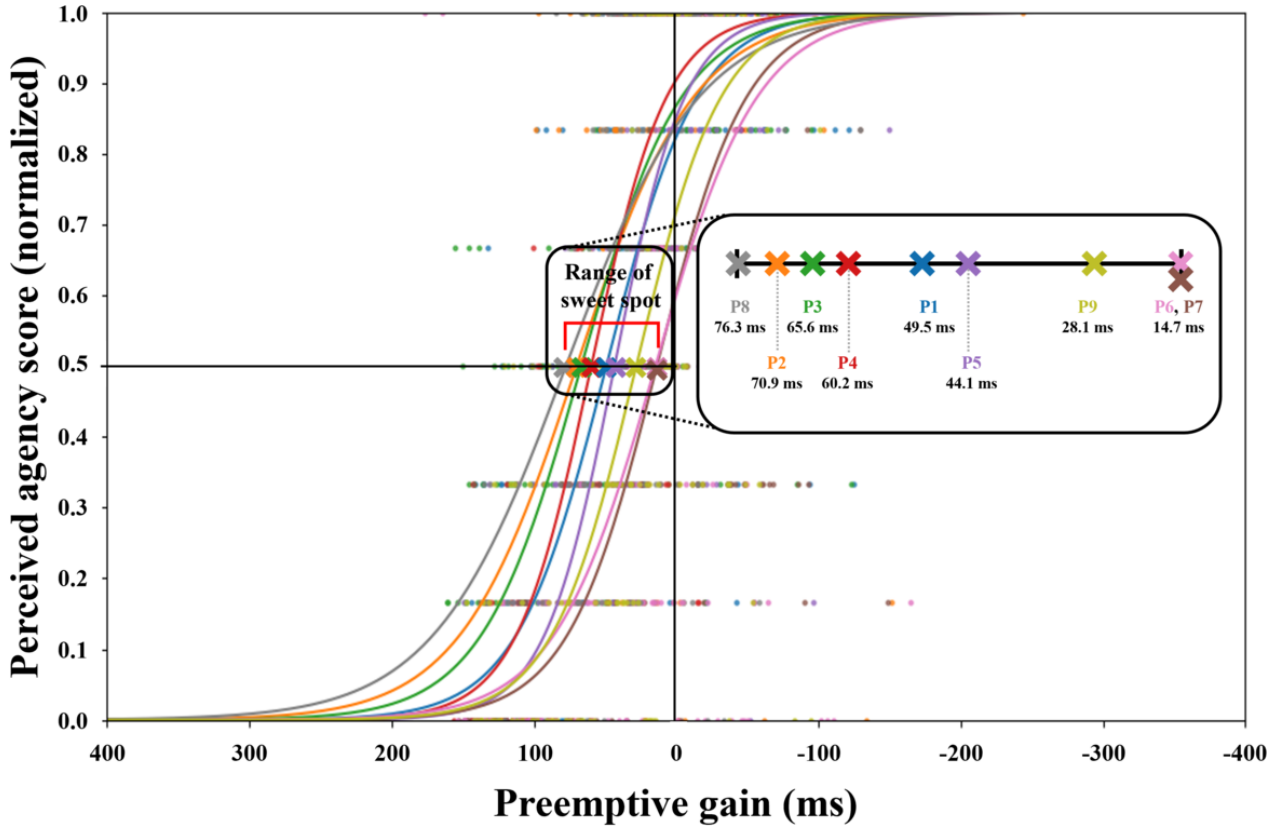
Participant	Sex	Baseline (ms)	Sweet spot (ms)	Group
1	Female	262.7	49.5	FTA[b]
2	Male	208.0	70.9	FTA
3	Male	225.6	65.6	FTA
4	Male	252.3	60.2	FTA
5	Male	259.5	44.1	STA[c]
6	Male	243.6	14.7	STA
7	Male	206.9	14.7	STA
8	Female	246.8	76.3	FTA
9	Male	226.1	28.1	STA

[a]All participants first measured baseline response time to visual stimuli in the absence of electrical muscle stimulation (EMS). Based on the individually measured baseline, stimulus timings ranging from −200 to +100 ms EMS were determined. Subsequently, the sweet spot was measured. If the sweet spot was faster than the average value of Study 1, the participants were assigned to the FTA group, and if it was slower, the participants were assigned to the STA group.

[b]FTA: faster than average.

[c]STA: slower than average.

Figure 7. Result of logistic regression analysis between preemptive gain and sense of agency in Study 2.



XSL•FO
RenderX

Response Time

A significant main effect of the EMS conditions (F (1.805, 14.437)=27.528, P<.001, η^2=0.79) was observed. However, no significant main effect of the group and interaction effect ($F_{1,7}$=0.175, P=.69, η^2=0.024; $F_{1.856,\ 12.989}$=0.554, P=.58, η^2=0.073, respectively) was observed. Table 4 shows response time analysis results in each EMS condition and group.

The averagely embodied EMS condition was slower than the immediate-EMS condition (t_8=3.881, P=.03) and faster than the no-EMS condition (t_8=3.776, P=.04). No significant difference was observed between the individually embodied and averagely embodied EMS conditions (t_8=–0.986, P>.99). The response time in the individually embodied EMS condition was also slower than that in the immediate-EMS condition (t_8=5.792, P=.002) and faster than that in the no-EMS condition (t_8=6.753, P=.001). The response time in the immediate-EMS condition was significantly shorter than that in the no-EMS condition (t_8=–7.970, P=.001).

Table . Results of the response time[a].

Response time, mean (SD)	Averagely embodied EMS	Individually embodied EMS	Immediate EMS (baseline 1)	No EMS (baseline 2)
Faster than average (FTA) group	202.1 (38.7)	218.2 (54.4)	131.5 (79.0)	273.5 (63.8)
Slower than average (STA) group	209.6 (22.3)	214.0 (29.4)	96.7 (44.8)	259.0 (41.5)

[a]A total of 4 electrical muscle stimulation (EMS) conditions were included: averagely embodied EMS, individually embodied EMS, immediate EMS, and no EMS. Participants were grouped into 2 groups: the faster than average (FTA) group and slower than average (STA) group.

Sense of Agency

Significant main effects of the EMS condition and interaction effects between the EMS condition and groups ($F_{3,\ 24}$=63.350, P<.001, η^2=0.900; $F_{3,\ 21}$=3.786, P=.03, η^2=0.351, respectively) were observed. However, no significant difference was observed depending on the group ($F_{1,\ 7}$=1.753, P=.23, η^2=0.200). Table 5 shows the sense of agency analysis results in each EMS condition and group.

Table . Results of the sense of agency[a].

Response time, mean (SD)	Averagely embodied EMS	Individually embodied EMS	Immediate EMS (baseline 1)	No EMS (baseline 2)
Faster than average (FTA) group	4.7 (0.7)	4.1 (0.9)	3.0 (1.0)	6.7 (0.7)
Slower than average (STA) group	3.3 (1.6)	4.3 (0.9)	1.8 (0.4)	7.0 (0.0)

[a] A total of 4 electrical muscle stimulation (EMS) conditions were included: averagely embodied EMS, individually embodied EMS, immediate EMS, and no EMS. Participants were grouped into 2 groups: the faster than average (FTA) group and slower than average (STA) group.

Sense of agency in the averagely embodied EMS condition was higher than in the immediate-EMS condition (t_8=5.807, P=.002) and lower than in the no-EMS condition (t_8=5.736, P=.003). No significant difference was observed between the individually embodied and averagely embodied EMS conditions (t_8=–0.315, P>.99). Sense of agency in the individually embodied EMS condition was also higher than in the immediate-EMS condition (t_8=4.495, P=.01) and lower than in the no-EMS condition (t_8=–7.419, P<.001). Sense of agency in the no-EMS condition was significantly higher than in the immediate-EMS condition (t_8=–12.452, P<.001).

The sense of agency across the 4 conditions manifested differently in the 2 groups. In the FTA group, the averagely embodied EMS condition was higher than the immediate-EMS condition (t_4=8.064, P=.001) and lower than the no-EMS condition (t_4=–5.706, P=.005). Agency in the individually embodied EMS condition was significantly lower than in the no-EMS condition (t_4=–4.466, P=.01). However, no significant difference was observed with immediate EMS (t_4=2.125, P=.10). Agency in the immediate- and no-EMS conditions was significantly different (t_4=–9.105, P=.001).

Unlike the FTA group, in the STA group, the sense of agency in the individually embodied EMS condition was significantly higher than in the immediate-EMS condition (t_3=–8.996, P=.003) and significantly lower than in the no-EMS condition (t_3=–6.063, P=.009). No significant difference was observed with the averagely embodied EMS condition (t_3=2.284, P=.12). The averagely embodied EMS condition was not significantly different from the immediate-EMS condition (t_3=2.414, P=.095). However, the agency was lower than the no-EMS condition (t_3=–4.680, P=.02). Agency in the immediate-EMS was significantly lower than in the no-EMS condition (t_3=–24.981, P<.001).

XSL•FO
RenderX

Discussion

Principal Findings

The findings offer valuable insights into the use of embodied EMS for enhancing rapid reaction performance in serious games and related contexts. Averagely and individually embodied EMS significantly improved response times over baseline conditions while preserving the sense of agency. Individually embodied EMS provided further benefits, highlighting the value of personalization in EMS applications.

Study 1 identified a specific EMS timing sweet spot that significantly improved reaction time while preserving the sense of agency. EMS applied within this window (average preemptive gain: 46.8 ms) enhanced responses without diminishing perceived control. Notable interindividual variability highlighted the need for personalized EMS settings, as uniform configurations may yield suboptimal results owing to physiological differences.

Study 2 applied these findings to a practical pistol-shooting game requiring rapid and accurate reactions. Averaged and individually calibrated EMS significantly reduced reaction times, confirming that the identified sweet spot improves physical response in applied contexts. Individually embodied EMS provided additional benefits, reinforcing the value of personalization. This aligns with prior research [5,6] showing that customized EMS enhances performance while minimizing fatigue.

The comparative analysis between the FTA and STA groups revealed important individual differences. Both groups reported an intermediate sense of agency under the individually embodied EMS condition, indicating that the sweet spots were effectively personalized. However, differences in agency levels between the groups suggest variation in how participants perceived control over their actions. Although sweet spot calibration was conducted using a nonjudgmental task, it was subsequently applied to a pistol-shooting game involving decision-making (ie, whether to shoot or not), introducing additional cognitive load. This may explain why participants in the FTA group reported a stronger sense of agency in the averagely embodied EMS condition. In contrast, the STA group appeared particularly sensitive to the sense of agency under individually embodied EMS and exhibited markedly low agency scores in the immediate-EMS condition. This suggests heightened sensitivity to externally induced control and a stronger aversive reaction to forced interventions.

Additionally, the STA group may require more time to form a sense of agency. Their average sweet spot was 25.4 ms, compared to 64.5 ms in the FTA group, indicating that their reaction speed improved only by 39.4% relative to the FTA group. Despite this discrepancy, both groups reported a similar sense of agency under the individually embodied EMS condition, suggesting that the STA group, while showing smaller performance gains, achieved comparable subjective control—likely through more gradual interventions.

The findings from Study 2 underscore the importance of personalized EMS settings in practical applications. Participants exhibited considerable variability in optimal stimulation timing, and the impact of individually embodied EMS differed across groups. Thus, applying a uniform average sweet spot is insufficient. Instead, tailoring EMS to individual profiles is essential to maximize performance benefits and maintain user agency.

Comparison to Prior Works

This study extends the work of Kasahara et al [5,6] by demonstrating how agency-preserving EMS can enhance motor performance in more ecologically valid scenarios. Although prior studies employed simplified reaction-time tasks (eg, tapping in response to a visual cue), we applied EMS in a dynamic pistol-shooting game requiring visual discrimination and decision-making. This context more closely reflects real-world applications and cognitive demands. Furthermore, we refined the concept of the “sweet spot” by introducing individualized EMS timing rather than relying on fixed or averaged values. This personalization accounts for interindividual variability in agency perception and motor responsiveness, and our findings suggest that tailored EMS delivery offers additional performance and agency-related benefits. Finally, we present a practical system pipeline—from EMS calibration to real-time actuation and feedback—providing a replicable framework for deploying EMS in interactive applications.

Beyond reaction enhancement, prior research has shown that EMS can improve muscle strength and posture, illustrating its broader applicability [24,25]. Strojnik et al [18] and Takahashi et al [3] demonstrated that EMS can effectively target specific muscle groups to support motor function improvements. These findings indicate that EMS-enhanced training can be extended to muscle conditioning and injury prevention, in addition to reflex-based tasks. Integrating EMS into structured strength and rehabilitation programs may provide neuromuscular reinforcement, enhance postural control, and improve performance during high-intensity movements. This also aligns with rehabilitation protocols where EMS has been used successfully to restore muscle function in patients recovering from injury [11,12]. By incorporating EMS for performance enhancement and injury prevention, athletic training programs and professional organizations can leverage its potential to support short-term gains and long-term physical resilience.

The individually embodied EMS system demonstrated its capacity to enhance motor performance within a game-based context, and this capability can be extended to a range of real-world applications. Athletes in high-speed disciplines—such as fencing or shooting—require rapid reflexes and high levels of precision. The individually embodied EMS can be calibrated to improve reaction times during training, thereby enhancing competitive performance without compromising the athlete’s sense of control. This aligns with findings in strength training research, where EMS has been shown to increase maximal voluntary contraction through neural adaptations rather than muscle hypertrophy [14,28-30]. In high-stakes environments such as battlefield simulations or tactical decision-making tasks, the need for rapid responses and precise actions is equally critical. Our results suggest that individually embodied EMS

XSL•FO
RenderX

can optimize reaction time in such contexts, potentially reducing errors in high-pressure scenarios. Notably, the ability to preserve a user's sense of agency while enhancing speed is essential for maintaining situational awareness and cognitive control required in these fields.

Integrating findings from the current and prior research, we propose a 3-phase implementation roadmap for deploying EMS enhancements in real-world scenarios. First, a calibration phase should be conducted to identify each user's individual sweet spot using nonjudgmental tasks, ensuring that stimulation improves response speed without disrupting the sense of agency. Second, EMS should be incorporated into structured training programs—such as athletic drills or immersive game-based environments—to develop reaction speed and motor precision. Third, a real-time feedback and adaptation phase should adjust EMS parameters dynamically based on the user's performance and physiological state (eg, fatigue), thereby sustaining training effectiveness and reducing the risk of overstimulation. This roadmap offers a feasible pathway for implementing embodied EMS in commercial and professional settings, emphasizing personalization, adaptability, and user-centered design.

Limitations and Future Works

Although this study offers valuable insights into the use of embodied EMS in practical scenarios, several limitations should be acknowledged. First, although our sample size was comparable to previous studies [5,6], it remained relatively small owing to the inherent challenges associated with EMS research. Future studies should replicate these findings with larger and more diverse participant populations to enhance generalizability. Second, although the pistol-shooting game was selected for its high demands on speed and precision, future research should investigate EMS applications in other complex, real-world tasks that require a balance between rapid responses and decision-making—such as surgical procedures or athletic competitions. Third, EMS in this study was applied to the flexor digitorum profundus, a commonly targeted muscle in EMS research [5,6,21,22]. However, more recent work has begun to explore EMS applications on other muscle groups [24]. Therefore, the effectiveness of embodied EMS across a wider range of muscles should be examined to determine its broader applicability. Finally, although this study highlighted the effectiveness of personalized EMS settings, it also underscored the need for more advanced EMS systems that can dynamically adapt to individual user conditions, such as fatigue or stress. Future research should optimize EMS systems to accommodate these fluctuations, ensuring sustained performance improvements across varying contexts.

Beyond immediate performance gains, an important question remains: Does EMS-based reaction training produce lasting improvements? Previous research offers some promising insights. Kasahara et al [5] reported that EMS training can yield long-term improvements in reaction time, even after the EMS is removed—likely due to neuromuscular plasticity and enhanced motor learning. These results are consistent with other studies involving repeated electrical stimulation, which show strengthened sensorimotor integration and improved voluntary control over time [2,14]. These findings suggest that EMS training may offer enduring benefits beyond its immediate effects. To fully explore this potential, future studies should conduct longitudinal assessments of participants' reaction performance over extended periods.

Although EMS offers significant advantages in reaction training and human augmentation, its ethical implications warrant careful consideration in future research. The induction of involuntary muscle movement raises concerns about user autonomy, informed consent, and potential misuse, particularly in competitive contexts [1,7,8]. Future studies should prioritize systems that allow users to maintain full control over EMS intensity and activation parameters, ensuring that all movements remain voluntary. The application of EMS in competitive gaming or professional sports introduces ethical and regulatory concerns. Organizations should establish clear guidelines, distinguishing between training enhancement and real-time performance augmentation. Previous research by Faltaous et al [1] indicated that user acceptance of EMS depends largely on the degree of control, highlighting the need for ethical design considerations. Overuse of EMS could lead to muscle fatigue, adaptation limitations, or cognitive overload. Future research should focus on safety thresholds. Developing ethical EMS guidelines and regulatory policies will ensure responsible adoption in athletics, gaming, and human augmentation fields.

Conclusion

The results of this study demonstrate the potential of embodied EMS in practical game scenarios that require rapid reactions and precise decision-making. By identifying and applying individualized sweet spots, we enhanced participants' performance without compromising their sense of agency. Moreover, the individually embodied EMS configuration provided an additional benefit. This study contributes to the broader field of EMS research by showing how embodiment-preserving stimulation can effectively augment human capabilities in realistic, performance-driven contexts.

Acknowledgments

This work was supported by the National Research Foundation of Korea grant funded by the Korean government (Ministry of Science and ICT) (No. RS-2024 - 00355411) and Culture, Sports and Tourism R&D Program through the Korea Creative Content Agency grant funded by the Ministry of Culture, Sports and Tourism in 2024 (No. RS-2024 - 00399136).

Data Availability

All relevant data are included in this published article and its supplementary information files in Multimedia Appendix 1.

XSL•FO
RenderX

Authors' Contributions

Conceptualization: Jihwan Kim (lead), KK (supporting), MK (equal)
Data curation: Jihwan Kim (lead), MK (equal)
Formal analysis: Jihwan Kim (lead), KK (supporting), MK (equal),
Funding acquisition: KK
Investigation: Jihwan Kim (lead), MK (equal)
Methodology: Jihwan Kim (lead), MK (equal)
Project administration: KK
Resources: KK
Supervision: KK (lead), Jejoong Kim (supporting)
Validation: Jihwan Kim (lead), MK (equal)
Visualization: Jihwan Kim (lead), MK (equal)
Writing – original draft: Jihwan Kim (lead), MK (equal), Jejoong Kim (supporting), KK (supporting)
Writing – review & editing: KK (lead), Jihwan Kim (supporting), MK (supporting), Jejoong Kim (supporting)

Conflicts of Interest

None declared.

Multimedia Appendix 1
Measurement data.
[XLSX File, 15 KB - games_v13i1e69330_app1.xlsx]

References

1. Faltaous S, Williamson JR, Koelle M, Pfeiffer M, Keppel J, Schneegass S. Understanding user acceptance of electrical muscle stimulation in human-computer interaction. 2024 May 11 Presented at: CHI '24; May 11-16, 2024; Honolulu, HI, USA URL: https://dl.acm.org/doi/proceedings/10.1145/3613904 [doi: 10.1145/3613904.3642585]
2. Raisamo R, Rakkolainen I, Majaranta P, Salminen K, Rantala J, Farooq A. Human augmentation: past, present and future. Int J Hum Comput Stud 2019 Nov;131:131-143. [doi: 10.1016/j.ijhcs.2019.05.008]
3. Takahashi A, Brooks J, Kajimoto H, Lopes P. Increasing electrical muscle stimulation's dexterity by means of back of the hand actuation. Presented at: CHI '21: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems; May 8-13, 2021; Yokohama, Japan. [doi: 10.1145/3411764.3445761]
4. Correia A, Carvalho S, Santos CP. Literature review of functional electrical stimulation systems. Presented at: 2019 IEEE 6th Portuguese Meeting on Bioengineering (ENBENG); Feb 22-23, 2019; Lisbon, Portugal URL: https://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=8686245 [doi: 10.1109/ENBENG.2019.8692456]
5. Kasahara S, Takada K, Nishida J, Shibata K, Shimojo S, Lopes P. Preserving agency during electrical muscle stimulation training speeds up reaction time directly after removing EMS. Presented at: CHI '21: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems; May 8-13, 2021; Yokohama Japan. [doi: 10.1145/3411764.3445147]
6. Kasahara S, Nishida J, Lopes P. Preemptive action: accelerating human reaction using electrical muscle stimulation without compromising agency. Presented at: CHI '19: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems; Glasgow, Scotland, UK. [doi: 10.1145/3290605.3300873]
7. Jeunet C, Albert L, Argelaguet F, Lecuyer A. "Do you feel in control?": towards novel approaches to characterise, manipulate and measure the sense of agency in virtual environments. IEEE Trans Visual Comput Graphics 2018;24(4):1486-1495. [doi: 10.1109/TVCG.2018.2794598]
8. Nishida J, Kasahara S, Lopes P. Demonstrating preemptive reaction: accelerating human reaction using electrical muscle stimulation without compromising agency. Presented at: SIGGRAPH '19: ACM SIGGRAPH 2019 Emerging Technologies; Jul 28, 2019; Los Angeles, CA, USA. [doi: 10.1145/3305367.3327997]
9. Kono M, Ishiguro Y, Miyaki T, Rekimoto J. Design and study of a multi-channel electrical muscle stimulation toolkit for human augmentation. 2018 Feb 6 Presented at: AH2018; Feb 7-9, 2018; Seoul, Republic of Korea URL: https://dl.acm.org/doi/proceedings/10.1145/3174910 [doi: 10.1145/3174910.3174913]
10. Kashyap N, Baranwal VK, Basumatary B, Bansal R, Sahani A. A systematic review on muscle stimulation techniques. IETE Technical Review 2023 Jan 2;40(1):76-89. [doi: 10.1080/02564602.2022.2030815]
11. Doucet BM, Lam A, Griffin L. Neuromuscular electrical stimulation for skeletal muscle function. Yale J Biol Med 2012 Jun;85(2):201-215. [Medline: 22737049]
12. Gerovasili V, Stefanidis K, Vitzilaios K, et al. Electrical muscle stimulation preserves the muscle mass of critically ill patients: a randomized study. Crit Care 2009;13(5):R161. [doi: 10.1186/cc8123] [Medline: 19814793]
13. Kemmler W, Shojaa M, Steele J, et al. Efficacy of whole-body electromyostimulation (WB-EMS) on body composition and muscle strength in non-athletic adults. a systematic review and meta-analysis. Front Physiol 2021;12:640657. [doi: 10.3389/fphys.2021.640657] [Medline: 33716787]

XSL•FO
RenderX

14. Hortobágyi T, Maffiuletti NA. Neural adaptations to electrical stimulation strength training. Eur J Appl Physiol 2011 Oct;111(10):2439-2449. [doi: 10.1007/s00421-011-2012-2] [Medline: 21643920]
15. Tanaka Y, Nishida J, Lopes P. Demonstrating electrical head actuation: enabling interactive systems to directly manipulate head orientation. 2022 Apr 27 Presented at: CHI '22; Apr 29 to May 5, 2022; New Orleans, LA, USA URL: https://dl.acm.org/doi/proceedings/10.1145/3491101 [doi: 10.1145/3491101.3519904]
16. Lopes P, You S, Cheng LP, Marwecki S, Baudisch P. Providing haptics to walls & heavy objects in virtual reality by means of electrical muscle stimulation. Presented at: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17); May 6-11, 2017; New York, USA. [doi: 10.1145/3025453.3025600]
17. Cheng CY, Chen Y, Handani SW, Balabantaray A, Chen MY. Paired-EMS: enhancing electrical muscle stimulation (EMS)-based force feedback experience by stimulating both muscles in antagonistic pairs. 2024 May 11 Presented at: CHI '24; May 11-16, 2024; Honolulu, HI, USA URL: https://dl.acm.org/doi/proceedings/10.1145/3613904 [doi: 10.1145/3613904.3642841]
18. Strojnik P, Kralj A, Ursic I. Programmed six-channel electrical stimulator for complex stimulation of leg muscles during walking. IEEE Trans Biomed Eng 1979 Feb;26(2):112-116. [doi: 10.1109/tbme.1979.326520] [Medline: 761930]
19. Nishida J, Kasahara S, Suzuki K. Wired muscle: generating faster kinesthetic reaction by inter-personally connecting muscles. Presented at: ACM SIGGRAPH 2017 Emerging Technologies (SIGGRAPH '17); Jul 30 to Aug 3, 2017; Los Angeles, USA. [doi: 10.1145/3084822.3084844]
20. Ebisu A, Hashizume S, Suzuki K, Ishii A, Sakashita M, Ochiai Y. Stimulated percussions: method to control human for learning music by using electrical muscle stimulation. Presented at: Proceedings of the 8th Augmented Human International Conference (AH '17); Mar 16-18, 2017; California, USA. [doi: 10.1145/3041164.3041202]
21. Lopes P, Ion A, Mueller W, Hoffmann D, Jonell P, Baudisch P. Proprioceptive interaction. 2015 Apr 18 Presented at: CHI '15; Apr 18-23, 2015; Seoul, Republic of Korea URL: https://dl.acm.org/doi/proceedings/10.1145/2702123 [doi: 10.1145/2702123.2702461]
22. Lopes P, Jonell P, Baudisch P. Affordance++: allowing objects to communicate dynamic use. Presented at: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15); Apr 18-23, 2015; New York, USA. [doi: 10.1145/2702123.2702128]
23. Nishida J, Suzuki K. BioSync: a paired wearable device for blending kinesthetic experience. Presented at: CHI '17: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems; May 6-11, 2017; Denver, CO, USA. [doi: 10.1145/3025453.3025829]
24. Hwang S, Oh J, Kang S, Seong M, Elsharkawy A, Kim S. ErgoPulse: electrifying your lower body with biomechanical simulation-based electrical muscle stimulation haptic system in virtual reality. 2024 May 11 Presented at: CHI '24; May 11-16, 2024; Honolulu, HI, USA URL: https://dl.acm.org/doi/proceedings/10.1145/3613904 [doi: 10.1145/3613904.3642008]
25. Kruijff E, Schmalstieg D, Beckhaus S. Using neuromuscular electrical stimulation for pseudo-haptic feedback. 2006 Nov Presented at: VRST06; Nov 1-3, 2006; Limassol, Cyprus URL: https://dl.acm.org/doi/proceedings/10.1145/1180495 [doi: 10.1145/1180495.1180558]
26. Jeannerod M. The mechanism of self-recognition in humans. Behav Brain Res 2003 Jun 16;142(1-2):1-15. [doi: 10.1016/s0166-4328(02)00384-4] [Medline: 12798261]
27. Wegner DM. The mind's best trick: how we experience conscious will. Trends Cogn Sci 2003 Feb;7(2):65-69. [doi: 10.1016/s1364-6613(03)00002-0] [Medline: 12584024]
28. Collins DF. Central contributions to contractions evoked by tetanic neuromuscular electrical stimulation. Exerc Sport Sci Rev 2007 Jul;35(3):102-109. [doi: 10.1097/jes.0b013e3180a0321b] [Medline: 17620928]
29. Baldwin ERL, Klakowicz PM, Collins DF. Wide-pulse-width, high-frequency neuromuscular stimulation: implications for functional electrical stimulation. J Appl Physiol (1985) 2006 Jul;101(1):228-240. [doi: 10.1152/japplphysiol.00871.2005] [Medline: 16627680]
30. Han BS, Jang SH, Chang Y, Byun WM, Lim SK, Kang DS. Functional magnetic resonance image finding of cortical activation by neuromuscular electrical stimulation on wrist extensor muscles. Am J Phys Med Rehabil 2003 Jan;82(1):17-20. [doi: 10.1097/00002060-200301000-00003] [Medline: 12510180]
31. Carè M, Chiappalone M, Cota VR. Personalized strategies of neurostimulation: from static biomarkers to dynamic closed-loop assessment of neural function. Front Neurosci 2024;18:1363128. [doi: 10.3389/fnins.2024.1363128] [Medline: 38516316]
32. van Duinen H, Gandevia SC, Taylor JL. Voluntary activation of the different compartments of the flexor digitorum profundus. J Neurophysiol 2010 Dec;104(6):3213-3221. [doi: 10.1152/jn.00470.2010] [Medline: 20881206]
33. Kilbreath SL, Gorman RB, Raymond J, Gandevia SC. Distribution of the forces produced by motor unit activity in the human flexor digitorum profundus. J Physiol 2002 Aug 15;543(Pt 1):289-296. [doi: 10.1113/jphysiol.2002.023861] [Medline: 12181299]
34. Kim J, Kim J, Jung M, Kwon T, Kim KK. Individualized foveated rendering with eye-tracking head-mounted display. Virtual Real 2024 Mar;28(1). [doi: 10.1007/s10055-023-00931-8]
35. HYMAN R. Stimulus information as a determinant of reaction time. J Exp Psychol 1953 Mar;45(3):188-196. [doi: 10.1037/h0056940] [Medline: 13052851]

XSL•FO
RenderX

Abbreviations

EMS: electrical muscle stimulation
FTA: faster than average
STA: slower than average

Edited by A Coristine; submitted 27.11.24; peer-reviewed by M Ferrara, M Ghobaei-Arani, V Sarlis; revised version received 26.08.25; accepted 26.08.25; published 22.10.25.

Please cite as:
Kim J, Kang M, Kim J, Kim K(
Becoming a Rapid Shooter in a Game Using Embodied Electrical Muscle Stimulation: Development and Usability Study
JMIR Serious Games 2025;13:e69330
URL: https://games.jmir.org/2025/1/e69330
doi:10.2196/69330

© Jihwan Kim, Mingyu Kang, Jejoong Kim, Kwanguk (Kenny) Kim. Originally published in JMIR Serious Games (https://games.jmir.org), 22.10.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Identifying Key Principles and Commonalities in Digital Serious Game Design Frameworks: Scoping Review

Raluca Ionela Maxim[1*], PhD; Joan Arnedo-Moreno[1*], PhD.

Universitat Oberta de Catalunya, Barcelona, Spain

[*]all authors contributed equally

Corresponding Author:
Raluca Ionela Maxim, PhD
Universitat Oberta de Catalunya
Rambla del Poblenou, 154-156
Sant Martí
Barcelona, 08018
Spain
Phone: 34 724346819
Email: rmaxim@uoc.edu

Abstract

Background: Digital serious games (DSGs), designed for purposes beyond entertainment and consumed via electronic devices, have garnered attention for their potential to enhance learning and promote behavior change. Their effectiveness depends on the quality of their design. Frameworks for DSG design can guide the creation of engaging games tailored to objectives such as education, health, and social impact.

Objective: This study aims to review, analyze, and synthesize the literature on digital entertainment game design frameworks and DSG design frameworks (DSGDFWs). The focus is on conceptual frameworks offering high-level guidance for the game creation process rather than component-specific tools. We explore how these frameworks can be applied to create impactful serious games in fields such as health care and education. Key goals include identifying design principles, commonalities, dependencies, gaps, and opportunities in the literature. Suggestions for future research include empathic design thinking, artificial intelligence integration, and iterative improvements. The findings culminate in a synthesized 4-phase design process, offering generic guidelines for designers and developers to create effective serious games that benefit society.

Methods: A 2-phase methodology was used: a scoping literature review and cluster analysis. A targeted search across 7 databases (ACM, Scopus, Springer, IEEE, Elsevier, JMIR Publications, and SAGE) was conducted using PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 guidelines. Studies included academic or industry papers evaluating digital game design frameworks. Cluster analysis was applied to categorize the data, revealing trends and correlations among frameworks.

Results: Of 987 papers initially identified, 25 (2.5%) met the inclusion criteria, with an additional 22 identified through snowballing, resulting in 47 papers. These papers presented 47 frameworks, including 16 (34%) digital entertainment game design frameworks and 31 (66%) DSGDFWs. Thematic analysis grouped frameworks into categories, identifying patterns and relationships between design elements. Commonalities, dependencies, and gaps were analyzed, highlighting opportunities for empathic design thinking and artificial intelligence applications. Key considerations in DSG design were identified and presented in a 4-phase design baseline with the outcome of a list of design guidelines that might, according to the literature, be applied to an end-to-end process of designing and building future innovative solutions.

Conclusions: The main benefits of using DSGDFWs seem to be related to enhancing the effectiveness of serious games in achieving their intended objectives, such as learning, behavior change, and social impact. Limitations primarily seem to be related to constraints associated with the specific contexts in which the serious games are developed and used. Approaches in the future should be aimed at refining and adapting existing frameworks to different contexts and purposes, as well as exploring new frameworks that incorporate emerging technologies and design principles.

(JMIR Serious Games 2025;13:e54075) doi:10.2196/54075

XSL•FO
RenderX

KEYWORDS

entertainment game design frameworks; serious game design frameworks; design principles; empathic design thinking; artificial intelligence

Introduction

Overview

Digital serious games (DSGs) have gained substantial attention and investigation in recent years, with a continuously expanding collection of scholarly works dedicated to this topic [1-3]. There are several definitions of serious games [4], but for this paper, the one proposed by Abt [5] is used: "we are concerned with serious games in the sense that these games have an explicit and carefully thought-out educational purpose and are not intended to be played primarily for amusement." These games can be used for education, training, behavioral change, exergaming or therapeutic purposes, and other applications by using learning objectives in game mechanics to engage learners in an immersive learning experience [6,7].

Having a plan is the key when it comes to designing DSGs, particularly when the game is meant to serve an educational purpose beyond fun. Game design can be complex and cannot be improvised and that is why evaluating and improving design frameworks is essential. The purpose of such frameworks is to organize the design process into various phases or categories with specific guidelines and principles for each area, such as game design in education. This design flow is adjustable to different steps that involve evaluating the expected performance of a solution and identifying significant design challenges. The design process tackles each design issue and refines solutions based on information collected during the design flow, with each step informed by different guidelines and perspectives that impact the system. Therefore, the general goal of a framework design is to create solutions from preexisting components (analysis phase), to describe the critical decomposition problems, and to deliver a concrete example, not abstract (design and test phases) [8,9]. Because of their importance, there is a vast interest and amount of literature on DSGs with a variety of design frameworks that evolved to better serve experiences of users and learners. Frameworks and best practices exist to ensure that DSGs are designed appropriately [10-12].

The initial inspiration for the design of serious games derives from general fun or entertainment games. Things such as game mechanics (point systems and levels), player interaction (competition and collaboration), and visual design elements (appealing graphics and sound effects) are borrowed from games people play for enjoyment. Entertainment games prioritize fun, often featuring elaborate narratives and fantastical settings. However, serious games focus on achieving a specific learning objective. In essence, serious games borrow the engaging aspects of their entertainment counterparts to make learning more effective, as part of what Deterding et al [13] define as "the ludification of culture." Serious games demonstrate the power of applying game design principles to achieve goals beyond pure entertainment. However, they are not simply educational resources dressed up with game mechanics. Instead, they are full-fledged games that integrate educational principles into their core design, as well as fostering motivation through game design approaches, making the learning experience not only effective but also engaging and enjoyable [14].

It is important to remark that serious games, despite their engaging nature, are designed to prioritize teaching or training players in specific skills or knowledge, encompassing fields such as science, arts, or even health promotion. Serious games leverage mechanics and technologies typically found in games to facilitate learning and potentially even influence behavior changes [15]. In contrast, entertainment games focus solely on providing players with enjoyable experiences. While the design considerations differ between these categories, both serious and entertainment games rely on engaging gameplay and immersive experiences to achieve success. In a purely digital context, video games have become a powerful force, encompassing both the realm of pure entertainment and the world of education and training [16]. Thus, as games become increasingly influential across various industries, understanding the design principles and frameworks behind both types, and their similarities and differences, becomes important [17].

The main objective of this research is a scoping review into the existing design frameworks in the literature that guide the creation of these diverse games. The focus lies on 2 distinct categories: digital entertainment game design frameworks (DEGDFWs) and DSG design frameworks (DSGDFWs). By grouping similar elements together, this analysis aimed to reveal commonalities and interdependencies between the 2 categories. This clustering process facilitated the identification of patterns, gaps, and opportunities within current design practices. This exploration serves 2 key purposes: first, to provide aspiring game designers with a clear understanding of existing frameworks, and second, to contribute valuable insights that can be used to develop more robust and versatile design tools for serious games in the future.

This paper is divided into several sections—first, establishing theoretical and conceptual foundations on digital games' background and debriefing on entertainment and serious game design frameworks, next elaborating on the scoping review methods and qualitative or quantitative analyses conducted to evaluate relationships between game frameworks, then discussing key results found between serious and entertainment games, and finally synthesizing the results in a 4-phase design model for serious game development, which summarizes the main aspects identified in the literature review. The synthesis of findings into a 4-phase design visualization offers a structured approach to navigating the current serious game design space, fulfilling the need for accessible and informative resources in this domain. This baseline visualization aims to provide clarity to advance future applied initiatives and research by incorporating collective guidelines to drive the creation of serious games for scalable impact.

XSL•FO
RenderX

Theoretical Background

DSGs are active and entertaining digital learning environments, accessible on computers and smartphones. Unlike pure entertainment games, their primary purpose is education and learning, "by combining with consistency, both serious aspects such as non-exhaustive and nonexclusive, teaching, learning, communication, or the information, with playful springs from the video game" as stated by Alvarez and Damien [18]. Through engaging gameplay and in-game challenges, players learn and develop practical concepts by applying their skills to overcome obstacles within the game. DSGs have been used in many different fields, such as health care, culture, training, business, and even to address social and political issues [4,17]. The variety of applications reflects the wide range of goals serious games can achieve, from raising awareness, to cultivating knowledge, to changing behavior [5,19].

Serious games can be in analog and digital format, whose differentiation lies in the platform they use to deliver the learning experience. Analog serious games typically involve physical components such as board games, card games, or role-playing activities. They offer tangible, hands-on experiences that promote collaboration, communication, and critical thinking skills. In contrast, DSGs are delivered through electronic platforms such as computers, consoles, or mobile devices. They often leverage multimedia elements such as graphics, sound, and interactivity to create immersive environments that enhance engagement and learning outcomes [7,20].

Creating impactful serious games involves additional challenges compared to entertainment games due to necessitating expertise spanning both game design fundamentals and instructional design strategies to guide outcomes. While serious games and entertainment games share certain surface-level characteristics such as interactivity, rules, challenges, and narrative elements, their fundamental purposes diverge [21,22]. Entertainment games primarily aim to provide captivating, rewarding experiences focused on having fun, emotional immersion, and intrinsic player motivation to progress further for self-actualization [23]. In contrast, serious games have explicit extrinsic goals beyond the gameplay itself, intending to educate users on new skills, catalyze developmental outcomes, or bring about behavioral changes—objectives tied to real-world usefulness. Overall, the research landscape on serious games applications and frameworks remains complex.

Applications of DSGs

Although DSGs may be perceived as derived from entertainment games, their design necessitates additional expertise beyond traditional game design skills. Specifically, designing games for learning purposes demands proficiency in instructional design, student guidance, and assessment of learning outcomes [19,24]. Nevertheless, there is a connection between both types of games, which some researchers and designers explain as a combination of story, art, and software. Besides the game mechanics, player interaction, and visual design elements, the incorporation of educational principles into the game's story is another significant characteristic of serious games. By integrating these principles into the story, the serious games transform learning from a passive activity into an active and engaging experience. Players become participants in the story, motivated by the narrative to explore, discover, and apply educational concepts within the immersive world of the game [25,26].

This connection enables the entertainment aspect of the game to effectively communicate the intended educational or training message. As a result, serious games provide a more engaging and effective experience than traditional methods of education and training, because they are more immersive and interactive [4,10]. For instance, in health care, serious games are used for patient education, training medical professionals, and rehabilitation [27]. In education, they are used to enhance student engagement and learning outcomes [28]. The military field uses serious games for training and simulation purposes. In the corporate world, they are used for employee training, leadership development, and team-building exercises [29]. In addition, serious games have also been used for public policy initiatives, such as disaster preparedness and environmental awareness campaigns [30].

In that regard, the use of serious games has been particularly effective in the health care sector, where complex medical concepts can be taught using games to both children and adults. Interactive games have also been proven to be an effective tool for educating young individuals about their health situations and medications [31,32]. The purpose of serious games' design in the health care field is to support the gain of a general understanding of the previous work done, such as the use of games to reduce knowledge gaps on various topics, which can aid in improving the effectiveness of game development [33]. For instance, a recent serious game has been designed to make the medical hematology topic more engaging for health care students, encouraging them to analyze blood tests to provide better care for their patients. This game-based approach has proved to enhance learners' knowledge of hematology, providing an effective tool for education and skills development [34].

All those mentioned fields underline the usefulness of serious games, whose main purpose is to teach learners different topics in joyful and interactive experiences. An important component of serious game design is represented by the learning theories mixed with hedonic elements. Learning theories used in serious games can vary and are dependent on the game's specific objectives. For instance, some serious games may be based on constructivism, which highlights the learner's role in constructing knowledge through active exploration and discovery [35]. Alternatively, some games may be based on behaviorism, which concentrates on the observable behavior of the learner and the application of rewards and punishments to influence behavior. Moreover, other games may use cognitive behavioral theories, which focus on the relationship among thoughts, emotions, and behaviors [36].

To create a great DSG, designers need to turn goals into a clear picture of what the game will be. This involves understanding what kind of learning experience they want to create. The team then asks a series of questions to define the game's mechanics: what specific skills or knowledge will it teach, what emotions and feelings should players have while learning, and how will

XSL•FO
RenderX

the game itself (objectives, rules, and story) achieve this. Designing DSGs requires considering various aspects of playful learning, including how people think (cognitive), feel (affective), and are motivated (motivational), alongside social and cultural factors (sociocultural). The learners' emotional engagement is powerful and strongly correlated to their deeper cognitive effort [26,37,38].

In essence, mixing all hedonic and pragmatic elements is essential for the development of various serious games solutions. However, their effective and successful creation should start from a formal design process and methodology that provide the essential guidance for development teams. However, designing serious games is not a straightforward process, as there is no single framework or theory that can be applied to every type of serious game.

Game Design Frameworks

Generally speaking, design frameworks are systematic approaches used to design and develop products. They provide a structure for designers and developers to follow, helping to ensure that the final designed product meets the desired objectives and user needs. The concept of a framework is defined as a skeleton structure of theories and guiding principles that shape and organize the components of a phenomenon or identified problem [39,40]. However, creating a design framework that satisfies the complexity of different approaches remains a challenge for designers and developers [41].

There are different adapted approaches applied to both entertainment and serious games, starting with their main focus and goals. Entertainment games target the creation of engaging and immersive gameplay, using components such as game mechanics, esthetic experiences in gameplay, and reward systems to create a fun and enjoyable experience for players with the goal of commercial success on one hand [42-44]. On the other hand, serious games are designed with the educational purpose of creating effective and motivating learning experiences, even though not everyone agrees. One study argues that serious games are more effective for learning and retention than traditional teaching but not more motivating than conventional methods [45]. Other studies state that serious games cultivate greater learner motivation through incorporating interactivity, competition, social connections, achievement-based progression systems, and freeform experimentation appeal unavailable in conventional coursework [46].

Serious games often incorporate elements such as instructional design that turns education into game-based learning. These elements are tailored to specific learning objectives and curriculum standards, while also considering both accessibility and user-centered design to ensure ease of use and usefulness for diverse users. The process of education can be viewed as a game-like journey with challenges and goals. Motivation is a key aspect of engagement in both education and games, leading to an interest in applying motivational strategies from games to education. However, a challenge in doing so is creating a game scenario that feels natural and not artificial. Game designers emphasize that the game is a tool for achieving an experience, rather than the experience itself. Therefore, it is important to understand game design principles when applying game elements to education [47].

In addition, most existing serious games frameworks are designed for specific types of educational games. This is due to the diversity of serious games and the various learning paradigms and contexts in which they are used. Existing design frameworks for serious games also differ in their objectives. Some focus more on the design elements of serious games, while others emphasize the design and development process. Some frameworks cover both design elements and process, while others are more concerned with the theoretical foundations of game-based learning [39,48]. Therefore, when starting a new serious game development plan, the design team must make a decision on a specific framework to work with [48-50].

Designing Digital Entertainment Games

Digital entertainment games (DEGs) are typically designed for the purpose of providing enjoyment to players. They often involve fictional worlds, characters, and storylines and can take the form of action games, adventure games, puzzle games, and more. They can be played on various platforms and interfaces, including console, PC, or mobile [6]. The game industry allows for a wide range of options in game design and creation, with practical considerations such as time constraints, financial budgeting, and available personnel often dictating the process. While there is no one-size-fits-all formula for creating a game, certain practices have emerged over time through trial and error and successful results. These practices include the use of frameworks in design, but information about these frameworks is inconsistent from one proposal to another [51].

Inconsistencies in game design frameworks can arise from various factors, including the lack of standardized guidelines or universally accepted best practices, the rapidly evolving nature of the industry, diverse design goals and objectives, and varying target audiences. It is important to note that these inconsistencies are not necessarily negative, as they can offer a diverse range of perspectives and approaches to game design, enabling designers to select and adapt frameworks that best align with their specific needs and design goals. These inconsistencies also highlight the dynamic and ever-evolving nature of the game design field, which constantly adjusts to changing technologies, player preferences, and design trends. For instance, these frameworks may differ in their focuses, such as player motivations (Bartle taxonomy) [43], types of fun (Four Keys to Fun) [52], player personalities (Engines of Play) [53], or player involvement (Player Involvement Model) [54]. They may also vary in their scopes, with some frameworks concentrating on specific aspects of gameplay, such as mechanics, dynamics, and aesthetics [55], while others adopt a more holistic approach that considers multiple elements, such as formal, dramatic, and dynamic elements, esthetics, and story (Layered Tetrad) [56].

Entertainment game frameworks can influence educational serious games by providing structures for creating engaging and interactive gameplay, adapting well-designed user interfaces and navigation systems, and incorporating well-developed storytelling and narrative elements to create immersive and engaging educational experiences. In addition, entertainment

XSL•FO
RenderX

games have a wide range of mechanics and gameplay elements that can be used to create educational games tailored to different subjects and learning styles [57].

Designing DSGs

DSGs are designed to provide a learning experience that balances both educational and fun aspects equally. Motivation is a very important factor in their development, as it is essential for engagement in education. Incorporating motivational strategies from DEG has become a growing interest in education [53]. Collaboration and immersion are important aspects that designers should consider when creating DSGs. They allow players to learn and experience together in a digital or virtual environment [58,59]. Despite their effectiveness, for example, in the computer science area, designing them requires additional effort during the design process to cater to learners with varying abilities and needs [60].

In fact, education can be seen as a journey toward acquiring knowledge, similar to the experience of playing a game. Tests and practical work can be thought of as challenges, and the desired learning outcomes can be thought of as the goals to be achieved. The purpose of this journey is to change the student or learner through innovative experiences such as some DSG with virtual reality or augmented reality that combines multisensory experiences [61]. Game designers often argue that the game itself is not the experience but rather a tool for achieving an experience. Therefore, it is essential to have the know-how of game design when creating game scenarios to avoid creating a forced or artificial game environment [62]. A typical method of validating serious games is analyzing individual game mechanics to understand their impact on player behavior and the overall gaming experience [63].

Though there is not much research on how individual aspects of game design contribute to the effectiveness of serious games, it is important to thoroughly test individual aspects of their design to ensure their effects. In addition, their purpose and perceived value may also play a role in its success, as players may have a more positive attitude toward a game if they consider it to be beneficial. Another important aspect for more useful and usable serious game design would be the incorporation of empathic design principles by involving lecturers and learners in the design process [64].

Learning Theories for DSGs

Effective DSG design is grounded in an understanding of various learning theories that can be applied in game development [65]. Constructivism, for example, emphasizes the learner's active role in constructing their understanding of the world through open-ended problem-solving tasks, experiences, and digital interactions [66]. Cognitive load theory is another important learning theory applied in serious game design, which posits that excessive cognitive load can interfere with learning [66]. To minimize extraneous cognitive load, immersive game environments that use multimedia and visual aids to support learning are used.

Social learning theory is yet another important learning theory applied in serious game design. This theory emphasizes learning through observation and imitation of others, and collaborative and multiplayer game environments that allow learners to work together and learn from one another are often used to implement this theory [67]. Self-determination theory, which supports the fact that learners are more motivated and engaged when they have autonomy and control over their learning experiences, is also applied in serious game design [68]. Personalization and adaptive game mechanics that allow learners to set their own goals and progress at their own pace are used to support self-directed learning.

In addition, behaviorism is a learning theory that underlines the role of environmental factors in shaping behavior, and it is applied in serious game design by providing immediate feedback to players and rewarding or punishing certain behaviors [69]. Goal-oriented instruction is another learning theory that focuses on the achievement of specific goals or objectives in a learning environment, which aligns well with the objectives of educational games [35]. Problem-based learning is another interesting learning theory that emphasizes solving real-world problems to learn, which aligns well with the interactive and hands-on nature of games [68]. Video games can be valuable educational tools when effectively integrated with learning theories. It is important to examine both the advantages and obstacles in aiding learning. This includes exploring strategies for integrating video games into educational settings and understanding their impact on student engagement, motivation, and cognitive development [70].

Overall, effective serious game design needs to be rooted in an understanding of various learning theories that inform game development. Constructivism highlights the learner's active role in constructing an understanding through problem-solving tasks and digital interactions, while cognitive load theory addresses the impact of excessive cognitive load on learning. Social learning theory emphasizes learning through observation and collaboration, while self-determination theory supports learner motivation through autonomy. Behaviorism is applied through feedback and reinforcement, while goal-oriented instruction and problem-based learning align with the objectives of educational games. When integrated effectively, serious games have the potential to be valuable educational tools, impacting student engagement, motivation, and cognitive development.

Methods

A comprehensive literature review was conducted to establish a foundational understanding as a first phase, followed by a cluster analysis to categorize data into meaningful groups and identify trends and correlations within the design frameworks as a second phase in this study.

Literature Review

As a first step, a comprehensive literature review was conducted to gather existing knowledge and insights on the topic, laying the foundation for the subsequent cluster analysis.

While there are frameworks for other game formats such as hands-on analog play, instructor-guided play, or e-books, the focus of this study is to specifically explore frameworks tailored to the unique characteristics of DEGs and DSGs. The

XSL•FO
RenderX

widespread integration of digital games and gamification into education highlights the importance of grasping the foundations and design principles essential for creating impactful game-based learning encounters [25]. This paper is analyzing frameworks that can be used in digital games but not only necessarily for digital games. Specifically, the focus of the review is on both DEG and DSG design frameworks, including their various design elements and principles. The study aims to answer several questions, such as which digital game design frameworks are available in the literature and what their main features are and how entertainment game design principles might influence serious game design frameworks.

The identification and reporting process of the studies included in the review were guided by the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist [71]. The PRISMA statement with the completed checklist provided in Multimedia Appendix 1, allowed to better organize the reporting of the systematic review with some steps: database search of the papers based on all desired search terms, duplicate papers removal, screening of titles and abstracts and excluding the papers that are not relevant for our research questions, full-text screening to assess the eligibility and exclude again the papers that did not contain prototypes with needed design details, and final mention of the included studies. No preregistration or protocol was completed before the literature search.

This literature review aims to analyze the details related to digital entertainment and serious games design frameworks using a systematic mapping approach. The papers selected were searched on 6 electronic databases: ACM, Scopus, Springer, IEEE, Elsevier, JMIR Publications, and SAGE. The search keywords included “game design frameworks,” “serious game design frameworks,” “game design guidelines,” and “serious game design guidelines.”

The search strategy, presented in Multimedia Appendix 2, focused on two key aspects: (1) serious games and (2) entertainment games (the search keywords used were “design” OR “frameworks” OR “guidelines”). Papers that used similar but not these exact terms were identified during the abstract reading phase, ensuring a comprehensive capture of relevant literature. To ensure the relevance and quality of the selected works, inclusion and exclusion criteria were rigorously applied. The inclusion criteria encompassed academic works (not papers only) explicitly identified “as a framework” in peer-reviewed literature. This research seeks academic publications from the years 2015 to 2023, which outline established frameworks and recommendations for structuring the design and development of games. The focus is on understanding how these frameworks provide a systematic approach to game creation, while guidelines offer best practices for developers. The analysis also examines key design elements common to successful games and identify the intended learning outcomes associated with them.

Conversely, the exclusion criteria ruled out papers that mention game design without detailed phase definitions, contain models as simplified process representations, or reference serious games without defining framework details. Moreover, the exclusion criteria eliminate the works that never explicitly were identified as “frameworks” when referenced by other authors in the peer-reviewed literature. The study is open to non–peer-reviewed works but only if peer-reviewed works explicitly consider them frameworks.

Duplicate papers were removed, and a manual selection process was used to narrow down the number of papers. The titles and abstracts of the papers were carefully examined to determine whether they met the inclusion criteria. After the initial manual selection, the selected papers were read in detail to ensure they met the inclusion criteria, and further elimination of irrelevant papers was done to narrow down the papers to those most relevant to the research questions.

Moreover, a snowball process was performed in addition to the initial search strategy to ensure a comprehensive scoping review [72]. This process involved identifying and analyzing papers that were referenced by the previously identified design frameworks. It should be noted that this snowball process was limited to papers that referenced design frameworks, even if they were published before 2015. This was done to ensure that all relevant papers were included in the review, regardless of their publication date, by revealing additional key works that are influential in the field and leading to a more robust understanding of the topic. This process allowed the exploration of the relationship between different design frameworks and identification of any additional design guidelines or good practices that were not captured in the initial search.

The synthesis of the selected studies was carried out using a thematic analysis approach, where the design frameworks identified in the reviewed papers were first classified into 2 broad categories: DEGDFWs and DSGDFWs. Each framework was assessed for its relevance to the characteristics of digital entertainment and serious games, and the subsequent cluster analysis for commonalities and differences was applied in the dedicated Results section of this paper.

Clustering

Following the comprehensive literature review, which provided a foundational understanding of the existing research landscape, we proceeded with a more granular examination through cluster analysis. In this second phase of the research journey, we applied clustering techniques to categorize the data into meaningful groups, allowing us to explore specific trends and correlations that were not immediately apparent in the earlier stages.

General Clustering

To categorize the frameworks, clustering through thematic analysis coding was done with the MAXQDA (VERBI GmbH) software for qualitative data analysis. Clustering refers to the task of grouping a set of objects in such a way that objects in the same group are more similar to each other than to those in other groups. The frameworks were clustered into distinct conceptual themes based on their focus areas. This process involved coding the frameworks’ key elements, analyzing the relationships between codes, and ultimately grouping them into broader themes that form the final clusters.

The clustering process, conducted collaboratively by 2 coders, involved several steps. First, each framework was reviewed to

XSL•FO
RenderX

create initial codes based on its primary focus and objectives (game mechanics, learning objectives, etc). An inductive approach was used in this process, where codes and themes were identified directly from the data based on observed patterns rather than predefined categories. This method allowed new insights to emerge and themes to be developed that accurately reflected the frameworks' focus area [73]. By using MAXQDA, the coders were able to create, assign, and visualize codes related to the primary focus areas of each framework. This software enabled efficient identification of patterns and relationships between the codes, helping to reveal commonalities and cooccurrences across the frameworks.

After coding, commonalities and relationships (frequency of codes and their cooccurrences) between frameworks were identified to develop higher-level themes. For example, the codes "learning objectives" and "educational content" were grouped into the theme "learning-oriented frameworks," while "gameplay elements" and "game mechanics" were grouped under "game design–oriented frameworks" (the full list of the identified codes is discussed in the Clustering Results section). The coders largely agreed on the final clusters, which are described in Multimedia Appendix 3, for distribution of the frameworks into each corresponding cluster.

The discussions among coders were structured around collaborative analysis sessions, where they reviewed the initial codes and debated their meanings and implications. This iterative process involved each coder presenting their interpretations and justifications for specific codes, allowing for a comprehensive examination of the frameworks in question. During these discussions, the coders identified common themes by highlighting overlaps and distinctions among the codes. They focused on the context and nuances of each code, fostering a dialogue that encouraged critical thinking and reflection on the broader implications of their findings. Moreover, the fit of clustering (intercoder agreement) was calculated using Cohen κ [74], which accounts for agreement by chance of coders. The Cohen κ value was 0.85, indicating a high level of agreement between the coders. This suggests that the frameworks were well categorized within the identified clusters.

Dependencies and Similarities Clustering

To analyze the potential dependencies and similarities between DSGDFWs and DEGDFWs, a comprehensive clustering procedure was also used. This process combined qualitative coding and quantitative network analysis using MAXQDA and R (R Foundation for Statistical Computing) to uncover the frameworks' conceptual links and dependencies.

In the first phase, data on framework dependencies were coded in MAXQDA, and a code cooccurrence matrix was created to quantify conceptual associations. For example, the "activist-casual framework" was identified as depending on the "Bartle taxonomy" to create games focused on delivering a customized player experience based on performance levels. In MAXQDA, qualitative coding was applied to these data. Codes such as "custom player experience" and "performance level dependence" were created and applied to segments describing these dependencies. Furthermore, in the second phase, a code cooccurrence matrix was derived in MAXQDA. For instance, the matrix indicates that "custom player experience" and "performance level dependence" frequently cooccur.

This matrix was then exported to the statistical software R, where a similarity matrix was derived to reflect the degree of relatedness between frameworks. For example, frameworks "activist-casual framework" and "Bartle taxonomy," sharing high similarity scores, were shown to be closely related. A distance matrix was derived from this similarity matrix, indicating how similar or different the frameworks were in terms of their theoretical dependencies. Clustering algorithms were applied to group frameworks into clusters based on these similarities. For example, let us presume that the frameworks such as "activist-casual framework" and "Bartle taxonomy" were assigned to clusters based on these analyses. In the third phase, a concept network graph was generated in R to visualize these clusters and the dependency flows between frameworks.

The final part of the clustering procedure involved interpreting the clusters to define overarching principles. Specifically, once the clustering algorithms had grouped the frameworks based on their conceptual dependencies, each cluster was analyzed to identify the shared characteristics and design philosophies of the frameworks within it. Consequently, these insights were used to associate each cluster with a specific principle. For instance, the principle of enjoyable education emerged from the analysis of a cluster where frameworks commonly emphasized balancing fun and educational content. This process ensured that each principle was grounded in the actual design practices and theoretical foundations reflected in the clusters.

Overall, by organizing these codes and analyzing their frequencies and relationships, we could highlight framework domains or categories and assess commonalities and differences across them. This enabled us to uncover conceptual clusters across the game design frameworks. The following section highlights the results of the clustering analysis.

Risk of Bias Assessment

To ensure a reliable and accurate analysis, the management of risk of bias was conducted using the MAXQDA [75] tool for qualitative data analysis and synthesis. This was done to refine and organize the relevant information. In addition to the tool, a manual assessment was performed by 2 reviewers to further verify the findings. This double approach was taken to ensure that the analysis was as thorough and accurate as possible.

Results

The results are structured into 2 main parts: findings from the literature review, which provide an overview of existing game design frameworks, and insights from the clustering analysis, which categorize and highlight patterns and relationships within the frameworks.

Literature Review Results

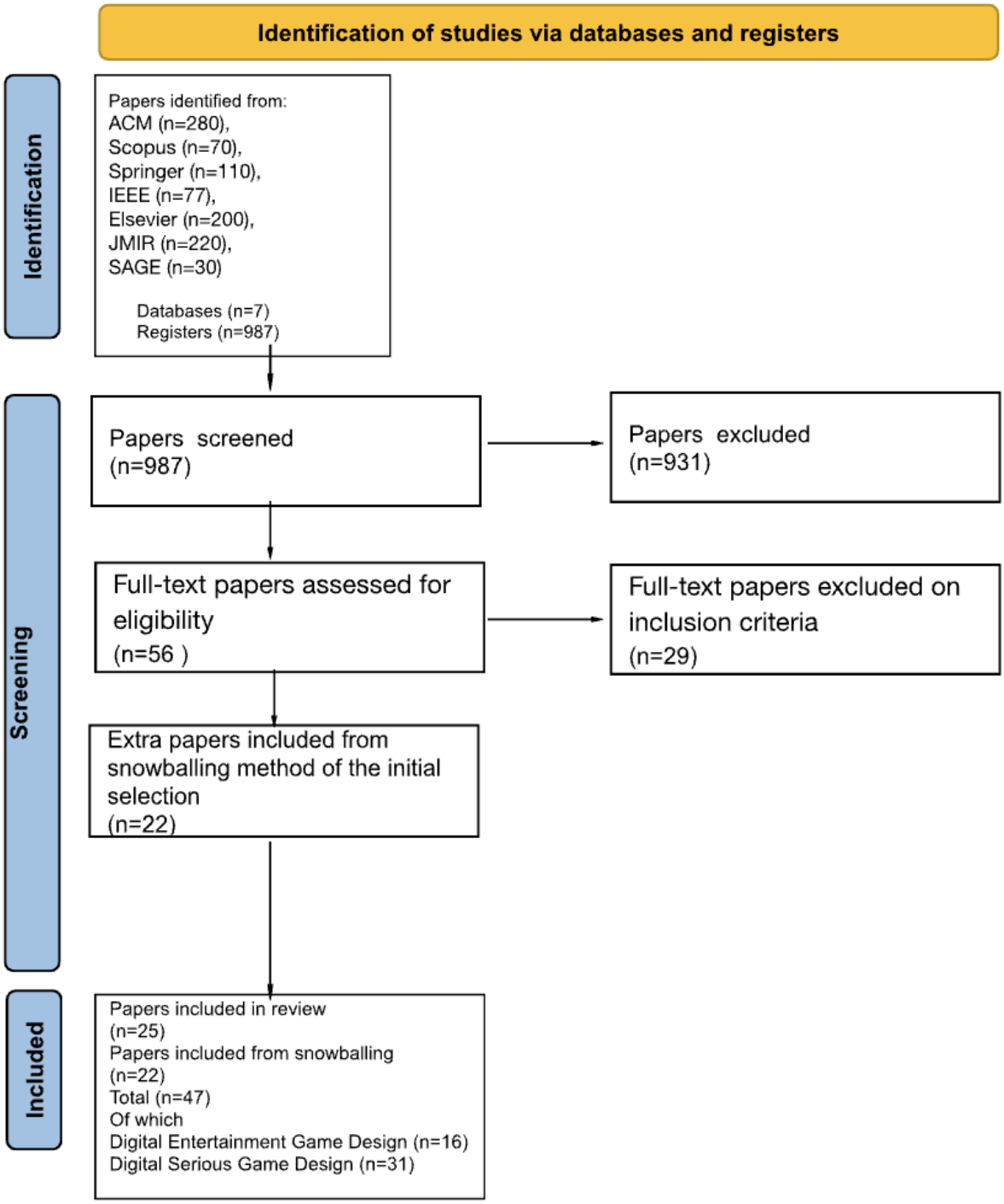
This review focused specifically on analyzing high-level conceptual frameworks that provide overarching guidance on the entire game design and development process and not narrower design lenses. The goal was to elucidate broader theoretical scaffolding around core phases, methodologies, needs

XSL•FO
RenderX

analysis factors, and evaluative pillars that inform practice foundations. Hence, the review concentrated on understanding process-oriented knowledge generalizable across use cases rather than tactical formulas for crafting individual interactions. In summary, the scope covered conceptual frameworks directing holistic developmental sequences rather than granular tools optimizing precise design decisions.

Figure 1 presents a PRISMA 2020 diagram outlining the study identification process. The research strategy initially identified 987 papers, of which 54 (5.5%) were selected based on the inclusion and exclusion criteria. The search parameters yielded the following number of results for each database: ACM (280/987, 28.4%), Scopus (70/987, 7.1%), Springer (110/987, 11.1%), IEEE (77/987, 7.8%), Elsevier (200/987, 20.3%), JMIR Publications (220/987, 22.3%), and SAGE (30/987, 3%).

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) diagram.



XSL•FO
RenderX

Following a second review, 25 papers were chosen for full reading and synthesis, focusing on design framework components. To ensure a comprehensive review, a snowballing method was applied in addition to the initial search. This involved analyzing papers cited by the identified design frameworks, including an additional 22 papers.

Ultimately, 2 types of frameworks were identified from a total of 47 papers: 25 (53%) from the initial selection and 22 (47%) from the snowballing process. The study synthesized 47 frameworks overall, with 16 (34%) related to DEGDFW (Table 1) and 31 (66%) related to DSGDFW (Table 2). The details of the frameworks can be found in Multimedia Appendices 4 and 5.

Table 1. Summary of the reviewed digital entertainment game design frameworks (DEGDFWs).

ID code	DEGDFW
EG[a]1	The Bartle taxonomy [43]
EG2	The Four Keys to Fun [52]
EG3	The Engines of Play [53]
EG4	The Player Involvement Model [54]
EG5	MDA (Mechanics, Dynamics, Aesthetics) [55]
EG6	The Layered Tetrad [56]
EG7	Design, Dynamics, Experience (DDE) [76]
EG8	Elemental Tetrad [77]
EG9	MTDA+N Narratives framework [78]
EG10	Integrated Framework for Game Design [79]
EG11	The 5-part Model [80]
EG12	Risk and Reward Model, (difficulty balancing) [44]
EG13	Game Design Patterns Model [81]
EG14	Game design framework [23]
EG15	Game design workshop [82]
EG16	Game Element and Mechanic framework [83]

[a]EG: entertainment game.

XSL•FO
RenderX

Table 2. Summary of the reviewed digital serious game design frameworks (DSGDFWs).

ID code	DSGDFW
SG[a]1	Activist-Casual framework [84]
SG2	Educational Games Design framework [21]
SG3	Design, Play, Experience framework [85]
SG4	Conceptual framework [86]
SG5	Four Dimensional Framework [87]
SG6	Triadic game design evaluation framework [88]
SG7	RETAIN framework [89]
SG8	Computational Puzzle Design framework [90]
SG9	Serious educational games framework [91]
SG10	FRACH framework [58]
SG11	Augmented intelligence framework [92]
SG12	Lu Lu framework [93]
SG13	Learning Mechanics-Game Mechanics [94]
SG14	Digital Game–Based Learning–Instructional Design Model [95]
SG15	Immersive Educational Games Model [96]
SG16	Design, Dynamics, Experience (DDE) framework [76]
SG17	MECONESIS methodology [97]
SG18	Framework with game design, learning content modeling, and pedagogy [21]
SG19	Methodology based on graphic notation and interactive narrative [98]
SG20	Methodology based on cognitive behavioral techniques [99]
SG21	EMERGO methodology [100]
SG22	ATMSG framework based on Activity Theory [22]
SG23	Methodology based on problem-based learning [101]
SG24	Six-dimensional framework in a reviewed paper [59]
SG25	Baseline content-centric framework [102]
SG26	GAMED framework [103]
SG27	User-centered design methodology [104]
SG28	Collaborative Learning Game framework [105]
SG29	Conceptual Model for Serious Games [106]
SG30	Adaptive Learning Game Design Model [107]
SG31	Intervention Mapping Framework [108]

[a]SG: serious game.

Following the literature review, the Clustering Results section presents the cluster analysis conducted to further explore and categorize the identified frameworks, providing additional insights into their relationships and design elements.

Clustering Results

Building on the findings from the literature review, the cluster analysis delves deeper into the relationships and patterns between DSGDFWs and DEGDFWs, offering a detailed examination of their commonalities, dependencies, and key design principles.

Purpose of Game Design Frameworks in Practice

Frameworks are essential tools that help designers and developers create effective and engaging educational games. These frameworks come in different types (Textbox 1), each with its own purpose, as shown in Multimedia Appendix 3.

XSL•FO
RenderX

Textbox 1. Types of frameworks.

- Learning-oriented frameworks: these frameworks focus on the integration of educational content and learning objectives into the game design process. Examples include the Conceptual Model for Serious Games, the Dimensional Framework, the Educational Games Design Framework, and the Serious Educational Games Framework.
- Game design–oriented frameworks: these frameworks focus on the game mechanics, game design, and gameplay elements that are necessary to create engaging and effective educational games. Examples include the Learning Mechanics—Game Mechanics model; the Immersive Educational Games Model; the Design, Dynamics, Experience framework; the Computational Puzzle Design framework; and the Collaborative Learning Game framework.
- User-centered frameworks: these frameworks focus on understanding the needs and preferences of the users and providing a personalized experience for them. Examples include user-centered design and the mild cognitive impairment game therapy experience (MCI-GaTE) frameworks.
- Game development–oriented frameworks: these frameworks provide a structured approach for overcoming the complexity of game development and provide a guide for the development process. Examples include the GAMED methodology and the MCI-GaTE frameworks.
- Immersivity and collaboration-oriented frameworks: these frameworks focus on the importance of immersivity and collaboration when promoting group learning. Examples include the FRACH and the serious educational games frameworks.
- Mobile and ubiquitous learning–oriented frameworks: these frameworks focus on the various components of mobile game–based learning to provide learners with a mobile or ubiquitous learning experience. Examples include the six-dimensional framework.

Commonalities in DEGDFWs

Despite their differences in scope and focus, many of these game design frameworks share some commonalities (Textbox 2).

Textbox 2. Commonalities among entertainment game design (DEG-C) frameworks.

1. Emphasis on player experience (DEG-C1): the player experience is a key focus in many game design frameworks. Creating an enjoyable and engaging experience for players is seen as essential in keeping them motivated and invested in the game. By prioritizing player experience, designers can create games that are not only educational but also enjoyable to play.
2. Consideration of game mechanics (DEG-C2): game mechanics are essential in game design frameworks, as they play a significant role in creating a fun and challenging game. The various game mechanics used in a game are useful in determining the level of challenge, reward, and engagement the game provides. By emphasizing game mechanics, designers can create games that are not only fun to play but also educational and informative.
3. Interconnectivity of game components (DEG-C3): many entertainment game design frameworks acknowledge the interconnectivity of game components such as mechanics, esthetics, story, and technology. The game mechanics, esthetics, and story all work together to create an immersive and cohesive game that can transport players to another world. By recognizing this interconnectivity, designers can create games that provide a seamless and enjoyable experience for players.
4. Importance of play testing (DEG-C4): several entertainment game design frameworks emphasize the importance of play testing in refining game design and improving the player experience. Play testing allows designers to identify areas for improvement in game mechanics, difficulty, and engagement. By iterating on the game design based on feedback from play testing, designers can create games that are more enjoyable and effective.
5. Need for iteration (DEG-C5): iterating on game design is recommended in many game design frameworks as a means to improve the player experience. By adjusting difficulty, refining game mechanics, and play testing, designers can create games that are more engaging and effective. Iteration is also essential in keeping up with technological advancements and changing user preferences.
6. Focus on understanding player motivations (DEG-C6): some game design frameworks suggest that understanding player motivations and personalities can help tailor game design to appeal to specific player types. By identifying what motivates players to play, designers can create games that are more engaging and enjoyable. Understanding player motivations also helps designers create games that meet specific learning objectives and goals.
7. Acknowledgment of the importance of storytelling (DEG-C7): it is emphasized in some game design frameworks as a means of creating an immersive game world and engaging players. By incorporating a compelling story into the game, designers can transport players to another world and provide a more enjoyable and engaging experience. Storytelling also helps designers create games that are more educational and informative by conveying information in a narrative format.

Commonalities in DSGDFWs

The analyzed serious game design frameworks underline some commonalities that are important because they provide a foundation for creating effective serious games (Textbox 3).

XSL•FO
RenderX

Textbox 3. Commonalities among serious game design (DSG-C) frameworks.

1. Educational content with gameplay mechanics (DSG-C1): by incorporating educational content into games, designers can create immersive and engaging learning experiences for players. Serious games have the potential to enhance learning outcomes by using gameplay mechanics to motivate and engage learners in educational content.
2. Instructional design principles (DSG-C2): serious game development follows a systematic and structured approach that includes instructional design principles and phases. The process involves defining learning objectives, analyzing the target audience, designing the game mechanics, developing the game, testing it, and evaluating its effectiveness.
3. Immersive gameplay elements (DSG-C3): immersive elements help designers to create a more engaging and compelling learning experience for players. Immersive elements, such as virtual reality, audio and visual effects, and realistic scenarios, help create a sense of presence and realism that can enhance the learning experience.
4. Player experience and feedback (DSG-C4): designers need to consider the player's experience throughout the game development process to ensure that the game is engaging and effective. Collecting player feedback helps identify areas for improvement, which can be addressed to improve the game's usability, ease of use, functionality, and effectiveness.
5. Pedagogy and learning content modeling (DSG-C5): the integration of these 2 elements helps ensure that the game is aligned with the desired learning outcomes. Pedagogy helps designers create effective instructional strategies, while learning content modeling helps them organize and structure the educational content.
6. Cognitive and psychological aspects of players (DSG-C6): designers need to consider the cognitive and psychological aspects of the player, such as attention, memory, emotion, and motivation, to create an engaging and emotional gameplay experience. By incorporating these elements into the game design, designers can create a more personalized and effective learning experience for the player.
7. Collaboration and communication (DSG-C7): the emphasis on collaboration and communication among team members is essential to facilitate the exchange of ideas and artifacts. This collaborative approach helps ensure that the game is designed effectively, aligns with the learning objectives, and meets the needs of the learners.

Dependencies Between DSGDFWs and DEGDFWs

DEGs are popular because they provide players with enjoyable and engaging experiences that allow them to escape from reality and immerse themselves in a different world. DEGs are designed to be fun and pleasing, with challenging gameplay mechanics and compelling storylines that keep players coming back for more. Therefore, DSGs have extended from DEGs because they recognize the potential of games as a tool for education, training, and other serious applications.

There are several dependencies that exist between DSGDFWs and DEGDFWs, as both types of games share some commonalities. For instance, both types of digital game design frameworks emphasize the importance of player experience and game mechanics. In DEGs, game mechanics are important in creating a winsome and challenging game, while in DSGs, game mechanics are essential in delivering an immersive and attractive learning experience. Similarly, both types of games acknowledge the importance of play testing and iteration to refine game design and improve the end-to-end player experience.

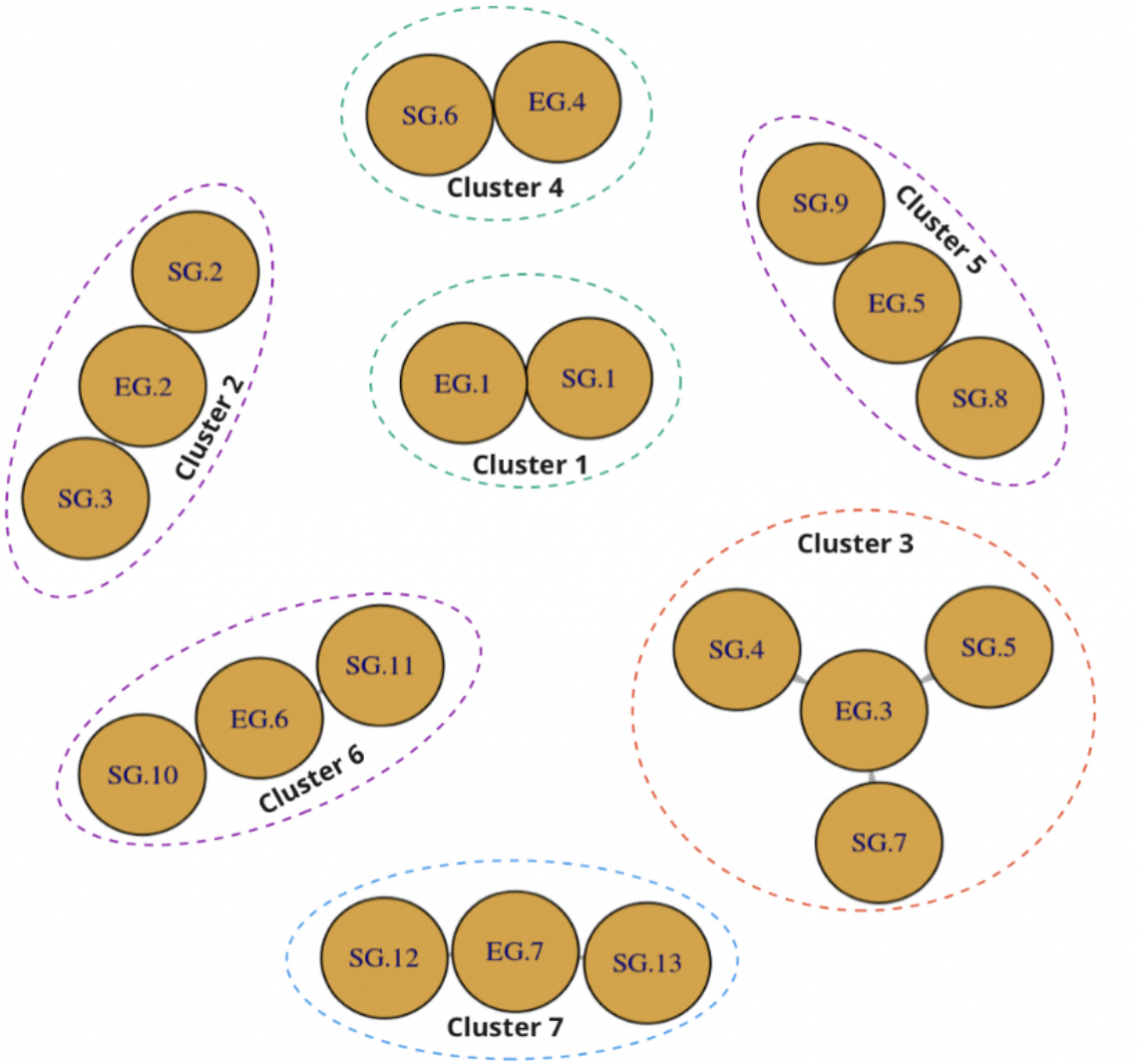
Another common characteristic between DSGDFWs and DEGDFWs is the emphasis on understanding player motivations. In DEGs, understanding player motivations can help tailor game design to appeal to specific player types, while in DSGs, they are pillars to create games that meet specific learning objectives and outcomes.

Storytelling is also an important design element identified in both DSGs and DEGs, as it supports an immersive game world and captivation of players. In DEGs, storytelling is used to provide a more compelling experience, while in DSGs, it underlines the conveyance of educational content in a narrative format. Overall, both serious and entertainment games design frameworks share several commonalities and dependencies, highlighting the importance of considering both types of games when designing effective and engaging games.

The scoping review identified 2 main types of frameworks from the 47 papers analyzed: 16 (34%) DEGDFWs (Table 1) and 31 (66%) DSGDFWs (Table 2). Of the 47 frameworks, 25 (53%) were initially selected, and an additional 22 (47%) were gathered through snowballing techniques. Figure 2 illustrates a graph showing the dependencies between 13 DSGDFWs and 7 DEGDFWs, organized into 7 clusters. These specific frameworks were chosen based on their prominence and relevance in the literature, as they demonstrated significant interdependencies and were most representative of the key principles identified in the study. Either the remaining frameworks did not show significant relationships, often forming isolated or single-node clusters, or their inclusion did not contribute substantially to the overarching themes and trends we aimed to analyze.

XSL•FO
RenderX

Figure 2. Graph with clusters on dependencies between SG.1-13 and DG.1-7 coded design frameworks.



A subset of the design frameworks discussed in the Literature Review section were clustered with the associated ID code. These ID codes correspond to the papers listed in Tables 1 and 2, where each paper is assigned a unique identifier for easier reference. The selected frameworks for clustering were chosen based on their relevance to the research objectives, while other papers either formed single-node clusters or did not yield significant patterns in this analysis. The clusters are visually represented in Figure 2, allowing for clear identification and comparison (Textbox 4).

XSL•FO
RenderX

Textbox 4. Clusters of design frameworks for serious games (SGs) and entertainment games (EGs).

1. Cluster 1: the first dependency shown in this graph is the Activist-Casual framework (SG.1), which depends on the Bartle taxonomy (EG.1) to create games that focus on delivering a custom experience to players based on performance level.
2. Cluster 2: the Serious Educational Games (SG.2) and Design, Play, and Experience (SG.3) frameworks both depend on the Four Keys to Fun (EG.2) framework to create games that are fun and engaging while also providing educational content.
3. Cluster 3: the Conceptual Model (SG.4) and Four Dimension (SG.5) frameworks both rely on the Engines of Play (EG.3) framework to create games that are challenging and immersive while also incorporating compelling storylines. The RETAIN framework (SG.7) also relies on the Engines of Play framework (EG.3) to create games that are designed to be played repeatedly and encourage players to a continuous learning experience.
4. Cluster 4: the Triadic Game Design (SG.6) framework depends on the Player Involvement Model (EG.4) to create games that are tailored to the individual learning styles and preferences of each player.
5. Cluster 5: the Computational Puzzle Design (SG.8) and SG-LMI (SG.9) frameworks both depend on the Mechanics, Dynamics, and Aesthetics (EG.5) framework, which prioritizes the player's experience and journey throughout the game. SG.8 uses the EG.5 framework to create mechanics that facilitate problem-solving through programming, while SG.9 uses it to create an immersive and engaging game experience that supports effective learning outcomes.
6. Cluster 6: the FRACH Serious Game (SG.10) and Augmented Intelligence (SG.11) frameworks both depend on the Layered Tetrad Model (EG.6) to create games that are designed to be immersive and engaging while also incorporating elements of augmented reality and artificial intelligence. SG.10 applies the EG.6 model to incorporate game mechanics and learning objectives, while SG.11 uses the EG.6 model to create an augmented intelligence system that adapts to the player's learning needs.
7. Cluster 7: the LU LU framework (SG.12) and mild cognitive impairment game therapy experience framework (SG.13) both depend on the Design, Dynamics, and Experience (EG.7) framework to create games that are designed to be fun and engaging while also incorporating educational content and promoting positive behaviors.

Principles (PR1-6) derived from the clusters of dependencies between frameworks (Textbox 5).

Overall, while DSGDFWs are focused on delivering educational content to players, they rely heavily on DEGDFWs to ensure that the games they create are enjoyable and engaging for players. DEGDFWs provide DSG designers with a set of tools and techniques that they can use to make their games more immersive and impactful. By leveraging the commonalities and dependencies between DSGDFWs and DEGDFWs, designers can create games that are both educational and entertaining, which is essential to attracting and retaining players. As the field of DSG continues to evolve, it will be interesting to see how designers continue to leverage DEGDFWs to create impactful and meaningful experiences for players.

XSL•FO
RenderX

Textbox 5. Principles (PR) derived from the clusters.

- Challenge tailoring (PR1): design gameplay mechanics and challenges that are tailored to the abilities and progress of players, providing an engaging experience. By adjusting the difficulty level or introducing adaptive elements, players are motivated to push their boundaries and feel a sense of accomplishment as they overcome challenges suitable for their skill level. Cluster 1 focuses on creating games that offer a customized experience based on performance levels. This aligns with the principle of challenge tailoring, which emphasizes designing gameplay mechanics and challenges tailored to players' abilities and progress.
- Enjoyable education (PR2): balancing enjoyment and educational content is crucial. Games should provide enjoyable experiences that captivate players' attention and interest while effectively delivering educational material. This can be achieved through incorporating interactive learning elements, gamified approaches, and engaging narratives that make the educational content more enjoyable and memorable. Cluster 2 balances fun and educational content. This corresponds to the principle of Enjoyable Education, which stresses the importance of making educational content both engaging and enjoyable.
- Immersive experience (PR3): create gameplay mechanics and features that immerse players in the game world, enhancing the overall experience. Creating gameplay mechanics and features that immerse players in the game world enhances the overall experience. By focusing on details such as realistic graphics, immersive audio, intuitive controls, and captivating storytelling, players can become fully absorbed in the game, fostering a sense of presence and enhancing their emotional connection with the game environment. Cluster 3 relies on creating immersive and challenging games. This maps to the principle of immersive experience, focusing on enhancing player immersion through engaging gameplay mechanics and features.
- Player engagement (PR4): designing games that cater to specific player preferences and learning needs increases player engagement. This can involve providing customizable options, diverse gameplay styles, and adaptive mechanisms that allow players to approach the game in ways that align with their interests and individual learning styles. Cluster 4 highlights its focus on tailoring games to individual learning styles and preferences. This aligns with the principle of player engagement, which involves designing games that cater to specific player interests and learning needs.
- Motivate continuous learning (PR4): incorporating gameplay mechanics and features that motivate players to continue playing and learning is crucial. This can be achieved through elements such as rewards, achievements, leaderboards, progression systems, and unlocking new content. By offering tangible incentives and a sense of progression, players are encouraged to keep playing, exploring, and learning. Cluster 5 prioritizes player experience and problem-solving. This corresponds to the principle of motivate continuous learning, emphasizing the importance of encouraging ongoing engagement and learning through game mechanics.
- Effective learning outcomes (PR5): designing a game experience that supports effective learning outcomes involves aligning educational objectives with gameplay elements. This includes incorporating interactive challenges, problem-solving scenarios, and knowledge- or application-based tasks that promote active learning and meaningful engagement. Cluster 6 is related to unique or innovative elements such as augmented reality. This aligns with the principle of effective learning outcomes, which involves aligning educational objectives with gameplay elements to support meaningful learning not only playing for fun.
- Positive behaviors (PR6): provide a gaming experience while encouraging desirable behaviors. This involves rewards for completing educational tasks or promoting prosocial behaviors within the game context. By fostering positive behaviors and attitudes, games can have a broader impact on players' lives beyond the game itself. Cluster 7 contains fun and engaging games with educational content. This maps to the principle of positive behaviors, focusing on incorporating positive reinforcement and promoting desirable behaviors through game design.

Discussion

Principal Findings

The field of serious games has emerged as an effective digital solution for traditional education, training, and behavior change. However, designing effective DSGs that achieve their intended outcomes is a complex process that requires a thorough understanding of the design frameworks used in both entertainment games (DEGDFWs) and serious games (DSGDFWs). The purpose of this scoping review was to investigate the design frameworks used in both types of games and explore how they can inform the design and development of serious games.

The review involved identification and analysis of the recurring steps that appeared across the selected game design frameworks. The synthesis of the reviewed frameworks involved several steps. First, the frameworks were analyzed to identify key elements, structures, and processes relevant to game development. Common components and phases across them were then extracted, such as game mechanics, player engagement, learning outcomes, and iterative design. Next, frameworks with overlapping concepts were grouped together, highlighting similarities in approach, such as those focusing on player behavior and personality, game mechanics and dynamics, or the integration of educational content.

This section discusses the key findings of the scoping review by presenting a visualization of the 4-phase design. It must be noted that this is not a design framework proposal but a conceptual and visual summary of the reviewed design space for DSG design. In addition, a set with guiding design principles are mapped on each proposed design phase to better inform the future design initiatives.

The synthesis of common steps into the 4 design phases baseline directly cross-references insights from the literature review and cluster analysis, which emphasized the structured concepts essential for effective educational game design. For instance, the importance of aligning game mechanics with educational objectives, aligning with the conceptualization phase identified in the current analysis. Similarly, the previous sections discussed the necessity of detailed planning in preproduction, echoing the importance of narrative and educational goals in game development.

Although each framework was unique in terms of its specific components, several key steps were consistently noted. These steps included conceptualization—where the initial idea or game

XSL•FO
RenderX

concept is formed; preproduction planning—involving the detailed planning of game mechanics, narrative, and educational goals; prototyping iterations—where initial versions of the game are developed, tested, and refined; and evaluation—a process of assessing the game's effectiveness, either through playtesting or measuring educational outcomes after its deployment in production. By recognizing these common steps, the analysis distilled the frameworks into a more abstract structure, which pointed to an underlying shared design flow as a 4 design phases baseline. This synthesis was further refined by grouping these steps into 4 overarching phases: exploration, design, development, and assessment.

In essence, the commonalities and dependencies identified in the previous sections of the paper directly align with key considerations in each phase of the 4-phase design baseline flow: initial exploration should be guided by an understanding of who the players are; design should integrate engagement, learning, and outcomes; development should be iterative, incorporating user input; and assessment should be continuous to enable refinement. The reliance on entertainment models for engagement shows the importance of player-centered research on motivations and interests in the initial exploration while still scoping learning goals. With many frameworks leveraging entertainment concepts such as narrative and challenge for enjoyment, the design phase should focus on seamlessly integrating both fun and pedagogy into cohesive mechanics. The adaptation and customization approaches demonstrate the need for iterative, user-centered development that continually gathers feedback to refine serious games to be more responsive to learners. The goal of continuous improvement highlighted in the clusters requires embedding assessment throughout the process to evaluate learning and track engagement, enabling real-time identification of issues through analytics.

Furthermore, this section presents some detected gaps as highlights for future research. By understanding these gaps, researchers and designers can focus on addressing them for the goal to enhance the effectiveness of serious games in real practice.

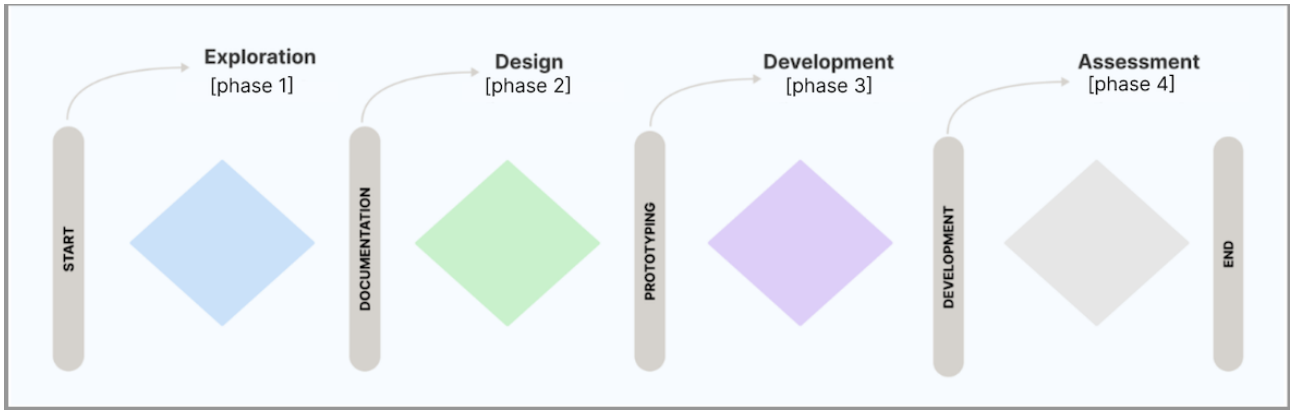
DSGDFW Baseline With 4 Design Phases

Overview

As a result of the analysis performed during the literature review, several high-level phases that were common across both entertainment and serious games, despite differences in their intended outcomes, were revealed. For example, key phases such as conceptualization, preproduction planning, prototyping iterations, and evaluation processes were noted consistently. These common phases pointed to a shared design thinking structure, even when specific tools and techniques differed between entertainment and instructional contexts.

These common phases inform an integrated design flow encapsulating exploration, design, development, and assessment, which have been synthesized in this section, summarized in Figure 3. While specifics within each phase need tailoring or change depending on each paper included in the review, the macrolevel sequencing of the design process itself provides an intuitive skeleton. This enabled simplicity in the baseline visualization focused on connecting sequential design stages shared across frameworks. Moving forward, further layers of complexity can be added within phases. The visualization therefore aims to provide clarity to serious game creators on the core high-level workflow derived from common practices identified through analysis.

Figure 3. Overview of the 4-design phase.



The exploration phase sets the stage for integration by interweaving learning principles and content considerations with gameplay factors from the outset, thereby ensuring the entertainment and education elements synergize.

The identified commonalities and dependencies principles such as emphasis on player experience, game mechanics consideration, challenge tailoring, and immersive experience, from previous sections, informed this first phase. During this phase, understanding player motivations and preferences is central. Both DEG and DSG frameworks stress the need for designers to focus on creating enjoyable and engaging player experiences. The game mechanics must be conceptualized here, as they influence challenge levels, engagement, and the player's interaction with the game's content.

Exploration Phase

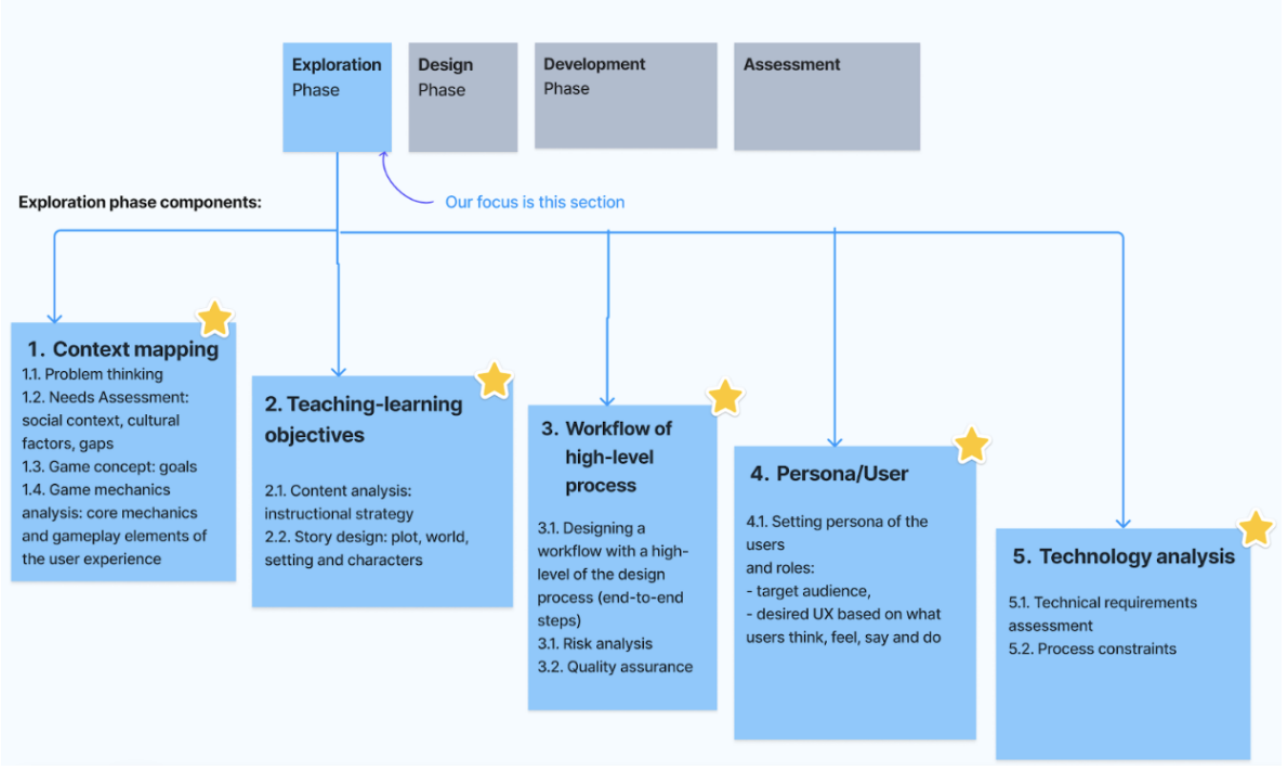
In the exploration phase, the key elements are defining the main idea, intended audience, learning goals, and identifying user profiles and risks. The Conceptual Model for Serious Games provides a framework for designing serious games by defining the key design elements of the game, such as goals, challenges,

XSL•FO
RenderX

rewards, and feedback. For instance, GAMED methodology focuses on identifying user needs and preferences by considering the game's content, structure, and user interface, while digital game–based learning—instructional design (DGBL-ID) model helps to determine the learning objectives and target users of the proposed game by following a structured approach to designing and developing educational games. This phase is important as it lays the foundation for the entire serious game development process.

The exploration phase of serious game design requires careful consideration of various components as shown in Figure 4. Context mapping involves problem thinking, identifying the social context, cultural factors, and gaps through a needs assessment, while the game concept is established by defining the game goals. Game mechanics analysis is used to analyze the core mechanics and gameplay elements of the user experience, while teaching-learning objectives are set by conducting a content analysis to determine the instructional strategy. Story design is developed to establish the plot, world, setting, and characters. Workflow design is essential for developing games systematically and efficiently, and involves risk analysis and quality assurance. Setting the persona of the users and roles is critical to identifying the target audience and designing the desired user experience through empathy map modeling. Finally, technology analysis involves assessing technical requirements and process constraints to ensure that the game is developed within the necessary technological limitations.

Figure 4. Exploration phase. UX: user experience.



When designing serious games during the exploration phase, it is important to follow these guidelines to ensure thoughtful consideration of various aspects. The EG1-6 guidelines are correlated with the principles and commonalities (eg, DEG-C1-7, DSG-C1-7, and PR1-6) outlined in the previous sections (Textbox 6).

XSL•FO
RenderX

Textbox 6. EG1-6 guidelines.

- EXP-G1: understand the problem and educational needs—focus on identifying the learning gaps and instructional challenges to be addressed, ensuring the game design supports effective learning outcomes (PR5 and PR6). Take into account player experience (DEG-C1) and feedback (DSG-C4) to create a customized solution that meets both educational and motivational needs (DSG-C5).
- EXP-G2: define game goals and overall concept—set clear learning objectives aligned with the educational content, incorporating instructional design principles (DSG-C2) and pedagogical modeling (DSG-C5). Ensure the balance between educational and entertainment content (PR2) to create enjoyable learning experiences (DEG-C1 and DSG-C1).
- EXP-G3: outline core gameplay mechanics and dynamics—design adaptive and immersive gameplay mechanics (PR1 and PR3) that integrate educational content (DSG-C1). Tailor the challenges to player abilities (DEG-C2) and ensure mechanics are interconnected with esthetics and narrative elements (DEG-C3) to promote continuous engagement (PR4).
- EXP-G4: determine instructional strategies for teaching players—analyze and design the content delivery approach by blending entertainment game elements with instructional strategies. Focus on gameplay mechanics that reinforce learning objectives (PR5), and integrate psychological aspects such as motivation and memory (DSG-C6) to enhance player engagement and retention (PR4).
- EXP-G5: develop an engaging narrative and world—create a compelling story (DEG-C7) and immersive game environment (DSG-C3) that supports the educational goals. The narrative should guide players through their learning journey, ensuring the plot and game tasks align with instructional outcomes (PR3 and PR2), while maintaining the focus on motivation and enjoyment (PR2).
- EXP-G6: conduct risk analysis and establish quality standards—assess potential risks and ensure the game adheres to quality assurance standards. Play testing (DEG-C4) and iteration (DEG-C5) are key for refining the game design to maintain balance between difficulty, engagement, and educational content (PR1 and PR5).
- EXP-G7: identify and customize for the target audience—research and define the characteristics of the target learners, including their knowledge levels, learning styles, and motivational drivers (DSG-C5 and DSG-C6). Customize the game design to meet these needs by offering adaptive learning paths (PR4) and immersive experiences tailored to player preferences (DEG-C6).

Overall, the exploration phase takes a broad focus on laying the strategic groundwork through investigative research and early concept ideation to identify well-grounded opportunities where purposeful games can deliver substantial educational value. The phase delivers clarity on the needs, goals, possibilities, and priorities to focus the global initiative.

The design phase serves as the critical foundation that informs all subsequent stages of serious game creation by merging educational and entertainment elements into a unified vision. It is during design that learning goals get translated into concrete gameplay mechanics, dynamic challenges, and a narrative interweaving obstacle progression with instructional scaffolding. Established pedagogical frameworks should provide guidance to map activities supporting knowledge construction, mastery through application, evaluative feedback, and metacognitive reflection.

The interconnectivity of game components, importance of play testing, immersive experience, and effective learning outcomes commonalities and principles are linked to this design phase. The narrative, esthetics, and game mechanics must work together harmoniously, aligning with both DEG and DSG frameworks. The iterative process begins here, allowing the game design to evolve based on user feedback, enhancing the player experience. Incorporating immersive elements, such as realistic graphics or compelling storylines, is essential. DSG frameworks emphasize aligning gameplay with educational objectives, meaning that the mechanics designed must not only be engaging but also facilitate learning.

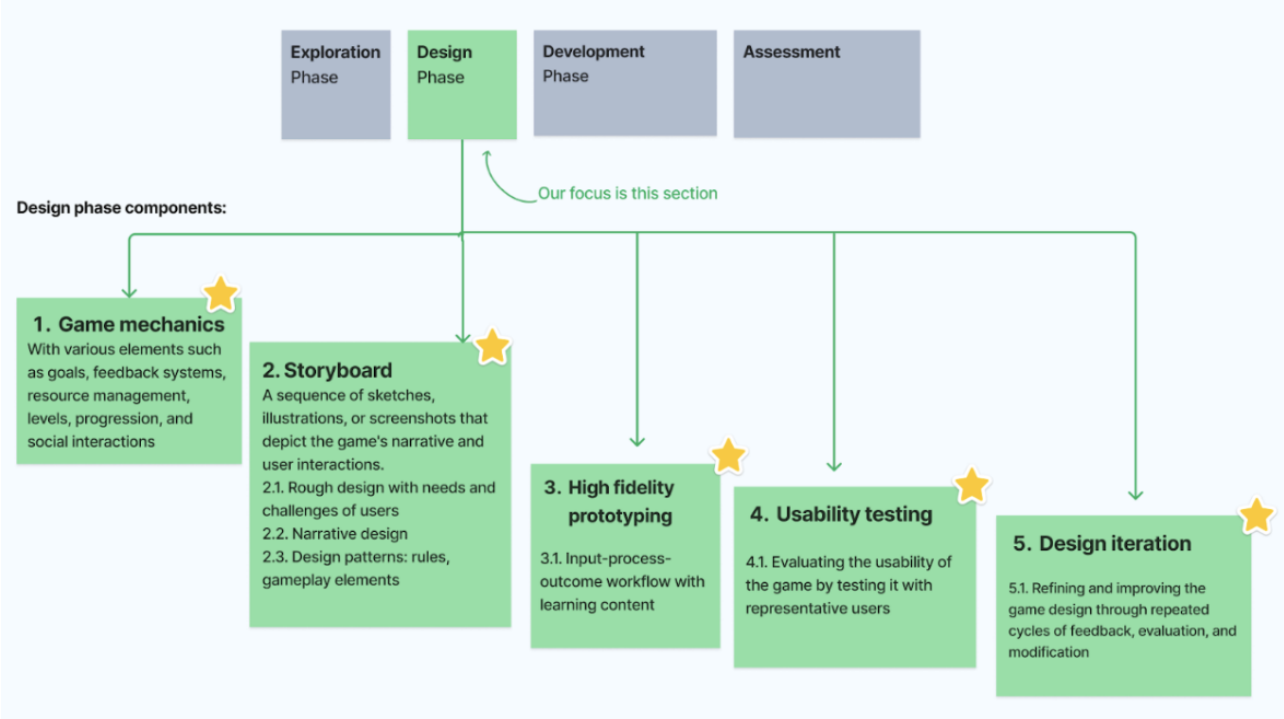
Design Phase

In the design phase, the key elements are designing the game environment, mechanics, scenarios, objects, and architecture, and capturing and defining the framework of the game process. For instance, the Conceptual Model for Serious Games provides a framework for designing serious games by defining the key design elements of the game. The DGBL-ID model focuses on creating the graphical user interface and designing the game's fantasy and story context. The RETAIN model emphasizes the transfer of knowledge and repetition of use in designing the game's fantasy and story context. This phase is important as it encompasses the creative and artistic elements of the serious game, shaping its visual, audio, and interactive aspects. Well-designed game elements are essential in creating an engaging, immersive, and enjoyable experience for players, which can enhance their motivation, learning outcomes, and overall experience with the serious game.

The design phase of serious game design consists of several components as shown in Figure 5. The first component involves the actual design of game mechanics, which has been agreed upon in the previous phase of analysis. The second component is the creation of a storyboard for the serious game solution, which includes a rough design with the needs and challenges of users, a narrative design, and agreed design patterns such as rules and gameplay elements. The third component involves creating a high-fidelity prototype, which includes an input-process-outcome workflow with learning content. The fourth component is usability testing, which involves testing the usability of the serious game solution with users. The final component is DSG design iteration before development, which allows for feedback and improvements to be made before the development phase begins.

XSL•FO
RenderX

Figure 5. Design phase.



The guidelines listed are important in the design phase as they provide a foundation for decision-making. The Dg1-5 guidelines are aligned with the principles and commonalities (eg, DEG-C1-7, DSG-C1-7, and PR1-6) from earlier sections, ensuring that the design phase is informed by established frameworks (Textbox 7).

Textbox 7. Dg1-5 guidelines.

- DES-G1: align mechanics with analysis—map game mechanics (DEG-C2) to player engagement (DEG-C1 and PR4) and learning objectives (DSG-C1, DSG-C5, and PR5), ensuring that challenges (PR1) and positive behaviors (PR6) are tailored to players' cognitive abilities and motivations (DEG-C6 and DSG-C6). Instructional design principles (DSG-C2) should guide this alignment to ensure the educational content is effectively delivered.
- DES-G2: create a comprehensive storyboard—develop a rich narrative (DEG-C7) that integrates educational content with storytelling (PR2) and immersive elements (DEG-C3, DSG-C3, and PR3). The storyboard should map out how the player's progression and learning activities (DEG-C2 and DSG-C1) become essential for advancing through challenges (PR1) and achieving effective learning outcomes (PR5).
- DES-G3: build high-fidelity prototypes—build prototypes that focus on immersive gameplay mechanics (DEG-C3, DSG-C3, and PR3) and collaboration features (DSG-C7), ensuring they are realistic and engaging (PR2 and PR4). These prototypes should test the integration of educational content with learning loops and performance feedback (DSG-C1 and PR5), fostering early insights into player engagement and behaviors (PR4 and PR6).
- DES-G4: test usability with users in the early phase—conduct usability testing with target learners to gather feedback (DEG-C4 and DSG-C4) on engagement, usability, and learning effectiveness (PR5). This early feedback loop helps refine the alignment of game mechanics with player motivations and cognitive aspects (DEG-C6, DSG-C6, and PR4), ensuring the gameplay supports effective learning outcomes (PR5).
- DES-G5: iterate design before development—iterate on the game design (DEG-C5) based on playtesting (DEG-C4 and DSG-C4) and feedback on engagement, challenge tailoring (PR1), and positive behaviors (PR6). This iterative process ensures that the game's educational content and mechanics are refined to enhance both player experience and learning results (PR5).

Overall, effective serious games arise not from simply layering a thin game on top of educational content, but from baking deep learning principles into design thinking itself. So the blend of engagement and outcomes begins here by thoughtfully transitioning high concepts around pedagogy and theories into an emergent unified practice.

Development Phase

In the development phase, the key element is developing the game using programming software and graphic editing tools. For example, the DGBL-ID model provides a framework for developing educational games by following a structured approach to designing and developing the game. This phase is very important as it brings the game to life by incorporating the designed elements into a functional game. Efficient development processes, coding, and deployment are essential in ensuring that the game functions as intended and meets the design requirements.

The development phase considers elements from the mapped commonalities and dependencies principles such as play testing in the real built game where the improvements area of the games can be deepened as it was in the prototyping phase with limited clickable actions. The importance of storytelling is another key

XSL•FO
RenderX

principle that guides the coding of game mechanics, ensuring a smooth, end-to-end flow in narrative sequences and also influences the development of graphics, audio, and video assets. Collaboration and communication are elements that interrelate all 4 design phases, but their importance is even higher in this phase in which designers from the previous 2 phases of exploration and design should concisely share the strategy with deep details to both game front- and backend engineers.

The development phase is critical for serious games because it determines how well the conceptual design gets translated into an actual playable, educational experience. With so many technical components to coordinate, it is easy for the learning goals to get lost in the coding efforts and asset creation. That is why the highest focus during development should remain on manifesting the learning theories and instructional strategies that were outlined earlier.

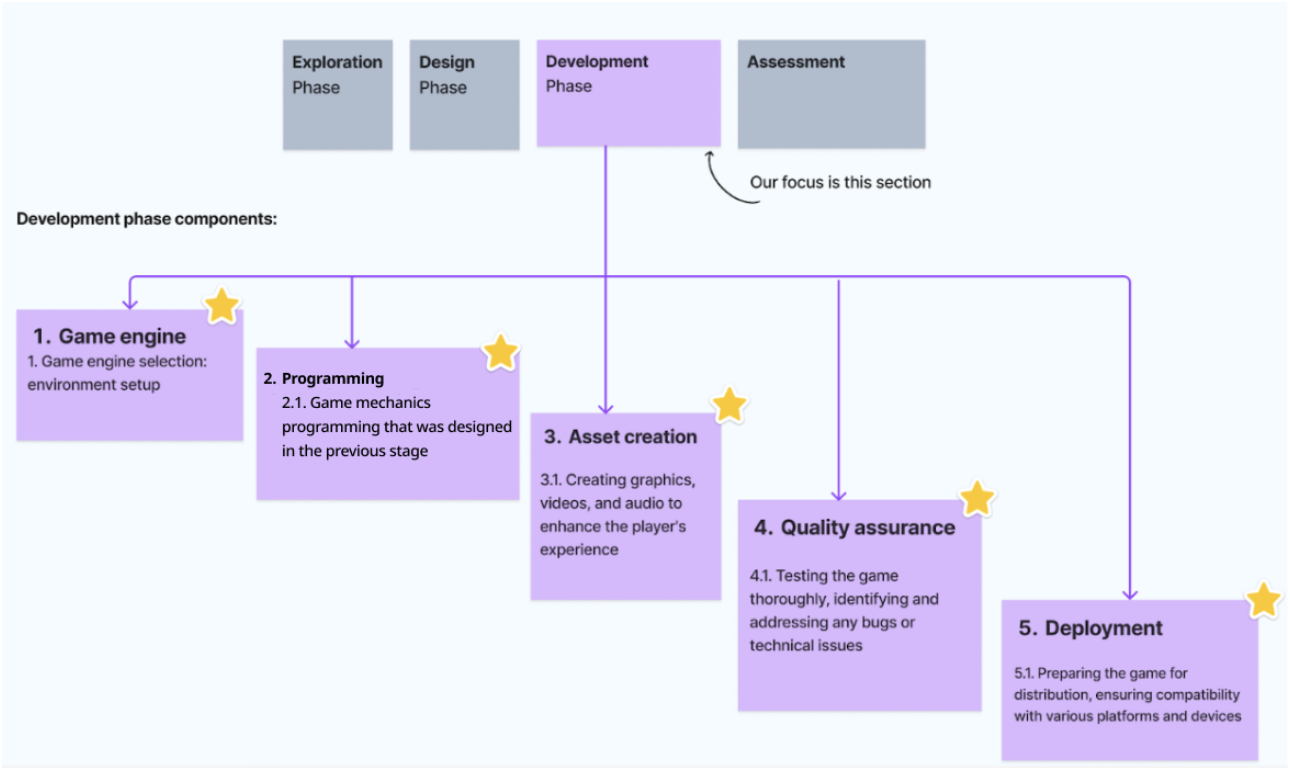
Keeping the target learners at the center, developers need to constantly question how each new mechanic and gameplay element reinforces the intended skill-building instead of solely crafting an entertaining experience. The educational content should drive the creation of media assets to correctly illustrate target knowledge instead of generic placeholders. Programming adaptive levels requires understanding how to model evolving mastery. Iterative playtesting verifies that embedded assessments provide useful diagnostic data to tell how well learners are acquiring expertise.

In essence, the biggest risk is cleanly designed learning principles not making an intact transition into the implemented product. By emphasizing the instructional elements as the true measures of quality assurance over graphics polish or clever moderately fun interactions lacking constructive pedagogical alignment, serious game developers keep advancing toward the vision of fusing engagement with substantive outcomes originally conceived. The highest focus stays on the target users achieving growth.

On top of the development phase components, artificial intelligence (AI) can play a very important role in supporting serious game personalization. AI machine learning algorithms can adapt game content for learners' motivation. AI techniques, such as procedural content generation, can automate the creation of game content, levels, and scenarios. AI-powered natural language processing techniques can enable communication and interaction with game characters through speech or text. This opens up possibilities for language learning, dialogue-based decision-making, or simulated conversations within serious games [109,110].

The components for the development phase of serious game design represent the technical implementation of the game as shown in Figure 6. Game engine selection involves choosing an appropriate game engine for the development of the game, and setting up the environment for development. Game mechanics programming involves coding the game mechanics that were designed in the previous phase. Asset creation involves the creation of graphics, videos, and audio that are required for the game, by building the game's characters, environments, and animations, as well as creating sound effects and background music. Quality assurance ensures that the game functions correctly, is free of bugs and glitches, and meets all necessary standards. Deployment and distribution involves making the game available for users to access and play.

Figure 6. Development phase.



XSL•FO
RenderX

In the development phase, designers are encouraged to incorporate some guidelines alongside the specified components. The development guidelines (DEV-G) are aligned with the principles and commonalities (eg, DEG-C1-7, DSG-C1-7, and PR1-6) from earlier sections (Textbox 8).

Textbox 8. DEV-G guidelines.

- DEV-G1: choose an appropriate game engine for development—select a game engine that supports embedding assessment analytics (DSG-C5), allowing the integration of mechanics tailored to player skills (PR1) and immersive educational experiences (PR5). The engine should support both seamless gameplay (DEG-C3) and educational outcomes (DSG-C1).
- DEV-G2: set up the development environment—create a modular environment that facilitates iterative refinement (DEG-C5 and DSG-C2) and supports adaptive algorithms for customizing challenges (PR1). This setup helps balance player experience (DEG-C1) and learning content (DSG-C5), allowing continual adjustments based on learner feedback (PR2 and PR4).
- DEV-G3: code the game mechanics based on the design—develop game mechanics that integrate learning objectives (DSG-C1 and DSG-C5) and player experience elements (DEG-C1 and DEG-C2), ensuring the mechanics are tailored to challenge players progressively (PR1) while delivering educational content in an engaging way (PR2). Gameplay mechanics should also reinforce immersive experiences (PR3).
- DEV-G4: build game characters, environments, and animations—design characters and environments that provide an immersive experience (DEG-C3 and DSG-C3), ensuring the game world is engaging and supports storytelling (DEG-C7 and PR5). Adaptive gameplay should respond to player emotions and motivations (DEG-C6 and DSG-C6), fostering engagement and promoting continuous learning (PR4).
- DEV-G5: develop sound effects and background music—use sound and music to enhance immersion (PR3 and DEG-C3) and motivate positive behaviors (PR6). Well-designed audio cues can support both storytelling (DEG-C7) and the reinforcement of educational tasks (DSG-C5), contributing to an overall engaging player experience (DEG-C1 and DSG-C4).
- DEV-G6: conduct quality assurance to ensure bug-free and glitch-free gameplay—thorough testing (DEG-C4 and DSG-C4) should be conducted with target users to identify usability issues affecting engagement (PR4) and ensure the gameplay supports continuous learning (PR2). Iteration based on playtesting ensures refined mechanics (PR1) and better player experience (DEG-C1).
- DEV-G7: meet all necessary standards for the game—ensure the game complies with relevant industry and educational standards (DSG-C2 and DSG-C5), integrating structured instructional design principles (DSG-C2) and mechanics that align with effective learning outcomes (PR5). The game must also maintain player engagement (DEG-C1) while meeting technical standards for performance and interactivity (DEG-C3).
- DEV-G8: deploy and distribute the game for user access and play—ensure that the final game deployment supports seamless access (DEG-C3), providing a smooth user experience (DEG-C1) and aligning with learning objectives (DSG-C1 and DSG-C5). After launch, continue gathering feedback for further iterations (DEG-C5) to maintain engagement and foster learning (PR2 and PR4).

Overall, the purpose of the development phase in serious game design is to effectively translate the learning theories and instructional strategies from the conceptual stages into a functional, playable experience that meaningfully impacts the target learners. It is when the foundations in educational content and gameplay principles finally converge into integrated software centering on the knowledge and skill acquisition goals that propelled this endeavor. By prioritizing learner needs through iterative testing and refinement, serious game developers fulfill the promise of leveraging engagement to achieve substantive growth and performance gains—the core vision that galvanized the exploration, design, and now development workflow. The end goal remains focused on players completing the experience with measurable, demonstrable educational gains at scale.

Assessment Phase

The assessment phase is the fourth phase in this visualization of the design space for serious games with the objective to show evidence on whether the game is achieving the pedagogical goals that justified its development in the first place. Compared to entertainment games where success is measured purely by player enjoyment and engagement, serious games necessitate more complex evaluation based on learning principles.

The assessment phase is informed by commonalities and principles such as catering the game experience to player-learners' motivations for long-term satisfaction, effective learning outcome, player engagement and positive behaviors. This highlights the need to ensure the final product aligns with various player types and objectives. The game's mechanics, narrative, educational content, and user experience are measured against the initial design goals, ensuring that both entertainment value and learning outcomes are achieved.

Both quantitative and qualitative data gathered from embedded assessments, gameplay analytics, standardized tests, interviews, surveys, and other instruments should tie back to the original learning objectives, content integration blueprints, and target competencies identified during the exploration phase. Statistical analysis combined with input from expert educators helps determine alignment and efficacy.

Assessment is not a one-time event, either. The promise of serious games lies in their capacity to be continually refined and adapted at scale based on learner performance patterns. So systematic assessment should feed back into development iterations that improve instructional strategy integrations for existing players while also informing subsequent game designs.

In the assessment phase, the key elements are measuring the quality of the game and testing the prototype for effectiveness and usability by the target users. For example, GAMED methodology focuses on evaluating the game's quality by indicators such as acceptability, clarity, effectiveness, engageability, enjoyability, interactivity, localizability, rewardability, simplicity, transformativeness, and usability. The six-dimensional framework examines the various components of mobile game–based learning to provide learners with a mobile

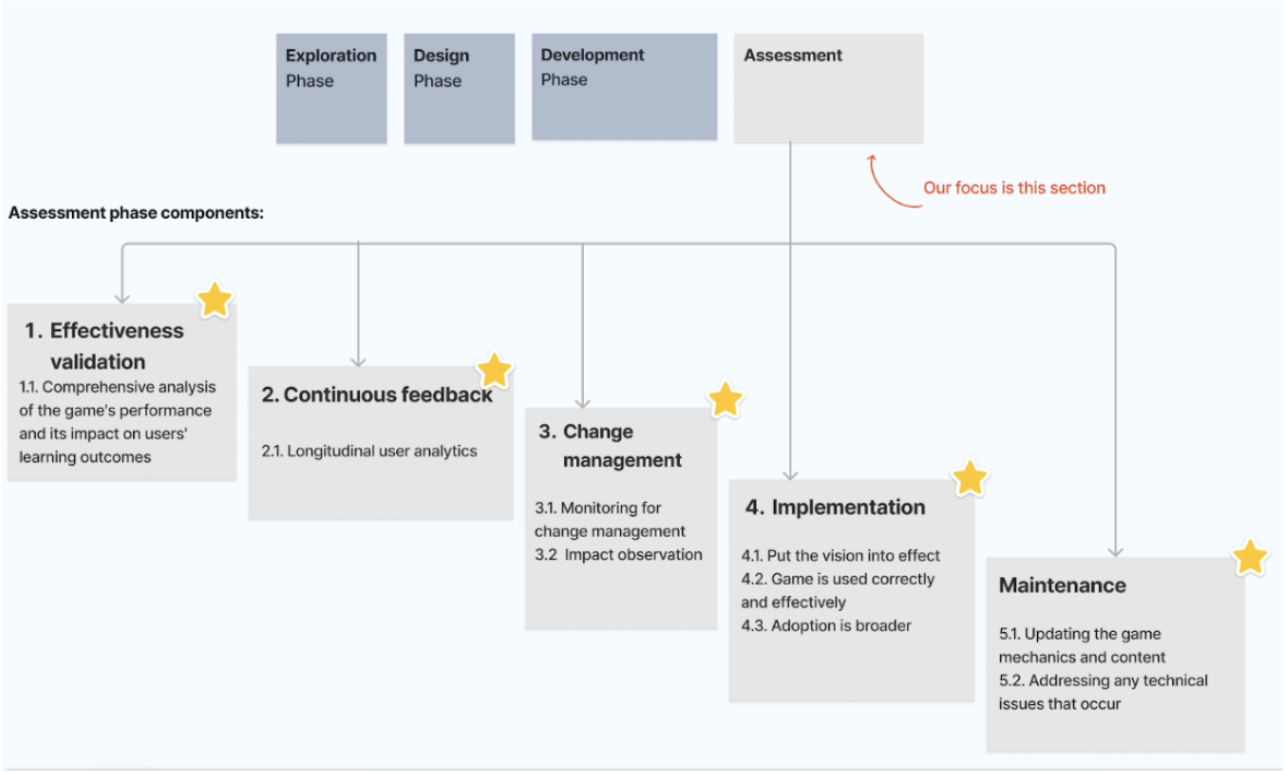
XSL•FO
RenderX

or ubiquitous learning experience. The DGBL-ID model emphasizes testing the prototype and real game for effectiveness and usability by the target users.

This phase is required in assessing the effectiveness of the serious game in achieving its intended learning goals and identifying areas for improvement. Even in this phase, AI can be supportive by analyzing large volumes of player data generated during gameplay to extract valuable insights. By examining player behavior, performance patterns, and learning outcomes, AI algorithms can provide valuable feedback to game designers and educators. These insights can inform the iterative design process, allowing for continuous improvement and optimization of serious games [110].

The components of the assessment phase, as shown in Figure 7, outline the steps necessary to achieve the general goals and objectives initially established. The effectiveness validation is the first component of this phase, where designers conduct a comprehensive analysis of the game's performance and its impact on users' learning outcomes. The continuous user feedback collection is another vital component, which involves gathering feedback from users through various channels to identify areas for improvement and enhance the user experience. The third component is monitoring for change management, where designers keep track of the game's performance and the broader context to adapt the game's design to better meet changing user needs. Implementation is the fourth component, where designers ensure that the game is used correctly and effectively by providing training to users or incorporating it into a broader training program. Finally, maintenance is the fifth component, which involves regularly updating the game's content, fixing bugs, and addressing any technical issues to ensure that it remains effective in achieving its intended goals and continues to provide a positive user experience.

Figure 7. Assessment phase.



During the assessment phase, designers should consider the guidelines alongside the specified components. These guidelines (ASM-G1-G6) are mapped to the relevant principles, or commonalities (eg, DEG-C1-7, DSG-C1-7, and PR1-6) from the previous sections (Textbox 9).

Overall, the assessment phase should use empirical data from a pedagogical lens to validate and continuously enhance learning outcomes. In essence, the validation needed for serious games to proliferate relies on demonstrating tangible learning gains.

While assessment is codified as a key overarching phase, evaluation and user feedback mechanisms should be embedded across the design and development cycle. Effective serious game design necessitates agile, continual review of educational effectiveness beyond entertainment engagement. Frequent playtesting research and qualitative user input must directly shape iterative refinement of instructional content and game mechanics.

XSL•FO
RenderX

Textbox 9. Assessment guidelines (ASM-G) G1-G6.

- ASM-G1: achieve general goals and objectives initially set up—align the game design with player experience goals (DEG-C1) and learning objectives (DSG-C1), ensuring the gameplay mechanics challenge players appropriately (DEG-C2 and PR1). By focusing on effective learning outcomes (PR5), the game should support both entertainment and educational objectives.
- ASM-G2: validate effectiveness through comprehensive analysis—leverage gameplay metrics to assess the interconnectivity of components (DEG-C3) and ensure immersive experiences (PR3) that align with instructional design principles (DSG-C2). Conduct analysis using embedded assessments (DSG) and data mining to validate both engagement and learning outcomes (PR4 and PR5).
- ASM-G3: collect continuous user feedback for improvement—use playtesting (DEG-C4) to gather qualitative feedback on player experience (DEG-C1) and gameplay mechanics (DEG-C2). Ensure continuous iterations based on player feedback (DEG-C5) to improve motivation and learning engagement (PR4). Collect and refine feedback to align with educational content (DSG-C1 and PR2).
- ASM-G4: monitor for change management and adapt design accordingly—ensure adaptive algorithms align with player skill levels (PR1), and adjust immersive elements (DEG-C3, PR3) to maintain engagement. Refine the game mechanics and pedagogical strategies (DSG-C5 and PR5) based on continuous monitoring and feedback from players (PR4).
- ASM-G5: ensure correct and effective use through implementation—develop analytics dashboards to track player progression and learning outcomes (DSG-C5 and PR5). Align game mechanics with cognitive aspects (DSG-C6) to ensure that gameplay not only engages players (DEG-C1) but also reinforces desired behaviors (PR6) through positive feedback mechanisms.
- ASM-G6: maintain games through regular updates, bug fixes, and addressing technical issues—regularly update game mechanics (DEG-C2) and immersive elements (PR3) to align with advancing technology and changing player preferences. Ensure the continued collection of feedback (DEG-C4) and maintain player engagement through iterative improvements (DEG-C5 and PR4).

Assessment seems like a final phase but it is strongly related to previous phases of design (usability testing) and development (quality testing), as a continuous goal during this user-centered design process flow. During the initial exploration phase, rapid crowdsourced needs validation surveys help establish user criteria for the subsequent evaluation stages. Usability testing in the design phase prioritizes educational obstacles alongside entertainment engagement early on. Quality assurance procedures throughout the development gather user perspectives on improving learning outcomes alongside gameplay.

Ongoing analysis of interaction telemetry, knowledge assessments, qualitative feedback, and field pedagogy insights provides continuous signals to refine serious game effectiveness. By embedding ongoing reciprocal user input opportunities, the design space phases foster natural critical review cycles rather than one-way static reporting.

Ultimately, quality user research interactions allow designers and developers to dynamically adapt game elements on the fly to better achieve target learning goals. Assessment interwoven across phases provides that tight, iterative development feedback loop for optimization.

Detected Gaps in DSGDFWs

Serious game development frameworks have made significant strides in addressing various aspects of the field. However, there remain gaps that could be addressed to further enhance the effectiveness of serious game development. One such gap is the consideration of cultural diversity. Many existing frameworks may not fully take into account the impact of cultural differences in learning styles or preferences. Some researchers highlight the importance of cultural differences in game performance with pointers that might inspire how this gap can be reduced. Future frameworks could explore ways to incorporate cultural diversity considerations into serious game design to ensure that games are inclusive and appealing to a broader range of learners [111].

Another area that requires more attention in serious game development frameworks is ethical considerations. With the increasing use of serious games in educational settings, it is crucial to address issues such as data privacy, user consent, and potential biases. Ethical considerations related to serious game development, including the responsible handling of user data and ensuring unbiased representation of diverse groups, should be explicitly addressed in frameworks to ensure the ethical use of serious games in educational contexts. One pointer for this gap would be for designers to ensure that ethical considerations are at the forefront of their decision-making process [112].

Sustainability is another gap in many serious game development frameworks. While current frameworks often focus on the development of individual games, there is limited consideration of the long-term impact of serious games and their sustainability over time. DSG developers could incorporate approaches for creating serious games that have a longer life span, can be updated or scaled, and remain relevant and effective in the face of technological advancements and changing educational needs.

Interdisciplinary collaboration is recognized as an influential aspect of serious game development, but existing frameworks may not provide detailed guidance on how to facilitate this effectively. Collaboration between different stakeholders, such as game designers, educators, and researchers, is essential for creating high-quality serious games. Future frameworks could include more explicit guidance on fostering interdisciplinary collaboration and communication, including strategies for effective team dynamics and communication among diverse stakeholders throughout the serious game development process.

In addition, some serious game development frameworks may not explicitly incorporate empathic design principles, even though they emphasize user-centered design. Empathic design, which involves understanding the user's needs, emotions, and experiences, is crucial in creating meaningful serious games. Addressing the gap of empathic design principles in serious game design involves prioritizing the understanding of lecturers and players' emotions, needs, and experiences. Designers could

XSL•FO
RenderX

explicitly integrate empathic design principles, recognizing the importance of designing serious games that truly resonate with the needs and experiences of the lecturers-players-learners [113].

Overall, while DSGDFW have made significant progress, there remain gaps that could be addressed to further enhance the field. Consideration of cultural diversity, ethical considerations, sustainability, interdisciplinary collaboration, and empathic design are areas that could benefit from more explicit inclusion in serious game development frameworks, ensuring that serious games are effective, inclusive, and ethically responsible educational tools.

Conclusions

This study explored DEGs and DSGs with the aim of identifying the approaches used in the frameworks, their commonalities and gaps, as well as the fundamental design elements and phases that can inform better design strategies for future DSGs. This scoping review has identified key findings that can inform the future design and development of serious games. By leveraging the design frameworks used in both DEGs and DSGs, designers and developers can create effective games that meet their intended outcomes and expectations of learners. DEGs design elements such as adaptive challenge levels based on player's skill level or performance [51,83], cultivation of continuous curiosity through exploration, provision of clear game objectives for confusion reduction and motivation increase [79], creation of deep engaging experience shared in collaboration and competition between players [76], and immediate reward tailored to player achievements [83] are inspirational for serious game design.

This review of established entertainment and serious game design frameworks discerned common phases in development processes, alongside recurring interdependencies between engagement models and instructional components. Synthesizing framework research revealed the design space visualization in the 4 design phases outlined as one of the outcomes in this paper—exploration, design, development, and assessment—as a baseline for DSG development, with each phase playing an essential role in the overall success of the game. The exploration phase involves defining the main idea, intended audience, learning goals, and identifying user profiles and risks. This phase sets the foundation for the entire serious game development process and allows for informed decision-making and strategic planning. The design phase focuses on designing the game environment, mechanics, scenarios, objects, and architecture, and capturing and defining the framework of the game process. Well-designed game elements are essential in creating an engaging, immersive, and enjoyable experience for players. The development phase involves developing the prototype of the game using programming software and graphic editing tools, ensuring that the game functions as intended and meets the design requirements. Finally, the assessment phase involves measuring the quality of the game and testing the prototype for effectiveness and usability by the target users. The resulting visualization aims to give clarity, so developers can efficiently craft serious games across domains based on research-driven best practices. This design process underscores that the lines separating games and learning are far more connected at their core.

The contribution of this review that other papers have not addressed in serious game design frameworks lies in several aspects. First, it explores the dependencies between DEGDFWs and DSGDFWs, providing insights into the approaches applied in both frameworks and identifying design elements that can inform future better design strategies for serious games. This broader perspective allows designers and developers to leverage the design frameworks used in entertainment games to create more effective serious games that meet their intended outcomes and learners' expectations. Second, the study proposes 4 key design phases (exploration, design, development, and assessment) that serve as a baseline for serious game development, providing a structured foundation for decision-making and strategic planning and execution. These phases address the need for a holistic approach to serious game design and emphasize the importance of considering various aspects throughout the end-to-end buildout process.

The guiding design heuristics derived from the principles of dependency analysis can be easily mapped to the generic 4-design-phase flow. The exploration phase should focus on assessing learning needs, player preferences, and target outcomes grounded in principles of enjoyment, engagement, and behavioral change. The design phase aims to translate this knowledge into an experiential vision, seamlessly integrating education and immersion through balanced challenges and incentivized mastery. The development phase uses continuous feedback and analytical tracking to iterate real-time adaptations, supporting personalized progress with positive reinforcement. Finally, the assessment phase prioritizes measurable learning impacts verified empirically while gathering player perspectives to improve engagement quality by recalibrating enjoyment and difficulty. Across all stages, identified principles guide purposeful experiential configurations, inform tradeoffs, set data collection priorities, direct prototyping efforts, and shape evaluation standards for optimizing serious gameplay efficacy.

Finally, the study highlights the significance of some gaps such as empathic design approach in serious game design, recognizing the potential of understanding users' needs and expectations to create more engaging and effective games. Empathic design thinking can revolutionize serious game design in fields such as general education and health care training, resulting in games that are tailored to the unique needs, preferences, and motivations of the target users, which are lecturers and learners.

In summary, the expansive literature scoping, multimodal analysis, theoretical model contribution, actionable guidelines for practice, and overarching conceptual connections orient this paper to deliver unique significance advancing the field. The serious game design frameworks have shown promise in a variety of fields, including education, health care, and training. Overall, although this study makes a valuable contribution to the field of DSGs by proposing a well-defined design foundation, continuous research is necessary to fully realize the potential of creating serious games that effectively address the evolving needs of future generations.

XSL•FO
RenderX

Future Work

There are several future trends in serious game design frameworks that might emerge from recent research. One trend is the use of data analytics and AI to personalize serious game experiences for individual users, based on their preferences, abilities, and health status. This can improve the effectiveness of serious games by tailoring them to the needs of each user. Finally, there is a need for application of empathic design thinking approach for a greater collaboration between game designers and end users (lecturers and learners) that serious games are designed and evaluated in a way that is scientifically rigorous and meets their needs.

In future research, we might explore the potential of AI mixed with empathic design thinking approaches for serious game design and development. For example, an AI-based game design framework that uses facial recognition and machine learning to personalize the game content and difficulty level based on the player's emotions and cognitive effort.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 checklist.
[DOCX File , 23 KB - games_v13i1e54075_app1.docx]

Multimedia Appendix 2
Full search strategy for included databases.
[PDF File (Adobe PDF File), 59 KB - games_v13i1e54075_app2.pdf]

Multimedia Appendix 3
Purposes of digital serious game design frameworks.
[PNG File , 289 KB - games_v13i1e54075_app3.png]

Multimedia Appendix 4
Results with papers with digital entertainment game design frameworks (n=16).
[PDF File (Adobe PDF File), 140 KB - games_v13i1e54075_app4.pdf]

Multimedia Appendix 5
Results with papers with digital serious game design frameworks (n=31).
[PDF File (Adobe PDF File), 183 KB - games_v13i1e54075_app5.pdf]

References

1. Djaouti D, Alvarez J, Jessel JP, Rampnoux O. Origins of serious games. In: Ma M, Oikonomou A, Jain LC, editors. Serious Games and Edutainment Applications. Cham, Switzerland: Springer; 2011:25-43.
2. Granic I, Lobel A, Engels RC. The benefits of playing video games. Am Psychol 2014 Jan;69(1):66-78 [FREE Full text] [doi: 10.1037/a0034857] [Medline: 24295515]
3. Mueller FF, Gibbs MR, Vetere F. Design influence on social play in distributed exertion games. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2009 Presented at: CHI '09; April 4-9, 2009; Boston, MA p. 4-9 URL: https://dl.acm.org/doi/10.1145/1518701.1518938 [doi: 10.1145/1518701.1518938]
4. Zyda M. From visual simulation to virtual reality to games. Computer 2005 Sep;38(9):25-32. [doi: 10.1109/mc.2005.297]
5. Abt CC. Serious Games. New York, NY: Viking Press; 1970.
6. Heidig S, Kapp F, Narciss S. Serious games: a new domain for pedagogical agents research? 2018 Presented at: Annual Meeting of the American Educational Research Association; Apr 13-17, 2018; New York, NY p. 25.
7. Baptista G, Oliveira T. Gamification and serious games: a literature meta-analysis and integrative model. Comput Human Behav 2019 Mar;92:306-315. [doi: 10.1016/j.chb.2018.11.030]
8. Chun Tie Y, Birks M, Francis K. Grounded theory research: a design framework for novice researchers. SAGE Open Med 2019 Jan 02;7:2050312118822927 [FREE Full text] [doi: 10.1177/2050312118822927] [Medline: 30637106]
9. Churchill R, Ferguson P, Godinho S, Johnson N, Keddie A, Letts W, et al. Teaching: Making a Difference. 6th edition. Hoboken, NJ: John Wiley & Sons; 2011.
10. Connolly TM, Boyle EA, MacArthur E, Hainey T, Boyle JM. A systematic literature review of empirical evidence on computer games and serious games. Compute Educ 2012 Sep;59(2):661-686. [doi: 10.1016/j.compedu.2012.03.004]

XSL•FO
RenderX

11. Vugts MA, Zedlitz AM, Joosen MC, Vrijhoef HJ. Serious gaming during multidisciplinary rehabilitation for patients with chronic pain or fatigue symptoms: mixed methods design of a realist process evaluation. J Med Internet Res 2020 Mar 09;22(3):e14766 [FREE Full text] [doi: 10.2196/14766] [Medline: 32149720]
12. Gurbuz SC, Celik M. Serious games in future skills development: a systematic review of the design approaches. Comput Appl Eng Educ 2022 Aug 08;30(5):1591-1612. [doi: 10.1002/cae.22557]
13. Deterding S, Dixon D, Khaled R, Nacke L. From game design elements to gamefulness: defining gamification. In: Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments. 2011 Presented at: MindTrek '11; September 28-30, 2011; Tampere, Finland p. 9-15 URL: https://dl.acm.org/doi/10.1145/2181037.2181040 [doi: 10.1145/2181037.2181040]
14. Williams D. Utilising game design to create engaging education: a framework for Gameful learning. In: Proceedings of the 2019 ACM Conference on Innovation and Technology in Computer Science Education. 2019 Presented at: ITiCSE '19; July 15-17, 2019; Aberdeen, UK p. 351-352 URL: https://dl.acm.org/doi/10.1145/3304221.3325594 [doi: 10.1145/3304221.3325594]
15. Bahreini K, Nadolski R, Westera W. Communication skills training exploiting multimodal emotion recognition. Interact Learn Environ 2016 Oct 27;25(8):1065-1082. [doi: 10.1080/10494820.2016.1247286]
16. Cabellos B, Sánchez DL, Pozo JI. Do future teachers believe that video games help learning? Tech Know Learn 2021 Dec 16;28(2):803-821. [doi: 10.1007/S10758-021-09586-3]
17. Kumar L, Herger M. Serious games vs entertainment games. In: Herger M, editor. Gamification in Business. Cham, Switzerland: Springer; 2021:125-139.
18. Alvarez J, Damien D. An introduction to serious games - definitions and concepts. In: Proceedings of the 2011 International Conference on Serious Games & Simulation Workshop. 2011 Presented at: SGS '11; October 10-15, 2011; Paris, France p. 10-15 URL: https://www.ludoscience.com/files/ressources/Proceedings-SGS-Wkshp-2011-ind-0.pdf
19. Westera W. The devil's advocate: identifying persistent problems in serious game design. Int J Serious Games 2022 Sep 09;9(3):115-124. [doi: 10.17083/ijsg.v9i3.547]
20. Altamirano A, Ortega B, Loranca A, Olmos L. Enredados: an analogue-digital board game to rethink gender violence in social networks. In: Proceedings of the IX Latin American Conference on Human Computer Interaction. 2019 Presented at: CLIHC '19; September 30-October 4, 2019; Panama City, PA p. 1-6 URL: https://dl.acm.org/doi/abs/10.1145/3358961.3358994 [doi: 10.1145/3358961.3358994]
21. Ibrahim R, Jaafar A. Educational games (EG) design framework: combination of game design, pedagogy and content modeling. In: Proceedings of the 2009 International Conference on Electrical Engineering and Informatics. 2009 Presented at: ICEEI '09; August 5-7, 2009; Bangi, Malaysia p. 293-298 URL: https://ieeexplore.ieee.org/abstract/document/5254771 [doi: 10.1109/iceei.2009.5254771]
22. Carvalho MB, Bellotti F, Berta R, De Gloria A, Sedano CI, Hauge JB, et al. An activity theory-based model for serious games analysis and conceptual design. Comput Educ 2015 Sep;87:166-181 [FREE Full text] [doi: 10.1016/j.compedu.2015.03.023]
23. Barakat NH. A framework for integrating software design patterns with game design framework. In: Proceedings of the 8th International Conference on Software and Information Engineering. 2019 Presented at: ICSIE '19; April 9-12, 2019; Cairo, Egypt p. 47-50 URL: https://dl.acm.org/doi/10.1145/3328833.3328871 [doi: 10.1145/3328833.3328871]
24. Saveski GL, Westera W, Yuan L, Hollins P, Fernández MB, Moreno GP, et al. What serious game studios want from ICT research: identifying developers' needs. In: Proceedings of the 4th International Conference on Games and Learning Alliance. 2015 Presented at: GALA '15; December 9-11, 2015; Rome, Italy p. 32-41 URL: https://dl.acm.org/doi/10.1007/978-3-319-40216-1_4 [doi: 10.1007/978-3-319-40216-1_4]
25. Nima U, Wangdi R, Hauge JB. A serious game for competence development in internet of things and knowledge sharing. In: Proceedings of the 2018 IEEE International Conference on Industrial Engineering and Engineering Management. 2018 Presented at: IEEM '18; December 16-19, 2018; Bangkok, Thailand p. 1786-1790 URL: https://ieeexplore.ieee.org/document/8607737 [doi: 10.1109/ieem.2018.8607737]
26. Plass JL, Homer BD, Mayer RE, Kinzer CK. Theoretical foundations of game-based and playful learning. In: Plass JL, Mayer RE, Homer BD, editors. Handbook of Game-Based Learning. New York, NY: MIT Press; Feb 06, 2016:258-283.
27. Xu Y, Tong M, Ming W, Lin Y, Mai W, Huang W, et al. A depth camera-based, task-specific virtual reality rehabilitation game for patients with stroke: pilot usability study. JMIR Serious Games 2021 Mar 24;9(1):e20916 [FREE Full text] [doi: 10.2196/20916] [Medline: 33759795]
28. Hays R. The effectiveness of instructional games: a literature review and discussion. Acad Med 2005;80(7):14-27. [doi: 10.21236/ada441935]
29. Lim H, Chun S, Soung M, Kim J, Kim S. Antimicrobial efficacy of granulysin-derived synthetic peptides in acne vulgaris. ResearchGate. 2015 Jul. URL: https://www.researchgate.net/publication/263843445_Serious_Games_in_Business_Field_the_Challenges_and_the_Perspectives [accessed 2024-04-29]
30. Pflanzl N, Classe T, Araujo R, Vossen G. Designing serious games for citizen engagement in public service processes. In: Cruz-Cunha MM, editor. Handbook of Research on Entrepreneurial Development and Innovation Within Smart Cities. Hershey, PA: IGI Global; 2017:266-285.

XSL•FO
RenderX

31. Pereira P, Duarte E, Rebelo F, Noriega P. A review of gamification for health-related contexts. In: Marcus A, editor. Design, User Experience, and Usability: User Experience Design for Diverse Interaction Platforms and Environments. Cham, Switzerland: Springer; 2014:742-753.
32. Thomas TH, Sivakumar V, Babichenko D, Grieve VL, Klem ML. Mapping behavioral health serious game interventions for adults with chronic illness: scoping review. JMIR Serious Games 2020 Jul 30;8(3):e18687 [FREE Full text] [doi: 10.2196/18687] [Medline: 32729836]
33. Abraham O, LeMay S, Bittner S, Thakur T, Stafford H, Brown R. Investigating serious games that incorporate medication use for patients: systematic literature review. JMIR Serious Games 2020 Apr 29;8(2):e16096 [FREE Full text] [doi: 10.2196/16096] [Medline: 32347811]
34. Perrin J, Meeus A, Broseus J, Morieux P, Di Ceglie V, Gravoulet J, et al. A serious game about hematology for health care workers (SUPER HEMO): development and validation study. JMIR Serious Games 2023 Feb 13;11:e40350 [FREE Full text] [doi: 10.2196/40350] [Medline: 36780215]
35. VanLehn K. The behavior of tutoring systems. Int J Artif Intell Educ 2006;16:227-265. [doi: 10.1016/j.artint.2005.09.004]
36. Savery JR. Comparative pedagogical models of problem-based learning. In: Moallem M, Hung W, Dabbagh N, editors. The Wiley Handbook of Problem-Based Learning. Hoboken, NJ: Wiley-Blackwell; 2019:81-83.
37. Lau SY, Agius H. A framework and immersive serious game for mild cognitive impairment. Multimed Tools Appl 2021 May 31;80(20):31183-31237. [doi: 10.1007/s11042-021-11042-4]
38. Ke F, Shute V, Clark KM, Erlebacher G. Interdisciplinary Design of Game-based Learning Platforms. Cham, Switzerland: Springer; 2019.
39. Rooney P. A theoretical framework for serious game design: exploring pedagogy, play and fidelity and their implications for the design process. Int J Game Based Learn 2012;2(4):41-60. [doi: 10.4018/ijgbl.2012100103]
40. O'Shea Z, Bartle R, Pan X, Freeman J. Apples and oranges: a study of ‘tend and befriend’ as a phenomenon in digital games. In: Proceedings of the 4th International Conference, HCI-Games 2022, Held as Part of the 24th HCI International Conference. 2022 Presented at: HCII '22; June 26-July 1, 2022; Virtual Event p. 269-288 URL: https://link.springer.com/chapter/10.1007/978-3-031-05637-6_17 [doi: 10.1007/978-3-031-05637-6_17]
41. Ávila-Pesántez DF, Rivera LA, Alban MS. Approaches for serious game design: a systematic literature review. Comput Educ J 2017;8(3):5-14 [FREE Full text]
42. Junior R, Silva F. Redefining the MDA framework—the pursuit of a game design ontology. Information 2021 Sep 26;12(10):395. [doi: 10.3390/info12100395]
43. Bartle RA. Hearts, clubs, diamonds, spades: players who suit MUDs. J Virtual Environ 1996;1:1-16 [FREE Full text]
44. Wang H, Sun CT. Game reward systems: gaming experiences and social meanings. Proceedings of DiGRA Conference. 2011. URL: https://dl.digra.org/index.php/dl/article/view/594/594 [accessed 2025-02-26]
45. Wouters P, van Nimwegen C, van Oostendorp H, van der Spek ED. A meta-analysis of the cognitive and motivational effects of serious games. J Educ Phycol 2013 May;105(2):249-265. [doi: 10.1037/a0031311]
46. Yamoul S, Ouchaouka L, Radid M, Moussetad M. Implementing a serious game as a learner motivation tool. Procedia Comput Sci 2022;10(4):351-357 [FREE Full text] [doi: 10.2196/SG.12345]
47. Almeida FL. Adoption of a serious game in the developing of emotional intelligence skills. Eur J Investig Health Psychol Educ 2019 Jul 20;10(1):30-43 [FREE Full text] [doi: 10.3390/ejihpe10010004] [Medline: 34542467]
48. Verschueren S, Buffel C, Vander Stichele G. Developing theory-driven, evidence-based serious games for health: framework based on research community insights. JMIR Serious Games 2019 May 02;7(2):e11565 [FREE Full text] [doi: 10.2196/11565] [Medline: 31045496]
49. Bunt L, Greeff J, Taylor E. Enhancing serious game design: expert-reviewed, stakeholder-centered framework. JMIR Serious Games 2024 May 31;12:e48099 [FREE Full text] [doi: 10.2196/48099] [Medline: 38820585]
50. Jaccard D, Suppan L, Sanchez E, Huguenin A, Laurent M. The co.LAB generic framework for collaborative design of serious games: development study. JMIR Serious Games 2021 Jul 02;9(3):e28674 [FREE Full text] [doi: 10.2196/28674] [Medline: 34255655]
51. O'Shea Z, Freeman J. Game design frameworks: where do we start? In: Proceedings of the 14th International Conference on the Foundations of Digital Games. 2019 Presented at: FDG '19; August 26-30, 2019; San Luis Obispo, CA p. 1-10 URL: https://dl.acm.org/doi/10.1145/3337722.3337753 [doi: 10.1145/3337722.3337753]
52. Lazzaro N. Why we play games: four keys to more emotion without story. XEODesign. 2004. URL: https://gamemodworkshop.com/readings/xeodesign_whyweplaygames.pdf [accessed 2024-04-29]
53. Zimmerman E. The engines of play. In: Salen K, Zimmerman E, editors. The Game Design Reader: A Rules of Play Anthology. Cambridge, MA: MIT Press; 2006:633-645.
54. Bartle R. Designing Virtual Worlds. Berkeley, CA: New Riders; 2004.
55. Hunicke R, LeBlanc M, Zubek R. MDA: A formal approach to game design and game research. In: Proceedings of the 2004 AAAI Workshop on Challenges in Game AI. Washington, DC: Association for the Advancement of Artificial Intelligence; 2004.
56. Bond JG. Introduction to Game Design, Prototyping, and Development: From Concept to Playable Game with Unity and C#. New York, NY: Pearson Education; 2014.

XSL•FO
RenderX

57. Lazzaro N. Why we play games: four keys to more emotion in player experiences. In: 2004 Game Developers Conference Proceedings. San Jose, CA: Game Developers Conference; 2004:1-8.
58. Andreoli R, Corolla A, Faggiano A, Malandrino D, Pirozzi D, Ranaldi M, et al. A framework to design, develop, and evaluate immersive and collaborative serious games in cultural heritage. J Comput Cult Herit 2017 Dec 07;11(1):1-22. [doi: 10.1145/3064644]
59. Giannakas F, Kambourakis G, Papasalouros A, Gritzalis S. A critical review of 13 years of mobile game-based learning. Educ Tech Res Dev 2017 Dec 12;66(2):341-384. [doi: 10.1007/s11423-017-9552-z]
60. Miljanovic MA. Enhancing computer science education with adaptive serious games. In: Proceedings of the 2019 ACM Conference on International Computing Education Research. 2019 Presented at: ICER '19; August 12-14, 2019; Toronto, ON p. 194-202 URL: https://dl.acm.org/doi/10.1145/3291279.3339438 [doi: 10.1145/3291279.3339438]
61. Lee LK, Wei X, Chui KT, Cheung SK, Wang FL, Fung YC, et al. A systematic review of the design of serious games for innovative learning: augmented reality, virtual reality, or mixed reality? Electronics 2024 Feb 26;13(5):890. [doi: 10.3390/electronics13050890]
62. Dörner R, Göbel S, Effelsberg W, Wiemeyer J. Serious Games: Foundations, Concepts and Practice. Cham, Switzerland: Springer; 2016.
63. Kniestedt I, Gómez Maureira MA, Lefter I, Lukosch S, Brazier FM. Dive deeper: empirical analysis of game mechanics and perceived value in serious games. Proc ACM Hum Comput Interact 2021 Oct 06;5:1-25. [doi: 10.1145/3474663]
64. Maxim IR, Arnedo-Moreno J. Assessing the usefulness of empathic design in mobile game-based learning (mGBL). In: Proceedings of the 9th International Conference on Technological Ecosystems for Enhancing Multiculturality. 2021 Presented at: TEEM '21; October 26-29, 2021; Barcelona, Spain p. 157-161 URL: https://dl.acm.org/doi/10.1145/3486011.3486438 [doi: 10.1145/3486011.3486438]
65. Ritterfeld U, Cody MJ, Vorderer P. Serious Games: Mechanisms and Effects. Abingdon, UK: Routledge; 2009.
66. Koedinger KR, Brunskill E, Baker RS, McLaughlin EA, Stamper J. New potentials for data‐driven intelligent tutoring system development and optimization. AI Mag 2013 Sep;34(3):27-41. [doi: 10.1609/aimag.v34i3.2484]
67. Lee JJ, Hammer J. Gamification in education: what, how, why bother? Acad Exch Q 2011:15 [FREE Full text]
68. Barab S, Arid A, Jackson C. Eat your vegetables and do your homework: a design-based investigation of enjoyment and meaning in learning. In: Proceedings of the 2005 International Conference on Education and Technology.: Calgary, AB; 2005 Presented at: ICET '05; July 4-6, 2005; Virtual Event p. 1-15 URL: https://api.semanticscholar.org/CorpusID:56118619
69. Skinner BF. Science and Human Behavior. New York, NY: Macmillan; 1953.
70. Squire K. Video games in education. Int J Intell Simul Gaming 2003 Oct;1(1):49-62. [doi: 10.1145/950566.950583]
71. Rethlefsen ML, Kirtley S, Waffenschmidt S, Ayala AP, Moher D, Page MJ, PRISMA-S Group. PRISMA-S: an extension to the PRISMA statement for reporting literature searches in systematic reviews. Syst Rev 2021 Jan 26;10(1):39 [FREE Full text] [doi: 10.1186/s13643-020-01542-z] [Medline: 33499930]
72. Wohlin C. Guidelines for snowballing in systematic literature studies and a replication in software engineering. In: Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering. 2014 Presented at: EASE '14; May 13-14, 2014; London, UK p. 1-10 URL: https://dl.acm.org/doi/10.1145/2601248.2601268 [doi: 10.1145/2601248.2601268]
73. Wicks D. The coding manual for qualitative researchers (3rd edition). Qual Res Organ Manag 2017 Jun 12;12(2):169-170. [doi: 10.1108/QROM-08-2016-1408]
74. Rau G, Shih YS. Evaluation of Cohen's Kappa and other measures of inter-rater agreement for genre analysis and other nominal data. J Engl Acad Purp 2021 Sep;53:101026. [doi: 10.1016/j.jeap.2021.101026]
75. Godau R. Qualitative data analysis software: MAXqda. Qual Res J 2004;4(1):66-72 [FREE Full text]
76. Walk W, Görlich D, Barrett M. Design, dynamics, experience (DDE): an advancement of the MDA framework for game design. In: Blatz M, Korn O, Amato A, Walk W, editors. Game Dynamics: Best Practices in Procedural and Dynamic Game Content Generation. Cham, Switzerland: Springer; 2017:27-45.
77. Schell J. The Art of Game Design: A Book of Lenses. New York, NY: Morgan Kaufmann; 2008.
78. Monu K, Ralph P. Beyond gamification: implications of purposeful games for the information systems discipline. arXiv Preprint posted online August 2, 2013 [FREE Full text]
79. Mora-Zamora R, Brenes-Villalobos E. Integrated framework for game design. In: Proceedings of the IX Latin American Conference on Human Computer Interaction. 2019 Presented at: CLIHC '19; September 30-October 4, 2019; Panama City, PA p. 1-6 URL: https://dl.acm.org/doi/10.1145/3358961.3358984 [doi: 10.1145/3358961.3358984]
80. Costikyan G. I have no words and I must design: toward a critical vocabulary for games. In: Proceedings of the 2002 Conference on Computer Games and Digital Cultures Conference. 2002 Presented at: CGDC '02; June 6-8, 2002; Tampere, Finland p. 1-25 URL: https://dl.digra.org/index.php/dl/article/view/29/29 [doi: 10.4324/9781003140184-10]
81. Björk S, Zagal JP. Game design and role-playing games. In: Zagal JP, Deterding S, editors. Role-Playing Game Studies: Transmedia Foundations. New York, NY: Routledge; 2018:323-336.
82. Fullerton R. Game Design Workshop: A Playcentric Approach to Creating Innovative Games. Boca Raton, FL: CRC Press; 2018.

XSL•FO
RenderX

83. Ferro LS. The game element and mechanic (GEM) framework: a structural approach for implementing game elements and mechanics into game experiences. Entertain Comput 2021 Jan;36:100375. [doi: 10.1016/j.entcom.2020.100375]
84. King D. Integrating serious and casual game design approaches: a framework for activist-casual game design. In: Proceedings of the 16th International Conference on the Foundations of Digital Games. 2021 Presented at: FDG '21; August 3-6, 2021; Montreal, QC p. 1-9 URL: https://dl.acm.org/doi/10.1145/3472538.3472568 [doi: 10.1145/3472538.3472568]
85. Winn BM. The design, play, and experience framework. In: Winn BM, editor. Handbook of Research on Effective Electronic Gaming in Education. New York, NY: IGI Global; 2008:1010-1024.
86. Yusoff A, Crowder R, Gilbert L, Wills G. A conceptual framework for serious games. In: Proceedings of the 9th IEEE International Conference on Advanced Learning Technologies. 2009 Presented at: ICALT '09; July 15-17, 2009; Riga, Latvia p. 15-17 URL: https://ieeexplore.ieee.org/document/5194153 [doi: 10.1109/icalt.2009.19]
87. De FS, Jarvis S. A framework for developing serious games to meet learner needs. In: Proceedings of the 2006 International Conference on Interservice/Industry Training, Simulation & Education. 2006 Presented at: I/ITSEC '06; December 4-7, 2006; Orlando, FL p. 5-27 URL: https://researchportal.murdoch.edu.au/esploro/outputs/conferencePaper/A-framework-for-developing-serious-games/991005544482907891
88. Harteveld C. Triadic game evaluation: a framework for assessing games with a serious purpose. In: Proceedings of the 2010 Design and Engineering of Game-like Virtual and Multimodal Environments Workshop. 2010 Presented at: DEGVME '10; June 20, 2010; Berlin, Germany p. 5-11 URL: https://www.researchgate.net/publication/319449032_Triadic_Game_Evaluation_A_Framework_for_Assessing_Games_with_a_Serious_Purpose
89. Gunter GA, Kenny RF, Vick EH. Taking educational games seriously: using the RETAIN model to design endogenous fantasy into standalone educational games. Educ Tech Res Dev 2007 Oct 16;56(5-6):511-537. [doi: 10.1007/S11423-007-9073-2]
90. Jiang X, Harteveld C, Huang X, Fung AY. The computational puzzle design framework: a design guide for games teaching computational thinking. In: Proceedings of the 14th International Conference on the Foundations of Digital Games. 2019 Presented at: FDG '19; August 26-30, 2019; San Luis Obispo, CA p. 1-11 URL: https://dl.acm.org/doi/10.1145/3337722.3337768 [doi: 10.1145/3337722.3337768]
91. Annetta LA. The “I's” have it: a framework for serious educational game design. Rev Gen Psychol 2010 Jun 01;14(2):105-113. [doi: 10.1037/a0018985]
92. Golestan S, Mahmoudi-Nejad A, Moradi H. A framework for easier designs: augmented intelligence in serious games for cognitive development. IEEE Consumer Electron Mag 2019 Jan;8(1):19-24. [doi: 10.1109/mce.2018.2867970]
93. Daylamani-Zad D, Angelides MC, Agius H. Lu-Lu: a framework for collaborative decision making games. Decis Support Syst 2016 May;85:49-61. [doi: 10.1016/j.dss.2016.02.011]
94. Proulx JN, Romero M, Arnab S. Learning mechanics and game mechanics under the perspective of self-determination theory to foster motivation in digital game based learning. Simul Gaming 2016 Nov 02;48(1):81-97. [doi: 10.1177/1046878116674399]
95. Mat Zin NA, Jaafar A, Wong SY. Digital game-based learning (DGBL) model and development methodology for teaching history. WSEAS Trans Comput 2009;8:323-333 [FREE Full text] [doi: 10.1163/9789004719118_030]
96. Baker J, Wanick V, Asiri M, Wills G, Ranchhod A. Immersion and narrative design in educational games across cultures. In: Ma M, Oliveira MF, Petersen S, Hauge JB, editors. Serious Games and Edutainment Applications. Volume II. Cham, Switzerland: Springer; 2017:605-621.
97. Cano S, Munoz Arteaga J, Collazos CA, Gonzalez CS, Zapata S. Toward a methodology for serious games design for children with auditory impairments. IEEE Latin Am Trans 2016 May;14(5):2511-2521 [FREE Full text] [doi: 10.1109/tla.2016.7530453]
98. de Lope RP, López Arcos JR, Medina-Medina N, Paderewski P, Gutiérrez-Vela F. Design methodology for educational games based on graphical notations: designing Urano. Entertain Comput 2017 Jan;18:1-14. [doi: 10.1016/j.entcom.2016.08.005]
99. Szczesna A, Tomaszek M, Wieteska A. The methodology of designing serious games for children and adolescents focused on psychological goals. In: Proceedings of the 3rd International Conference on Information Technologies in Biomedicine. 2012 Presented at: ITIB '12; June 11-13, 2012; Gliwice, Poland p. 151-162 URL: https://link.springer.com/chapter/10.1007/978-3-642-31196-3_24 [doi: 10.1007/978-3-642-31196-3_24]
100. Nadolski RJ, Hummel HG, van den Brink HJ, Hoefakker RE, Slootmaker A, Kurvers HJ, et al. EMERGO: a methodology and toolkit for developing serious games in higher education. Simul Gaming 2007 Jul 31;39(3):338-352 [FREE Full text] [doi: 10.1177/1046878108319278]
101. Barbosa AF, Pereira PN, Dias JA, Silva FG. A new methodology of design and development of serious games. Int J Comput Games Technol 2014;2014:1-8. [doi: 10.1155/2014/817167]
102. Tsikinas S, Xinogalos S. Designing effective serious games for people with intellectual disabilities. In: Proceedings of the 2018 IEEE Global Engineering Education Conference. 2018 Presented at: EDUCON '18; April 17-20, 2018; Santa Cruz de Tenerife, Spain p. 1-6 URL: https://ieeexplore.ieee.org/document/8363467 [doi: 10.1109/educon.2018.8363467]
103. Aslan S, Balci O. GAMED: digital educational game development methodology. Simulation 2015 Mar 02;91(4):307-319. [doi: 10.1177/0037549715572673]

XSL•FO
RenderX

104. Moosa AM, Al-Maadeed N, Saleh M, Al-Maadeed SA, Aljaam JM. Designing a mobile serious game for raising awareness of diabetic children. IEEE Access 2020;8:222876-222889. [doi: 10.1109/access.2020.3043840]
105. Wang C, Huang L. A systematic review of serious games for collaborative learning: theoretical framework, game mechanic and efficiency assessment. Int J Emerg Technol Learn 2021 Mar 30;16(06):88. [doi: 10.3991/ijet.v16i06.18495]
106. Avila-Pesantez D, Delgadillo R, Rivera LA. Proposal of a conceptual model for serious games design: a case study in children with learning disabilities. IEEE Access 2019;7:161017-161033. [doi: 10.1109/access.2019.2951380]
107. Lamb RL, Annetta L, Firestone J, Etopio E. A meta-analysis with examination of moderators of student cognition, affect, and learning outcomes while using serious educational games, serious games, and simulations. Comput Human Behav 2018 Mar;80:158-167. [doi: 10.1016/j.chb.2017.10.040]
108. Pouls BP, Bekker CL, van Dulmen S, Vriezekolk JE, van den Bemt BJ. A serious puzzle game to enhance adherence to antirheumatic drugs in patients with rheumatoid arthritis: systematic development using intervention mapping. JMIR Serious Games 2022 Feb 18;10(1):e31570 [FREE Full text] [doi: 10.2196/31570] [Medline: 35179510]
109. Westera W, Prada R, Mascarenhas S, Santos PA, Dias J, Guimarães M, et al. Artificial intelligence moving serious gaming: presenting reusable game AI components. Educ Inf Technol 2019 Jul 30;25(1):351-380. [doi: 10.1007/S10639-019-09968-2]
110. Rafiq A, Abdul Kadir TA, Ihsan SN. A review of artificial intelligence in serious game for public health. J Phys Conf Ser 2021 Apr 01;1830(1):012001. [doi: 10.1088/1742-6596/1830/1/012001]
111. Lukosch H, Kurapati S, Groen D, Verbraeck A. Gender and cultural differences in game-based learning experiences. Electron J E Learn 2017;15:1479-1319.
112. El-Nasr MS, Kleinman E. Data-driven game development: ethical considerations. In: Proceedings of the 15th International Conference on the Foundations of Digital Games. 2020 Presented at: FDG '20; September 15-18, 2020; Bugibba, Malta p. 1-10 URL: https://dl.acm.org/doi/10.1145/3402942.3402964 [doi: 10.1145/3402942.3402964]
113. Maxim RI, Arnedo-Moreno J. Programming games as learning tools: using empathic design principles for engaging experiences. In: Proceedings of the 2024 IEEE Gaming, Entertainment, and Media Conference. 2024 Presented at: GEM '24; June 5-7, 2024; Turin, Italy p. 1-6 URL: https://ieeexplore.ieee.org/document/10585562 [doi: 10.1109/gem61861.2024.10585562]

Abbreviations

AI: artificial intelligence
DEG: digital entertainment game
DEGDFW: digital entertainment game design framework
DGBL-ID: digital game–based learning—instructional design
DSG: digital serious game
DSGDFW: digital serious game design framework
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by T Leung, A Coristine; submitted 28.10.23; peer-reviewed by S Downie, F Adell, E Braad; comments to author 26.02.24; revised version received 30.05.24; accepted 24.12.24; published 05.03.25.

Please cite as:
Maxim RI, Arnedo-Moreno J
Identifying Key Principles and Commonalities in Digital Serious Game Design Frameworks: Scoping Review
JMIR Serious Games 2025;13:e54075
URL: https://games.jmir.org/2025/1/e54075
doi:10.2196/54075
PMID:40053743

©Raluca Ionela Maxim, Joan Arnedo-Moreno. Originally published in JMIR Serious Games (https://games.jmir.org), 05.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Smartwatch-Based Tailored Gamification and User Modeling for Motivating Physical Exercise: Experimental Study With the Maximum Difference Scaling Segmentation Method

Jie Yao[1], PhD; Di Song[1], MA; Tao Xiao[2], PhD; Jiali Zhao[1], BSc

[1]School of Economics and Management, Harbin Institute of Technology (Shenzhen), Shenzhen, China

[2]School of Mathematical Sciences, Shenzhen University, Shenzhen, China

Corresponding Author:
Jie Yao, PhD
School of Economics and Management
Harbin Institute of Technology (Shenzhen)
Nanshan District Xili University Town
Shenzhen, 518055
China
Phone: 86 0755 26033876
Email: yaojiejulie@hit.edu.cn

Abstract

Background: Smartwatch-based gamification holds great promise for enhancing fitness apps and promoting physical exercise; however, empirical evidence on its effectiveness remains inconclusive, partly due to “one-size-fits-all” design approaches that overlook individual differences. While the emerging research area of tailored gamification calls for more accurate user modeling and better customization of game elements, existing studies have relied primarily on rating scale–based measures and correlational analyses with methodological limitations.

Objective: This study aimed to improve smartwatch-based gamification through an innovative user modeling approach to better motivate physical exercise among different user groups with tailored solutions. It incorporated both individual preferences and needs for game elements into the user segmentation process and used the maximum difference scaling (MaxDiff) technique, which can overcome the limitations of traditional methods.

Methods: With data collected from 2 MaxDiff experiments involving 378 smartwatch users and latent class statistical models, the relative power of each of the 16 popular game elements was examined in terms of what users liked and what motivated them to exercise based on which distinct user segments were identified. Prediction models were also proposed for quickly classifying future users into the right segments to provide them with tailored gamification solutions on smartwatch fitness apps.

Results: We identified 3 segments of smartwatch users based on their preferences for gamification. More importantly, we uncovered 4 segments motivated by goals, immersive experiences, rewards, or social comparison. Such user heterogeneity confirmed the susceptibility of the effects of gamification and indicated the necessity of accurately matching gamified solutions with user characteristics to better change health behaviors through different mechanisms for different targets. Important differences were also observed between the 2 sets of user segments (ie, those based on preferences for game elements vs those based on the motivational effects of the elements), indicating the gap between what people enjoy using on smartwatches and what can motivate them for physical exercise engagement.

Conclusions: To our knowledge, this study is the first to investigate MaxDiff-based user segmentation for tailored gamification on smartwatches promoting physical exercise and contributes to a detailed understanding of preferences for, and the effectiveness of, different game elements among different groups of smartwatch users. As existing tailored gamification studies continue to explore ways of user modeling with mostly surveys and questionnaires, this study supported the adoption of MaxDiff experiments as an alternative method to better capture user heterogeneity in the health domain and inform the design of tailored solutions for more application types beyond smartphones.

(JMIR Serious Games 2025;13:e66793) doi:10.2196/66793

XSL•FO
RenderX

KEYWORDS

tailored gamification; user segmentation; maximum difference scaling; MaxDiff; smartwatch; physical exercise

Introduction

Background

The COVID-19 pandemic and its widespread consequences have brought public health issues and related preventive measures into sharper focus. Increasing efforts are being made to educate people on the harms of unhealthy lifestyles, which are closely associated with the leading global risks for mortality and chronic diseases such as high blood glucose and cholesterol levels [1,2]. One of the most prevalent health concerns in modern society is sedentary behavior, whereas adequate physical exercise can produce positive physical and mental effects [3,4]. Although many people are aware of such benefits, motivational techniques and systems are needed to push people to take action [5] and sustain long-term exercise engagement.

Gamification, a form of persuasive design [6] and a useful behavioral change support system [7], has gained increasing popularity in recent years [8] owing to its theoretical bases in motivation research (such as self-determination theory [9] and the behavior model proposed by Fogg [10]) and wide practical applications integrating motivational features such as goal setting and feedback. By incorporating common game design elements into nongame contexts to make activities more game like [11], gamification is believed to hold great promise in the area of health behavior change [12], particularly for promoting physical activity [13]. Despite the increasingly widespread use of gamification on many mobile fitness apps such as Nike+ Running (subsequently rebranded as Nike Run Club) and Zombies, Run! [14], the evaluation of such commercial apps and their metrics has been rare and inconclusive; for example, Direito et al [15] compared the use of Zombies, Run! with a nonimmersive app as well as the control group and did not find any significant intervention effect on the adolescent participants' fitness and activity levels. In general, empirical evidence on the motivational effects of game elements has been accumulating and promising, although with heterogeneity across the relatively small number of high-quality experimental studies [16,17]. According to 2 most recent meta-analyses on gamification and physical activity [18,19], gamified interventions compared to a nongamified control promoted small increases in step count but not in more intense physical exercises; significant differences in the effect size were also observed among the reviewed studies, and it was suggested to provide tailored experiences for improving gamification effectiveness across diverse user populations. Overall, gamification in physical activity as a young area of research still faces many open questions that need to be rigorously studied, such as the effectiveness of various game elements, as well as the consideration of different application contexts and notable user differences [20].

Tailored Gamification and User Modeling

One important reason for the discrepancies detected in existing gamification research is the reliance on the "one-size-fits-all" approaches adopted in many current apps. These approaches either did not consider variability across individuals due to user motivations and personal characteristics [21,22] or implemented too many game elements simultaneously [23], therefore yielding less effective results. Besides, the majority of prior studies focused on testing gamified systems with multiple elements together, providing limited information on the unique effects of individual elements [20,24].

To address these issues, there is an emerging area of interest called "tailored gamification," which focuses on customizing game elements in given contexts to suit individual needs and preferences for the gamification experience [25]. Accordingly, a robust and accurate classification of users is a critical first step before any effective tailored gamification can be implemented. Meanwhile, current user modeling approaches rely primarily on rating scale–based measures as the basis for profiling, such as personality traits or player typologies [20], the latter being developed either in the context of games (eg, Bartle and BrainHex) or specifically for gamification (ie, Hexad) [26].

However, these popular measurement instruments possess methodological limitations for user segmentation and tailored gamification design. The most common issue inherent in rating scales is scale use bias, such as small ranges of mean item scores, a tendency for respondents to favor the top end of the scale, and scalar inequivalence across countries and cultures [27]. As rating scales do not require people to make trade-offs, they provide less accurate assessments of the relative importance of each item, making it harder to clearly differentiate one segment from another. A key example is the widely used gamification user types Hexad scale [28,29], which is a framework for categorizing individuals based on their motivations and preferences when interacting with gamified systems. The scale consists of 6 distinct user types: philanthropist, socializer, achiever, free spirit, player, and disruptor. Each is motivated by different aspects of gamification: purpose, relatedness, competence, autonomy, extrinsic rewards, and the triggering of change, respectively. Despite the helpfulness of the Hexad scale for understanding user diversity, the issues with rating scales are still unavoidable; for example, although users can be classified into 1 of the 6 types based on their highest scores, the score differences across the dimensions might not be conspicuous or practically meaningful, and there are also strong correlations between certain types, such as those between the philanthropist and socializer types [29]. Further evidence has shown that the Hexad user orientations are not necessarily stable and can change over time, suggesting that gamification design based on a 1-time measurement of the dominant Hexad user type might not be adequate [30].

In addition, while standard practices of correlation analysis can be used to match user types with their preferred game elements based on ratings [29], the accuracy of such methods is not guaranteed, given the limitations of rating scales again, as well as the challenge of determining the most powerful game elements from a large pool of suitable candidates for a given

XSL•FO
RenderX

user type. As summarized by Klock et al [25], >20 different game elements were preferred by achiever, player, and socializer types, and 11 game elements were proposed even for the least common type, disruptor. More research is needed to tease out the unique effects of individual game elements to optimize their use for different user groups, particularly for tailored design on limited interfaces.

Maximum Difference Scaling: An Alternative User Modeling Method

Given the aforementioned issues with existing user modeling approaches, alternative methods are needed to advance the field of tailored gamification. If the purpose of tailoring is to make gamification better fit what the users will enjoy or be motivated by, a natural way would be directly using user preferences and needs for various game elements as the basis for profiling. However, there is a lack of research using this intuitive approach, which might be due to the limitations of traditional rating and segmentation methods, such as scale use bias and statistical deficiencies [27]. Besides, the problem with asking users about game design elements directly, according to Tondello et al [29], is that users might not be aware of their game preferences and thus be unable to rate each element accurately. It is also important to examine psychological factors such as motivation in a gamified context beyond what users like, which can help determine what type of gamification works best for stimulating behavior change rather than merely pleasing the users.

Accordingly, the maximum difference scaling (MaxDiff) method, a widely used technique in marketing research, provides a useful alternative for classifying gamification users because it is "a rating method that does not experience scale use bias, forces trade-offs, and allows each scale point to be used once and only once" [27]. As a classic pairwise comparative analysis method [31], MaxDiff requires respondents to select 2 items from a given set—their most and least preferred—that can indicate the maximum difference in their preferences. This approach has been shown to solve the problem of scalar inequivalence across countries [32] and demonstrate greater discrimination among items and between respondents than Likert scales [33]. MaxDiff can also be conveniently integrated into scenario-based experiments that simulate real-world contexts and improve experimental results [34]. In addition, it generates valuable data for segmenting users into distinct groups.

The MaxDiff technique provides a great opportunity for tailored gamification research, which is still a new area at the very early stage of development, with the majority of research conducted in the education domain [35,36]. Limited evidence exists on the usefulness of MaxDiff for teasing out user preferences for game design elements [37,38]; however, its full potential for user segmentation remains underused. In general, user modeling for tailored gamification needs to "consider many characteristics simultaneously to evaluate how much one may influence the other" [25] to achieve better segmentation of gamification users. It should also consider the context of use that can affect the effectiveness of game elements (eg, task, domain, and device), with most existing research limited to education and smartphone apps [25]. More inquiry into the health domain and other new application types will be particularly valuable.

Smartwatch-Based Gamification for Physical Activity

One promising health-related application of tailored gamification is its integration into popular wearable activity trackers such as smartwatches, which were predicted to be used by approximately 225 million people globally in 2024 [39] and hold great promise for promoting healthy lifestyles through useful functions such as health monitoring, reminders, and sharing [40]. In particular, a recent umbrella review confirmed that using wearable devices could significantly increase physical activity [41] and suggested leveraging motivational features to enhance the effectiveness of wearables. However, near half of the users abandoned their devices within the first 6 months [42,43]; yet, most related research has only focused on technical issues rather than user perceptions and preferences [44]. A small number of studies did point out the importance of providing tailored feedback and intrinsically motivating users [45], as well as optimizing design elements on the limited interfaces of smartwatches [46].

Meanwhile, gamification is believed to enhance personal informatics apps and sustain user engagement [47]. As an effective feedback mechanism [48], gamification can aid the interpretation of health monitoring data for users in a straightforward and motivating way. Moreover, unlike mostly self-reported data from mobile phone apps, smartwatches offer the distinct advantage of providing immediate feedback based on real-time objective health data. Their wearable nature also makes them more suitable for exercise contexts, thereby improving the effectiveness of gamification in a more accurate and timely fashion [16]; for example, gamification on smartwatches has been found effective for obesity control [49,50], while a systematic review of mobile health–based gamification interventions suggested the need for more empirical research to explore the efficacy of combining gamification and wearables for promoting physical activity in various populations [17]. In particular, tailored gamification has the potential to further improve engagement and motivation for exercising among all users [22].

As smartwatch users vary in their needs and behaviors, accurate user modeling is required to categorize them accurately and tailor different game elements for different groups to better motivate users for maintaining regular exercise engagement and lifelong healthy lifestyles. From a methodological perspective, the MaxDiff method is particularly useful for optimizing gamification design on the small smartwatch screen, which requires a nuanced understanding of trade-offs that different users make among various game elements (rather than the absolute importance of each element). Only on the basis of a clear understanding of these subtle elements and heterogeneous priorities can smartwatch users be accurately classified, who will then be offered tailored design solutions promoting both user enjoyment and health benefits through a carefully selected set of functions and displays on the limited interfaces of smartwatches.

XSL•FO
RenderX

This Study

This study aimed to understand how smartwatch-based gamification should be tailored for different user groups to effectively promote physical exercise based on a more accurate and innovative user modeling approach. We incorporated both user preferences and needs for game elements from smartwatch users into the segmentation process and adopted the MaxDiff experimental method requiring users to make trade-offs among various game elements, followed by a quantified categorization of users into distinct groups using the latent class technique. As formulating theory-based hypotheses was not possible, given the limited prior research in the young area of tailored gamification, we proposed the following research questions (RQs):

- RQ1: What is the relative importance of each game element in smartwatch fitness apps based on the preferences of smartwatch users?
- RQ2: Which game elements in smartwatch fitness apps are more effective than others for motivating physical exercise among smartwatch users?
- RQ3: Are there distinct user groups with unique preferences for smartwatch-based gamification? If so, how can such segment membership be predicted among smartwatch users?
- RQ4: Are there distinct user groups who will be differentially motivated by smartwatch-based gamification? If so, how can such segment membership be predicted among smartwatch users?

Methods

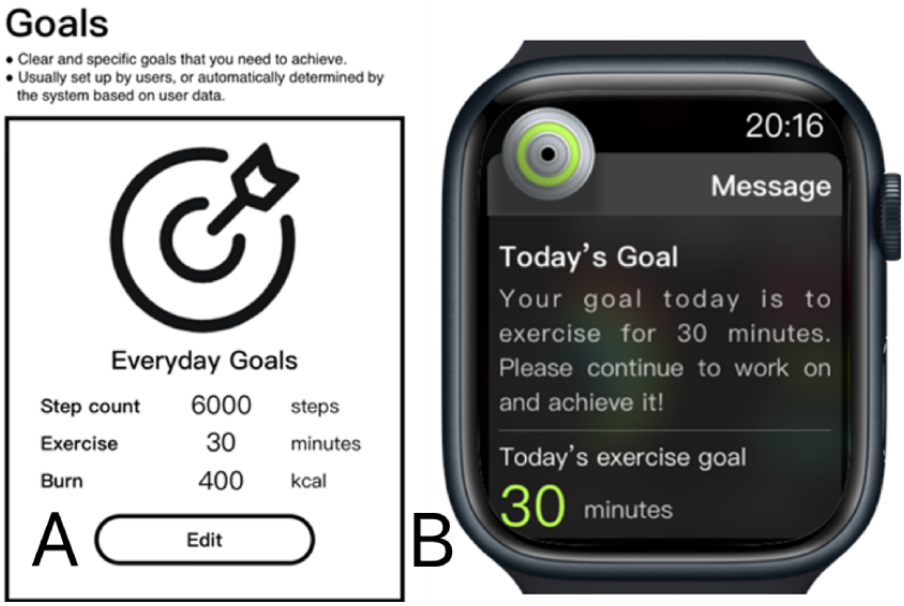
Overview

As explained earlier, this study sought to compare the differences in user preferences and needs for gamification; therefore, using a simple sorting method was not appropriate due to the potential order effect in ranking that might bias the experimental results [27]. Meanwhile, if a Likert rating scale was used, respondents might assign similar preference scores to each game element, based on which the elements could not be clearly distinguished from one another. Therefore, we implemented the MaxDiff method and conducted 2 experiments to investigate user preferences for different game elements in smartwatch gamification apps. In addition, we compared the motivational effects of these elements on user willingness to exercise.

Materials for the MaxDiff Experiments

On the basis of a comprehensive literature review and a preliminary study [51], we synthesized 16 commonly used game elements for the MaxDiff experiments: goals, progress, feedback, overview, points, levels, badges, leaderboards, community, sharing, cooperation, competition, challenges, narrative, avatars, and digital currency. These can be categorized into 4 groups: goal-related, reward-related, socialization-related, and immersion-related elements. After carefully reviewing related studies and market apps for the visualization and verbal explanation of each element, we designed 2 sets of experimental materials for the 2 experiments: a low-fidelity version for studying user preferences and a high-fidelity version for measuring the motivational effects of gamification (eg, the 2 versions of “goals” in Figure 1). The low-fidelity version adopted black-and-white wireframes to explain the concept of each game element in a simple manner to avoid the influence of color and design on user preferences. In comparison, the motivation-focused experiment was designed to simulate the real-life situation of using a smartwatch for exercising; therefore, the Apple Watch S6 prototype was used to represent the game elements in the form of various reminders displayed on a smartwatch screen. It should be noted that the text font and color were kept consistent across all materials for each game element to exclude the potential effects of such design factors on user responses.

Figure 1. Samples of gamification materials for the MaxDiff experiments on (A) user preferences and (B) motivational effects.



XSL•FO
RenderX

Other Instruments

In addition to conducting the MaxDiff experiments, we collected background data from users through survey questions as well as instruments for measuring gamification user types and motivation for smartwatch use (Multimedia Appendix 1). The Hexad scale [29] included 24 items to be rated on a 7-point scale ranging from 1 (*strongly disagree*) to 7 (*strongly agree*) to determine gamification user type, while the user motivation scale for smartwatch use [52] used a 6-point Likert scale and included 6 items, with 3 items measuring intrinsic motivation and 3 items measuring extrinsic motivation.

Procedures

The study consisted of 3 parts: background information of users was collected first, followed by the preference-focused and motivation-focused MaxDiff experiments. Respondents were asked to provide basic information about themselves, including demographics, smartwatch use, exercise and gaming habits, and gamification user types. In particular, we collected important data about their attitudes and behaviors regarding smartwatch use, physical exercise, playing games, and gamification, which provided valuable complementary information for user segmentation.

The MaxDiff experiments started with a short sample MaxDiff exercise that helped familiarize the participants with the process so that they could quickly and accurately make their choices in the formal experiments that followed. In the first MaxDiff experiment, simple pictures and text descriptions about the 16 game elements related to a smartwatch fitness app (ie, the low-fidelity version) were presented to the participants based on an experimental design. Each time, participants were asked to repeatedly select their favorite and least favorite elements from the group of game elements presented. Specifically, the parameters of the MaxDiff experimental design were as follows: (1) 16 total items (game elements); (2) 4 items per question, following convention; (3) 12 questions per respondent to ensure that each item appeared at least 3 times; (4) 300 versions to meet the web-based questionnaire criteria; and (5) a minimum sample size of 200 to ensure that each item appeared at least 500 times. We followed the principle of balanced incomplete block design in selecting the MaxDiff experimental parameters, and the experimental design passed the merit criteria test, indicating that the design was reasonable and could be formally investigated.

For the second, motivation-focused MaxDiff experiment, the parameters of the experimental design were exactly the same as those of the first experiment because the number of game elements selected were identical. The main difference lay in the contents and form of presentation: high-fidelity experimental materials were used, and the game elements were presented as context-specific reminders on a simulated smartwatch screen. We hypothesized the context for the participants as follows: “You are very tired after a long day of work and finally come home. You don’t want to exercise and just want to play with your smartphone, whereas a message alert suddenly pops up on your smartwatch.” In the 12 scenarios that followed, each composed of 4 messages representing gamification elements (refer to Figure 2 for a sample scenario), participants were asked to choose the messages that motivated them the most to start exercising and those that motivated them the least, which measured the motivational effects of different smartwatch-based game elements in the context of exercising.

Figure 2. A sample scenario with 4 messages representing gamification elements from the motivation-focused MaxDiff experiment.



XSL•FO
RenderX

Sample and Data Collection

We used the specialized software Lighthouse Studio (Sawtooth Software) to design the MaxDiff experiments, program the measurement scales and survey questions, and collect data on the web. The study targeted smartwatch and smart band users aged 18 to 35 years in China. We recruited participants through invitations posted on popular social networking apps (eg, WeChat and QQ). A total of 529 people responded, of whom 378 (71.5%) passed the screener and provided valid responses. The whole study took 10 to 15 minutes to complete. Screening and data cleaning criteria included not meeting the background requirements (such as age not suitable and not having a smartwatch or smart band), incomplete responses, completion time <10 minutes, and straight-lining answers.

Ethical Considerations

This study was approved by the ethics committee of Shenzhen University (2022010710). Informed consent was obtained electronically before participants began the web-based study, and participants could opt out at any time. All data collected were anonymous. Each participant received a cash incentive of CN ¥10 (US $1.5).

Data Analysis

SPSS software (version 26.0; IBM Corp) was used for data cleaning and conventional statistical analysis, while data from the MaxDiff experiments were analyzed using Lighthouse Studio. Specifically, the utility scores for all users' preferences for each game element were calculated using both logit analysis and hierarchical Bayes estimation. We also obtained the overall ranking of users' preferences for the 16 game elements based on the rescaled preference scores which summed up to 100. Similarly, we analyzed the motivation-focused MaxDiff experiment data using the same approach, calculating the utility scores and rankings of the elements in terms of their motivational effects on users.

For the user segmentation, we performed latent class multinomial logit analysis on the MaxDiff data. As a powerful user segmentation method, latent class multinomial logit analysis is more suitable for analyzing MaxDiff data than other classification methods (eg, cluster analysis) because it is not solely data driven but based on statistical models. Users with similar choices were grouped into categories, and the part-worth utilities and the probability of each respondent belonging to each category were estimated for each category. After obtaining the preference-based and motivation-based segmentation schemes, we enhanced them by adding demographic information, smartwatch use, exercise and gaming habits, and gamification user types by means of chi-square tests and multivariate analysis of variance (MANOVA) to describe the specific characteristics of each user segment. Finally, significant factors that could help predict the future classification of users were also determined through multinomial logistic regression models.

Results

Descriptive Statistics

As shown in Table 1, our final sample consisted of 378 participants (male: n=204, 54%; female: n=174, 46%) with an average age of 23 (SD 2.916) years; 82 (21.7%) were employed, and the rest (n=296, 78.3%) were students. As shown in Table 2, of the 378 participants, 156 (41.3%) had prior experience using smartwatches, and 222 (58.7%) had prior experience using smart bands. The majority (293/378, 77.5%) had worn the device for >3 months with daily use. The main purpose of using the wearable varied from health monitoring and fitness to messaging and mobile assistance, with health monitoring and fitness reported by 224 (59.3%) of the 378 participants. Regarding physical exercise, 70.9% (268/378) of the participants did not meet the World Health Organization's recommendation of exercising at least 3 times per week, with exercises limited to walking or aerobics. By contrast, the distributions of game playing preferences and behaviors were fairly even, suggesting notable individual differences in this area.

Table 1. Demographics of participants (n=378).

Characteristics	Participants, n (%)
Age (y)	
18-22	154 (40.7)
23-35	224 (59.3)
Sex	
Male	204 (54)
Female	174 (46)
Occupation	
Student	296 (78.3)
Employed	82 (21.7)

XSL•FO
RenderX

Table 2. Self-reported smartwatch use, exercise habits, and gaming behaviors of participants (n=378).

Variables	Participants, n (%)
Device	
Smartwatch	156 (41.3)
Smart band	222 (58.7)
Duration of use	
<3 mo	85 (22.5)
3-12 mo	128 (33.9)
1-2 y	106 (28)
≥3 y	59 (15.6)
Frequency of use	
Wear occasionally	138 (36.5)
Wear daily	240 (63.5)
Main purpose of use	
Health monitoring	136 (36)
Sports and fitness	88 (23.3)
Messaging	68 (18)
Mobile assistance	86 (22.8)
Most frequently chosen exercises	
Walking	145 (38.4)
Aerobics	135 (35.7)
Other	98 (25.9)
Exercise frequency	
≤3 times per mo	131 (34.7)
1 to 2 times per wk	137 (36.2)
≥3 times per wk	110 (29.1)
Frequency of playing games	
Seldom	72 (19)
Occasionally	133 (35.2)
Frequently	99 (26.2)
Daily	74 (19.6)
Enjoyment of games	
Relatively dislike	219 (57.9)
Relatively like	159 (42.1)

We also analyzed the scale measures. Table 3 presents the results for each Hexad user type and each of the 4 dimensions of gamification drive that we hypothesized, as well as motivation for smartwatch use. Among the Hexad user types, player, free spirit, and achiever had the highest average scores, followed by philanthropist and socializer, whereas disruptor had substantially lower average scores. Meanwhile, respondents reported higher scores in the goal-driven and achievement-driven dimensions in comparison with the immersion-driven and socialization-driven dimensions. Interestingly, intrinsic motivation for smartwatch use was higher than extrinsic motivation. It should be noted that all measures for each item in the scale were tested for normality for further statistical modeling.

XSL•FO
RenderX

Table 3. Scores for the Hexad user types, dimensions of gamification drive, and motivation for smartwatch use.

Items	Scores, mean (SD)
Hexad user type	
Philanthropist	5.02 (0.915)
Socializer	4.69 (1.071)
Free spirit	5.48 (0.782)
Achiever	5.26 (0.854)
Disruptor	4.16 (1.031)
Player	5.56 (0.901)
Dimensions of gamification drive	
Goal driven	5.67 (0.930)
Achievement driven	5.50 (0.975)
Socialization driven	5.04 (1.159)
Immersion driven	5.13 (1.153)
Motivation for smartwatch use	
Intrinsic	4.68 (0.903)
Extrinsic	3.98 (0.954)

Overall MaxDiff Results

User Preferences for Gamification Elements

The results from the 2 analyses of the preference-focused MaxDiff data—logit analysis and hierarchical Bayes estimation—are compared in Table 4, showing overall consistency and negligible differences. It should be noted that we used the results of the logit analysis for further segmentation modeling to simplify the analyses and also because logit analysis is more commonly used in estimating mean utility scores and more applicable to studies involving a larger number of items [53].

Table 4. User preferences for each gamification element based on the preference-focused MaxDiff experiment.

Gamification element	Best (number of times)	Worst (number of times)	Logit analysis		Hierarchical Bayes estimation	
			Aggregate-level coefficient (rescaled)	Rank	Individual-level coefficient averaged (rescaled)	Rank
Goals	328	179	7.775	2	7.644	2
Progress	523	85	11.656	1	11.361	1
Feedback	263	220	6.561	7	6.205	7
Overview	338	244	7.126	4	7.188	3
Points	262	259	6.114	8	5.845	9
Levels	279	221	6.724	6	6.504	6
Badges	314	225	7.094	5	6.926	5
Leaderboards	268	301	5.748	10	5.874	8
Community	205	403	4.290	15	4.409	16
Sharing	251	275	5.848	9	5.506	11
Cooperation	254	289	5.716	11	5.766	10
Competition	234	353	4.959	12	5.162	13
Challenges	312	199	7.354	3	6.940	4
Narrative	271	450	4.432	14	5.244	12
Avatars	223	364	4.772	13	4.941	14
Digital currency	211	469	3.831	16	4.485	15

XSL•FO
RenderX

Motivational Effects of Gamification Elements

Similarly, based on the 2 aforementioned statistical approaches, we analyzed the average utility scores and rankings of the 16 elements for motivating users, as shown in Table 5. Interestingly, the order of importance among the elements differed from that in the first experiment, suggesting that user preferences did not necessarily equal the motivational power of gamification elements or predict actual behavior change; for example, it was surprising to see that cooperation as the 11th preferred element could indeed motivate users quite effectively, with a ranking of 4 among all 16 elements.

Table 5. Motivational effects of each gamification element based on the motivation-focused MaxDiff experiment.

Gamification elements	Best (number of times)	Worst (number of times)	Logit analysis		Hierarchical Bayes estimation	
			Aggregate-level coefficient (rescaled)	Rank	Individual-level coefficient averaged (rescaled)	Rank
Goals	389	120	9.181	2	8.876	2
Progress	541	71	11.985	1	11.663	1
Feedback	138	428	3.470	13	2.921	16
Overview	358	165	8.151	3	8.062	3
Points	289	214	6.738	7	6.521	9
Levels	283	211	6.721	8	6.583	8
Badges	324	207	7.275	5	7.188	5
Leaderboards	326	250	6.741	6	7.153	6
Community	142	474	3.177	15	3.005	15
Sharing	223	284	5.347	11	4.985	12
Cooperation	368	232	7.462	4	7.754	4
Competition	317	272	6.454	9	6.882	7
Challenges	247	220	6.233	10	5.762	10
Narrative	224	544	3.243	14	4.556	13
Avatars	140	494	3.049	16	3.073	14
Digital currency	227	350	4.773	12	5.015	11

Segmentation Results Based on MaxDiff User Preferences

Model Estimation Results

To determine the fittest segmentation model with the best number of segments, multiple rounds of modeling were performed, with 7 metrics plus their log-likelihood values reported in Table 6 as measures of model fit. Specifically, users were classified into 2, 3, 4, and 5 categories according to their MaxDiff preferences for game elements, and the results were then compared based on a balanced evaluation of all 7 metrics in the following way. If relative chi-square was used as the fit-level criterion, 2 categories with the largest value should be selected as the classification result, whereas the optimal solution became 5 categories if percentage certainty and chi-square were used as the criteria. Similarly, if Akaike information criterion, consistent Akaike information criterion, Bayesian information criterion, and adjusted Bayesian information criterion were used as the criteria, 5 categories with the smallest value would be the best classification result. However, as 2 categories and 5 categories seemed to be extreme, we also found that the values of Akaike information criterion, consistent Akaike information criterion, Bayesian information criterion, and adjusted Bayesian information criterion decreased differently as the number of categories increased, with the greatest decrease occurring from 2 to 3 categories; therefore, it seemed most appropriate to classify users into 3 groups. In addition, the selection of the best number of categories should also consider other important factors, for example, whether the differences among the categories are obvious and whether the market sizes of each category are balanced. Therefore, based on the aforementioned criteria altogether, it was ultimately determined to best divide the users into 3 categories according to their distinct preferences for game elements, which was also supported by further analysis later.

XSL•FO
RenderX

Table 6. Model fitting tests for MaxDiff preference-based segmentation.

Metrics	Number of categories			
	2	3	4	5
Log-likelihood	–11,672.12	–11,338.98	–11,102.67	–10,930.04
Percentage certainty	7.19	9.84	11.72	13.09
AIC[a]	23,406.24	22,771.96	22,331.33	22,018.08
CAIC[b]	23,657.74	23,153.27	22,842.45	22,659.00
BIC[c]	23,626.74	23,106.27	22,779.45	22,580.00
ABIC[d]	23,528.23	22,956.91	22,579.24	22,328.95
Chi-square (*df*)	1808.7 (1)	2475.0 (2)	2947.6 (3)	3292.9 (4)
Relative chi-square (*df*)	58.3 (1)	52.7 (2)	46.8 (3)	41.7 (4)

[a]AIC: Akaike information criterion.

[b]CAIC: consistent Akaike information criterion.

[c]BIC: Bayesian information criterion.

[d]ABIC: adjusted Bayesian information criterion.

Preference Scores and Market Sizes by Segment

On the basis of the 3-segment model and the distinctive characteristics of each group, we categorized the 3 segments as goal-preferred, immersion-preferred, and reward-preferred segments. Table 7 shows by segment the market sizes and the logit preference coefficients for each gamification element, the latter being obtained from the latent class multiple logit analysis for each segment.

Clearly, the 3 segments differed significantly in their preference scores. Users in the goal-preferred segment scored notably higher on progress, goals, challenges, overview, and feedback than the other 2 segments but with much lower preferences for immersive elements such as avatars, narrative, and digital currency. In comparison, users in the immersion-preferred segment had much higher preference scores on narrative, avatars, badges, and challenges than those in the other 2 segments, who showed notably lower preferences for social elements such as community, sharing, competition, and leaderboards. By contrast, users in the reward-preferred segment loved reward-related elements and had the highest preference scores on digital currency, which was the least preferred by the other 2 segments. Not surprisingly, the reward-preferred users showed much lower preferences for social elements (eg, community, cooperation, and competition) as well as immersive elements (eg, avatars and narrative).

XSL•FO
RenderX

Table 7. Segment-level user preference coefficients for each gamification element and market sizes[a].

Gamification elements	Goal-preferred segment	Immersion-preferred segment	Reward-preferred segment
Goals	9.529	6.322	6.175
Progress	13.977	9.324	9.806
Feedback	7.309	5.761	5.053
Overview	8.447	5.412	6.240
Points	4.058	4.283	11.426
Levels	6.982	5.470	6.572
Badges	7.307	6.998	5.825
Leaderboards	6.704	3.545	5.904
Community	2.887	5.026	4.654
Sharing	6.408	4.606	5.373
Cooperation	5.891	6.264	4.031
Competition	5.886	3.882	3.800
Challenges	9.471	6.355	4.703
Narrative	1.545	14.840	2.330
Avatars	2.243	9.823	4.123
Digital currency	1.356	2.090	13.986

[a]Market size (n=378): goal-preferred segment, n=161 (42.6%); immersion-preferred segment, n=113 (29.9%); and reward-preferred segment, n=104 (27.5%).

Demographic and Behavioral Characteristics of Each Segment

On the basis of the segmentation scheme, we also summarized the characteristics of each segment using descriptive statistics and chi-square tests, focusing on user demographics, physical exercise habits, and game playing. The goal-preferred segment had a higher proportion of male individuals than the other 2 segments, whereas the reward-preferred segment had a higher proportion of female individuals, although the sex differences were not statistically significant. Meanwhile, there was indeed a significant difference in age distribution among the 3 segments (χ^2_2=3.8; *P*=.049): of the 104 reward-preferred segment members, 72 (69.2%) were aged 23 to 35 years, a significantly higher proportion than in the other 2 segments. Occupations also differed significantly across the 3 segments (χ^2_2=7.6; *P*=.02), with a much smaller proportion of nonstudents (24/161, 14.9%) in the goal-preferred segment.

Furthermore, significant between-segment user differences were also observed regarding the most frequent exercise type (χ^2_4=12.1; *P*=.02), exercise frequency (χ^2_4=9.6; *P*=.048), and level of interest in exercising (χ^2_4=17.1; *P*=.002); for example, the goal-preferred segment had the smallest percentage of users (48/161, 29.8%) who identified walking as their most frequently chosen exercise, while the other 2 segments had higher percentages of users who preferred aerobic exercise. Meanwhile, the frequency of exercise was higher for the goal-preferred group and lower for the immersion-preferred group. Not surprisingly, there were more sports enthusiasts in the former segment but more users who hardly exercised in the latter segment. By contrast, no significant difference was detected among the segments in terms of liking games, although there was a higher percentage of game lovers in the immersion-preferred group.

In addition, we used 1-way MANOVA to analyze the differences among the segments in terms of Hexad user types and the 4 dimensions of gamification drive. First, we found that scores for the philanthropist (*P*<.001), socializer (*P*=.003), free spirit (*P*=.004), and achiever (*P*<.001) dimensions differed significantly across the segments, but there were no significant differences for the disruptor (*P*=.58) and player (*P*=.05) dimensions. As further detailed in Table 8, both the immersion-preferred and reward-preferred segments had significantly lower scores on the philanthropist dimension compared to the goal-preferred group, while on the socializer dimension, only the immersion-preferred segment scored significantly lower (*P*=.005). Similarly, on the free spirit and achiever dimensions, both the immersion-preferred and reward-preferred segments had significantly lower scores than the goal-preferred group.

XSL•FO
RenderX

Table 8. Multiple comparisons of the Hexad user type scores across the 3 preference-based user segments (where significant differences were detected; n=378).

Dependent variables, reference group, and segments	Participants, n (%)	Scores, mean (SD)	*P* value
Philanthropist			
Goal-preferred segment			
Goal preferred	161 (42.6)	5.23 (0.837)	—[a]
Immersion preferred	113 (29.9)	4.94 (0.936)	.03
Reward preferred	104 (27.5)	4.76 (0.940)	<.001
Socializer			
Goal-preferred segment			
Goal preferred	161 (42.6)	4.89 (1.031)	—
Immersion preferred	113 (29.9)	4.45 (1.163)	.005
Reward preferred	104 (27.5)	4.63 (0.974)	.11
Free spirit			
Goal-preferred segment			
Goal preferred	161 (42.6)	5.64 (0.645)	—
Immersion preferred	113 (29.9)	5.39 (0.795)	.02
Reward preferred	104 (27.5)	5.34 (0.918)	.02
Achiever			
Goal-preferred segment			
Goal preferred	161 (42.6)	5.56 (0.691)	—
Immersion preferred	113 (29.9)	5.05 (0.855)	<.001
Reward preferred	104 (27.5)	5.01 (0.936)	<.001

[a]Not applicable.

We also conducted multiple comparisons on the 4 dimensions of gamification drive, which showed that scores for the goal-driven (*P*<.001), achievement-driven (*P*<.001), socialization-driven (*P*=.02), and immersion-driven (*P*<.001) dimensions all differed significantly across segments. Specifically, as presented in Table 9, the immersion-preferred and reward-preferred segments had significantly lower scores than the goal-preferred segment on the goal-driven dimension. Not surprisingly, the goal-preferred and reward-preferred segments scored significantly lower on the immersion-driven dimension than the immersion-preferred group.

XSL•FO
RenderX

Table 9. Multiple comparisons of the dimensions of gamification drive across the 3 preference-based user segments (n=378).

Dependent variables, reference group, and segments	Participants, n (%)	Scores, mean (SD)	*P* value
Goal driven			
Goal-preferred segment			
Goal preferred	161 (42.6)	5.93 (0.778)	—[a]
Immersion preferred	113 (29.9)	5.52 (0.892)	<.001
Reward preferred	104 (27.5)	5.43 (1.084)	<.001
Achievement driven			
Goal-preferred segment			
Goal preferred	161 (42.6)	5.87 (0.826)	—
Immersion preferred	113 (29.9)	5.26 (0.894)	<.001
Reward preferred	104 (27.5)	5.19 (1.084)	<.001
Socialization driven			
Goal-preferred segment			
Goal preferred	161 (42.6)	5.23 (1.126)	—
Immersion preferred	113 (29.9)	4.89 (1.118)	.04
Reward preferred	104 (27.5)	4.91 (1.218)	.09
Immersion driven			
Immersion-preferred segment			
Goal preferred	161 (42.6)	4.89 (1.076)	<.001
Immersion preferred	113 (29.9)	5.59 (1.110)	—
Reward preferred	104 (27.5)	5.01 (1.179)	.001

[a]Not applicable.

Predicting Preference-Based User Segments by Logistic Regressions

To develop a useful method for quickly segmenting future users (ie, to assign them to a particular segment based on their answers to a few simple questions), we conducted rounds of multiple logistic regression modeling that included different combinations of variables from our survey. The final prediction model was composed of 2 dimensions of gamification drive (represented by achievement-driven and immersion-driven scores) and age group (divided into users aged 18-22 y and 23-35 y), with a good model fit (Table 10). Using the reward-preferred segment as the reference, age, achievement-driven scores, and immersion-driven scores could be significant predictors of particular user segments; for example, users aged between 18 and 22 years scoring high on the achievement-driven dimension or low on the immersion-driven dimension are more likely to belong to the goal-preferred segment rather than the reward-preferred segment. In comparison, users scoring high on the immersion-drive dimension are more likely to belong to the immersion-preferred segment.

Table 10. The final logistic regression models that predict preference-based user segments.

Segments[a] and predictors	B	SE	*P* value	Exp(B)
Goal-preferred segment				
Immersion driven scores	–0.283	0.126	.02	0.753
Achievement driven scores	0.915	0.158	<.001	2.497
Age (y): 18-22 (vs 23-35)	0.707	0.281	.01	2.027
Immersion-preferred segment				
Immersion driven scores	0.509	0.137	<.001	1.663

[a]The reference category is the reward-preferred segment.

XSL•FO
RenderX

Segmentation Results Based on MaxDiff Motivational Effects

Model Estimation Results

Similar to the segmentation process based on user preferences, we analyzed the 7 model-fit metrics with log-likelihood values when users were classified into 2, 3, 4, and 5 categories based on the MaxDiff results from the motivation-focused experiment (Table 11). Different models were compared, and we finally determined that dividing the users into 4 categories provided the best fit, following the same criteria and process as detailed in the previous subsection.

Table 11. Model fitting tests for MaxDiff motivation-based segmentation.

Metrics	Number of categories			
	2	3	4	5
Log-likelihood	–11,149.17	–10,803.67	–10,587.67	–10,432.31
Percentage certainty	11.35	14.10	15.81	17.05
AIC[a]	22,360.34	21,701.33	21,301.35	21,022.63
CAIC[b]	22,611.84	22,082.64	21,812.46	21,663.55
BIC[c]	22,580.84	22,035.64	21,749.46	21,584.55
ABIC[d]	22,482.32	21,886.28	21,549.26	21,333.50
Chi-square (*df*)	2854.6 (1)	3545.6 (2)	3977.6 (3)	4288.3 (4)
Relative chi-square (*df*)	92.1 (1)	75.4 (2)	63.1 (3)	54.3 (4)

[a]AIC: Akaike information criterion.

[b]CAIC: consistent Akaike information criterion.

[c]BIC: Bayesian information criterion.

[d]ABIC: adjusted Bayesian information criterion.

Motivation Scores and Market Size by Segment

On the basis of the 4-segment model and the distinctive characteristics of each group, we categorized the 4 segments as goal-motivated, immersion-motivated, reward-motivated, and socially motivated segments. Table 12 shows the market sizes of the 4 segments, together with the average utility scores of each game element's motivational effect, which were calculated using logit scores, following the same approach as in the preference-based models.

The 4 segments also differed significantly in their motivation scores across the 16 game elements. Users in the goal-motivated segment were significantly more likely to be motivated by progress, goals, and overview than those in the other 3 segments. This segment, however, had much lower motivation scores on social gamification elements such as community, sharing, competition, and leaderboards. In comparison, users in the immersion-motivated segment scored significantly higher on narrative, badges, and avatars than the other 3 segments, suggesting that gameplay and fun elements were more effective in engaging them. Meanwhile, users in the reward-motivated segment were effectively motivated by reward-related elements such as digital currency, points, and levels, whereas immersive game elements such as avatars and narrative had less influence on them. It should be particularly noted that compared with the 3-segment scheme from the preference-based classification, a new segment was discovered in terms of the motivational effects of gamification, named the socially motivated segment. Users in this unique segment had significantly higher motivation scores than the other 3 segments on social elements such as leaderboards, competition, cooperation, and sharing but were the least influenced by immersive elements such as narrative and avatars.

XSL•FO
RenderX

Table 12. Segment-level motivational coefficients for each gamification element and market sizes[a].

Gamification elements	Goal-motivated segment	Immersion-motivated segment	Reward-motivated segment	Socially motivated segment
Goals	13.981	6.214	6.983	10.171
Progress	15.228	9.572	9.754	13.171
Feedback	6.265	3.058	1.787	2.192
Overview	11.360	6.147	5.569	9.563
Points	7.241	5.809	12.106	3.630
Levels	9.009	5.670	7.141	4.886
Badges	9.969	8.069	5.163	4.611
Leaderboards	2.077	5.151	9.086	11.231
Community	1.354	4.007	4.408	1.687
Sharing	1.958	4.804	5.256	8.021
Cooperation	3.418	8.210	5.913	9.746
Competition	1.569	5.914	7.101	11.225
Challenges	9.384	4.423	4.363	6.699
Narrative	0.847	13.031	0.658	0.676
Avatars	2.510	5.225	1.605	1.132
Digital currency	3.830	4.695	13.108	1.359

[a]Market size (n=378): goal-motivated segment, n=79 (20.9%); immersion-motivated segment, n=128 (33.9); reward-motivated segment, n=64 (16.9%); and socially motivated segment, n=107 (28.3%).

Demographic and Behavioral Characteristics of Each Segment

On the basis of the segmentation scheme, we also summarized the characteristics of each segment using descriptive statistics and chi-square tests, focusing on user demographics, physical exercise habits, and game playing. The socially motivated segment had a higher proportion of male individuals (63/107, 58.9%) than the other 3 segments, whereas the reward-motivated segment had a higher proportion of female individuals, although the gender differences were not statistically significant. Meanwhile, there was indeed a significant difference in age distribution among the 4 user groups (χ^2_3=10.9; *P*=.01), with more users aged 23 to 35 years in the reward-motivated segment and more users aged 18 to 22 years in the socially motivated segment. The 4 segments also showed significant differences in the type of device used (χ^2_3=8.9; *P*=.03) and level of interest in exercising (χ^2_6=23.3; *P*=.001); for example, there were more smart band users in the socially motivated segment (75/107, 70.1%), which also featured more sports enthusiasts compared to many more users who hardly exercised in the immersion-motivated segment. Although there was no significant difference in terms of game playing, more users in the immersion-motivated segment preferred games compared to the other 3 groups.

In addition, we used 1-way MANOVA to analyze the differences among the segments in terms of Hexad user types and the 4 dimensions of gamification drive. First, we found that scores for the socializer (*P*=.001), free spirit (*P*=.02), and achiever (*P*<.001) dimensions differed significantly across the 4 segments, but there were no significant differences for the philanthropist (*P*=.08), disruptor (*P*=.40), and player (*P*=.12) dimensions. As further shown in Table 13, the goal-motivated and immersion-motivated segments had significantly lower scores than the socially motivated segment on both the socializer and achiever dimensions, while on the free spirit dimension, only the immersion-motivated segment scored significantly lower.

XSL•FO
RenderX

Table 13. Multiple comparisons of the Hexad user type scores across the 4 motivation-based user segments (where significant differences were detected; n=378).

Dependent variables, reference group, and segments	Participants, n (%)	Scores, mean (SD)	*P* value
Socializer			
Socially motivated segment			
Goal motivated	79 (20.9)	4.44 (0.954)	.007
Immersion motivated	128 (33.9)	4.53 (1.156)	.04
Reward motivated	64 (16.9)	4.92 (0.873)	.07
Socially motivated	107 (28.3)	4.93 (1.090)	—[a]
Free spirit			
Socially motivated segment			
Goal motivated	79 (20.9)	5.60 (0.795)	.99
Immersion motivated	128 (33.9)	5.34 (0.788)	.02
Reward motivated	64 (16.9)	5.39 (0.870)	.40
Socially motivated	107 (28.3)	5.62 (0.677)	—
Achiever			
Socially motivated segment			
Goal motivated	79 (20.9)	5.20 (0.914)	.01
Immersion motivated	128 (33.9)	5.03 (0.839)	<.001
Reward motivated	64 (16.9)	5.24 (0.917)	.06
Socially motivated	107 (28.3)	5.58 (0.682)	—

[a]Not applicable.

We also conducted multiple comparisons on the 4 dimensions of gamification drive, which showed that scores for the goal-driven (*P*<.001), achievement-driven (*P*<.001), socialization-driven (*P*=.009), and immersion-driven (*P*<.001) dimensions all differed significantly across the 4 segments. Specifically, as presented in Table 14, the goal-motivated and socially motivated segments had significantly higher scores than the immersion-motivated segment on both the goal-driven and achievement-driven dimensions. On the socialization-driven dimension, only the socially motivated group scored significantly higher than the immersion-motivated segment. However, the immersion-motivated segment had a much higher score on the immersion-driven dimension than almost all other segments, with strong statistical significance, except for a borderline significance (*P*=.05) compared to the reward-motivated segment.

XSL•FO
RenderX

Table 14. Multiple comparisons of the dimensions of gamification drive across the 4 motivation-based user segments (n=378).

Dependent variables, reference group, and segments	Participants, n (%)	Scores, mean (SD)	*P* value
Goal-driven dimension			
Immersion-motivated segment			
Goal motivated	79 (20.9)	5.79 (0.949)	.03
Immersion motivated	128 (33.9)	5.41 (0.938)	—[a]
Reward motivated	64 (16.9)	5.68 (1.021)	.39
Socially motivated	107 (28.3)	5.89 (0.774)	<.001
Achievement-driven dimension			
Immersion-motivated segment			
Goal motivated	79 (20.9)	5.58 (1.020)	.03
Immersion motivated	128 (33.9)	5.18 (0.980)	—
Reward motivated	64 (16.9)	5.46 (0.989)	.33
Socially motivated	107 (28.3)	5.85 (0.796)	<.001
Socialization-driven dimension			
Immersion-motivated segment			
Goal motivated	79 (20.9)	4.90 (1.197)	.99
Immersion motivated	128 (33.9)	4.85 (1.196)	—
Reward motivated	64 (16.9)	5.11 (1.121)	.59
Socially motivated	107 (28.3)	5.33 (1.055)	.008
Immersion-driven dimension			
Immersion-motivated segment			
Goal motivated	79 (20.9)	4.92 (1.321)	.009
Immersion motivated	128 (33.9)	5.49 (1.101)	—
Reward motivated	64 (16.9)	5.05 (1.101)	.05
Socially motivated	107 (28.3)	4.91 (1.012)	<.001

[a]Not applicable.

Predicting Motivation-Based User Segments by Logistic Regressions

With the same aim of quickly assigning future users to a particular motivation-based segment using a few simple questions, we again conducted rounds of multiple logistic regression modeling that included different combinations of variables from our survey. The final prediction model was composed of scores from 3 Hexad dimensions (ie, socializer, free spirit, and achiever), the score on the immersion-driven question, age group (divided into users aged 18-22 y and 23-35 y), and level of interest in exercising (1=hardly any exercise, 2=general sports crowd, and 3=sports enthusiasts), with a good model fit (Table 15). Using the reward-motivated segment as the reference group, the aforementioned variables could be significant predictors of particular user segments; for example, users aged between 18 and 22 years scoring high on the achiever dimension are more likely to belong to the socially motivated segment rather than the reward-motivated segment. In comparison, users scoring high on the immersion-driven dimension or having at least some interest in exercising are more likely to belong to the immersion-motivated segment.

XSL•FO
RenderX

Table 15. The final logistic regression models that predict motivation-based user segments.

Segments[a] and predictors	B	SE	*P* value	Exp(B)
Socially motivated segment				
Achiever	0.600	0.260	.02	1.822
Age (y): 18-22 (vs 23-35)	1.185	0.354	.001	3.272
Goal-motivated segment				
Free spirit	0.841	0.290	.004	2.318
Socializer	–0.613	0.196	.002	0.542
Immersion-motivated segment				
Socializer	–0.367	0.183	.04	0.693
Immersion driven scores	0.468	0.163	.004	1.596
General sports crowd	1.103	0.555	.047	3.013
Sports enthusiasts	0.922	0.428	.03	2.515

[a]The reference category is reward-motivated segment.

Comparison of MaxDiff Results Among the Hexad Types

In addition to the 2 sets of MaxDiff-based segmentation results, we also examined the 6 Hexad user groups in terms of their gamification preferences as well as motivators for behavioral change among the 16 elements, given the popularity of the Hexad scale in prior research for assessing gamification user types. Compared to the MaxDiff-based segments, the Hexad types were less clearly differentiated in both user preferences and the motivational power of game elements; for example, "progress" always appeared as the top-selected element, with similar scores across all 6 Hexad types (Table 16), whereas the immersion-preferred or reward-preferred segments exhibited much stronger preferences for other elements (Table 7). These results highlight the limitations of using Hexad types alone as the basis for user modeling and tailored gamification, suggesting that the MaxDiff method may be more effective in teasing out the relative importance of gamification elements and accurately matching them with the right users.

Table 16. User preferences for, and the motivational effects of, the gamification element "progress" among the 6 Hexad types (n=378).

Hexad type	Participants, n (%)	Hierarchical Bayes estimation scores, mean (SD)	
		Preference	Motivation
Philanthropist	43 (11.4)	11.49	12.06
Socializer	91 (24.1)	11.28	11.37
Free spirit	33 (8.7)	11.63	12.78
Achiever	47 (12.4)	12.00	11.78
Disruptor	113 (29.9)	11.07	11.29
Player	51 (13.5)	11.30	11.83

Discussion

Principal Findings

On the basis of 2 MaxDiff experiments and the latent class statistical technique, this study contributes to the growing body of research on user modeling for tailored gamification. Specifically, the relative power of each of the 16 popular game elements was examined in terms of what users preferred to see on smartwatch-based fitness apps as well as what motivated them to exercise, with interesting differences observed between the 2 approaches. More importantly, these factors formed the basis for identifying and predicting user groups with distinct characteristics as future targets of tailored gamification on wearables.

When smartwatch users were making trade-offs among different game elements based on their preferences (ie, RQ1), we found that progress, goals, challenges, overview, and badges were the top 5 favored elements, whereas digital currency, community, narrative, avatars, and competition were the least preferred. This ranking suggests the overall importance of goal setting and constant feedback for attracting users to gamification, who seem to care less about immersive features such as narrative and avatars commonly found in digital games. These overall preferences were also mostly consistent with the ranking of game elements as motivators for exercising (ie, RQ2), where immersion-related gamification was still less effective, while progress, goals, overview, and badges remained in the top 5. However, cooperation rose to be a powerful motivator (ie, ranked fourth) despite being less preferred by the users (ie, ranked 11th), indicating the potential of social gamification for

XSL•FO
RenderX

stimulating behavioral change regardless of user preference. Similar to the contradictory results on the gap between the experiences of enjoyment and the motivational benefits of social gamification on an educational platform [54], this finding suggested that the theoretically presumed relationship between gamification preferences and motivators might not hold in the health context either.

While it is already useful to quantify the overall importance of each individual game element, more valuable findings arose from the MaxDiff-based segmentation models, showing huge heterogeneity among the smartwatch users in terms of both user preferences (ie, RQ3) and the motivational effects of gamification (ie, RQ4), which have confirmed the problem of the one-size-fits-all gamification approach found in many prior studies [22]. Each user segment displayed distinct personal characteristics as well as preferences and needs for gamification; for example, although the overall rankings of immersive game elements were low, the immersion-preferred segment scored particularly high on narrative as well as avatars, 2 main features for engaging players in digital games. This segment also differed significantly from the goal-preferred segment in user characteristics and behaviors, with a much lower frequency of exercise, a much lower proportion of sports enthusiasts, and much higher scores on the immersion-driven scale, suggesting the need for developing more immersive gamification solutions on smartwatches to attract this group. Furthermore, the motivation-based segmentation results showed narrative as the key element prompting the immersion-motivated segment to exercise, whereas the other 3 segments unanimously ranked this element as the least effective, with extremely low scores. Such notable differences further supported the necessity of tailored gamification based on accurate user modeling [25] to motivate health behavior change through different mechanisms for different targets.

In addition, important differences were also observed between the 2 sets of user segmentation results (ie, 3 preference-based segments vs 4 motivation-based segments). Although users in general displayed a weak preference for socialization-related elements and did not form a socialization-preferred group, such elements did play a key role in prompting the socially motivated user segment to exercise, which suggested the importance of tailored design with elements such as competition and leaderboards for those who do not necessarily like social comparison but can be motivated by such pressure. In comparison, goals and progress were ranked the highest by both the goal-preferred and goal-motivated segments, while the reward-preferred and reward-motivated segments were influenced the most by digital currency and points. Interestingly, although digital currency was the least preferred element by the goal-preferred and immersion-preferred segments, it still exerted some motivating influences on the goal-motivated and immersion-motivated segments (ie, at least not among the bottom 5 elements). Such discrepancies further indicate the gap between what people like to use and what motivates them subconsciously, which contributes to the research needed on understanding "whether the motivationability and benefits from gamification coincide with experiences of enjoyment as is theoretically presumed" [54] and can provide valuable guidance for developing apps aimed at changing behaviors rather than merely entertaining users.

Implications for Research and Practice

As the majority of prior studies focused on gamified systems with multiple game elements combined, and the scope of elements examined was limited [20], this study applied an innovative approach to quantify the individual effects of major game elements in the context of smartwatch-based fitness apps, which can advance our understanding of what might or might not work in health gamification; for example, for the widely adopted trinity of gamification design elements—points, badges, and leaderboards [55]—only badges appeared in the top 5 elements in terms of either user preference or motivational influence. In comparison, progress and goals were overall more favored and effective, supporting the importance of goal setting and feedback in health gamification [8]. In addition, smartwatch users overall seemed to care the least about common components of digital games—such as narrative, avatars, and digital currency—which have been inadequately examined in existing gamification research [16].

However, this was only part of the picture. The overall rankings of game elements did not hold across different types of users who indeed displayed quite distinct preferences and needs for gamification. As such, the key contribution of this study is to the emerging research area of tailored gamification [25], which hinges on addressing the problem of "one-size-fits-all" gamification and capturing the types of individual differences to improve health outcomes such as physical activity. Prior systematic reviews [18,19] have pointed out the small, positive effect of gamification on step count but not on moderate to vigorous physical activity, with limited existing research on the latter. This study, given the findings from the second motivation-based MaxDiff experiment that measured users' tendency to exercise, provided new evidence on the potential of gamification to effectively promote physical exercise beyond steps, particularly so with tailored solutions for different user segments. As gamification in physical activity is still a young area of inquiry, more field research should target the challenging but important outcome of regular physical exercise and leverage the principles of personalization and tailoring [56,57] for enhancing intervention effectiveness across diverse user populations. Regarding user modeling, researchers have suggested that incorporating more precise and relevant information, such as motivational and behavioral factors, to determine user types will help better explain heterogeneity among users and interventions [58,59], while this study clearly demonstrated the power of MaxDiff-based segmentation for tailoring smartwatch-based gamification.

In addition, both experiential and instrumental outcomes are needed in the context of health gamification design [60], that is, users should not only like the gamification experience but also be motivated to engage in a target activity because motivation is a major factor for health behavior change. Taking both user preferences and motivations into account, this study proposed a new categorization of smartwatch users in relation to gamification design elements on wearable fitness apps and developed prediction models for easily classifying future users

XSL•FO
RenderX

into the right segments and accordingly designing more accurate, targeted solutions.

From a methodological perspective, our research also supported the usefulness of MaxDiff experiments for teasing out the relative importance of each individual game element [37,38] and confirmed the feasibility of MaxDiff-based user segmentation in future studies on tailored gamification. The MaxDiff method has been widely used in marketing research and proven advantageous over traditional rating scale questions [61,62], yielding more accurate responses. Moreover, such responses can be easily integrated into user segmentation models (such as latent class models) and contribute to higher classification efficiency. Therefore, MaxDiff experiments hold great promise in helping researchers quantify individual differences more efficiently and informing designers on the development of optimized solutions for different users. Specifically, this study demonstrated that MaxDiff segmentation achieved better differentiation than popular scales, such as Hexad [29], for assessing gamification user types, offering an alternative user modeling approach for better tailored gamification. An additional benefit of the MaxDiff method for future studies is to enable further analysis using the total unduplicated reach and frequency technique [53] to obtain an optimized set of game elements that simultaneously considers preferences from all user segments and maximizes the utilities for all users. This can help alleviate the problem of one-size-fits-all approaches when tailored solutions are not readily available.

Limitations

MaxDiff experiments and the latent class segmentation technique for classifying smartwatch users are both novel and promising, but the results of this study should be interpreted in the context of a few limitations. With the burgeoning research attention devoted to gamification in recent years, new innovative game elements might be quickly developed and implemented on a wide range of apps on various platforms, making it challenging to come up with an exhaustive list of all elements for the quantification of individual effects. There are also limitations regarding the maximum number of attributes allowed to be tested in MaxDiff experiments. Therefore, based on a comprehensive literature review and a preliminary study [51], we tried our best to synthesize a total of 16 popular and important game elements, which hopefully could represent current mainstream trends in gamification research and practice and were particularly applicable to smartwatch-based fitness apps. If needed, future studies can continue to investigate novel game elements not considered in this research that might rise in popularity and deserve more attention.

Another limitation is that, given the exploratory nature of this study and to achieve a straightforward categorization of smartwatch users, our user segmentation modeling scheme did not include user characteristics beyond their preferences and needs for gamification. While more complex user models can be developed in the future, we did evaluate each user segment in terms of related attitudes and behaviors, which added details to the characteristics of each segment and also helped us determine key variables for predicting segment membership. This information provides the basis for the next key step in tailored gamification, that is, to transform the theoretical understanding of individual differences to apps that can detect them efficiently and provide tailored solutions accordingly in real practice.

Conclusions

Despite the increasingly widespread use of gamification in digital health apps, empirical evidence on its effectiveness remains promising but inconclusive due to a lack of methodological rigor and the insufficient consideration of individual differences in many studies. To our knowledge, this study is the first to investigate MaxDiff-based user segmentation for tailored gamification on smartwatches promoting physical exercise—one of the most frequently encouraged yet inadequately adopted health behaviors worldwide. By quantifying the unique importance of each individual game element and improving the accuracy of user segmentation, our research identified 3 segments of smartwatch users based on their preferences for gamification in fitness apps. More importantly, we uncovered 4 segments motivated by goals, immersive experiences, rewards, or social comparison. Such user heterogeneity confirmed the susceptibility of the effects of gamification [8] and informed researchers and practitioners of the need to match gamified solutions with user characteristics for ensuring the effectiveness of gamification. As existing tailored gamification studies continue to explore ways of user modeling with mostly surveys and questionnaires [25], this study supported the adoption of MaxDiff experiments as an alternative method and contributed to the methodological rigor of this emerging area of inquiry, particularly in the health domain and for more application types beyond smartphones.

Acknowledgments

This study was funded by the National Social Science Fund of China (20BTY120), the Guangdong Philosophy and Social Science Fund (GD22CJY05), the Guangdong Adolescent Research Fund (2024WT007), and the Shenzhen Social Science Fund (SZ2022B008).

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request.

XSL•FO
RenderX

Authors' Contributions

JY, DS, and TX conceptualized the study. JY and TX acquired funding and designed the methodology. DS and JZ provided project administration, collected data, and created the visualizations. DS, JY, and TX performed the formal analysis. JY and DS wrote the original draft of the manuscript. JY, DS, TX, and JZ reviewed and edited the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Smartwatch-based gamification questionnaire.
[DOCX File , 481 KB - games_v13i1e66793_app1.docx]

References

1. Stevens G, Mascarenhas M, Mathers C. Global health risks: progress and challenges. Bull World Health Organ 2009 Sep 1;87(9):646 [FREE Full text] [doi: 10.2471/blt.09.070565] [Medline: 19784438]
2. Physical activity fact sheet. World Health Organization. 2021 Oct 12. URL: https://www.who.int/publications/i/item/WHO-HEP-HPR-RUN-2021.2 [accessed 2025-02-12]
3. Chekroud SR, Gueorguieva R, Zheutlin AB, Paulus M, Krumholz HM, Krystal JH, et al. Association between physical exercise and mental health in 1·2 million individuals in the USA between 2011 and 2015: a cross-sectional study. Lancet Psychiatry 2018 Sep;5(9):739-746. [doi: 10.1016/S2215-0366(18)30227-X] [Medline: 30099000]
4. Elbe AM, Lyhne SN, Madsen EE, Krustrup P. Is regular physical activity a key to mental health? Commentary on "Association between physical exercise and mental health in 1.2 million individuals in the USA between 2011 and 2015: a cross-sectional study", by Chekroud et al., published in. J Sport Health Sci 2019 Jan;8(1):6-7 [FREE Full text] [doi: 10.1016/j.jshs.2018.11.005] [Medline: 30719377]
5. Segar M. No Sweat: How the Simple Science of Motivation Can Bring You a Lifetime of Fitness. New York, NY: AMACOM; 2015.
6. Werbach K. (Re)defining gamification: a process approach. In: Proceedings of the 9th International Conference on Persuasive Technology - Persuasive, Motivating, Empowering Videogames. 2014 Presented at: Persuasive '14; May 21-23, 2014; Padua, Italy. [doi: 10.1007/978-3-319-07127-5_23]
7. Oinas-Kukkonen H. A foundation for the study of behavior change support systems. Pers Ubiquit Comput 2012 Jul 19;17(6):1223-1235. [doi: 10.1007/s00779-012-0591-5]
8. Koivisto J, Hamari J. The rise of motivational information systems: a review of gamification research. Int J Inf Manag 2019 Apr;45:191-210. [doi: 10.1016/j.ijinfomgt.2018.10.013]
9. Deci EL, Ryan RM. Intrinsic Motivation and Self-Determination in Human Behavior. New York, NY: Springer; 1985.
10. Fogg BJ. A behavior model for persuasive design. In: Proceedings of the 4th International Conference on Persuasive Technology. 2009 Presented at: Persuasive '09; April 26-29, 2009; Claremont, CA.
11. Deterding S, Dixon D, Khaled R, Nacke L. From game design elements to gamefulness: defining "gamification". In: Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments. 2011 Presented at: MindTrek '11; September 28-30, 2011; Tampere, Finland. [doi: 10.1145/2181037.2181040]
12. Cugelman B. Gamification: what it is and why it matters to digital health behavior change developers. JMIR Serious Games 2013 Dec 12;1(1):e3 [FREE Full text] [doi: 10.2196/games.3139] [Medline: 25658754]
13. Alahäivälä T, Oinas-Kukkonen H. Understanding persuasion contexts in health gamification: a systematic analysis of gamified health behavior change support systems literature. Int J Med Inform 2016 Dec;96:62-70. [doi: 10.1016/j.ijmedinf.2016.02.006] [Medline: 26944611]
14. Huang CK, Chen CD, Liu YT. To stay or not to stay? Discontinuance intention of gamification apps. Inf Technol People 2019 Dec 02;32(6):1423-1445. [doi: 10.1108/itp-08-2017-0271]
15. Direito A, Jiang Y, Whittaker R, Maddison R. Apps for IMproving FITness and increasing physical activity among young people: the AIMFIT pragmatic randomized controlled trial. J Med Internet Res 2015 Aug 27;17(8):e210 [FREE Full text] [doi: 10.2196/jmir.4568] [Medline: 26316499]
16. Johnson D, Deterding S, Kuhn KA, Staneva A, Stoyanov S, Hides L. Gamification for health and wellbeing: a systematic review of the literature. Internet Interv 2016 Nov;6:89-106 [FREE Full text] [doi: 10.1016/j.invent.2016.10.002] [Medline: 30135818]
17. Xu L, Shi H, Shen M, Ni Y, Zhang X, Pang Y, et al. The effects of mHealth-based gamification interventions on participation in physical activity: systematic review. JMIR Mhealth Uhealth 2022 Feb 03;10(2):e27794 [FREE Full text] [doi: 10.2196/27794] [Medline: 35113034]
18. Mazeas A, Duclos M, Pereira B, Chalabaev A. Evaluating the effectiveness of gamification on physical activity: systematic review and meta-analysis of randomized controlled trials. J Med Internet Res 2022 Jan 04;24(1):e26779 [FREE Full text] [doi: 10.2196/26779] [Medline: 34982715]

XSL•FO
RenderX

19. Nishi SK, Kavanagh ME, Ramboanga K, Ayoub-Charette S, Modol S, Dias GM, et al. Effect of digital health applications with or without gamification on physical activity and cardiometabolic risk factors: a systematic review and meta-analysis of randomized controlled trials. EClinicalMedicine 2024 Oct;76:102798 [FREE Full text] [doi: 10.1016/j.eclinm.2024.102798] [Medline: 39764571]
20. Nacke LE, Deterding S. The maturing of gamification research. Comput Hum Behav 2017 Jun;71:450-454. [doi: 10.1016/j.chb.2016.11.062]
21. de-Marcos L, Garcia-Lopez E, Garcia-Cabot A. On the effectiveness of game-like and social approaches in learning: comparing educational gaming, gamification and social networking. Comput Educ 2016 Apr;95:99-113. [doi: 10.1016/j.compedu.2015.12.008]
22. Zhao Z, Arya A, Orji R, Chan G. Effects of a personalized fitness recommender system using gamification and continuous player modeling: system design and long-term validation study. JMIR Serious Games 2020 Nov 17;8(4):e19968 [FREE Full text] [doi: 10.2196/19968] [Medline: 33200994]
23. Schmidt-Kraepelin M, Thiebes S, Sunyaev A. Investigating the relationship between user ratings and gamification: a review of mHealth apps in the Apple App Store and Google Play Store. In: Proceedings of the 52nd Hawaii International Conference on System Sciences. 2019 Presented at: HICSS 2019; January 8-11, 2019; Maui, HI. [doi: 10.24251/HICSS.2019.181]
24. Seaborn K, Fels DI. Gamification in theory and action: a survey. Int J Hum Comput Stud 2015 Feb;74:14-31. [doi: 10.1016/j.ijhcs.2014.09.006]
25. Klock AC, Gasparini I, Pimenta MS, Hamari J. Tailored gamification: a review of literature. Int J Hum Comput Stud 2020 Dec;144:102495. [doi: 10.1016/j.ijhcs.2020.102495]
26. Krath J, Palmquist A, Jedel I, Barbopoulos I, Helmefalk M, Munkvold RI. Does behaviour match user typologies? An exploratory cluster analysis of behavioural data from a gamified fitness platform. In: Proceedings of the 6th International GamiFIN Conference. 2022 Presented at: GamiFIN 2022; April 26-29, 2022; Tampere, Finland URL: http://ceur-ws.org/Vol-3147/paper11.pdf
27. Cohen S, Neira L. Measuring preference for product benefits across countries: overcoming scale usage bias with maximum difference scaling. In: Proceedings of the ESOMAR Conference 2003. 2003 Presented at: ESOMAR 2003; January 1, 2003; Prague, Czech Republic.
28. Tondello GF, Wehbe RR, Diamond L, Busch M, Marczewski A, Nacke LE. The gamification user types hexad scale. In: Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play. 2016 Presented at: CHI PLAY '16; October 16-19, 2016; Austin, TX. [doi: 10.1145/2967934.2968082]
29. Tondello GF, Mora A, Marczewski A, Nacke LE. Empirical validation of the gamification user types hexad scale in English and Spanish. Int J Hum Comput Stud 2019 Jul;127:95-111. [doi: 10.1016/j.ijhcs.2018.10.002]
30. Santos AC, Oliveira W, Hamari J, Joaquim S, Isotani S. The consistency of gamification user types: a study on the change of preferences over time. Proc ACM Hum Comput Interact 2023 Oct 04;7(CHI PLAY):1253-1281. [doi: 10.1145/3611068]
31. Thurstone LL. A law of comparative judgment. Psychol Rev 1994 Apr;101(2):266-270. [doi: 10.1037/0033-295X.101.2.266]
32. Auger P, Devinney TM, Louviere JJ. Using best–worst scaling methodology to investigate consumer ethical beliefs across countries. J Bus Ethics 2006 Oct 17;70(3):299-326. [doi: 10.1007/s10551-006-9112-7]
33. Heo CY, Kim B, Park K, Back RM. A comparison of Best-Worst Scaling and Likert Scale methods on peer-to-peer accommodation attributes. J Bus Res 2022 Sep;148:368-377. [doi: 10.1016/j.jbusres.2022.04.064]
34. Flynn TN, Huynh E. BWS profile case application: preferences for quality of life in Australia. In: Marley AA, Louviere JJ, Flynn TN, editors. Best-Worst Scaling: Theory, Methods and Applications. Cambridge, UK: Cambridge University Press; 2015.
35. Oliveira W, Hamari J, Shi L, Toda AM, Rodrigues L, Palomino PT, et al. Tailored gamification in education: a literature review and future agenda. Educ Inf Technol 2022 Jun 29;28(1):373-406. [doi: 10.1007/s10639-022-11122-4]
36. Hong Y, Saab N, Admiraal W. Approaches and game elements used to tailor digital gamification for learning: a systematic literature review. Comput Educ 2024 Apr;212:105000. [doi: 10.1016/j.compedu.2024.105000]
37. Schöbel S, Söllner M, Leimeister JM. The agony of choice – analyzing user preferences regarding gamification elements in learning management systems. In: Proceedings of the International Conference on Information Systems. 2016 Presented at: ICIS 2016; December 11-14, 2016; Dublin, Ireland URL: https://aisel.aisnet.org/icis2016/HumanBehavior/Presentations/10/
38. Schmidt-Kraepelin M, Thiebes S, Schöbel S, Sunyaev A. Users' game design element preferences in health behavior change support systems for physical activity: a best-worst-scaling approach. In: Proceedings of the 40th International Conference on Information Systems. 2019 Presented at: ICIS 2019; December 15-18, 2019; Munich, Germany. [doi: 10.5445/IR/1000120472]
39. Smartwatches - statistics and facts. Statista. URL: https://www.statista.com/topics/4762/smartwatches/#topicOverview [accessed 2025-02-12]
40. Nelson EC, Verhagen T, Noordzij ML. Health empowerment through activity trackers: an empirical smart wristband study. Comput Hum Behav 2016 Sep;62:364-374. [doi: 10.1016/j.chb.2016.03.065]

XSL•FO
RenderX

41. Longhini J, Marzaro C, Bargeri S, Palese A, Dell'Isola A, Turolla A, et al. Wearable devices to improve physical activity and reduce sedentary behaviour: an umbrella review. Sports Med Open 2024 Jan 14;10(1):9 [FREE Full text] [doi: 10.1186/s40798-024-00678-9] [Medline: 38219269]
42. Marakhimov A, Joo J. Consumer adaptation and infusion of wearable devices for healthcare. Comput Hum Behav 2017 Nov;76:135-148. [doi: 10.1016/j.chb.2017.07.016]
43. Talukder MS, Chiong R, Bao Y, Hayat Malik B. Acceptance and use predictors of fitness wearable technology and intention to recommend. Ind Manag Data Syst 2019 Feb 04;119(1):170-188. [doi: 10.1108/imds-01-2018-0009]
44. Niknejad N, Ismail WB, Mardani A, Liao H, Ghani I. A comprehensive overview of smart wearables: the state of the art literature, recent advances, and future challenges. Eng Appl Artif Intell 2020 Apr;90:103529. [doi: 10.1016/j.engappai.2020.103529]
45. Attig C, Franke T. Abandonment of personal quantification: a review and empirical study investigating reasons for wearable activity tracking attrition. Comput Hum Behav 2020 Jan;102:223-237. [doi: 10.1016/j.chb.2019.08.025]
46. Canhoto AI, Arp S. Exploring the factors that support adoption and sustained use of health and fitness wearables. J Mark Manag 2016 Oct 26;33(1-2):32-60. [doi: 10.1080/0267257X.2016.1234505]
47. Morschheuser B, Hassan L, Werder K, Hamari J. How to design gamification? A method for engineering gamified software. Inf Softw Technol 2018 Mar;95:219-237. [doi: 10.1016/j.infsof.2017.10.015]
48. Patel MS, Asch DA, Volpp KG. Wearable devices as facilitators, not drivers, of health behavior change. JAMA 2015 Feb 03;313(5):459-460. [doi: 10.1001/jama.2014.14781] [Medline: 25569175]
49. Kniestedt I, Maureira MA. Little fitness dragon: a gamified activity tracker. In: Proceedings of the 15th IFIP TC 14 International Conference on Entertainment Computing. 2016 Presented at: ICEC 2016; September 28-30, 2016; Vienna, Austria. [doi: 10.1007/978-3-319-46100-7_18]
50. Zhao Z, Etemad SA, Whitehead A, Arya A. Motivational impacts and sustainability analysis of a wearable-based gamified exercise and fitness system. In: Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts. 2016 Presented at: CHI PLAY Companion '16; October 16-19, 2016; Austin, TX. [doi: 10.1145/2968120.2987726]
51. Song D, Yao J. User preferences for gamification elements in smartwatch fitness applications. In: Proceedings of the 2022 2nd International Conference on Computer Technology and Media Convergence Design. 2022 Presented at: CTMCD 2022; May 13-15, 2022; Dali, China. [doi: 10.2991/978-94-6463-046-6_22]
52. Attig C, Franke T. I track, therefore I walk – exploring the motivational costs of wearing activity trackers in actual users. Int J Hum Comput Stud 2019 Jul;127:211-224. [doi: 10.1016/j.ijhcs.2018.04.007]
53. MaxDiff results and simulator. Sawtooth Software. URL: https://sawtoothsoftware.com/help/maxdiff-analyzer/manual/ [accessed 2025-02-13]
54. Hassan L, Rantalainen J, Xi N, Pirkkalainen H, Hamari J. The relationship between player types and gamification feature preferences. In: Proceedings of the 4th International GamiFIN Conference. 2020 Presented at: GamiFIN2020; April 1-3, 2020; Levi, Finland.
55. Werbach K, Hunter D. For the Win. Philadelphia, PA: Wharton Digital Press; 2012.
56. Chen J, Mullins CD, Novak P, Thomas SB. Personalized strategies to activate and empower patients in health care and reduce health disparities. Health Educ Behav 2016 Feb 06;43(1):25-34 [FREE Full text] [doi: 10.1177/1090198115579415] [Medline: 25845376]
57. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb 17;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
58. Chen XS, Changolkar S, Navathe AS, Linn KA, Reh G, Szwartz G, et al. Association between behavioral phenotypes and response to a physical activity intervention using gamification and social incentives: secondary analysis of the STEP UP randomized clinical trial. PLoS One 2020 Oct 14;15(10):e0239288 [FREE Full text] [doi: 10.1371/journal.pone.0239288] [Medline: 33052906]
59. Rhodes RE, Janssen I, Bredin SS, Warburton DE, Bauman A. Physical activity: health impact, prevalence, correlates and interventions. Psychol Health 2017 Aug 30;32(8):942-975. [doi: 10.1080/08870446.2017.1325486] [Medline: 28554222]
60. Liu D, Santhanam R, Webster J. Toward meaningful engagement: a framework for design and research of gamified information systems. MIS Q 2017;41(4):1011-1034. [doi: 10.25300/misq/2017/41.4.01]
61. Kuklys W. Stated choice methods: analysis and application, Jordan J. Louviere, David A. Hensher and Joffre D. Swait, Cambridge University Press, ISBN: 0 - 521 - 78830 - 7. J Appl Econom 2002 Dec 09;17(6):701-704. [doi: 10.1002/jae.701]
62. Orme BK. Getting Started with Conjoint Analysis: Strategies for Product Design and Pricing Research. Madison, WI: Research Publishers; 2006.

Abbreviations

MANOVA: multivariate analysis of variance
MaxDiff: maximum difference scaling

XSL•FO
RenderX

RQ: research question

Edited by A Coristine; submitted 24.09.24; peer-reviewed by J Ou, M Fraser; comments to author 16.12.24; revised version received 12.02.25; accepted 10.03.25; published 18.04.25.

Please cite as:
Yao J, Song D, Xiao T, Zhao J
Smartwatch-Based Tailored Gamification and User Modeling for Motivating Physical Exercise: Experimental Study With the Maximum Difference Scaling Segmentation Method
JMIR Serious Games 2025;13:e66793
URL: https://games.jmir.org/2025/1/e66793
doi:10.2196/66793
PMID:40067118

©Jie Yao, Di Song, Tao Xiao, Jiali Zhao. Originally published in JMIR Serious Games (https://games.jmir.org), 18.04.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Digital Catalysts for Noncommunicable Disease Prevention Serious Games and Gamified Applications: Framework Design Study

Christoph Aigner[1], MSc; René Baranyi[1], PhD; Thomas Grechenig[1], Prof Dr

Research Group for Industrial Software, INSO, TU Wien, Wien, Austria

Corresponding Author:
Christoph Aigner, MSc
Research Group for Industrial Software, INSO
TU Wien
Wiedner Hauptstraße 76/2
Wien, 1040
Austria
Phone: 43 1 5872197
Fax: 43 1 5872198
Email: christoph.aigner@inso-world.com

Abstract

Background: Unhealthy behaviors can cause so-called noncommunicable diseases (NCDs), which are on the rise. Notable examples include chronic respiratory diseases, diabetes, cardiovascular diseases, and various types of cancer. They are responsible for approximately 41 million deaths annually, which accounts for a staggering 74% of all global deaths. Major risk factors include physical inactivity, the use of tobacco, unhealthy diets, the harmful use of alcohol, and poor mental health, which can be classified as modifiable behavioral risk factors. Other factors include metabolic and environmental risk factors, such as air pollution. Many individuals struggle to make informed decisions about their health, which contributes to the risk factors mentioned earlier and, ultimately, can lead to the development of one or more NCD.

Objective: This research presents design and standardization considerations to enable the exchange of medical and game data to maximize their impact and usefulness. Serious games and gamified applications that strategically use behavior change techniques and educational content can help users change their behavior on a lasting basis, thereby reducing the aforementioned NCD risk factors. Still, each of them is currently independently designed and cannot interact with other applications.

Methods: We previously developed serious games and gamified applications to prevent NCDs. These served as the foundation of an interoperable framework for NCD prevention games and applications. On the basis of a comprehensive analysis, 6 key areas were identified, ultimately leading to a framework definition that was then evaluated against the already-developed games and applications.

Results: This paper presented a novel interoperable framework to support the design and development of serious games and gamified applications that enable individuals to achieve sustainable behavior change and improve their overall health and well-being by defining 6 key areas, emphasizing interoperability, and exchanging meaningful medical and game data.

Conclusions: The framework presented in this study covers the major design and implementation aspects of NCD prevention games and applications in 6 key areas. Therefore, researchers should consider these guidelines when creating novel serious games and applications in those areas. The framework also intensively encourages the use of standards in the domain of medical informatics to ensure the semantic interoperability of patients' data produced. Thus, it promotes the exchange of meaningful data to improve patient care and anonymous data use for research.

(JMIR Serious Games 2025;13:e69246) doi:10.2196/69246

KEYWORDS

serious games; gamification; framework; interoperability; noncommunicable disease prevention

XSL•FO
RenderX

Introduction

Problem Description

Many people find it hard to make conscious decisions about their health, including their physical and mental well-being. Whether it is about eating healthy food, quitting smoking, reducing alcohol consumption, exercising, or maintaining good mental health, changing one's behavior for the long term can be challenging. According to Bouton [1], behavior change is difficult because new behavior does not erase the old behavior, and new learnings can be relatively specific to the context in which they were learned. Coorevits and Coenen [2] stated that there is a high attrition rate among consumers who no longer wear their fitness trackers. The study by Alshawmar et al [3] had similar findings regarding fitness apps. They observed that, despite the potential benefits, many downloads, and high use rates of fitness apps, studies indicate a significant dropout rate after a short period of use [3]. Even if someone manages to make substantial changes, they often fail to last. Therapies such as psychotherapy and dietary treatment are promising in helping people achieve lasting behavior change, but they can be expensive and may only sometimes be covered by health insurance. For instance, there is compelling evidence of income-based disparities in access to psychotherapy and other mental health services, as seen in current conditions in Australia and Canada [4].

Noncommunicable diseases (NCDs) are on the rise. They are not contagious but can be caused by unhealthy behaviors [5]. Some major examples include cardiovascular diseases, cancers, diabetes, and chronic respiratory diseases. These diseases are responsible for an estimated 41 million deaths per year, which is about 74% of all deaths globally [6]. Major risk factors include physical inactivity, the use of tobacco, unhealthy diets, the harmful use of alcohol, and poor mental health. These can be classified as modifiable behavioral risk factors. Other factors include metabolic factors, such as increased blood pressure, overweight, hyperglycemia, and hyperlipidemia, and environmental risk factors, such as air pollution. To effectively combat NCDs, it is crucial to reduce these risk factors [6]. In the work of Zhao et al [7], it was shown that while increasing age remains a nonmodifiable risk factor for cancer, and thus for an NCD, the incidence for patients aged <50 years is rising. Early-onset cancer deaths increased by a tremendous 79.1% between 1990 and 2019. According to Zhao et al [7], risk factors identified for early-onset cancers are very similar to the ones specified by the World Health Organization for NCDs.

Zhao et al [7] also projected that the global incidence would increase by 31% in 2030. Their work clearly shows that NCD prevention would be very beneficial, even for younger age groups. Another aspect that was long overlooked is mental health disorders in combination with NCD prevention. O'Neil et al [8] showed that NCD prevention and control have historically focused on cancer, cardiovascular diseases, diabetes, and chronic respiratory diseases. Despite recent developments in the integration of common mental illnesses, the prevention and control of these disorders remain largely separate and independent; therefore, O'Neil et al [8] proposed a shared framework for disease prevention and control. An article on linking mental health and NCD alliance campaign priorities, created for the 2018 United Nations High-Level meeting on the Prevention of NCDs, also proposed that governments should address the common risk factors and systems barriers to reducing premature and preventable distress and death by integrating mental health into the response to NCDs [9]. According to Jacka et al [10], there is evidence that could indicate that unhealthy diets can also be risk factors for mental health disorders, particularly depression and dementia, and therefore can be considered a multiplier for the risk of developing NCDs.

Dörner et al [11] described serious games as digital games designed to entertain and achieve additional goals, known as characterizing goals. These additional goals can be related to learning, training, education, health-related use cases, or other applications. Numerous categories of serious games exist within the digital health domain, including exergames promoting physical activity, medical education games and simulations, lifestyle behavior change games, and medical diagnosis learning games. Serious games can take various forms, such as simulations, role-playing, and educational games [12]. They incorporate textual, graphical, haptic, and audio elements to create immersive gaming experiences [13,14]. It is important to note that serious games are distinguished from profession simulators and virtual reality apps due to their goal-oriented nature [12].

It is also essential to differentiate between a serious game and the terms gamification or gamified application. Gamification uses game elements within nongaming systems to enhance user experience and engagement. This concept involves adding game-like elements to nongaming applications and platforms, a process known as "gamifying." According to Dörner et al [11], gamified applications are typically, but not necessarily, less like a game than full serious games, as they may only incorporate specific game elements or mechanics. Table 1 presents a short overview, distinguishing the 2 terms.

Table 1. Comparison of the terms "serious game" and "gamified application."

Term	Definition	Key characteristics
Serious game	A game that is designed for a primary purpose other than entertainment; used in settings such as classrooms, workplaces, or health care [12]	Digital game of any particular game genre; categorization according to the characterizing goal [11]
Gamified application	Transfer of game methodologies or elements to nongame applications and processes [11]	Adds game elements to a nongame area; the result of gamification is not necessarily a game [11]

XSL•FO
RenderX

A key way for individuals to lower the risk of NCDs is by modifying their lifestyle. Although people are becoming more aware of medical conditions and risk factors involved with NCDs and therefore are more inclined to change their lifestyles, which can be partially attributed to the impact of mass media, most are still finding it hard to change their behavior and maintain it for an extended period. The so-called behavior change techniques (BCTs) can be used to overcome the problems described earlier. Michie et al [15] define BCTs as structured interventions to modify behavior incorporated as active components in an intervention. For example, the goal-setting BCT can be used to define a daily walking goal or to specify how many fruits or vegetables one should consume per day. Another example of a BCT is the self-monitoring of behavior BCT. With this BCT, the individual could be asked to record, in a personal diary, whether they have engaged in a daily 2-minute toothbrushing routine before bedtime [16]. These BCTs were successfully used in serious games and gamified mobile apps. Baranowski et al [17] concluded that serious game–based behavior change is an exciting form of media-based intervention, and the many desirable outcomes warrant moving forward in this area.

Semantic interoperability can be described as the ability to automatically interpret information that is exchanged meaningfully and accurately to produce useful results as defined by the end users of both systems [18]. Health data produced by games and apps must be defined in an interoperable way. As a solution, standards within the domain of medical informatics, such as Health Level 7 (HL7), Integrating the Healthcare Enterprise (IHE), and openEHR, can be used. Interoperability is hard to achieve because it needs to be ensured that the understanding works horizontally across business processes and also vertically within computer system suppliers and between users and developers who speak different dialects [19]. Oemig and Snelick [20] emphasized that clear standards and adherence to these standards are essential for creating reliable, effective, usable, and interoperable health care information systems, highlighting the crucial role of standards. However, Peters et al [21] stated that patients and their caretakers would benefit considerably from integrating semantically rich and standardized health data into existing hospital information systems or electronic health records (EHRs).

We have already published various serious games and gamified applications for NCD prevention and showed that these can be useful and are effective in behavior change. Furthermore, we have already published a baseline for an interoperable framework, comprising a comparative analysis and the definition of a requirement catalog [22]. Each of the developed application's results, conclusions, findings, and additional analyzed projects were the foundation for constructing an interoperable application framework for behavior change–driven serious games and gamified mobile applications to combat NCDs presented in this paper. The framework includes best practices for designing and implementing games and applications specific to preventing NCDs. It also intensively encourages the use of standards within the domain of medical informatics to ensure the semantic interoperability of patients' data produced from the applications and games mentioned earlier.

Related Work

Frameworks for Serious Games and Gamification

Peters et al [21] proposed a framework for collecting health-related and game data from serious games and applications that cover the requirements to interpret data from various health and game sources using a decision support system (DSS)–based approach. The initial motivation for creating the framework was the Integrating Entertainment and Reaction Assessment into Child Cancer Therapy (INTERACCT) project, a serious game for young oncology patients. The game world consists of a 2.5D adventure game where the player has to complete procedurally generated levels filled with hidden treasures and hostile avatars. A huge part of the framework's design is a DSS that calculates structured health scores based on the medical and game data. Peters et al [21] stated that it was challenging to design a ubiquitous data model for serious games for health, mainly because of the wide variety of possible games, mechanics, and medical topics. This observation shows the importance of well-made implementation guides (IGs) focusing on well-defined medical areas, such as NCD prevention, and using current interoperability standards. The framework suggests using standards, but no IGs or interoperability profiles are provided.

Yusoff et al [23] developed a conceptual framework for serious learning games. It combines learning and pedagogy theory with gaming requirements. The framework includes components such as "capability," which refers to the cognitive, psychomotor, and affective skills players develop in the serious game; instructional content; intended learning outcomes; game attributes supporting engagement; learning activities, reflection on these outcomes; genre; game mechanics; and achievements, such as scores, resources collected, or time to reach goals.

The eAdventure platform is a tool for creating educational adventure games for various domains, including health sciences. It integrates into Moodle-like platforms using the Sharable Content Object Reference Model (SCORM) but solely focuses on one type of video game [24].

Another example is the work of Stanescu et al [25], which proposes a multidimensional interoperability framework for serious games that integrates 3 key dimensions: the core components, the ecosystem, and external factors. According to them, the research on serious game interoperability should focus on standardization, which means the creation of functionally interchangeable items; interchangeability, which refers to identifying methods that would make game components interchangeable; standards adoption, which involves creating adaptable solutions; open system architecture, which provides a modular design that defines key interfaces; and unique specifications and proprietary devices, which indicate that proprietary devices may be counterproductive but are often necessary to provide the needed functionality. Their work focuses on serious learning games.

Cowan and Kapralos [26] focused on technical insights into serious game development by conducting an extensive literature

XSL•FO
RenderX

review that examined frameworks and game engines, including a summary of their features. According to their study, the frameworks Second Life, Unity, Unreal, Flash, XNA Game Studio, and the Torque Game Engine were extensively used because they had the highest search results. They concluded that developers primarily used technical frameworks that were not originally designed for serious games but for commercial entertainment games.

Interoperability Within Serious Games and Gamified Applications

Digital health applications, especially mobile apps, can (as of January 2023) export their data in a semantically interoperable way into the German national EHR. The term "DiGA" (digital health applications; German: "Digital Gesundheitsanwendungen") is a German term [27]. The DiGA Toolkit project, currently released as version 1.1.0, allows storing data in the EHR (Elektronische Patientenakte) using HL7 Fast Healthcare Interoperability Resources (FHIR) as part of the so-called Medizinische Informationsobjekte [27]. According to Mittermaier et al [28], physicians can already prescribe authorized DiGAs to patients, which can be reimbursed by statutory health insurance. They also stated that DiGAs can support the field of internal medicine. Although the DiGA profiles already define useful data structures for digital health applications, such as diary entries, questionnaires, and lifestyle factors, they currently do not focus on segments for serious games or gamification for health.

The German Institute for Standardization Registered Association (DIN; German: Deutsches Institut für Normung eV) published a specification for serious game metadata in 2018. This specification defines a standardized format for developers and publishers to describe their games so that potential users can find them more easily [29].

SCORM is a standard that can be used with learning games. It outlines a method for delivering e-learning content across various platforms. Its key components include the content aggregation model, which establishes a framework for packaging learning content, and the runtime environment, which specifies an interface to facilitate communication between the learning content and the system that initiates it [30].

Research Gaps

The framework proposed in this paper defines guidelines for designing and developing games and applications in 6 key areas. It emphasizes semantic interoperability by incorporating state-of-the-art standards from the medical informatics domain. Furthermore, it will predominantly focus on behavior change–driven games and gamified mobile apps to prevent NCDs. Both are novel approaches within serious games and gamification in health. Although rudiment framework approaches already exist, they either focus more on serious games for education or e-learning, such as the SCORM standard, or general serious games for health and try to use DSS approaches without constructing actual IGs for the use within national or personal EHRs and thus not allowing medical and game data to be further used for primary care or secondary use within science.

Although IGs for the interoperable exchange of medical data exist for mobile health apps, such as the DiGA Toolkit in Germany [31], they do not specifically focus on serious health games or gamified health apps. They also do not currently use the latest version of HL7 FHIR, which includes newly added resources relevant to serious game data for NCD prevention.

Summary

Serious games for health, particularly within NCD prevention, have been shown to be effective; however, solutions are scattered, and standardization is still lacking. Because current frameworks for serious games and gamified health applications within this field notably lack IGs for integrating medical and game data into EHRs, and current mobile health app conventions do not address serious health games or use the latest medical informatics standards for NCD prevention, this study proposes a novel interoperable framework for serious games and gamified applications aimed at promoting sustainable behavior change and enhancing health, thereby preventing NCDs. It outlines best practices for designing and implementing applications, facilitating the structured exchange of data for personal or national EHRs, and supporting clinical research. Built on previously published serious games and gamified applications, the framework began with an analysis and a best-practice catalog, which was shared at the International Conference on Healthcare Service Management in mid-2024 [22]. These findings then served as the foundation for the analysis, design, and development of the technical framework, which is divided into 6 key areas considered relevant for NCD prevention serious games and applications. Our claim is to develop a high standard for this field that can be used as a starting point for others.

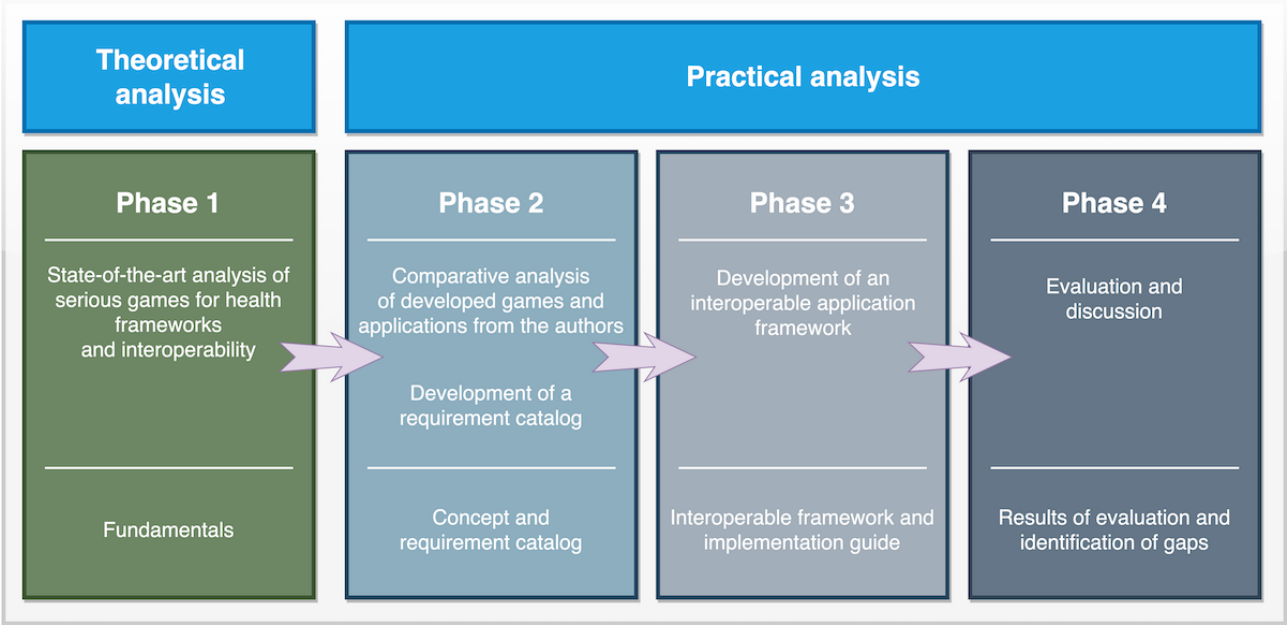
Methods

Overview

The methodological approach for this work consisted of 4 principal phases, as shown in Figure 1. The first 2 phases, the fundamental research, including the study of the state of the art, and the analysis of developed games and applications, which led to the definition of a requirement catalog, were partially completed within the initial work that had already been published as part of a conference paper [22]. The Results section refines and summarizes these 2 phases in the context of this particular work. In the third phase, the final interoperable framework was designed and developed based on the findings of the first 2 phases. The results were a collection of best practices and an IG using standards defined by HL7 International. In the final phase, the framework was evaluated against the analyzed games and apps, and gaps were identified.

XSL•FO
RenderX

Figure 1. Methodology used in this study.



Ethical Considerations

This work is based solely on the analysis and synthesis of previously published literature and publicly available information. As such, formal approval from an institutional review board or ethics committee was not required.

Results

Overview

This section provides comprehensive insights into designing and developing an interoperable framework for NCD prevention games and applications. The Analysis and Requirement Catalog subsection summarizes the comparative analysis and the definition of the requirement catalog that was made beforehand and has already been published by us. This study was extended with 2 further applications published by some of the authors of this study, MoodBooster, a gamified student mental health app [32], and DiaBeaThis, a gamified self-tracking portal to control the blood glucose level of patients with diabetes [33]. The Definition of the framework subsection describes every aspect of the framework in detail. The framework defines the best practices and design guidelines and provides an IG for the meaningful exchange of medical and game data. The Evaluation subsection evaluates the framework against the analyzed applications and tries to identify gaps.

Analysis and Requirement Catalog

Table 2 presents an overview of the analyzed applications.

Table 2. Summary of the comparative analysis.

ID	Analyzed applications	Type of application	Platform	Main noncommunicable disease risk factor	Secondary (clinical) factors	Studies
G1	BreathIn	Serious game	Android	Smoking	Well-being	Aigner et al [34,35]
G2	Food Pyramid Escape	Serious game	Android	Nutrition	Well-being and nutritional education	Aigner et al [36]
G3	Nutrition Garden	Gamified mobile app	Android	Nutrition	Well-being and nutritional education	Aigner et al [37]
G4	NutritionRush	Serious game	Android	Nutrition	Well-being and nutritional education	Baranyi et al [38]
G5	NutriMine	Serious game (modification)	PC	Nutrition	Well-being and nutritional education	Aigner et al [39]
G6	recoverApp	Gamified mobile app	Android and PC	Mental health	Depression, rehabilitation, injury, and pain	Aigner et al [40]
G7	MoodBooster	Gamified mobile app	Android	Mental health	Well-being and positive psychology	Aigner et al [32]
G8	DiaBeaThis	Gamified web application	Web (PC and mobile)	Hyperglycemia	Well-being	Baranyi et al [33]

XSL•FO
RenderX

The analyzed applications included BreathIn, a serious game for smoking cessation [34,35]; Food Pyramid Escape, a serious learning game for healthier eating [36]; Nutrition Garden, a gamified food tracker [37]; NutritionRush, a serious game for better nutrition [38]; NutriMine, a serious modification for Minecraft, also concerning nutrition [39]; and recoverApp, a gamified app for better mental health for patients in rehabilitation [40]. The newly published gamified app MoodBooster for student mental health [32] and DiaBeaThis, a gamified blood glucose tracking web application published in 2018, relevant to this framework [33], were added to the analysis after the publication of the first part of this work. The 8 selected applications provided illustrative examples and focused especially on nutrition, as overweight and obesity are significant NCD risk factors. According to the World Health Organization [41], 2.5 billion adults aged ≥18 years are overweight, including more than 890 million adults living with obesity, with high prevalence especially in Western countries—67% in the region of the Americas, for example. In addition, the model for the framework is considered an open model that is planned to be continually developed [41].

MoodBooster is a mobile, gamified app that supports good mental health for university students. The prototype runs on the Android (Google LLC) platform. The app tracks mood and stress-related symptoms and includes an educational hub that covers symptoms and the physical and cognitive impact of stress on the human body. The hub also gives students information on how to deal with stress-related symptoms, proper diet, and physical activity. The app also includes features for stress relief, including exercise guides [32]. Furthermore, the app consists of serious minigames. These are Just Breathe! (centered on focused breathing exercises); Draw! (a digitized coloring book to offer relaxation); Wipe Out Worry (a game where the player needs to destroy a castle of predefined negative thoughts); and Mind Your Senses (focusing on the psychological concept of mindfulness). The app includes badges for achieving goals within the minigames and suggested activities. The app produces patient data, mood-tracking data, stress level questionnaire entries, notes, and badges [32].

DiaBeaThis is a web portal accessible via mobile devices or PCs. Upon logging in, users can create diary entries that record their current blood glucose levels and track instances of hypoglycemia and hyperglycemia, carbohydrate intake, insulin injections, and physical activities. Each new diary entry contributes to the accumulation of experience points. These points directly affect the gameplay of the included game, a card game based on War. Two players face each other in multiple rounds and have to present cards. Cards with higher values lead to a win. By gaining experience points, the player levels up and gets access to higher-value cards, thus achieving advantages within the game. DiaBeaThis includes achievements and a leaderboard, which ranks all patients currently registered within the portal. The application records patient data, blood glucose level, amount of carbohydrates consumed, insulin dosage, physical activity duration and description, state of mind via an emoji, and additional notes [33].

A total of 8 applications were analyzed, and the subsequent comparative analysis revealed 14 requirements that should be considered when developing novel gamified apps and serious games for NCD prevention [22].

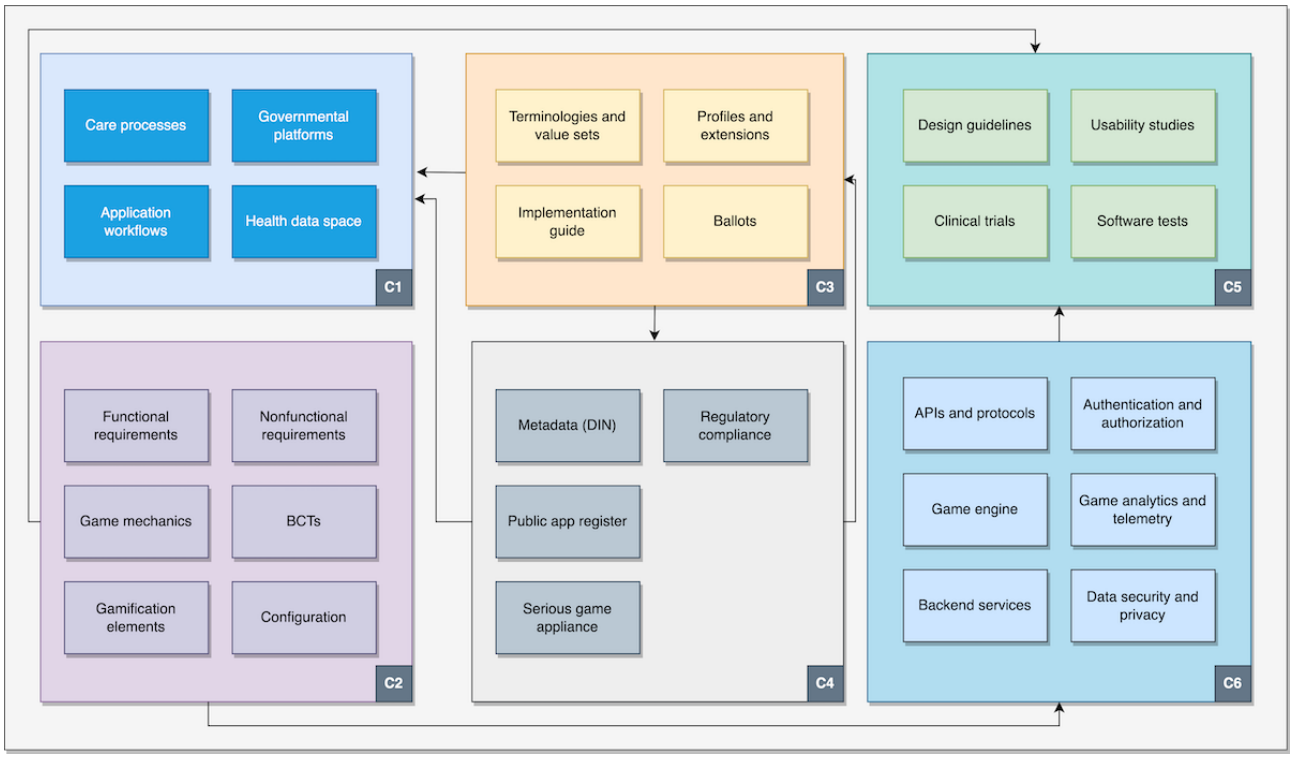
Definition of the Framework

Overview

The comparison study and the definition of basic requirements were the first steps in a process that combined a series of best practices into a framework focusing on the design, development, ecosystem, and interoperability of NCD prevention games and applications. Figure 2 shows the 6 key areas (organizational interoperability component [C1] to technical aspects component [C6]) of the envisioned framework.

XSL•FO
RenderX

Figure 2. Overview of the defined interoperable framework. API: application programming interface; BCT: behavioral change technique; C1: organizational interoperability component; C2: design and development best practices component; C3: semantic interoperability component; C4: serious games ecosystem component; C5: usability and quality aspects component; C6: technical aspects component; DIN: German Institute for Standardization Registered Association (German: Deutsches Institut für Normung eV); NCD: noncommunicable disease.



The framework is separated into 6 major, interconnected components. C1 details recommendations for care processes, application workflows, and governmental platforms and considers national and multinational health data spaces. The design and development best practices component (C2) focuses on best practices for designing and developing games and apps by refining the requirement catalog published beforehand, defining sets of functional and nonfunctional requirements and describing the use of game mechanics, gamification elements, BCTs, and configuration aspects. These best practices strongly connect to usability and quality aspects. The third component (semantic interoperability component) focuses on semantic interoperability and, thus, the interoperable definition and exchange of clinical and game data. It includes the definition of value sets and the meaningful use of medical terminologies, defines technical profiles and extensions, and suggests a basic ballot process for the participation of serious games and application developers. The serious games ecosystem component describes practical guidelines for the target group to find games and applications and use them easily. It includes best practices for using metadata, public app registers, and serious game appliances and also tries to give basic suggestions on regulatory compliance. The usability and quality aspects component (C5) covers usability and quality aspects by defining basic design guidelines and offering recommendations for executing usability and clinical studies. The final component (C6) covers technical aspects of serious game development, including application programming interfaces (APIs), game engine use, authentication and authorization, game analytics and telemetry, backend services, and data security and privacy.

Organizational Interoperability

Care Processes and Application Workflows

Games and apps must be integrated into the care process to support patient behavior change effectively. Health care workers need information about the availability of applications, their intended use, and guidelines for how patients can use them. Therapists and clinicians should be able to prescribe games and applications to patients who could benefit from them. For example, in Germany, applications can already be prescribed as so-called DiGAs [42]. This possibility should be extended to serious games for health. In addition, other countries must adapt so that more patients can benefit from similar opportunities. Patient care plans should determine at which stage of the treatment the application should be prescribed.

Application workflows are very important. BreathIn, for example, should be used as part of an ongoing smoking cessation therapy rather than as a stand-alone intervention. Food Pyramid Escape, by contrast, was designed as a learning app that can be used anytime to teach patients better nutrition and encourage them to change their eating behavior. NCD prevention games should always come with detailed instructions for clinical personnel so that applications can be perfectly integrated into the clinical daily routine.

Therefore, as part of the best practices defined within this framework, developers are strongly encouraged to collaborate with clinicians to define instructions, processes, and workflows during and after development. Because this framework promotes semantically interoperable data extraction into public and personal health records, it is also important to make actual use of the data after patients have finished playing a game or using

XSL•FO
RenderX

an application. These data can then be used within future steps in patients' care plans.

Governmental Platforms

Governments should develop platforms to make NCD prevention games and apps available to a wide audience and integrate them into national health promotion and prevention campaigns. This will contribute to NCD prevention and reduce the risk of developing NCDs.

Health Data Space

Health data space initiatives should include interoperable NCD prevention games and applications to promote the exchange of medical data across national borders. This gives patients greater control over their data and facilitates secondary use of clinical data for research, innovation, and policy making [43]. An example is the European Union (EU) initiative European Health Data Space (EHDS), which is currently being developed [44]. The EHDS regulation aims to establish a common framework for the use and exchange of electronic health data across the EU [44]. EHDS enables researchers to access large-scale health data through a clear and structured system that facilitates the identification of available data, its location, and its quality, ultimately leading to more cost-effective access to high-quality health information [44]. In the context of NCD prevention games and applications, this can include all health and game-related data elements outlined in the IG in an anonymized and secure manner. In addition, policy makers and regulators can benefit from easier, more transparent, and cost-efficient access to electronic health data for public health monitoring, enhanced health care system efficiency, and assured patient safety.

This framework proposed integrating NCD prevention games and applications into such health data spaces, especially because it encourages data exchange in a semantically interoperable way.

Summary

Within this key area, care processes; application workflows; and the integration of government platforms and health data initiatives, such as the EHDS, were explored, as they are essential for the adoption and effectiveness of NCD prevention applications in real-world clinical settings.

Design and Development Best Practices

Requirement Catalog

A requirement catalog for NCD prevention games has already been published as part of the preliminary work on this framework [22]. The catalog was enhanced for this framework's definition by adding the fulfillment level and categorizing the requirements into functional and nonfunctional, as presented in Table 3.

Table 3. Requirement catalog [22].

ID	Description	Requirement level	Requirement type
R01	Integration of behavior change techniques as core game mechanics	Must	Functional requirement
R02	Integration of engaging gamification elements for applications	Must	Functional requirement
R03	Learning and educational aspects	Should	Functional requirement
R04	Emphasizing good interaction design and usability	Should	Nonfunctional requirement
R05	Use of appealing game graphics	May	Nonfunctional requirement
R06	Integration of music and sound effects	May	Nonfunctional requirement
R07	Considering suitable game genres	May	Nonfunctional requirement
R08	Applying intuitive controls	Must	Nonfunctional requirement
R09	Integration of social support and interaction	May	Functional requirement
R10	Availability of games and applications	Must	Nonfunctional requirement
R11	Games and applications should come at a low cost for users	Should	Nonfunctional requirement
R12	Suitable (hardware) platform for low entry barrier	Should	Nonfunctional requirement
R13	Inclusion of therapists and clinicians	Must	Nonfunctional requirement
R14	Ability to extract data in a semantically interoperable way	Must	Functional requirement

For the level of fulfillment, the definition from Request for Comments 2119 was used [45]. "must" means that the definition is an absolute requirement of the specification; "should" means that a requirement should be implemented, but there may be valid reasons in particular circumstances to ignore it; and "may" means the requirement is truly optional. Additional definitions not used within this requirement catalog include "must not" and "should not," which can be understood as the opposite of "must" and "should" [45].

The specific requirement levels chosen for the catalog were self-determined, based on state-of-the-art research and the evaluations of the 8 analyzed games and applications.

Game Mechanics and Gamification Elements

Four projects included in the comparative analysis were applications with gamification elements rather than full-serious games. These applications primarily offer nongaming features, such as a nutrition tracker in Nutrition Garden, blood sugar level tracking in DiaBeaThis, mood tracking in recoverApp and MoodBooster, and gamification elements. Features such as

XSL•FO
RenderX

badges, achievements, challenges, and rewards were used successfully and should be implemented by novel gamified applications.

Some game mechanics are more effective for NCD prevention games than others. For instance, BreathIn and Nutrition Rush use a traditional side-scrolling platformer as their main gameplay. This approach is straightforward to grasp, as it draws from concepts popularized in the early days of video gaming during the 1970s and 1980s. While touch controls on mobile devices function adequately, attention should be given to them, as confusing controls could detract from the gaming experience. Food Pyramid Escape presents an adventure-like game where players directly navigate the main character with an on-screen directional pad. It also uses an inventory and an encyclopedia. NutriMine, a modification for Minecraft, was the only first-person game in the analysis. It was designed to work with the Java version of Minecraft for PC, making it easier to control than the mobile version using a mouse and a keyboard.

Overall, platformer and adventure-type games are best suited for serious NCD prevention games because they are easily accessible and can be controlled intuitively, even on small smartphone screens. In particular, 3D and first-person games may be more challenging to access and control on mobile devices, so they are not encouraged for NCD prevention games and applications.

BCT Guidelines

Serious games can effectively use BCTs to help with actual behavioral changes in their users. The BCT categories used within this framework came from the BCT Taxonomy (version 1), which describes 93 hierarchically clustered techniques [16]. For example, the BCT "information about health consequences" was implemented by all analyzed apps except recoverApp and DiaBeaThis, whether it was information about good nutrition, the hazards of smoking, or the various effects of bad mental health and how to achieve good mental health. Other important BCTs used by all games except Food Pyramid Escape and Nutrition Rush were the "self-monitoring of behavior" and "feedback on behavior" BCTs. They were mostly used for tracking cigarette consumption, food diaries, calorie tracking, blood sugar levels, and mood and stress tracking.

MoodBooster, recoverApp, Nutrition Garden, DiaBeaThis, and BreathIn used the "incentive (outcome)" BCT. It was mostly used in the form of achievements, badges, and (skill) points. In addition, recoverApp used the BCT "social support" by providing a chat feature and the BCT "graded tasks" by implementing the task feature, allowing the therapist to enter therapy-related tasks for their patients to perform. DiaBeaThis uses the BCT "social comparison" because its included game "War" requires 2 players facing each other.

Overall, the BCTs "information about health consequences," "self-monitoring of behavior," and "feedback on behavior" should be implemented in novel games and applications and are therefore recommended as part of this framework, as they were successfully implemented and have been shown to be useful. The "incentive (outcome)" BCT is a good addition and, depending on the game, should be strongly considered when designing new NCD prevention games and applications. Social BCTs such as "social support" and "social comparison" can be successful, but they strongly depend on the app's target group.

Configuration

Configuring certain aspects of serious games and gamified applications can be very beneficial and is therefore encouraged by this framework. This includes setting certain game or level settings, such as speed, time, the player's lives, achievements, and overall difficulty [46], as well as the general possibility of turning certain features on and off. Configuration options can be important regarding social features because some players can be negatively affected by them, especially if they involve social comparison, for example, a leaderboard, high scores, and badges and achievements, as was shown.

Summary

The previous subsections (in the Design and Development Best Practices section) explored the best practices for the design and development of NCD prevention games and applications, especially focusing on key requirements that need to be included when designing novel applications, the integration of game mechanics and gamification elements, how to effectively integrate BCTs, and aspects of configurational options that allow games and applications to adapt to each distinct user.

Semantic Interoperability

Overview

The ability to exchange data from NCD prevention games and applications meaningfully is one of this framework's key aspects. Using the data within public EHRs and personal health records can be very beneficial for patients to archive and record their data and help with ongoing treatment and therapy; therefore, it will be helpful for clinicians as well. The data can also be used anonymously for scientific purposes as part of secondary use. This section describes the developed IG that defines a set of profiles, extensions, and value sets using HL7 FHIR, a standard for exchanging electronic health care data created by the HL7 International organization [19]. The defined profiles encourage the use of current medical terminologies, such as Systematized Nomenclature of Medicine Clinical Terms Clinical Terms for general clinical terms [47] and Logical Observation Identifiers Names and Codes [48] for medical laboratory observations. HL7 FHIR was particularly chosen because of its wide adoption in the United States and Europe as well as its strong focus on implementers and web technology.

IG Structure

The developed IG, a collection of profiles, extensions, value sets, and code systems, defines the structure of documents, which are the units of exchange. It represents all relevant data from a specific serious game and gamified application for NCD prevention using HL7 FHIR. Extensions can be used for additional data elements not defined in the standard. Value sets represent actual elements from one or more code systems.

An "FHIR Document" is an immutable bundle with an attested narrative that serves as the unit of data that is exchanged between systems and persisted in health records [49], using integration profiles such as Cross-Enterprise Document Sharing

XSL•FO
RenderX

(XDSb) defined in the IT infrastructure technical framework by IHE [50]. XDSb is widely used, for example, within Elektronische Gesundheitsakte, the Austrian national health record, and Elektronische Patientenakte, the national health record in Germany, and provides a standards-based specification for the registration, distribution, and access of medical records [51]. It aims to offer a standards-based specification for sharing documents across various health care organizations, from private practices to large clinics. The XDSb profile presumes that each organization is part of one or more XDSb Affinity Domains—groups of health care entities that have mutually agreed to collaborate based on shared policies and infrastructure [51].

Technically, an FHIR document is an FHIR resource of type "bundle" that starts with an FHIR "composition" resource and contains various other resources of a set version that are referenced within that "composition." The HL7 FHIR standard defines a large set of resources that can be used for various clinical and administrative use cases. To further customize them, HL7 FHIR defines profiles that specify acceptable codes, contain new elements in the form of extensions, and specify constraints on existing elements and data types [52].

The collections of profiles, extensions, value sets, and code systems were designed using FHIR Shorthand (FSH). Shorthand is a domain-specific language for defining FHIR profiles. The first full version was released in March 2020. Many FHIR community projects, including the National Health Record Initiative in Austria (Elektronische Gesundheitsakte) [53], adopted FSH.

The transpiler "SUSHI" was used to process the defined FSH files and transform them into FHIR artifacts [54]. The HL7 FHIR IG Publisher tool was used to create a standardized FHIR IG. This exact combination of tools was chosen because the alternative would have been using complex Microsoft Excel spreadsheets or a graphical tool called "Forge," which is only free of charge for single users and only available for Microsoft Windows [52]. FSH has many advantages. It is concise, readable, and understandable, and changes can be made via text operations; because it is text based, it is suited for source code version control, has error checking, and incorporates best practices [52]. The basic requirement for an FSH-SUSHI setup is a recent computer with at least Java 17 and Node.js 18 installed.

Furthermore, the FHIR IG Publisher, written in Java, uses the Jekyll framework to render complete IGs as web pages. The actual shorthand files were created using Visual Studio Code (Microsoft Corporation) because it offered very useful plugins for HL7 FHIR, such as FHIR Tools [55] and FSH language support [56]. Version 5.0.0 of FHIR was used, which included new resources that were of particular interest for this IG, such as the NutritionIntake and NutritionProduct resources.

Each FSH file has the same structure. The first block contains the definition of the profile. The second block defines the parent resource, the title, and the resource description. Finally, the third block represents all the rules the profile should include. Elements can be marked as 'must support' or limited in their cardinality. The profile can also restrict references to specific other profiles. For example, within the NCD patient resource, the reference to the attending physician was limited to the "NCD Practitioner" profile.

Figure 3 gives an overview of the defined data model. All relevant data elements together form the FHIR document, which is technically represented as an FHIR Bundle resource. Each component was translated into an HL7 FHIR profile using FSH.

Figure 3. Data model of the implementation guide. D: dependent and mandatory element for particular games or applications; M: mandatory element; O: optional element; SG: serious game.



XSL•FO
RenderX

The main element is the "NCD Bundle," which contains the "NCD Composition" as its first element. The "NCD Composition" comprises 3 extensions: the time period of the exported data, the export date, and a free text field for metadata containing the DIN-specified metadata for the game (refer to the Serious Games Ecosystem section). The profile also includes slices for linking the practitioner and related persons. It contains 3 compositions that comprise clinical data, game data, and health-related and lifestyle data, as depicted in Figure 3. Two code systems, along with their value sets were created: "NCD Gamedata," which codes the different types of game data (achievements, high scores, and others), and "NCD Lifestyle data," which encodes the various types of lifestyle data (mood, alcohol consumption, tobacco use, and diary entry).

The IG Publisher rendered all profiles, code systems, and value sets as a web page. The web page includes an artifact summary page that overviews all defined resource profiles, extensions, value sets, and code systems. All artifacts can be exported as comma-separated values, Microsoft Excel spreadsheets, or Schematron files.

The IG is considered a 'draft standard for trial use' and has not yet been validated in real-world scenarios. For broader industry adoption, it must be validated by national HL7 technical committees and experts. This can be easily achieved because the IG was created with standardized and widely adopted HL7 tools and is publicly available.

Ballot Process

If changes to the NCD prevention IG need to be made, all stakeholders must be involved, especially serious game and EHR implementers. They should be able to give feedback and comment on proposed changes before publication. Therefore, the framework suggests implementing and following the HL7 balloting process, which is defined in a previous study [57] and is articulated. First, the stakeholders must sign up to participate in the ballot process. Then, the stakeholders review the proposed specification, provide feedback as ballot comments, and mark their comments as "negative" or "affirmative." After that, all comments are reviewed, and a decision is made about which changes to the specification will be incorporated. Stakeholders have the opportunity to "retract" or "withdraw" their "negative" comments. Agreed-upon changes to the specification will then be applied. Afterward, a final determination is made as to whether the specification has passed the ballot or needs additional changes. Finally, a recirculation ballot might be necessary in some cases.

Summary

The previous subsections outlined an IG for the interoperable exchange of medical and game data using the widely adopted HL7 FHIR standard. Change management was also addressed by defining a ballot process that follows the HL7 standard procedure.

Serious Games Ecosystem

Metadata (DIN)

Every NCD prevention game or gamified application should incorporate a metadata definition following the DIN standard established by Göbel et al [29]. The metadata is part of the FHIR document (refer to the IG Structure section). These data help categorize applications and make them easier to find. Potential users can also check the suitability for their needs beforehand. The DIN norm defines 3 levels of descriptive elements. In this case, level 1 (core) is the minimum that needs to be supported. The preferred option would be compliance with levels 1 and 2, which contain supplementary descriptive elements. Additional application profiles (level 3) are not required and therefore optional.

Public Register

Clinical personnel must be able to find appropriate applications that they can recommend or prescribe to patients (refer to the Care Process and Application Workflow section). This is only possible if public registers of accessible games and applications are available. Therefore, the framework proposes providing public registers that health care professionals can use to search for particular games and applications. Each application must provide specific metadata so that it can be easily found (refer to the Metadata [DIN]) section). Nonprofit organizations, governmental entities, or international entities are encouraged to host such registers.

Serious Games Appliance

Serious game appliances can incorporate NCD prevention games and applications. These are specific technical environments for serious games that can run on PCs, smartphones, or specialized hardware. They feature dedicated launchers to streamline game access and overall usability.

Regulatory Compliance

Approval for NCD prevention games to be used in clinical settings may be necessary. For example, digital health apps in Germany can be licensed as DiGAs [58], which requires a fast-track procedure according to section 139e of the German Fifth Book of the Social Code [42]. These applications must not currently be used for primary prevention and must be recognized as medical products. This framework encourages developers to undergo certification processes for games and applications in specific health care scenarios. Digital health application programs such as DiGAs in Germany should also include games and applications for primary prevention in the future.

Summary

This key area focused on creating a serious games ecosystem that enables building high-quality games and applications by considering a DIN norm centered on metadata, public registers, and applications where games can be found and run, along with regulatory and certification aspects.

Usability and Quality Aspects

Design Guidelines and Usability Studies

Design guidelines should be followed to provide consistent user experiences. Android and Apple publish official guidelines for mobile apps that are recommended to follow. For web applications, World Wide Web Consortium guidelines should be used. Specific design guidelines provided by game engines

XSL•FO
RenderX

or technical frameworks should be followed for games, in general.

Clinical Trials

Clinical trials are needed to fulfill certain certification requirements and provide valuable insights into the effectiveness of games and applications beyond smaller evaluations within respective target groups. Therefore, it is highly recommended that NCD prevention application developers participate in clinical trials.

Software Testing

Rigorous software testing methods are highly endorsed as part of this framework's best practices. NCD prevention games and applications can only be effective if they run well without noticeable errors and have concise user experiences.

Summary

The previous subsections outlined important software quality aspects by defining design guidelines, rigorous software testing, usability studies, and the meaningful use of clinical trials.

Technical Aspects

APIs, Protocols, Authentication, and Authorization

HL7 FHIR specifies not only how the various data elements are structured but also how they should be exchanged using an API. The IG, which was developed, also includes a capability statement that specifies exactly how to interact with the FHIR API. It consists of a specification in the "OpenAPI" standard [59] that developers can directly use.

However, if games use proprietary backend systems, this should also be allowed and will not be further specified by the IG. Communication should still be based on common standards, such as representational state transfer web services, and adhere to state-of-the-art security and privacy practices.

Game Engine

The analyzed games BreathIn and Food Pyramid Escape use the cross-platform game engine Unity [60]. NutritionGarden, NutritionRush, recoverApp, and Moodbooster were developed as native Android apps. The minigames in Moodbooster were developed with JavaScript. NutriMine, a Minecraft mod, was created with MinecraftForge. DiaBeaThis and the therapist application for recoverApp were implemented as web applications. Various engines and frameworks are available to develop NCD prevention games and applications. Native frameworks for mobile apps or cross-platform game engines such as Unity have been successfully used by analyzed games and applications and are recommended.

Game Analytics and Telemetry

Telemetry data can help developers improve games by allowing them to analyze user interactions. Therefore, this framework encourages the collection of telemetry data, provided data security and privacy aspects are considered.

Backend Services

Games and applications can use their proprietary backend systems if required for specific use cases. Because all analyzed applications did not use backend services, this framework neither encourages nor discourages the use of such systems. However, developers should adhere to the data security and privacy standards discussed within the framework.

Security and Privacy Aspects

Because sensitive data can be stored in proprietary backend systems, games and applications must adhere to common data privacy laws. This framework highly recommends integrating the General Data Protection Regulation for applications and games distributed in the EU [61]. The respective data protection laws must be followed if games and applications are provided in other countries.

Data should always be transmitted in an encrypted manner and, if possible, stored in an end-to-end encrypted form on any backend systems. Encryption ensures that only users have direct access to the data stored on a server on their behalf. Data transmitted to national or personal health records must be encrypted with transport-layer encryption. Data storage depends on the specific health record system's technical specifications and is not within this framework's scope.

Summary

The final key area focused on technical aspects, such as APIs, protocols, the use of game engines, backend services, and security and privacy concerns, which are especially important for personal and medical data.

Evaluation

This section compares the framework's 6 key areas, C1 to C6, with the actual state of the games and applications that were developed beforehand and served as the foundation for defining the best practices. Table 4 presents an overview of the degree of fulfillment in each key area.

XSL•FO
RenderX

Table 4. Comparison of the applications in terms of framework key areas and fulfillment (none or partial).

ID	Project	C1[a]	C2[b]	C3[c]	C4[d]	C5[e]	C6[f]
G1	BreathIn	None	Partial	None	None	Partial	Partial
G2	Food Pyramid Escape	None	Partial	None	None	Partial	Partial
G3	Nutrition Garden	None	Partial	None	None	Partial	Partial
G4	NutritionRush	None	Partial	None	None	Partial	Partial
G5	NutriMine	None	Partial	None	Partial	Partial	Partial
G6	recoverApp	None	Partial	None	None	Partial	Partial
G7	MoodBooster	None	Partial	None	None	Partial	Partial
G8	DiaBeaThis	None	Partial	None	None	Partial	Partial

[a]C1: organizational interoperability component.

[b]C2: design and development best practices component.

[c]C3: semantic interoperability component.

[d]C4: serious games ecosystem component.

[e]C5: usability and quality aspects component.

[f]C6: technical aspects component.

None of the analyzed games and applications currently fulfill the requirements regarding organizational and semantic interoperability (key areas C1 and semantic interoperability component) because they were only being evaluated in smaller clinical settings and are not yet widely available. Furthermore, the defined HL7 FHIR IG was unavailable when these applications were designed and implemented.

All analyzed applications fulfilled at least some of the defined best practices (key area C2) because they successfully implemented BCTs and included engaging gamification aspects and game mechanics.

Except for NutriMine, which was already successfully integrated into a prototypically implemented serious game appliance, the other applications still need to fulfill parts of the key area serious games ecosystem component.

All analyzed applications at least partially fulfilled the key areas of usability and technical aspects (areas C5 and C6). They were also playtested and evaluated with users from the target groups. However, as all analyzed applications are still at a prototypical stage, best practices regarding game engines, APIs, backend services, and security measures have yet to be fulfilled.

Overall, a comprehensive evaluation of all 8 applications revealed that none of them fully fulfilled any of the 6 key areas; some partially fulfilled certain areas, especially key areas C2, C5, and C6.

Discussion

Principal Findings

An interoperable framework was developed based on a comparative analysis of 8 serious games and gamified applications for NCD prevention published by or in part by the authors of this paper. The first part, the comparative analysis and basic requirements definition, was done beforehand and published at the International Conference on Healthcare Service Management in mid-2024 [22]. Six games were initially part of that analysis, but 2 games were added for this particular work. The defined requirements were refined and categorized. Table 5 compares the developed framework with work from the state of the art.

XSL•FO
RenderX

Table 5. Comparison of serious game (SG) frameworks.

Feature	The framework by Peters et al [21]	The framework by Yusoff et al [23]	The framework by Stanescu et al [25]	Proposed framework
Scope	Collection of health-related data from SGs with a DSS[a] approach	Conceptual model for serious games, including learning and pedagogy	Multidimensional interoperability framework	Interoperable technical framework for SGs and gamified applications
Target domain	SGs for health	Educational SGs	Educational and general SGs	NCD[b] prevention SGs and gamified applications
Interoperability	Recommends use of standards but lacks implementation guides	Conceptual only; does not specify interoperability mechanisms	Leverages interoperability between game components, SGs, and learning systems using SCORM[c] and LOM[d]	Full semantic interoperability
Implementation guidance	Moderate; focuses on the data model, SDK[e], and DSS	Low; mainly conceptual with few practical details	Medium; includes technical architecture suggestions	High; includes FHIR[f] profiles, implementation guides, and API[g] integration examples
Adaptability and extensibility	Inconclusive	Inconclusive	High; multiple technologies supported	Very high; modular FHIR-based resources allow for easy extension and ballot process for changes

[a]DSS: decision support system.

[b]NCD: noncommunicable disease.

[c]SCORM: Sharable Content Object Reference Model.

[d]LOM: learning object metadata.

[e]SDK: Software development kit.

[f]FHIR: Fast Healthcare Interoperability Resources.

[g]API: application programming interface.

The key finding of this work is the definition of an interoperable framework for NCD prevention that consists of 6 key areas. A central part of the framework is the definition of an IG for the interoperable extraction and exchange of medical and game data into EHRs based on the state-of-the-art medical informatics standards HL7 FHIR and FSH for the profile definition. The guide is available through technical artifacts (HTML representation, JSON, Microsoft Excel, and CSV). Moreover, it includes a capability statement with API definitions that implementers can use when designing novel, serious games. In practice, this enables serious games and gamified apps that follow the defined framework guidelines to bundle their collected data into HL7 FHIR documents, which can be represented as either an XML or JSON file, and store it on FHIR-enabled EHRs. This can be either an FHIR facade that receives FHIR documents over a representational state transfer interface and transforms them into the native format of the EHR, or an IHE- or XDSb-enabled EHR web service that allows for cross-enterprise document sharing.

Furthermore, the framework suggests best practices for designing and developing NCD prevention games and applications by providing key requirements and guidelines for BCTs, gamification elements, game mechanics, and general usability and technical aspects. It also addresses organizational aspects such as care processes, governmental platforms, health data spaces, and regulatory compliance.

The framework presented in this study can be a fundamental guideline for building new games and applications to prevent NCDs.

Comparison With Prior Work

Unlike the first study [22], which included only 6 games and applications in its comparative analysis, this work included 2 more. Furthermore, previous work presented only a catalog of requirements. This work presents a complete, interoperable framework comprising 6 key areas.

Future Work

The next steps of this work will involve refining the framework by including more applications and focusing on tools, documentation, reference implementations, and partnerships as well as using it in practice by storing and extracting actual data from EHRs. In addition, the use of large language models for HL7 FHIR data in NCD prevention is being evaluated because this field is emerging in semantic interoperability research [62].

Recommendations

This study's findings, based on the games and applications built beforehand, are valuable for the crucial area of NCD prevention. The use of semantically interoperable data formats presents an opportunity to enhance not only primary care for patients but also enables the use of data for scientific and regulatory purposes as part of secondary use, particularly within supranational health care projects such as the EHDS. This is especially relevant given

XSL•FO
RenderX

the rising use of large language models and artificial intelligence.

Limitations

Although we consider the framework's findings very valuable for the crucial area of NCD prevention, some limitations should be considered.

First, the most significant limitation arises from the number of applications analyzed. A total of 8 applications were evaluated. Most of the games and applications were created by us to understand the mechanisms and users involved, but additional applications were also used, and still they do not represent a full set of available applications and all categories. Future work should include more applications to refine the constructed interoperable framework further. As a result, the best practices established regarding the design and development of NCD prevention games and applications within the framework can be considered limited due to the narrow sample size. A more comprehensive approach that incorporates a broader range of NCD prevention games and applications could create a more effective interoperable framework for NCD prevention that does not inadvertently reflect the specific characteristics of the selected games and applications, potentially overlooking other successful strategies or innovations available in different applications. Furthermore, the evaluation was limited in depth. Therefore, future work should also focus on a more comprehensive assessment with pilot implementations, usability studies, and stakeholder interviews.

Second, the technical IG, developed as part of the interoperable framework, which uses HL7 FHIR and FSH to define extensions and profiles for NCD prevention games and clinical data, can be viewed as a prototype and will be used in practice as part of future work as well. For instance, some elements are currently represented as string values and must be accurately parsed by consuming systems. For example, the DIN metadata should be defined as an FHIR element instead of a string value.

Third, because the framework has a strong focus on technical and engineering aspects, developers require a specific background and knowledge of HL7 FHIR and other relevant technologies. To address this limitation, future work should focus on providing user-friendly tools and documentation as well as developing modular and open-source components, such as reference implementations, and establishing partnerships with academic institutions and national or local health networks.

Conclusions

This research was based on an investigation of the state of the art and a comparative analysis of 8 serious games and gamified applications. This led to the design and development of an interoperable framework for NCD prevention in 6 key areas. The framework defined a novel approach to best practices for developing serious NCD prevention games and gamified applications that should be considered when developing new projects in that area. This framework was then evaluated against the 8 games and applications mentioned earlier, and gaps were identified.

Acknowledgments

The authors would like to express their sincere gratitude to all who contributed to the framework's creation, with special acknowledgment to all involved colleagues and supervised students, whose significant contributions to developing serious games and gamified applications formed the foundation for this framework.

No funding was received for this paper.

For this paper, artificial intelligence–supported tools, such as Grammarly for writing support (including suggestions and rewrites), DeepL for translations, and scite.ai for research on the theoretical foundation and state of the art (including suggestions for text content), were used. The authors thoughtfully revised every piece of generated text, ensuring that no section was used in its original, unreflected form.

Data Availability

The datasets generated or analyzed during this study are available on the web [63].

Authors' Contributions

CA was responsible for conceptualization, methodology, formal analysis, investigation, writing the original draft, and reviewing and editing the manuscript. RB and TG contributed to the methodology and validation and provided supervision throughout the project.

Conflicts of Interest

RB serves as an editor for *JMIR Serious Games* at the time of this publication. All other authors declare no conflicts of interest.

References

1. Bouton ME. Why behavior change is difficult to sustain. Prev Med 2014 Nov;68:29-36 [FREE Full text] [doi: 10.1016/j.ypmed.2014.06.010] [Medline: 24937649]

XSL•FO
RenderX

2. Coorevits L, Coenen T. The rise and fall of wearable fitness trackers. Acad Manag Proc 2017;2016(1) [FREE Full text] [doi: 10.5465/ambpp.2016.17305abstract]
3. Alshawmar M, Tulu B, Hall-Phillips A. Influence of personality traits on the continued use of fitness apps. In: Proceedings of the 55th Hawaii International Conference on System Sciences. 2022 Presented at: HICSS 2022; January 4-7, 2022; Virtual Event URL: http://hdl.handle.net/10125/79814 [doi: 10.24251/hicss.2022.478]
4. Bartram M, Stewart JM. Income-based inequities in access to psychotherapy and other mental health services in Canada and Australia. Health Policy 2019 Jan;123(1):45-50. [doi: 10.1016/j.healthpol.2018.10.011] [Medline: 30424886]
5. Non-communicable diseases. International Federation of Red Cross and Red Crescent Societies. URL: https://www.ifrc.org/our-work/health-and-care/community-health/non-communicable-diseases [accessed 2024-12-02]
6. Noncommunicable diseases. World Health Organization. 2024. URL: https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases [accessed 2024-08-22]
7. Zhao J, Xu L, Sun J, Song M, Wang L, Yuan S, et al. Global trends in incidence, death, burden and risk factors of early-onset cancer from 1990 to 2019. BMJ Oncol 2023 Sep 05;2(1):e000049 [FREE Full text] [doi: 10.1136/bmjonc-2023-000049] [Medline: 39886513]
8. O'Neil A, Jacka FN, Quirk SE, Cocker F, Taylor CB, Oldenburg B, et al. A shared framework for the common mental disorders and non-communicable disease: key considerations for disease prevention and control. BMC Psychiatry 2015 Mar 05;15(1):15. [doi: 10.1186/s12888-015-0394-0] [Medline: 25652365]
9. Eaton J, O'Donnell K, Westerman L, Adshead F. Linking mental health and NCD priorities. Mental Health Innovation Network. 2018 Jan 1. URL: https://www.mhinnovation.net/resources/linking-mental-health-ncd-priorities [accessed 2024-12-03]
10. Jacka FN, Sacks G, Berk M, Allender S. Food policies for physical and mental health. BMC Psychiatry 2014 May 09;14:132 [FREE Full text] [doi: 10.1186/1471-244X-14-132] [Medline: 24884515]
11. Dörner R, Göbel S, Effelsberg W, Wiemeyer J. Serious Games: Foundations, Concepts and Practice. Cham, Switzerland: Springer; 2016. [doi: 10.1007/978-3-319-40612-1]
12. Damaševičius R, Maskeliūnas R, Blažauskas T. Serious games and gamification in healthcare: a meta-review. Information 2023 Feb 07;14(2):105. [doi: 10.3390/info14020105]
13. Ma M, Oikonomou A, Jain LC. Serious Games and Edutainment Applications. Cham, Switzerland: Springer; 2011. [doi: 10.1007/978-1-4471-2161-9]
14. Gómez B. Differences between e-learning, gamification and serious games. On Serious Games. 2017 Jun 5. URL: https://www.onseriousgames.com/differences-between-e-learning-gamification-and-serious-games/ [accessed 2024-10-14] [doi: 10.1007/978-3-319-40612-1]
15. Michie S, Johnston M, Carey R. Behavior change techniques. In: Gellman MD, editor. Encyclopedia of Behavioral Medicine. Cham, Switzerland: Springer; 2020. [doi: 10.1007/978-3-030-39903-0_1661]
16. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
17. Baranowski T, Buday R, Thompson DI, Baranowski J. Playing for real: video games and stories for health-related behavior change. Am J Prev Med 2008 Jan;34(1):74-82 [FREE Full text] [doi: 10.1016/j.amepre.2007.09.027] [Medline: 18083454]
18. ISO 13606-1:2008(en) health informatics — electronic health record communication — part 1: reference model. International Organization for Standardization. URL: https://www.iso.org/obp/ui/#iso:std:iso:13606:-1:ed-1:v1:en [accessed 2024-11-18]
19. Benson T, Grieve G. Principles of Health Interoperability: SNOMED CT, HL7 and FHIR. Cham, Switzerland: Springer; 2016. [doi: 10.1007/978-3-319-30370-3]
20. Oemig F, Snelick R. Healthcare Interoperability Standards Compliance Handbook. Cham, Switzerland: Springer; 2016. [doi: 10.1007/978-3-319-44839-8]
21. Peters K, Kayali F, Silbernagl M, Lawitschka A, Hlavacs H. A proposed framework for the collection of health-related data from serious games and apps. Int J Serious Games 2017 Jun;4(2):59-73. [doi: 10.17083/ijsg.v4i2.162]
22. Aigner C, Baranyi R, Grechenig T. Foundations of an interoperable framework for serious games and gamified mobile apps in non-communicable disease (NCD) prevention. In: Proceedings of the 2024 7th International Conference on Healthcare Service Management. 2024 Presented at: ICHSM '24; September 6-8, 2024; Istanbul, Turkiye. [doi: 10.1145/3704239.3704246]
23. Yusoff A, Crowder R, Gilbert L, Wills G. A conceptual framework for serious games. In: Proceedings of the 9th IEEE International Conference on Advanced Learning Technologies. 2009 Presented at: ICALT 2009; July 15-17, 2009; Riga, Latvia. [doi: 10.1109/icalt.2009.19]
24. Ger PM. eAdventure: serious games, assessment and interoperability. In: Proceedings of the 2014 International Symposium on Computers in Education. 2014 Presented at: SIIE 2014; November 12-14, 2014; Logroño, Spain. [doi: 10.1109/siie.2014.7017736]
25. Stanescu IA, Stepfan A, Kravcik M, Lim T, Bidarra R. Interoperability strategies for serious games development. In: Proceedings of the 8th International Scientific Conference eLearning and Software for Education. 2012 Presented at: eLSE 2012; April 26-27, 2012; Bucharest, Romania. [doi: 10.18278/il.2.1.4]

XSL•FO
RenderX

26. Cowan B, Kapralos B. A survey of frameworks and game engines for serious game development. In: Proceedings of the 14th International Conference on Advanced Learning Technologies. 2014 Presented at: ICALT 2014; July 07-10, 2014; Athens, Greece. [doi: 10.1109/icalt.2014.194]
27. Weber S, Heitmann KU. [Interoperability in healthcare: also prescribed for digital health applications (DiGA)]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2021 Oct;64(10):1262-1268 [FREE Full text] [doi: 10.1007/s00103-021-03414-w] [Medline: 34532746]
28. Mittermaier M, Sina C, Richter JG, Raspe M, Stais P, Vehreschild J, et al. [Practical use of digital health applications (DiGA) in internal medicine]. Internist (Berl) 2022 Mar 17;63(3):245-254. [doi: 10.1007/s00108-022-01264-5] [Medline: 35037948]
29. Göbel S, Caserman P, Hansen J, Bruder R, Abels S, Behrmann M, et al. DIN SPEC 91380:2018-06 serious games metadata format. DIN Media. 2018 Jun. URL: https://www.dinmedia.de/en/technical-rule/din-spec-91380/289947896 [accessed 2025-07-30]
30. Buendia FB, Hervás AH. Evaluating e-learning platforms through SCORM specifications. In: Proceedings of the IADIS Virtual Multi Conference on Computer Science and Information Systems 2006. 2006 Presented at: MCCSIS 2006; May 15-19, 2006; Online.
31. DiGA Toolkit 1.1.0. Kassenärztlichen Bundesvereinigung (KBV). URL: https://mio.kbv.de/display/DIGA1X1X0 [accessed 2024-10-28]
32. Aigner C, Moser M, Wolfsberger OC, Baranyi R, Hohenegger V, Grechenig T. MoodBooster - a gamified app to support stress reduction in university students. In: Proceedings of the 2024 IEEE International Conference on E-Health Networking, Application & Services. 2024 Presented at: HealthCom 2024; November 18-20, 2024; Nara, Japan. [doi: 10.1109/healthcom60970.2024.10880742]
33. Baranyi R, Willinger R, Lederer N, Walcher F, Grechenig T. DiaBeaThis — a gamified self-tracking portal to support people suffering from diabetes mellitus to control their blood glucose level. In: Proceedings of the IEEE 6th International Conference on Serious Games and Applications for Health. 2018 Presented at: SeGAH 2018; May 16-18, 2018; Vienna, Austria. [doi: 10.1109/segah.2018.8401352]
34. Aigner C, Zeillinger V, Baur K, Baranyi R, Grechenig T. BreathIn – a serious game to support patients with smoking cessation: analysis and design study for a mobile serious game to help patients quit smoking. In: Proceedings of the 2023 7th International Conference on Medical and Health Informatics. 2023 Presented at: ICMHI '23; May 12-14, 2023; Kyoto, Japan. [doi: 10.1145/3608298.3608342]
35. Aigner C, Baranyi R, Hoelbling D, Baur K, Zeillinger V, Grechenig T. Analysis, implementation, and assessment of a serious game for smoking cessation: investigating design and playtesting outcomes. In: Proceedings of the 11th International Conference on E-Health and Bioengineering. 2023 Presented at: EHB-2023; November 9-10, 2023; Bucharest, Romania. [doi: 10.1007/978-3-031-62502-2_60]
36. Aigner C, Resch EM, El Agrod A, Baranyi R, Grechenig T. Food Pyramid Escape - a serious escape game for the support of nutritional education in Austria and beyond. In: Proceedings of the IEEE 9th International Conference on Serious Games and Applications for Health. 2021 Presented at: SeGAH 2021; August 04-06, 2021; Dubai, United Arab Emirates. [doi: 10.1109/segah52098.2021.9551850]
37. Aigner C, Hofmann G, Winkler S, Baranyi R, Grechenig T. Nutrition Garden - a gamified mobile app for motivating people to eat specific food to prevent non-communicable diseases. In: Proceedings of the 2023 7th International Conference on Medical and Health Informatics. 2023 Presented at: ICMHI '23; May 12-14, 2023; Kyoto, Japan. [doi: 10.1145/3608298.3608336]
38. Baranyi R, Steyrer B, Lechner L, Agbektas GN, Lederer N, Grechenig T. NutritionRush - a serious game to support people with the awareness of their nutrition intake. In: Proceedings of the 5th International Conference on Serious Games and Applications for Health. 2017 Presented at: SeGAH 2017; April 02-04, 2017; Perth, WA. [doi: 10.1109/segah.2017.7939271]
39. Aigner C, Köck K, Baranyi R, Winkler S, Weindl K, Grechenig T. NutriMine - a serious game modification for Minecraft to support people keeping a healthy diet. In: Proceedings of the 12th International Conference on Serious Games and Applications for Health. 2024 Presented at: SeGAH 2024; August 07-09, 2024; Funchal, Portugal. [doi: 10.1109/segah61285.2024.10639569]
40. Aigner C, Eder M, Baranyi R, Grechenig T. recoverApp - a mobile health solution to support people in stationary rehabilitation. In: Proceedings of the 8th International Conference on Serious Games and Applications for Health. 2020 Presented at: SeGAH 2020; August 12-14, 2020; Vancouver, BC. [doi: 10.1109/segah49190.2020.9201839]
41. Obesity and overweight. World Health Organization. 2025 May 7. URL: https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight [accessed 2025-06-24]
42. DiGA-Leitfaden (stand: 28.12.2023, version 3.5). Bundesinstitut für Arzneimittel und Medizinprodukte. 2023. URL: https://www.bfarm.de/SharedDocs/Downloads/DE/Medizinprodukte/diga_leitfaden.html?nn=597198 [accessed 2024-10-20]
43. Hendolin M, European Observatory on Health Systems and Policies. Towards the European health data space: from diversity to a common framework. Eurohealth 2022;27(2):15-17 [FREE Full text]
44. European Health Data Space Regulation (EHDS). European Commission. URL: https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space_en [accessed 2024-10-18]

XSL•FO
RenderX

45. Bradner S. BCP 14 RFC 2119: key words for use in RFCs to Indicate Requirement Levels. RFC Editor. URL: https://www.rfc-editor.org/info/rfc2119 [accessed 2025-07-30]
46. Baranyi R. DeapSea: workflow-supported serious game design for stroke rehabilitation. Int J Comput Games Technol 2023 May 26;2023:1-15. [doi: 10.1155/2023/3169262]
47. SNOMED International homepage. SNOMED International. URL: https://www.snomed.org [accessed 2024-09-27]
48. LOINC homepage. LOINC. URL: https://loinc.org/ [accessed 2024-09-27]
49. FHIR documents. HL7 FHIR. URL: https://www.hl7.org/fhir/documents.html [accessed 2024-09-28]
50. IT Infrastructure (ITI) domain. Integrating the Healthcare Enterprise. URL: https://profiles.ihe.net/ITI/index.html [accessed 2024-09-28]
51. 10 cross-enterprise document sharing (XDS.b). Integrating the Healthcare Enterprise. URL: https://profiles.ihe.net/ITI/TF/Volume1/ch-10.html [accessed 2024-09-28]
52. Kramer M, Moesel C. Tutorial: create an implementation guide with FHIR shorthand. HL7 FHIR. 2021. URL: https://www.devdays.com/wp-content/uploads/2021/12/DD21US_20210608_Mark_Kramer_Create_an_Implementation_Guide_with_FHIR_Shorthand_.pdf [accessed 2024-09-29]
53. FHIR Shorthand. HL7 FHIR. URL: https://hl7.org/fhir/uv/shorthand/ [accessed 2024-09-27]
54. FHIR / sushi. GitHub. URL: https://github.com/FHIR/sushi [accessed 2024-09-27]
55. FHIR tools - Visual Studio Marketplace. Visual Studio Marketplace. URL: https://marketplace.visualstudio.com/items?itemName=Yannick-Lagger.vscode-fhir-tools [accessed 2024-11-18]
56. HL7 FHIR Shorthand. Visual Studio Marketplace. URL: https://marketplace.visualstudio.com/items?itemName=FHIR-Shorthand.vscode-fsh [accessed 2024-11-18]
57. HL7 Balloting. Confluence. URL: https://confluence.hl7.org/display/HL7/HL7+Balloting [accessed 2024-09-27]
58. Reichardt A. Serious Games: spielen um leben und tod. Deutsches Ärzteblatt 2021 Aug(5) [FREE Full text]
59. OpenAPI Initiative homepage. OpenAPI Initiative. URL: https://www.openapis.org/ [accessed 2024-10-07]
60. Unity homepage. Unity. URL: https://unity.com [accessed 2024-10-20]
61. Consolidated text: Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance). European Union. URL: http://data.europa.eu/eli/reg/2016/679/2016-05-04/eng [accessed 2024-10-21]
62. Li Y, Wang H, Yerebakan HZ, Shinagawa Y, Luo Y. FHIR-GPT enhances health interoperability with large language models. NEJM AI 2024 Jul 25;1(8). [doi: 10.1056/aics2300301]
63. Implementation guide for NCD-prevention games and apps. Fast Healthcare Interoperability Resources. URL: https://fhir.inso-world.com/ncd-prevention-games/ [accessed 2025-08-05]

Abbreviations

API: application programming interface
BCT: behavior change technique
C1: organizational interoperability component
C2: design and development best practices component
C5: usability and quality aspects component
C6: technical aspects component
DiGA: digital health application (German: Digital Gesundheitsanwendungen)
DIN: German Institute for Standardization Registered Association (German: Deutsches Institut für Normung eV)
DSS: decision support system
EHDS: European Health Data Space
EHR: electronic health record
EU: European Union
FHIR: Fast Healthcare Interoperability Resources
FSH: Fast Healthcare Interoperability Resources Shorthand
HL7: Health Level 7
IG: implementation guide
IHE: Integrating the Healthcare Enterprise
NCD: Noncommunicable Disease
SCORM: Sharable Content Object Reference Model
XDSb: Cross-Enterprise Document Sharing

XSL•FO
RenderX

Edited by A Coristine; submitted 25.11.24; peer-reviewed by E Oliveira, C Tziraki, E Maslowski; comments to author 02.06.25; revised version received 29.06.25; accepted 10.07.25; published 27.08.25.

Please cite as:
Aigner C, Baranyi R, Grechenig T
Digital Catalysts for Noncommunicable Disease Prevention Serious Games and Gamified Applications: Framework Design Study
JMIR Serious Games 2025;13:e69246
URL: https://games.jmir.org/2025/1/e69246
doi:10.2196/69246
PMID:

©Christoph Aigner, René Baranyi, Thomas Grechenig. Originally published in JMIR Serious Games (https://games.jmir.org), 27.08.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Haptic-Driven Serious Card Games for Older Adults: User Preferences Study

Xin Huang[1]; Nazlena Mohamad Ali[1]; Shafrida Sahrani[1]

Institute of Visual Informatics, Universiti Kebangsaan Malaysia, Bangi, Selangor, Malaysia

Corresponding Author:
Nazlena Mohamad Ali
Institute of Visual Informatics
Universiti Kebangsaan Malaysia
Aras 1, Akademia Siber Teknopolis
Universiti Kebangsaan Malaysia
Bangi, Selangor, 43600
Malaysia
Phone: 60 3 8927 2413
Email: nazlena.ali@ukm.edu.my

Abstract

Background: Haptic feedback technology is widely used, including in serious games. It is an additional sensory method that supplements audio and vision, provides a novel user experience through a physical control layer, and enhances the immersion of virtual reality, thereby improving the user's cognitive state while alleviating dependence on visual information. However, there is limited research on haptic feedback preferences for mobile serious card games designed for older adults.

Objective: The study aims to investigate older adults' preferences for haptic feedback in mobile serious card games.

Methods: This study recruited a total of 250 participants from Dangtu County, Anhui Province, China, with an average age of 64.78 (SD 4.23) years. A descriptive survey was conducted among older adults, structured questionnaires were distributed, and data were collected via the Wenjuanxing (Changsha Ranxing Information Technology Co Ltd) mobile app. Reliability and validity analyses were performed using SPSS (IBM Corp) software. The questionnaire investigated older adults' basic understanding of card games and serious games, the integration of card games with mobile devices, the potential impact of combining card games with haptic feedback technology, as well as suggestions and opinions.

Results: The results showed that 63.2% (158/250) of the older adults liked the slight haptic feedback mode, and 78.4% (196/250) of the participants believed that combining haptic feedback technology with mobile card games would help improve cognitive abilities. The study found that 73.6% (184/250) of the older adults believed that this technology could reduce their reliance on visual information. This confirms that the combination of serious card games and haptic feedback can alleviate sensory impairments in older adults. Qualitative analysis revealed the potential of haptic feedback to reduce visual fatigue and provide an engaging cognitive training experience.

Conclusions: Older adults have shown great interest in incorporating haptic feedback into mobile serious card games, believing that this could enhance their cognitive abilities while reducing their reliance on visual information. However, limitations include sample size and geographic restrictions, differences in digital literacy, self-reported data, and lack of longitudinal assessment. Longitudinal studies are recommended to evaluate the long-term effects of mobile serious card games with haptic feedback on cognitive abilities. Such investigations could provide valuable insights for game developers, rehabilitation institutions, and the development of cognitive training tools for older adults.

(JMIR Serious Games 2025;13:e73135) doi:10.2196/73135

KEYWORDS

serious games; older adults; haptic feedback; mobile card games; cognitive abilities

XSL•FO
RenderX

Introduction

Touch perception can be divided into skin perception and motion perception [1], so haptic perception is composed of 2 parts of sensory feedback: haptic feedback and force and torque feedback. Devices can be classified according to the different feedback they provide. Choose the appropriate device type based on the game concept and budget estimate. The device specifications usually considered include degrees of freedom, force feedback, workspace size, maximum force and torque, stiffness, software development kit compatibility, scalability, etc. Commercial games usually prioritize affordability. A low-cost solution is to embed vibrotactile actuators into gaming devices, such as game controllers, mobile phones, etc. In general, in terms of learning and training, the force torque and haptic perception provided by haptic technology can help users better understand their performance in tasks and help them improve their performance in a more intuitive and efficient way. Especially for tasks that rely mainly on haptic feedback, pure visual feedback is of little help in improving performance and may even lead to errors. For example, without haptic feedback, endoscopic surgery training is difficult to achieve the expected results [2], the same is true for applications that focus on improving motor skills. More specifically, visually impaired users or blind people primarily benefit from auditory and haptic input. Haptic technology, as a new way to interact with virtual reality, also benefits the gaming industry in expanding its market to previously unreachable users [3].

Haptic feedback provides an additional means of haptic perception and is combined with sight and sound for a more immersive user experience. The haptic experience provides patients with a new way of perceiving, thus enhancing the user's cognitive ability. Haptic technology is a kind of sensory feedback technology that simulates haptic perception through force, vibration, or motion. It captures the user's action and interaction information through sensors and generates corresponding haptic feedback through actuators so that users can feel the real physical interaction in the virtual reality. Haptic perception is a multimodal process involving many neural pathways such as mechanoreceptors and thermoreceptors. Haptic technology simulates the haptic sensation of different materials by accurately controlling vibration frequency and intensity, thus activating multiple neural pathways and enhancing the cognitive participation of users [1]. These techniques can be used to create virtual reality objects in computer simulations, control virtual reality objects, and enhance remote control of machines and devices. Haptic feedback technology provides an immersive experience for older people by combining their sense of touch with what they see, hear, or interact with [4]. At the same time, the use of haptic technology has huge advantages, as it can provide real and intuitive interaction with virtual reality objects, improve the usability and accessibility of applications, enhance the cognitive abilities of end users, and make human-computer interaction more interesting [5]. The application of haptic feedback technology not only improves user experience but also enriches user's entertainment activities. For example, the use of haptic suits that provide haptic feedback further enhances user immersion [6].

One promising application of haptic technology is its use in enhancing cognitive function [7]. In previous studies, performing somatosensory tasks without visual input, similar to the device DOT PAD (DOT Corporation), increased pain thresholds in the trained hand while decreasing pain thresholds in the untrained hand. In addition, sensory stimulation without visual feedback resulted in changes in cortical excitability as measured by transcranial magnetic stimulation [8]. This suggests that haptic devices can modulate cortical activity and could be used to enhance cognitive function. Haptic feedback mobilizes the somatosensory system, which is closely related to cognitive functions such as attention, memory, and spatial awareness [9-11]. By providing haptic stimulation, haptic training can enhance sensory processing and promote the integration of sensory information across multiple modalities, thereby strengthening neural connections [5]. This process can improve cognitive abilities, especially in tasks that require fine motor skills, spatial reasoning, or executive function. In addition, haptic training can stimulate brain regions involved in sensory and cognitive tasks, such as the parietal cortex and prefrontal cortex, thereby promoting more coordinated and efficient neural networks [12,13]. The repetitive and goal-oriented nature of haptic tasks may also promote the development of procedural memory and enhance attentional control, further promoting cognitive improvement [14-16]. Through these mechanisms, haptic training not only facilitates the acquisition of motor skills but also provides a promising avenue for enhancing cognitive abilities, especially for people with cognitive impairment or at risk of cognitive decline.

Haptic feedback has become an integral part of serious games in recent years. A serious game is a user-participatory game that allows players to carry out activities that allow them to practice their skills and achieve goals that go beyond just enjoying leisure activities [17]. A more modern definition of a serious game is “any form of computer-based interactive game software that can be used by single or multiple players on any platform and is developed for more than just entertainment” [18]. Serious games differ from traditional video games in that they focus on problem-solving and learning rather than pure entertainment. Nonetheless, they use the same media and elements as casual video games, making them engaging and enjoyable [19]. As a serious game with a long history, easy to simulate, and relatively simple rules, card games are loved by players of all ages [20]. In addition, card games facilitate the diagnosis and tracking of cognitive impairments, reducing the burden on patients and physicians, and thus potentially providing a more comprehensive and clearer cognitive profile [21]. Card games can also be used as an effective cognitive intervention to maintain certain cognitive functions [22].

Haptic interaction can add a new dimension to the development of serious games [23]. For visually impaired users, it is difficult for them to enjoy video games. Haptic serious games offer them a whole new game experience. Yuan et al [24] discuss the accessibility of games in detail and offer insights into the challenges of future related research. Chan and Black [25] studied the performance of mental acquisition based on 3 different forms of presentation (plain text, text plus static visuals, and text plus haptic animated games). They designed

XSL•FO
RenderX

experiments to teach middle school students Newtonian mechanics using each of the 3 forms of learning. The results showed that for difficult tasks, students who learned using the haptic animation serious game form performed better than other students. In addition, the user's learning experience is easier to understand than other forms. Haptic teaching helps students grasp abstract concepts more easily.

Although existing serious game research has provided valuable insights, there are still some unresolved issues. In particular, the impact of sensory decline on the interaction between serious games and older adults has not been fully explored, and there is limited research on older adults' preferences for haptic feedback in game-based training. The overall objective of this study was to explore the preference of older adults for haptic feedback in mobile card games and the potential for older adults to acquire information through haptic feedback. This study proposed the following hypotheses: first, older adults will have higher engagement and satisfaction with mobile card games with haptic enhancement. Second, older adults are more inclined to use haptic feedback to obtain information.

The rest of the paper is structured as follows. Section II outlines the research methodology, theoretical framework, and questionnaire designed based on the theoretical framework. Section III presents the research findings and details the integration of haptic feedback into a mobile card game. Section IV discusses the findings. Finally, Section V summarizes the main findings of the paper, discusses the limitations of the study, and suggests future research directions.

Methods

Ethical Considerations

The study was approved by the ethics committee before it was carried out and strictly adhered to the ethical guidelines for research involving human participants.

Research Ethics Approval

The institutional review board of the Institute of Visual Informatics, National University of Malaysia, formally reviewed and approved the study titled “Mobile Serious Gaming Technology for Enhancing Cognitive Abilities in Older Adults Using Haptic Feedback” on July 2, 2024 (approval reference: UKM.IVI.600/8/1-P136397). The study protocol, including the data collection methods, was considered to comply with the ethical standards for research involving human participants.

Informed Consent

Written informed consent was obtained from all participants before participating in the study. The consent form clearly outlined the purpose, procedures, risks, benefits, and voluntary nature of participating in the study. Participants were informed of their right to withdraw at any time without penalty. For secondary data analysis, the original consent form explicitly allowed the use of deidentified data for research purposes without the need for additional consent.

Privacy and Confidentiality

All participant data were anonymized and deidentified during collection and analysis. Personally identifiable information was removed from survey responses, and data were securely stored on a password-protected server accessible only to the research team. No identifiable information will be shared with external parties or institutions.

Compensation

Participants received no monetary or nonmonetary compensation for their participation in this study. Their contribution was voluntary and motivated by the potential societal benefits of the research.

Images and Identifiability

No images or identifiable personal information of any participant was collected or included in the manuscript or supplementary materials. All data reported were aggregated and anonymized to ensure participant confidentiality.

Compliance Statement

This study conformed to tenets of the Declaration of Helsinki and adhered to the ethical standards outlined by JMIR for research involving human participants.

Research Methods Integration and Overview

This study was conducted at Dangtu Old-age University, and 250 older adults aged 60-80 years were surveyed. Among the 250 participants aged 60-80 years, 209 (83.6%) older adults had symptoms of memory loss, 3 (1.2%) older adults had mild cognitive impairment, 1 (0.4%) older adult had Alzheimer disease, and 37 (14.8%) older adults chose other. Out of 250 participants, 213 (85.2%) older adults had obvious cognitive decline. The participating older adults have basically used mobile phone software with different haptic settings. Among them, sociodemographic characteristics are usually regarded as independent variables, such as gender, age, region, work status, etc. are independent variables, and the dependent variables include participants' views on card games, acceptance of new technologies, and their views on the potential benefits of combining card games with new technologies.

The researchers evaluated vibration intensity variation and pressure-based feedback. Vibration intensity variation is suitable for fine motor skill tasks [26], while pressure-based feedback is suitable for scenarios that require precise force application [27]. The researchers ultimately chose vibration haptic feedback for the following reasons: (1) vibration feedback can provide clear and immediate haptic cues, which is crucial for maintaining user engagement and attention in cognitive tasks [28]; (2) vibration intensity can be adjusted and customized according to user preferences, and surveys show that most participants prefer mild vibration modes [29]; (3) it is technically feasible and cost-effective, making it suitable for widespread adoption [30]; and (4) vibration haptic feedback can be well applied to mobile card games [31].

The theoretical framework of this study is the Technology Acceptance Model [32] and the Multisensory Experience [33] Theory. The Technology Acceptance Model explains the user's

XSL•FO
RenderX

acceptance of new technologies, including 2 key factors: perceived usefulness and perceived ease of use. The Multisensory Experience Theory emphasizes providing a rich user experience through multiple sensory channels, such as touch, vision, and hearing.

The questions in the questionnaire were designed based on the theoretical framework. The following questions are related to perceived usefulness. For example, “Do you think card games can exercise your brain?” reflects the perception of older adults on whether card games are useful in improving cognitive abilities. “Do you think card games combined with haptic feedback technology can help improve your cognitive abilities?” asks the older adults about their perception of the usefulness of haptic feedback technology in improving cognitive abilities. “What aspects of your cognitive abilities do you think card games combined with haptic feedback technology can improve?” asks older adults about their perception of the specific role of haptic feedback technology in improving cognitive abilities.

Questions related to perceived ease of use. For example, “Are you willing to try a mobile card game that reduces visual fatigue?” asks the older adults about their experience of whether mobile card games are easy to use. “Which mode of haptic feedback are you comfortable with?” asks the older adults about their comfort with haptic feedback modes and usage, which is related to perceived ease of use. Comfortable feedback modes are usually easier for users to accept and use. “Would you like to try a card game with haptic feedback technology?” asks older adults about their willingness to accept haptic feedback technology, which is often affected by their perception of the ease of use of the technology. These questions are related to perceived ease of use.

Questions related to multisensory stimulation. For example, “Which haptic feedback do you prefer?” directly involves the user's preference for haptic feedback, which is one of the multisensory experiences. “Do you think that card games incorporating haptic feedback technology could reduce the need to use vision to obtain information?” asks about the role of haptic feedback in replacing or reducing visual burden, reflecting the mutual complementation and collaboration between different senses in a multisensory experience. “Do you think card games combined with haptic feedback technology are a good combination?” asks the older adults about their overall evaluation of the combination of haptic feedback and card games, which can provide a multisensory interactive experience.

Questions related to experience richness: “Why do you like playing card games?” Understand the reasons why users like card games, which may include the multisensory experience brought by the game, such as visual patterns, auditory sound effects, and haptic operations. “What abilities of the brain do you think card games can train?” Although cognitive abilities are asked, from the perspective of multisensory experience, multisensory participation can often more comprehensively exercise the brain's multiple abilities, such as attention, memory, and thinking ability.

This study adopted a descriptive survey design [34] to investigate the preferences and opinions of older adults regarding haptic feedback in mobile card games. Data were collected through a structured questionnaire and the older adults using the Wenjuanxing [35,36] mobile app, aiming to quantitatively and qualitatively assess the respondents' acceptance of haptic-enhanced serious games. The questionnaire consisted of 4 parts: part I comprises basic cognition about card games and serious games, part II comprises a combination of card games with mobile devices, part III comprises the potential impact of combining card games with haptic feedback technology, and part IV comprises suggestions and opinions.

This study aims to quantitatively analyze and insight respondents' acceptance of new serious game designs by distributing questionnaires through the Wenjuanxing platform [35,36]. The questionnaire data were summarized as percentages and modularly analyzed using the platform to assess familiarity with card games and responses to other closed questions.

While this study provides valuable insights, certain limitations must be acknowledged. The relatively small sample size and regional focus on Dangtu County limit the generalizability of the investigation. In addition, respondents had varying levels of familiarity with digital technology, which may have influenced responses to haptic feedback.

Results

We divided the results based on a few aspects as described below.

Sociodemographic Characteristics

The respondents were selected by purposive sampling. A total of 250 older adults aged 60-80 years in Dangtu County were surveyed, 64 were male and 186 were female. Among them, 155 were 60-64 years old, accounting for 62% (155/250); 57 were 65-69 years old, accounting for 22.8% (57/250) of participants; 29 were 70-74 years old, accounting for 11.6% (29/250) of participants; 9 were 75-80 years old, accounting for 3.6% (9/250) of participants (Table 1).

XSL•FO
RenderX

Table 1. Sociodemographic characteristics of older adults aged 60-80 years in Dangtu County, China: a descriptive survey of 250 participants from October to December 2024.

Demographic characteristics	Proportion, n (%)
Gender	
Male	64 (25.6)
Female	186 (74.4)
Age group	
60-64	155 (62)
65-69	57 (22.8)
70-74	29 (11.6)
75-80	9 (3.6)
Area	
DangTu County	250 (100)
Work or retired	
Work	5 (2)
Retired	245 (98)

Reliability and Validity Analysis

The 14 questions in the questionnaire were analyzed for reliability and validity using SPSS data analysis software [37]. The coefficient obtained by Cronbach reliability [38] analysis was 0.729, which was between 0.7 and 0.8, indicating good reliability (refer to Table 2). The results of the validity analysis are shown in Table 3. The Kaiser-Meyer-Olkin (KMO) [39] value was 0.824, and the KMO value was greater than 0.8, indicating that the research data was very suitable for information extraction.

Table 2. A reliability analysis of the haptic user perception questionnaire among 250 older adults in Dangtu County, China from October to December 2024 was conducted using SPSS (Cronbach α coefficient=0.729; standardized Cronbach α coefficient=0.812).

Name	Corrected total correlation	α Coefficient with item deleted
Do you know serious games?	0.174	0.728
Are you familiar with card games?	0.307	0.719
Why do you like playing card games?	0.228	0.739
Do you think card games can exercise your brain?	0.301	0.726
What abilities of the brain do you think card games can train?	0.324	0.720
Would you like to try mobile card games?	0.414	0.712
Are you willing to try a mobile card game that reduces visual fatigue?	0.424	0.714
Which haptic feedback do you prefer?	0.402	0.714
Which mode of haptic feedback are you comfortable with?	0.438	0.703
Would you like to try a card game with haptic feedback technology?	0.520	0.706
Do you think that card games incorporating haptic feedback technology could reduce the need to use vision to obtain information?	0.520	0.707
Do you think card games combined with haptic feedback technology are a good combination?	0.557	0.705
Do you think card games combined with haptic feedback technology can help improve your cognitive abilities?	0.576	0.706
What aspects of your cognitive abilities do you think card games combined with haptic feedback technology can improve?	0.497	0.692

XSL•FO
RenderX

Table 3. A validity analysis[a] of the haptic user perception questionnaire among 250 older adults in Dangtu County, China from October to December 2024 was conducted using SPSS (IBM Corp).

Name	Factor 1	Factor 2	Factor 3	Factor 4	Commonality (common factor variance)
Do you know serious games? (Factor loading coefficient)	–0.001	0.207	–0.107	0.752	0.620
Are you familiar with card games? (Factor loading coefficient)	0.183	0.018	0.147	0.742	0.606
Why do you like playing card games? (Factor loading coefficient)	0.004	0.183	0.608	–0.185	0.437
Do you think card games can exercise your brain? (Factor loading coefficient)	0.151	0.668	–0.111	0.163	0.507
What abilities of the brain do you think card games can train? (Factor loading coefficient)	0.012	0.806	0.154	0.004	0.673
Would you like to try mobile card games? (Factor loading coefficient)	0.668	0.032	–0.022	0.269	0.520
Are you willing to try a mobile card game that reduces visual fatigue? (Factor loading coefficient)	0.671	0.098	0.089	0.023	0.468
Which haptic feedback do you prefer? (Factor loading coefficient)	0.175	–0.039	0.653	0.414	0.629
Which mode of haptic feedback are you comfortable with? (Factor loading coefficient)	0.273	0.080	0.684	0.060	0.553
Would you like to try a card game with haptic feedback technology? (Factor loading coefficient)	0.843	–0.018	0.139	0.082	0.738
Do you think that card games incorporating haptic feedback technology could reduce the need to use vision to obtain information? (Factor loading coefficient)	0.792	0.073	0.140	0.074	0.657
Do you think card games combined with haptic feedback technology are a good combination? (Factor loading coefficient)	0.850	0.147	0.112	0.002	0.757
Do you think card games combined with haptic feedback technology can help improve your cognitive abilities? (Factor loading coefficient)	0.809	0.166	0.165	0.012	0.710
What aspects of your cognitive abilities do you think card games combined with haptic feedback technology can improve? (Factor loading coefficient)	0.194	0.695	0.349	0.079	0.648
Eigenvalue (before rotation) (Factor loading coefficient)	4.509	1.661	1.250	1.103	—[b]
Variance explained (before rotation; %)	32.207	11.864	8.932	7.880	—
Cumulative variance explained (before rotation; %)	32.207	44.071	53.003	60.883	—
Eigenvalue (after rotation; factor loading coefficient)	3.812	1.727	1.542	1.442	—
Variance explained (after rotation; %)	27.228	12.338	11.014	10.303	—
Cumulative variance explained (after rotation; %)	27.228	39.565	50.580	60.883	—

[a]KMO: Kaiser-Meyer-Olkin. KMO value=0.824; Bart spherical value=1161.161; *df*=91; *P*<.001.

[b]Not applicable.

Familiarity With Card Games

In the study population, 49.6% (124/250) of the older adults were familiar with card games, with 31.2% (78/250) of those aged 60-64 years, 11.6% (29/250) of those aged 65-69 years, and further, the proportion of those aged 70-74 years familiar with card games was 6% (15/250), while for those aged 75-80 years, the proportion was 0.8% (2/250).

Willingness to Try Mobile Card Games

Within the scope of our survey, approximately 58.8% (147/250) of older adults indicated a willingness to try mobile card games. The distribution of willingness across age groups, from highest to lowest, is as follows: those aged 60-64 years accounted for 38.8% (97/250), those aged 65-69 years accounted for 12% (30/250), those aged 70-74 years accounted for 5.6% (14/250), and those aged 75-80 years accounted for 2.4% (6/250). Concurrently, 78.4% (196/250) of older adults expressed a willingness to try mobile card games designed to alleviate visual fatigue. Among them, the proportion of those aged 60-64 years was approximately 49.6% (124/250), those aged 65-69 years was approximately 17.6% (44/250), those aged 70-74 years was approximately 8.4% (21/250), and those aged 75-80 years was approximately 2.8% (7/250).

Motivations for Playing Card Games

In the survey regarding the reasons for enjoying card games, 23.6% (59/250) of older adults chose to play card games with friends, 43.2% (108/250) believed that playing card games could exercise their brains, 19.6% (49/250) selected playing card games to pass the leisure time, 7.2% (18/250) cited personal interest, and 6.4% (16/250) of the older adults selected other reasons (refer to Table 4).

XSL•FO
RenderX

Table 4. Survey results on motivation and haptic feedback preference for playing card games among 250 older adults in Dangtu County, China: A descriptive study from October to December 2024.

Survey project	Proportion (%)	Number
Motivations for playing card games		
Playing with friends	23.6	—[a]
Exercise brains	43.2	—
Passing leisure time	19.6	—
Personal interest	7.2	—
Other reasons	6.4	—
Preference for haptic feedback		
Haptic feedback akin to the vibration experienced when a smartphone receives a call in silent mode	30.8	77
Uncertainty regarding the type of haptic feedback	52.4	131
Tremor feedback, which resembles the involuntary hand tremors characteristic of Parkinson disease	9.6	24
Shaking feedback (similar to shivering in the cold)	4.8	12
Jiggle feedback (akin to the fine adjustments needed when attempting to open a lock with inappropriate force)	2.4	6
Haptic feedback mode		
Slight haptic feedback mode, analogous to the subtle vibration experienced when a smartphone receives a text message	63.2	158
Slightly and continuously haptic feedback mode, similar to the gentle, persistent vibration of an electric toothbrush	17.2	43
Strong haptic feedback mode, akin to the intense vibration of a high-speed spinning washing machine	4.0	10
Strongly and continuously haptic feedback, reminiscent of the strong and continuous vibrations felt when cycling or motorcycling over rough terrain	2.0	5

[a]Not applicable.

Preference for Haptic Feedback

In the third section of this paper, we focus on the preferences and perceptions of older adults toward haptic feedback in mobile card games. The study revealed that 77 respondents (comprising 30.8%, 77/250 of the total sample) exhibited a preference for haptic feedback akin to the vibration experienced when a smartphone receives a call in silent mode. Regarding age distribution, 42 respondents aged between 60 and 64 years accounted for 16.8% (42/250) of the total sample; 23 respondents aged between 65 and 69 years constituted 9.2% (23/250) of the total sample; 8 respondents aged between 70 and 74 years represented 3.2% (8/250) of the total sample, and 4 respondents aged between 75 and 80 years made up 1.6% (4/250) of the total sample. Notably, 131 respondents (52.4%, 131/250 of the total sample) expressed uncertainty regarding the type of haptic feedback. In total, 24 respondents (24/250, 9.6%) of the total sample) opted for tremor feedback, which resembles the involuntary hand tremors characteristic of Parkinson disease. In addition, 12 participants chose shaking feedback (similar to shivering in the cold) and 6 participants selected jiggle feedback (akin to the fine adjustments needed when attempting to open a lock with inappropriate force), representing 4.8% (12/250) and 2.4% (6/250) of the total sample, respectively (Table 4).

Haptic Feedback Mode

The survey findings indicate that 158 individuals (comprising 63.2%, 158/250 of the total sample) selected a slight haptic feedback mode, analogous to the subtle vibration experienced when a smartphone receives a text message. In terms of age distribution, there were 100 respondents aged 60-64 years, representing 40% (100/250) of the total sample. For those aged 65-69 years, there were 32 respondents, accounting for 12.8% (32/250) of the total sample. The 70- to 74-year age group had 21 respondents, constituting 8.4% (21/250) of the total sample, and the 75- to 80-year age group had 5 respondents, making up 2% (5/250) of the total sample. Furthermore, 43 out of 250 (17.2%) individuals opted for a slightly and continuously haptic feedback mode, similar to the gentle, persistent vibration of an electric toothbrush. In total, 10 participants (10/250, 4%) of the total sample) chose a strong haptic feedback mode, akin to the intense vibration of a high-speed spinning washing machine, while another 5 participants (5/250, 2%) of the total sample) selected a strong and continuously haptic feedback, reminiscent of the strong and continuous vibrations felt when cycling or motorcycling over rough terrain (Table 4).

Age-Based Preferences

Overall, 40% of older adults aged 60-64 years prefer slight haptic feedback more than older adults in other age groups. This is because they are more familiar with mobile phone vibrations and have better sensory abilities than older adults in other age

XSL•FO
RenderX

groups. Older adults aged 75-80 years (5/250, 2%) have a lower preference for haptic feedback, which may be due to their reduced sensitivity to haptic stimulation or greater reliance on visual information as they age.

Haptic Feedback and Card Games Integration

The survey revealed that 163 older adult individuals (163/250, 65.2% of the total sample) expressed a willingness to try this mobile card game. Notably, within the 60- to 64-year age group, 97 out of 250 (38.8%) respondents indicated a readiness to engage with the game. For the 65- to 69-year age group, 39 out of 250 (15.6%) respondents showed their willingness to participate. In the 70- to 74-year age bracket, 19 out of 250 (7.6%) respondents expressed an interest in playing the game. Finally, within the 75- to 80-year age range, 8 out of 250 (3.2%) respondents indicated their willingness to participate in the game. Regarding the integration of haptic feedback technology with mobile card games, 187 out of 250 (74.8%) older adults deemed this combination to be an excellent pairing. In the 60- to 64-year age group, 115 out of 250 (46%) respondents considered this combination to be very favorable. Among the 65- to 69-year age group, 41 out of 250 (16.4%) respondents held this view. In the 70- to 74-year age bracket, 24 out of 250 (9.6%) respondents agreed with this combination. Finally, within the 75- to 80-year age range, 7 out of 250 (2.8%) respondents concurred with this pairing.

Previous Gaming Experience

Older adults who had experience playing card games were more receptive to haptic feedback technology. Of the 49.6% (124/250) of older adults who were familiar with card games, 33.2% (83/250) of participants chose slight haptic feedback, indicating that familiarity with the gameplay of card games would enhance their acceptance of new technologies. This suggests that combining technology with familiar activities can improve acceptance.

Cognitive Status

Among 250 participants aged 60-80, 83.6% (209/250) had symptoms of memory loss, 1.2% (3/250) had mild cognitive impairment, 0.4% (1/250) had Alzheimer disease, and 14.8% (37/250) chose other. In total, 85.2% (213/250) of older adults had significant cognitive decline. These data suggest that haptic feedback technology may provide them with a friendlier cognitive training experience by reducing reliance on visual information.

Enhancing Cognitive Abilities

This study observed that 196 older adult individuals (comprising 78.4%, 196/250 of the sample) believed that the combination of haptic feedback technology with mobile card games is beneficial for enhancing cognitive abilities, a viewpoint that was uniformly recognized across all age groups. Notably, among respondents aged 75- to 80-year, 8 out of 250 (3.2%) individuals held this opinion. Following closely, 25 out of 250 (10%) respondents aged 70-74 years agreed with this combination, and 40 out of 250 (16%) respondents aged 65-69 years also believed that this integration could improve cognitive functions. Furthermore, 123 out of 250 (49.2%) respondents aged 60-64 years concurred that this combination could enhance cognitive performance.

Reducing Reliance on Visual Information

In the exploration of older adult individuals' perceptions toward mobile card games integrated with haptic feedback technology, our study found that 184 out of 250 (73.6%) older adult participants believed that this technology could reduce their dependence on visual information. Among them, 115 individuals aged 60-64 years, accounting for 46% (115/250) of the respondents in that age group, held this view. In the 65- to 69-year age group, 39 participants, representing 15.6% (39/250) of the respondents in that age group, shared this perspective. Within the 70- to 74-year age bracket, 22 older adult individuals agreed with this combination, constituting 8.8% (22/250) of the total sample. Finally, in the 75- to 80-year age group, 8 older adult individuals concurred with this combination, representing 3.2% (8/250) of the total sample.

Technological Integration

In the survey, the positive views of older adults individuals regarding the integration of haptic feedback technology in mobile card games for enhancing cognitive abilities were evident, with specific findings as follows: 108 older adults individuals (approximately 43.2%, 108/250) believed that the technology could improve their reaction capabilities, 66 older adults individuals (approximately 26.4%, 66/250) thought it could enhance their memory, and 47 older adults individuals (approximately 18.8%, 47/250) felt that it could improve their thinking abilities.

Long-Term Engagement and Effectiveness

Although this study focused on initial user preference, it is critical to evaluate the long-term engagement and effectiveness of haptic feedback technology in serious games. Qualitative analysis showed that older adults believed that haptic feedback made games attractive and could promote continued engagement. Haptic feedback technology provides another way to conduct long-term cognitive training. At the same time, serious card games incorporating haptic feedback can reduce visual dependence and may lead to sustained improvements in cognitive areas such as memory, attention, and problem-solving ability.

Opinions and Suggestions

In the final stage of the study, since this was not a mandatory question, the research team collected qualitative feedback from 31 older adult respondents to investigate their opinions and suggestions on incorporating haptic feedback technology into mobile card games. We categorized the qualitative responses into 5 distinct groups. The first category of participants emphasized the importance of regular engagement in card games for maintaining cognitive function, highlighting the need for such activities to “keep the mind sharp.” Another group expressed a willingness to explore methods that could enhance memory and improve their lifestyle, indicating an openness to new cognitive tools. A third group considered that serious games could enrich the daily lives of older adults, making their experiences “more colorful and enjoyable.” Furthermore, there was enthusiasm for the integration of haptic technology, with

XSL•FO
RenderX

the fourth category of participants noting their excitement about the potential fun it could bring. The final group of participants expressed a belief that these games have the potential to enhance memory and may help prevent dementia.

Discussion

Visual Accessibility and Multisensory Engagement

Card games have significant health benefits, particularly in improving sleep quality. A study of 4718 older adults showed that 62% (155/250) reported good sleep quality, and those who played Mahjong daily were more likely to report better sleep quality than those who did not play Mahjong [40]. The integration of new technologies, such as artificial intelligence and haptic feedback, has further enhanced the appeal of card games to older adults [41]. The multiplayer card game Dou Di Zhu [42] remains the most popular game in China with its optimal player characters, deck composition, and scoring system [43]. Innovations such as haptic feedback in mobile card games provide a more realistic experience, reduce visual fatigue, and encourage older players to adopt digital platforms. In virtual reality [44], immersive experiences such as CardsVR promote social interaction and provide realistic haptic sensations, significantly enhancing presence indicators such as action possibilities, realism, and touch [45].

User Preference for Haptic Feedback Intensity

In-depth interviews revealed that older adults were particularly receptive to vibration-based haptic feedback. Their familiarity with mobile phone vibrations (eg, call alerts and message notifications) also contributed to this preference. In addition, slight haptic feedback was perceived as comfortable and beneficial for relieving muscle stiffness and addressing age-related muscle and joint mobility decline [46]. These findings highlight the adaptability of older adults to new technologies when they combine them with familiar elements.

User Preferences (Immersive Benefit and Cognitive Effect)

Haptic feedback technology also enhances cognitive engagement in card games, offering great potential for cognitive intervention. Card games combined with haptic sensations simulate real-world interactions and increase realism and immersion. This innovation is consistent with research highlighting the role of card games in maintaining cognitive function and tracking cognitive impairment [47-49]. In addition, the slight haptic feedback in the haptic feedback can further support cognitive health by reducing physical discomfort and promoting sustained engagement.

The analysis of the research results based on the theoretical framework shows that 78.4% (196/250) of older adults believe that mobile card games combined with haptic feedback can improve their cognitive abilities. This shows that older adults have a high perceived usefulness of haptic feedback technology. Similarly, 63.2% (158/250) of the older adults prefer a slight haptic feedback mode, which shows that the older adults believe that this haptic feedback mode is more suitable for the older adults and that the older adults feel comfortable. In total, 73.6% (184/250) of the older adults believe that haptic feedback technology can reduce dependence on visual information, which shows that multisensory experience can reduce visual fatigue while enhancing user immersion and engagement. Overall, 74.8% (187/250) of the older adults believe that haptic feedback technology combined with card games is a good combination, which shows that multisensory experience can enhance user experience.

Design Recommendations

Older adults of all ages rated the combination of haptic feedback and mobile card games as a beneficial innovation. These technologies improve the gaming experience for visually impaired users [50] and help alleviate vision-related challenges associated with aging. By providing an alternative way to receive game information, haptic feedback enhances accessibility and cognitive stimulation. Overall, these innovations suggest that mobile card games with haptic feedback can have a positive impact on cognitive function and overall quality of life in older adults by promoting engagement and enjoyment.

Future serious games should provide rich sensory experiences through multimodal stimulation design to enhance user engagement and cognitive effects. The intensity and mode of haptic feedback should be personalized according to the haptic sensitivity and cognitive ability of older adults to ensure that each older person can obtain the best user experience and cognitive effects. In addition, game design should have tasks with progressive difficulty to ensure the challenge of the game, which can also be achieved by dynamically adjusting haptic feedback.

When designing haptic feedback technology, the accessibility and inclusiveness of the technology should be ensured, especially for older adults with sensory impairments. Overall, 73.6% (184/250) of older adults believe that haptic feedback technology can reduce dependence on visual information, which shows the importance of inclusive design. At the same time, the security and privacy protection of user data are also key to technology design. Technology design should also avoid exacerbating digital barriers for older adults and ensure that older adults with different digital literacy levels can use it.

The results provide valuable insights for game developers, indicating that slight haptic feedback is more popular with older adults. At the same time, the design of haptic technology should focus on multisensory experience to enhance user immersion and engagement. Based on the haptic preferences of older adults when using mobile card games, the researchers will design a serious mobile card game that includes slight haptic feedback and is suitable for older adults. This serious card game can improve older adults' cognition while alleviating their visual impairment.

This study extends the existing literature by focusing on older adults' preferences for haptic feedback in serious card games, an area that remains underexplored. Previous studies have highlighted the potential of serious games to improve cognitive abilities, but often overlook the specific needs and preferences of older adults [51]. Our findings are consistent with these studies, demonstrating the positive response of interactive

XSL•FO
RenderX

technologies on cognitive function [52]. However, this study uniquely identifies the potential of haptic feedback in reducing visual dependence in older adults, a new contribution to the field.

In addition, the high acceptance rate of haptic feedback technology by participants supports previous findings that older adults are more likely to accept new technologies when they are combined with familiar activities [53]. This study also builds on the research of Abd-Alrazaq et al [54]. They report that serious games can improve cognitive abilities in older adults with mild cognitive impairment, who experience memory loss as they age [55]. Our results suggest that haptic feedback technology can further enhance these benefits by providing a more immersive and accessible gaming experience.

Limitations

The limitations of this study are mainly reflected in sample size and geographical restrictions, differences in digital literacy, self-reported data, and lack of longitudinal evaluation. First, the sample size was limited to 250 older adults in Dangtu County, Anhui Province, China, which may affect the generalizability of the research results. Future studies should expand the sample size and research area. Second, the digital literacy of the participants varied, which may affect their response to haptic feedback, thereby limiting the ability of the research results to be generalized to different digital literacy populations. Third, this study relied on self-reported data, and future studies could supplement it with objective measurement methods such as cognitive performance tests and usability evaluations. Finally, this study did not evaluate the long-term effects of haptic feedback on cognitive ability in serious games. It is recommended that future longitudinal studies be conducted to evaluate its continued impact.

Future Directions

Haptic feedback technology has potential application value in cognitive training. Providing haptic stimulation can help older adults obtain more information, thereby promoting the improvement of cognitive function. The focus of future work should be to explore adaptive haptic feedback models based on user progress and cognitive response. Such a model can dynamically adjust the intensity, frequency, and mode of haptic feedback in a timely manner according to the users' real-time performance and cognitive state, providing personalized and more effective feedback to achieve better services for older adults.

Conclusions

This study explored the haptic preferences of older adults for haptic feedback integrated into a mobile serious card game. The results showed that older adults showed strong interest in this combination, believing it could improve cognitive abilities while reducing reliance on visual information. These findings highlight older adults' preference for haptic feedback in such games.

Acknowledgments

We would like to thank the participants from the Dangtu Old-age University. Funding was partly supported by a research grant from Universiti Kebangsaan Malaysia (grant GUP-2019-066).

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

XH led the conceptualization, with NMA and SS contributing equally. Data curation was also led by XH, with equal contributions from NMA and SS. Formal analysis was led by XH, with equal input from NMA. Funding acquisition was handled by NMA. XH led the development of the methodology, supported equally by NMA and SS. Resources were provided by XH. Supervision was led by NMA, with equal contribution from SS. Validation was led by XH, with equal involvement from NMA and SS. XH was responsible for visualization and for writing the original draft. The review and editing of the manuscript were led by XH, with equal contributions from NMA and SS.

Conflicts of Interest

None declared.

References

1. Deng S, Chang J, Zhang JJ. A survey of haptics in serious gaming. Lecture Notes in Computer Science 2013;8605:1-12. [doi: 10.1007/978-3-319-12157-4_11]
2. Kincaid JP, Westerlund KK. Simulation in education and training. 2009 Presented at: Proceedings of the 2009 Winter Simulation Conference (WSC); 2009 December 13-16; Austin, TX, USA p. 273-280. [doi: 10.1109/wsc.2009.5429337]
3. Derryberry A. Serious games: online games for learning. Adobe Whitepaper. 2007. URL: https://www.seriousgamemarket.com/2007/11/adobe-serious-games-whitepaper-by-anne.html [accessed 2007-11-14]

XSL•FO
RenderX

4. Li YX, Yoo Y, Weill-Duflos A, Cooperstock JR. Towards context-aware automatic haptic effect generation for home theatre environments. 2021 Presented at: VRST '21: Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology; 2021 December 8 - 10; Osaka Japan p. 1-10. [doi: 10.1145/3489849.3489887]
5. Kaklanis N, Votis K, Tzovaras D. Adding haptic Feedback to web applications towards improving end-users? Cognitive capabilities. 2015 Presented at: 6th IEEE International Conference on Cognitive Infocommunications (CogInfoCom); 2015 October 19-21; Gyor, Hungary p. 245-249. [doi: 10.1109/coginfocom.2015.7390599]
6. Saint-Louis C, Hamam A. Survey of haptic technology and entertainment applications. 2021 Presented at: In Proceedings of the 2021 IEEE SoutheastCon; 2021 March 10-13; Atlanta, GA, USA p. 493-499. [doi: 10.1109/southeastcon45413.2021.9401939]
7. Pacheco-Barrios K, Ortega-Márquez J, Fregni F. Haptic Technology: Exploring its underexplored clinical applications-a systematic review. Biomedicines 2024;12(12):2802 [FREE Full text] [doi: 10.3390/biomedicines12122802] [Medline: 39767709]
8. Volz MS, Suarez-Contreras V, Mendonca ME, Pinheiro FS, Merabet LB, Fregni F. Effects of sensory behavioral tasks on pain threshold and cortical excitability. PLoS One 2013;8(1):e52968 [FREE Full text] [doi: 10.1371/journal.pone.0052968] [Medline: 23301010]
9. Davis RL, Orta Martinez M, Schneider O, MacLean KE, Okamura AM, Blikstein P. The haptic bridge: Towards a theory for haptic-supported learning. 2017 Presented at: IDC '17: Proceedings of the 2017 Conference on Interaction Design and Children; 2017 June 27 - 30,; Stanford California USA p. 51-60. [doi: 10.1145/3078072.3079755]
10. Sun F, Lu Q, Hao M, Wu Y, Li Y, Liu L, et al. An artificial neuromorphic somatosensory system with spatio-temporal tactile perception and feedback functions. npj Flex Electron 2022;6(1):72. [doi: 10.1038/s41528-022-00202-7]
11. Weiss EJ, Flanders M. Somatosensory comparison during haptic tracing. Cereb Cortex 2011;21(2):425-434 [FREE Full text] [doi: 10.1093/cercor/bhq110] [Medline: 20542990]
12. Kitada R. The brain network for haptic object recognition. In Pervasive Haptics: Science, Design, and Application. Tokyo: Springer; 2016:21-37.
13. Millar S. Network models for haptic perception. Infant Behavior and Development 2005;28(3):250-265. [doi: 10.1016/j.infbeh.2005.05.003]
14. Zhang S, Wang D, Afzal N, Zhang Y, Wu R. Rhythmic haptic stimuli improve short-term attention. IEEE Trans. Haptics 2016;9(3):437-442. [doi: 10.1109/toh.2016.2531662]
15. List A, Iordanescu L, Grabowecky M, Suzuki S. Haptic guidance of overt visual attention. Atten Percept Psychophys 2014;76(8):2221-2228 [FREE Full text] [doi: 10.3758/s13414-014-0696-1] [Medline: 24935805]
16. Wang D, Zhang Y, Yang X, Yang G, Yang Y. Force control tasks with pure haptic Feedback promote short-term focused attention. IEEE Trans. Haptics 2014;7(4):467-476. [doi: 10.1109/toh.2014.2359007]
17. Muñoz JE, Montoya MF, Boger J. From exergames to immersive virtual reality systems: serious games for supporting older adults. In Services for Geriatric Rehabilitation 2022:141-204. [doi: 10.1016/b978-0-323-85173-2.00011-4]
18. King PH. Smart home technologies and services for geriatric rehabilitation. IEEE Pulse 2022;13(3):35-36. [doi: 10.1109/mpuls.2022.3175383]
19. Huang X, Ali NM, Sahrani S. Evolution and future of serious game technology for older adults. Information 2024;15(7):385. [doi: 10.3390/info15070385]
20. Gao YX, Li WX, Xiao YH, Khalid MNA, Iida H. Nature of attractive multiplayer games: Case study on China's most popular card game—DouDiZhu. Information 2020;11(3):141. [doi: 10.3390/info11030141]
21. Gielis K, Kennes J, De Dobbeleer C, Puttemans S, Abeele VV. Collecting digital biomarkers on cognitive health through computer vision and gameplay: an image processing toolkit for card games. 2019 Presented at: IEEE International Conference on Healthcare Informatics (ICHI); 2019 June 13; Xi'an, China p. 127-138. [doi: 10.1109/ichi.2019.8904511]
22. Estrada-Plana V, Montanera R, Ibarz-Estruga A, March-Llanes J, Vita-Barrull N, Guzmán N, et al. Cognitive training with modern board and card games in healthy older adults: two randomized controlled trials. Int J Geriatr Psychiatry 2021;36(6):839-850. [doi: 10.1002/gps.5484] [Medline: 33275804]
23. Hou XY, Sourina O, Klimenko S. Haptic-based serious games. 2014 Presented at: International Conference on Cyberworlds; 2014 October 06-08; Santander, Spain p. 39-46. [doi: 10.1109/cw.2014.14]
24. Yuan B, Folmer E, Harris FC. Game accessibility: a survey. Univ Access Inf Soc 2010;10(1):81-100. [doi: 10.1007/s10209-010-0189-5]
25. Chan MS, Black JB. In Annual Meeting of the American Educational Research Association (AERA). Learning Newtonian mechanics with an animation game: The role of presentation format on mental model acquisition. 2006. URL: https://ilt.tc.columbia.edu/publications/cognition-and-learning-lab/ [accessed 2025-05-01]
26. Kaaresoja T, Brewster S, Lantz V. Towards the temporally perfect virtual button: Touch-Feedback simultaneity and perceived quality in mobile touchscreen press interactions. ACM Trans. Appl. Percept 2014;11(2):1-25. [doi: 10.1145/2611387]
27. Pacchierotti C, Prattichizzo D, Pacchierotti C. Cutaneous/tactile haptic Feedback in robotic teleoperation: Motivation, survey, and perspectives. IEEE Trans. Robot 2024;40:978-998. [doi: 10.1109/tro.2023.3344027]
28. Rantala J, Salminen K, Raisamo R, Surakka V. Touch gestures in communicating emotional intention via vibrotactile stimulation. International Journal of Human-Computer Studies 2013;71(6):679-690. [doi: 10.1016/j.ijhcs.2013.02.004]

XSL•FO
RenderX

29. Choi S, Kuchenbecker KJ. Vibrotactile display: Perception, technology, and applications. Proc. IEEE 2013;101(9):2093-2104. [doi: 10.1109/jproc.2012.2221071]
30. Neacsu A. Exploring Haptic Technology?s Impact on User Experience. HyperSense. 2024. URL: https://hypersense-software.com/blog/2024/07/15/haptic-technology-user-experience/ [accessed 2024-09-19]
31. Neacsu A. Mobile Gaming UX: How Haptic Feedback Can Change the Game. HyperSense. 2024. URL: https://interhaptics.medium.com/mobile-gaming-ux-how-haptic-feedback-can-change-the-game3ef689f889bc [accessed 2025-05-01]
32. Ibrahim F, Münscher JC, Daseking M. The technology acceptance model and adopter type analysis in the context of artificial intelligence. Frontiers in Artificial Intelligence 2024:41. [doi: 10.2139/ssrn.4860202]
33. Velasco C, Obrist M. Multisensory experiences: A primer. Front. Comput. Sci 2021;3:614524. [doi: 10.3389/fcomp.2021.614524]
34. Höfers W, Jørgensen V, Sällström S, Vege KM, Strøm M, New PW, et al. Organisation of services and systems of care in paediatric spinal cord injury rehabilitation in seven countries: A survey with a descriptive cross-sectional design. Spinal Cord 2022;60(4):339-347. [doi: 10.1038/s41393-021-00726-1] [Medline: 34802054]
35. WJX.cn. (n.d.). About Us. URL: https://www.wjx.cn/html/aboutus.aspx [accessed 2025-05-01]
36. Wang W. Network Survey and Data Integration. China: Wuhan University Press; 2018.
37. Becker H, Bialasiewicz AA, Schaudig U, Schäfer H, von Domarus D. [Final clinical indications and etiology in 1,023 enucleations. Descriptive databank evaluation with SPSS software in variable response mode with dummy variables]. Ophthalmologe 2002;99(5):367-374. [doi: 10.1007/s003470100539] [Medline: 12043292]
38. Karakaya SPY, Alparslan ZN. Sample size in reliability studies: A practical guide based on cronbach's Alpha. PBS 2022;12(3):150. [doi: 10.5455/pbs.20220127074618]
39. Wikipedia. (n.d.). Kaiser?Meyer?Olkin test. URL: https://en.wikipedia.org/wiki/Kaiser%E2%80%93Meyer%E2%80%93Olkin_test [accessed 2025-05-01]
40. Lee YH, Chang YC, Shelley M, Liu CT. A panel analysis of the mahjong card game and social activity with sleep-related measurements among Chinese older adults. Sleep Biol. Rhythms 2020;18(2):109-119. [doi: 10.1007/s41105-019-00249-6]
41. Pookkuttath S, Abdulkader RE, Elara MR, Veerajagadheswar P. AI-enabled vibrotactile Feedback-based condition monitoring framework for outdoor mobile robots. Mathematics 2023;11(18):3804. [doi: 10.3390/math11183804]
42. Kong Y, Shi HY, Wu XC, Rui YF. Application of DQN-IRL framework in doudizhu’s sparse reward. Neural Process Lett 2023;55(7):9467-9482. [doi: 10.1007/s11063-023-11209-0]
43. Gao YX, Li WX, Xiao YH, Khalid MNA, Iida H. Nature of attractive multiplayer games: Case study on China’s most popular card game—DouDiZhu. Information 2020;11(3):141. [doi: 10.3390/info11030141]
44. Gao BY, Shao T, Tu HW, Ma QZ, Liu ZT, Han T, et al. Exploring bimanual haptic Feedback for spatial search in virtual reality. IEEE Trans. Visual. Comput. Graphics 2024;30(5):2422-2433. [doi: 10.1109/tvcg.2024.3372045]
45. Huard A, Chen MY, Sra M. CardsVR: A two-person VR experience with passive haptic Feedback from a deck of playing cards. 2022 Presented at: IEEE International Symposium on Mixed and Augmented Reality (ISMAR); 2022 October 17-21; Singapore p. 538-547. [doi: 10.1109/ismar55827.2022.00070]
46. Hadi R, Valenzuela A. Good vibrations: Consumer responses to technology-mediated haptic Feedback. Journal of Consumer Research, 2020;47(2):256-271. [doi: 10.1093/jcr/ucz039]
47. Gielis K, Kennes, J, De Dobbeleer C. Collecting digital biomarkers on cognitive health through computer vision and gameplay: an image processing toolkit for card games. 2019 Presented at: 2019 IEEE International Conference on Healthcare Informatics (ICHI); 2019 June 10-13; Xi'an, China p. 127-138. [doi: 10.1109/ichi.2019.8904511]
48. Estrada-Plana V, Montanera R, Ibarz-Estruga A, March-Llanes J, Vita-Barrull N, Guzmán N, et al. Cognitive training with modern board and card games in healthy older adults: two randomized controlled trials. Int J Geriatr Psychiatry 2021;36(6):839-850. [doi: 10.1002/gps.5484] [Medline: 33275804]
49. Hampshire A, Sandrone S, Hellyer PJ. A large-scale, cross-sectional investigation into the efficacy of brain training. Front Hum Neurosci 2019;13:221 [FREE Full text] [doi: 10.3389/fnhum.2019.00221] [Medline: 31338032]
50. Gay SL, Truong NT, Romeo K, Pissaloux E. Using force-Feedback haptic effects to develop serious and entertainment games accessible for visually impaired people. In ICT for Health, Accessibility and Wellbeing 2022;1538:185-197. [doi: 10.1007/978-3-030-94209-0_16]
51. Guzmán DE, Rengifo CF, García-Cena CE. Serious games for cognitive rehabilitation in older adults: A conceptual framework. MTI 2024;8(8):64. [doi: 10.3390/mti8080064]
52. Leung C, Wong KC, So WWY, Tse ZCK, Li D, Cao Y, et al. The application of technology to improve cognition in older adults: A review and suggestions for future directions. Psych J 2022;11(4):583-599 [FREE Full text] [doi: 10.1002/pchj.565] [Medline: 35675967]
53. Vaportzis E, Clausen MG, Gow AJ. Older adults perceptions of technology and barriers to interacting with tablet computers: A focus group study. Front Psychol 2017;8:1687 [FREE Full text] [doi: 10.3389/fpsyg.2017.01687] [Medline: 29071004]
54. Abd-Alrazaq A, Alhuwail D, Al-Jafar E, Ahmed A, Shuweihdi F, Reagu SM, et al. The effectiveness of serious games in improving memory among older adults With cognitive impairment: Systematic review and meta-analysis. JMIR Serious Games 2022;10(3):e35202 [FREE Full text] [doi: 10.2196/35202] [Medline: 35943792]

XSL•FO
RenderX

55. Md Fadzil NHM, Shahar S, Rajikan R, Singh DKA, Mat Ludin AFM, Subramaniam P, et al. A scoping review for usage of telerehabilitation among older adults with mild cognitive impairment or cognitive frailty. Int J Environ Res Public Health 2022;19(7):4000 [FREE Full text] [doi: 10.3390/ijerph19074000] [Medline: 35409683]

Abbreviations

KMO: Kaiser-Meyer-Olkin

Edited by A Coristine; submitted 25.02.25; peer-reviewed by A Baharum, E Monfort; comments to author 03.04.25; revised version received 21.04.25; accepted 28.04.25; published 27.05.25.

Please cite as:
Huang X, Ali NM, Sahrani S
Haptic-Driven Serious Card Games for Older Adults: User Preferences Study
JMIR Serious Games 2025;13:e73135
URL: https://games.jmir.org/2025/1/e73135
doi:10.2196/73135
PMID:

©Xin Huang, Nazlena Mohamad Ali, Shafrida Sahrani. Originally published in JMIR Serious Games (https://games.jmir.org), 27.05.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

A Gamified Assessment Tool for Antisocial Personality Traits (Antisocial Personality Traits Evidence-Centered Design Gamified): Randomized Controlled Trial

Yaobin Tang[1], MS; Yongze Xu[2,3], PhD; Qunli Zhou[4], PsyD; Ran Bian[1], PhD

[1]Faculty of Psychology, Beijing Normal University, Beijing, China

[2]Department of Psychology, Faculty of Arts and Sciences, Beijing Normal University, Zhuhai, China

[3]Beijing Key Laboratory of Applied Experimental Psychology, National Demonstration Center for Experimental Psychology Education, Faculty of Psychology, Beijing Normal University, Beijing, China

[4]Beijing Zhongce Kaiyuan Talent Assessment Technology Co., Ltd, Beijing, China

Corresponding Author:
Yongze Xu, PhD
Department of Psychology
Faculty of Arts and Sciences
Beijing Normal University
18 Jinfeng Road
Tangjiawan
Zhuhai, 519085
China
Phone: 86 13952716181
Email: yzxu@bnu.edu.cn

Abstract

Background: The traditional self-report instruments (eg, scales) used to measure antisocial personality traits are characterized by social desirability bias and fail to capture multidimensional behaviors (eg, manipulation and deception).

Objective: This study aimed to develop and validate an evidence-based design for a gamified assessment tool (Antisocial Personality Traits Evidence-Centered Design Gamified assessment tool; ASP-ECD-G) to measure 7 antisocial personality traits (manipulative, callous, deceptive, hostile, risk taking, impulsive, and irresponsible) as defined in the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* (*DSM-5*).

Methods: This research featured a 3-phase evidence-centered design framework. Ontology development (study 1): semistructured interviews were conducted with 9 workplace professionals to translate the *DSM-5* criteria into 24 observable workplace behaviors, which were integrated into a text-based game featuring 10 subscenarios, 34 interactive questions, and logic rooted in logical jumps to simulate real-world decision-making. Model construction (study 2): 6 machine learning models were trained by reference to a set of Personality Inventory for DSM-5 Short Form scores (n=286). The gated recurrent unit model, which uses 1-hot encoding to address nominal response data, was evaluated in terms of the root mean square error (RMSE), mean absolute error, criterion correlation (*r*), and test-retest reliability. Retest reliability was assessed using intraclass correlation coefficients based on 10 participants (1-month interval). Empirical validation (study 3): a 2×2 mixed design (n=148) was used to compare the gamified assessment with questionnaires under conditions involving incentives (ie, situations in which “rational results” led to increased payments).

Results: For model performance, the gated recurrent unit outperformed the alternatives, as indicated by the highest criterion correlation (r=0.850) and the lowest test RMSE (0.273); in particular, it excelled in moderate score ranges (1.5-3, RMSE≤0.377) and in resisting extreme value distortions (3.5-4, RMSE 0.854). Retest reliability was moderate to strong (intraclass correlation coefficients=0.776, *P*=.02). For validation findings, the gamified assessment was associated with higher levels of immersion (mean 7.628 vs 7.216; F147=14.259, *P*<.001) and interest (mean 7.095 vs 6.155; F147=47.940, *P*<.001), although it also elicited stronger negative emotions (mean 4.365 vs 2.473; F147=151.109, *P*<.001). Incentives reduced questionnaire scores (incentivized: 2.066 vs control: 2.201; F1=5.740, *P*=.02) but had no effect on gamified scores (*P*=.71), confirming resistance to manipulation.

XSL•FO
RenderX

Conclusions: By integrating evidence-centered design with gamified workplace simulations, ASP-ECD-G can provide more objective and ecologically valid measurements of antisocial personality traits, thereby supporting both research and organizational practice.

Trial Registration: Open Science Framework (OSF) Registries tvg6x; https://osf.io/tvg6x

(JMIR Serious Games 2025;13:e70453) doi:10.2196/70453

KEYWORDS

evidence-centered design; gamified assessment; antisocial personality; randomized controlled trial; psychometric validation

Introduction

Overview

Antisocial personality (ASP) traits, which are characterized by manipulativeness, callousness, and deceitfulness, entail significant threats to organizational trust, team dynamics, and ethical decision-making [1]. Individuals who exhibit these traits often exploit others for their personal gain, disregard their team responsibilities, and engage in risky behaviors that undermine the long-term organizational health [2,3]. Traditional assessment tools, such as the Psychopathy Checklist–Revised and the Personality Inventory for DSM-5–Short Form (PID-5-SF), rely heavily on self-reports or clinical interviews, which are susceptible to social desirability bias—especially in nonclinical settings, in which individuals may consciously or unconsciously underreport problematic behaviors [4,5]. For example, self-ratings on the PID-5-SF often fail to capture situational impulsivity or deceitfulness, as they lack the ecological validity of real-time behavioral data in high-stakes scenarios [6].

Even structured tools such as situational judgment tests (SJTs), which simulate workplace dilemmas, struggle to address the multidimensional nature of the antisocial traits defined by the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* (*DSM-5*). These tests typically assume item independence and 1-dimensionality, thereby ignoring the interactive effects among different traits (eg, the possibility of manipulativeness and risk taking co-occurring in decision-making) [7]. Specifically, these methods often lack operability and interest in the context of practical application, thus failing to engage participants effectively [1,3,8,9]; accordingly, they are susceptible to subjective influences and the manipulation of results [8,10,11].

Background

Gamified assessments represent a promising alternative in this context because they embed trait-relevant dilemmas in immersive contexts, in which participants' choices reflect genuine behavioral tendencies rather than reflective self-evaluations [12]. However, existing gamified tools have not yet been aligned with clinical diagnostic criteria, thus giving rise to a situation of fragmented validation and limited utility in organizational settings that require rigorous psychometric evidence [13-15]. We address this gap by using evidence-centered design (ECD), a systematic methodology that translates theoretical constructs into observable behavioral evidence through a process involving 3 iterative stages: capability modeling, evidence modeling, and task design [16]. In the *capability modeling* stage, we defined target traits (eg, manipulativeness or deceitfulness) based on DSM-5 diagnostic criteria, thus ensuring alignment with clinical and organizational contexts. In the *evidence modeling* stage, we identified behavioral indicators by using empirical methods (eg, expert interviews) to bridge abstract traits with observable decisions in simulated scenarios. In the *task design* stage, we constructed gamified tasks that elicit the target behaviors, particularly using dynamic logic (eg, question jump paths) to create immersive, consequence-laden environments in which genuine responses are more likely to emerge.

Although ECD offers significant potential in this context, its application in the field of dark personality traits has not yet been explored in full. This point is especially evident in the context of efforts to operationalize multidimensional constructs such as antisocial behavior.

The ECD framework is a systematic method that can be used to develop assessment tools to construct evidence arguments in game-based assessments beginning in the early stages of assessment by clarifying the evidence needed, and to guide the design and development of the assessment tool accordingly [17]. The ECD assessment has been described as a reasoning process that involves drawing inferences regarding participants' real-world knowledge and abilities from the limited evidence provided in the test environment. The core of ECD lies in the process of constructing a clear design framework that includes a capability model, an evidence model, and a task model [18]; thus, ensuring that the process of developing the assessment tool remains centered on explicit assessment objectives and evidence requirements [17].

The capability model describes and defines the personality traits to be measured, whereas the evidence model translates the capability model into observable behaviors and performances, including details on how participants' behaviors in given task contexts reflect the traits included in the capability model [19]. On the basis of the capability model and the evidence model, rules or models are established with the goal of constructing quantitative relationships between the models, which can range from simple scoring rules to complex logic trees or data-driven mathematical models, including machine learning models [20]. The task model describes the specific tasks that participants must complete as part of the assessment tool. These tasks should effectively elicit the target behaviors on the part of participants, thus providing useful evidence.

The ECD approach ensures the systematic and scientific development of assessment tools. This approach can enhance the reliability and validity of assessment tools and reduce assessment errors by establishing a clear design framework and

XSL•FO
RenderX

defining evidence needs, as noted in Mislevy et al [17]. Moreover, the ECD approach can respond flexibly to different assessment needs and application scenarios, thus rendering the resulting tools more diverse and broadly applicable. For example, in high-stakes recruitment and selection processes, ECD can facilitate the design of fairer and more effective assessment tools [7].

Our gamified assessment framework, which is anchored in the self-determination theory [21] and flow theory [22], functions based on 3 mutually reinforcing mechanisms: narrative immersion, responsive feedback, and dynamic flow induction. The narrative mechanism involves embedding questions in workplace scenarios (eg, "audit visits" or "human resources [HR] dismissal") with the aim of allowing participants to adopt a virtual role that is separate from their real-world identity, thus fulfilling the need for autonomy posited by the self-determination theory and reducing individuals' self-awareness of being assessed, thus mitigating the tendency toward socially desirable responses [23]. The feedback mechanism provides competence-relevant feedback via immediate consequences (eg, dismissal notices in response to dishonest decisions), thus prompting participants to internalize the goal of "surviving" the scenario and to align their responses with in-game logic rather than external judgment; thus, mitigating acquiescence bias [24]. The immersive flow mechanism induces cognitive flow via dynamic path selection (eg, branching storylines based on previous choices), in which context the high cognitive load that results from navigating these paths depletes the mental resources available for response manipulation and effectively "crowds out" deliberate distortion [22].

Objectives

This study aims to fill this gap by developing and validating the Antisocial Personality Traits Evidence-Centered Design Gamified assessment tool (ASP-ECD-G), which integrates the *DSM-5* criteria with workplace scenarios with the aim of measuring 7 core traits based on behavioral data.

This research was conducted in 3 phases: study 1 involved constructing the assessment ontology based on semistructured interviews; study 2 focused on the development of machine learning models aimed at mapping behavioral data to trait scores; and study 3 entailed validating the tool's resistance to manipulation and user experience based on a 2×2 mixed experimental design.

Methods

Overview

As part of this research, 3 studies were designed to develop the ASP-ECD-G: study 1 involved developing an assessment ontology for ASP as well as constructing the capability model, evidence model, and task model within the framework of ECD; study 2 involved constructing an assessment model that linked the response task model with the capability model based on study 1; and study 3 validated the assessment characteristics of the ASP-ECD-G through a 2×2 mixed experimental design.

Development of the Assessment Construct

Determining the Capability Model

This study used the alternative model for diagnosing ASP disorder from the *DSM-5* [25] as the capability model and integrated it with antisocial behaviors in organizational settings, thereby identifying the 7 behavioral characteristics of ASP provided in Textbox 1.

Textbox 1. Seven behavioral characteristics of antisocial personality.

1. Manipulativeness: frequent use of charm, glibness, or flattery to influence or control others for personal gain.
2. Callousness: a lack of empathy, which often involves disregarding others' feelings or problems. When the individual causes harm to others, they express no guilt or remorse and may engage in aggressive and abusive behaviors.
3. Deceitfulness: frequent engagement in fraud against others, misrepresentations of oneself, and embellishments or fabrication of information when it pertains to personal interests.
4. Hostility: persistent and frequent anger, feelings of anger in response to minor slights and insults, and retaliation with harsh, sarcastic, or vengeful behaviors.
5. Risk taking: engagement in potentially dangerous activities without fully considering the consequences, thereby often neglecting personal deficiencies and denying the reality of risks.
6. Impulsivity: rapid responses to immediate stimuli without planning or considering the consequences, and a feeling of difficulty in making and following plans.
7. Irresponsibility: a tendency to shirk one's duties, commitments, or agreements and to opt out of responsibilities when personal interests are at risk.

Constructing the Evidence Model

This study involved semistructured interviews with 9 professionals who had >3 years of work experience, including 1 senior manager from the retail industry; 2 midlevel managers from the telecommunications and smart hardware industries; and 6 frontline employees who were recruited from diverse sectors, such as smart hardware, traditional media, health care, finance, and business consulting. All the interviews were conducted on the web via a third-party platform, and each session lasted 30 to 60 minutes. The interview outline was developed based on the definitions of the 7 behavioral characteristics included in the capability model; its core content is presented in Table 1. Each interview started with warm-up

XSL•FO
RenderX

questions with the goals of establishing a trusting relationship, easing tension, and gradually facilitating in-depth discussions while simultaneously collecting relevant background information regarding the participants. The study used the situation, task, action, and result principle to formulate probing questions for each behavioral event, thereby ensuring the completeness and authenticity of the scenario events discussed during the interviews. The interviews concluded when all the questions had been answered, and the interviewees were asked whether they had anything else to add regarding the questions and answers to confirm that no further information was needed.

The interview contents were transcribed from audio to text form. Through manual screening, the interviews were categorized and organized based on the 7 behavioral characteristics outlined in the interview outline. Workplace behaviors often overlap, thereby reflecting multiple characteristics. This is the result of the complex nature of work environments and decision-making processes. Different perspectives and interpretations can lead to varied understandings of individual behavior. For example, a leader berating an employee for the employee's failure to complete a task because of illness might be viewed as hostile from the leader's perspective but callous from the employee's perspective. Accordingly, the study focused on specific scenarios, and 34 scenarios that effectively reflect ASP traits were derived from the 9 interviews. These scenarios are detailed in Multimedia Appendix 1. Specific behavioral characteristics were then extracted from each scenario based on the core content of the question design highlighted in Table 1, in which context, each scenario could include 1 to 7 ASP traits.

After the 34 scenarios were organized and duplicates removed, 15 unique scenarios remained. The behaviors associated with each role in these scenarios were then extracted and matched with the 7 ASP traits, thus facilitating the identification of 24 typical workplace behaviors. The evidence model was then constructed, as illustrated in Figure 1.

Table 1. Semistructured interview outline used to define antisocial personality behaviors according to the DSM-5a with workplace professionals (n=9).

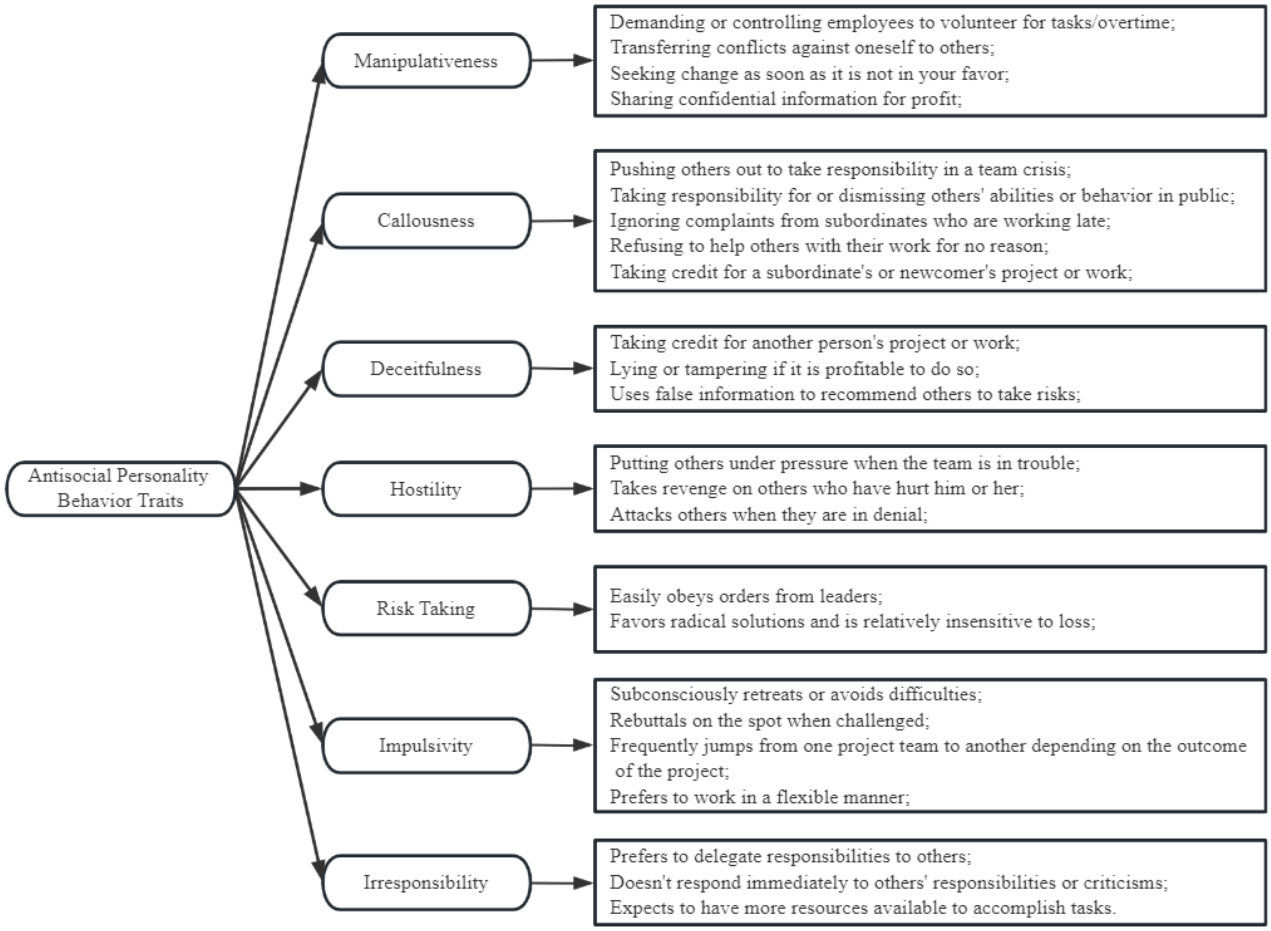
Behavioral characteristics	Core content of the question design
Warm-up question	"Briefly introduce your career experiences in chronological order"
Manipulativeness	Not following company rules, regulations, or laws
Deceitfulness	Intentionally hiding a great deal of information; delivering or reporting false information
Impulsivity	Making workplace decisions driven by emotions or motives without fully considering relevant information
Hostility	Expressing strong hostility, anger, or dissatisfaction, and proactively attacking others
Risk taking	Demonstrating indifference to workplace safety and regulations, including taking actions that pose potential risks to both individuals and teams
Irresponsibility	Displaying dereliction of duty or irresponsibility, leading to an inability to complete one's tasks on time
Callousness	Showing a lack of remorse for wrongdoings and an unwillingness to admit one's mistakes and improve
Probing questions	Probing and supplementing relevant information based on the STAR[b] principle

[a]DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition.

[b]STAR: situation, task, action, and result.

XSL•FO
RenderX

Figure 1. Capability model and evidence model for antisocial personality assessment.



Constructing Assessment Tasks

We operationalized these behavioral characteristics in a game setting by incorporating 34 representative workplace behaviors into an interactive game consisting of 10 detailed subscenarios. We assigned each virtual character a basic name, departmental affiliation, and personal background information to enhance the realism of the game. In addition, we carefully designed transitions between scenes to ensure that the narrative flow was coherent. The relationships between each game scenario and the corresponding behavioral characteristics are presented in Table 2.

Three psychology experts with more than 3 years of workplace experience reviewed and revised the questions, options, and jump relationships based on the capability and evidence models until they reached a consensus. The finalized game consisted of 34 questions with 115 options (2-5 options per question), including 13 questions featuring logical jump relationships across 6 scenarios. This jump logic took 2 forms: question jump (eg, a situation in which choosing Q1-A jumps to Q3, whereas other options proceed to Q2) and progressive (eg, a situation in which Q1/Q2-D jumps directly to Q4). The logical design is illustrated in Figure 2. Examples of the game's item presentation are provided in Multimedia Appendix 2.

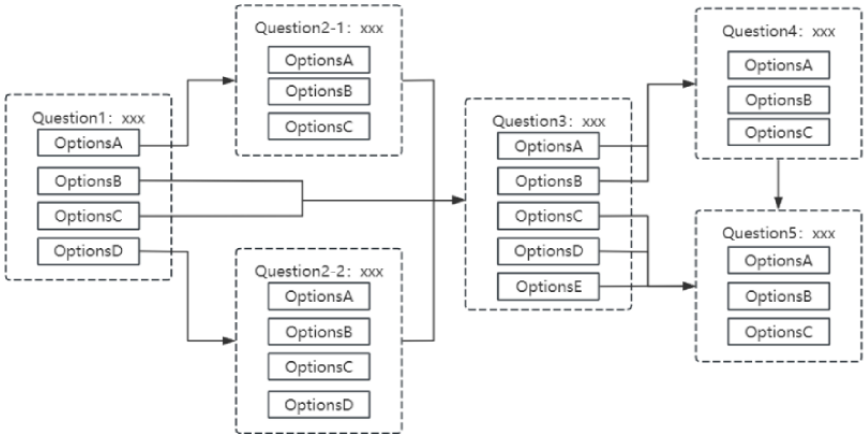
XSL•FO
RenderX

Table 2. Game scenarios and corresponding antisocial traits in the ASP-ECD-G[a] design (10 game scenarios and 34 workplace behavior characteristics).

Game scenario	Workplace behavior characteristics	Antisocial personality traits
Project and proposal decision	• Preference for radical proposals and relative insensitivity to losses	Risk taking
Proposal reporting	• Taking credit for others' proposals or achievements • Seizing subordinates' or newcomers' proposals or achievements	Callousness and deceitfulness
Project initiation meeting	• Retaliation against others who have harmed the focal individual • Proactive attacks when criticized • Immediate rebuttal when questioned • Publicly criticizing or denying others' abilities or actions • Shifting blame for personal conflicts to others	Hostility, impulsivity, and manipulativeness
Helping others	• Unreasonably refusing to help others at work	Callousness
Delivering results	• Demanding or requiring employees to take on tasks or overwork • Shifting blame for personal conflicts to others • Ignoring complaints regarding overtime from subordinates • Failing to respond to criticism during critical periods	Manipulativeness, callousness, and irresponsibility
Altering parameters	• Readily obeying orders from superiors • Lying or altering information for personal gain	Risk taking and deceitfulness
Emergency situations	• Frequently switching project groups or teams based on outcomes • A preference for flexible and changeable work • Pushing others to take responsibility during team crises • Demanding or requiring employees to take on tasks or overwork • Allowing others to bear pressure in situations featuring team difficulties • Expecting more resources to complete tasks • Instinctive retreat, or avoidance, in response to difficulties	Impulsivity, callousness, manipulativeness, and irresponsibility
Information leaks	• Sharing confidential information for personal gain • Lying or altering information for personal gain • Recommending that others should take risks based on false information	Manipulativeness and deceitfulness
Audit visits	• Retaliation against others who have harmed the focal individual • Sharing confidential information for personal gain • Shifting blame for personal conflicts to others • Pushing others to take responsibility during team crises • Lying, or altering information, for personal gain • Preference for radical proposals and relative insensitivity to losses • Expecting more resources to complete tasks	Hostility, manipulativeness, callousness, deceitfulness, impulsivity, and irresponsibility

[a]ASP-ECD-G: Antisocial Personality Traits Evidence-Centered Design Gamified assessment tool.

Figure 2. Logical roadmap for the gamified assessment tool.



XSL•FO
RenderX

Assessment Model Construction

Dependent Variable Acquisition

During the process of model training for the ASP-ECD-G, we used participants' scores on ASP items drawn from the simplified Chinese version of the DSM-5 Personality Inventory (PID-5-SF) [5] as labels for the training set (hereafter referred to as ASP scores). The PID-5-SF has exhibited good reliability and validity in previous research, as indicated by a Cronbach α coefficient of 0.916. The specific items are detailed in Multimedia Appendix 3.

Independent Variable Encoding Methods

We were inspired by traditional SJTs to assign scores to each option, although doing so exclusively based on fixed rules was impractical. The scoring rules were defined without consideration of the influences of prior questions, thus allowing them to serve as a comparative scheme to contrast this method with the 1-hot encoding approach in computer science.

Expert scoring involves assigning values to each option based on the number of ASP traits that it reflects (eg, 5 points for 1 option that reflects 5 out of 7 traits). Three experts assisted in the scoring process, and only the consensus scores obtained after multiple discussions were used. Because of varying question completion rates, the input of each participant was reshaped into a 34×2 matrix, in which context the figure of 34 represented the total number of questions, and the figure of 2 indicated question completion status and scores, for example, (0,0) for unanswered questions and (1,2) for a score of 2. The final training dataset consisted of a 286×34×2 matrix based on data obtained from 286 participants.

In this study, categorized option data that lacked ordinal information were processed via the 1-hot encoding approach. Specifically, each question was encoded as a 5-dimensional vector: for questions that featured fewer than 5 options (eg, Q1 featured only 4 options), zeros were added to the fifth dimension; when option A was selected, the first dimension was set to 1, and the remaining dimensions were set to 0; for unanswered questions, a 5D vector containing all zeros was generated. Ultimately, the training dataset is represented in the form of a 286×34×5 3D matrix.

Selection of Research Models

In our exploration of how game behaviors influence ASP traits, the linear regression (LR) model provides a preliminary framework to analyze the linear relationship between participants' behavioral data and their levels of antisocial traits. However, deeper insights into decision-making require more complex machine learning methods.

The random forest (RF) model, which involves constructing multiple decision trees and aggregating predictions, captures nonlinear data relationships more accurately but may not fully reveal the interactions among different options. Artificial neural networks (ANNs), which are capable of simulating brain processing, excel at extracting nonlinear patterns when each game option is treated as an independent variable featuring interactive effects.

Because of the sequential and skip-based nature of the questions considered in this context, participant choices form time-series data, thus rendering recurrent neural networks (RNNs) suitable for capturing dynamic changes. However, standard RNNs face gradient issues involving long sequences. The gated recurrent unit (GRU) and long short-term memory (LSTM) models address this issue via gating mechanisms: GRUs simplify the structure by merging hidden states with gates, whereas LSTM models use 3 gates to manage information flow.

Previous research has relied on statistical scoring rules, but the question structure of the ASP-ECD-G has made it challenging for experts to assign consistent scores to identical options or scenario paths. The complexity of behavioral data renders expert scoring impractical. Thus, this study prioritized statistical and machine learning models (ie, LR, RF, ANN, RNN, GRU, and LSTM) for the analysis of gamified data, thus establishing a balance between computational efficiency and interpretability.

Model Training Process

Model training was performed with the assistance of the Adam optimization algorithm and the dropout technique, with the aim of optimizing performance by adjusting the batch size, number of epochs, and fine-tuning model complexity. Standard hyperparameters were established for different models: in the RF approach, n_estimators and max_depth were tuned to balance performance and prevent overfitting; in ANN, the number of layers or neurons and activation function were optimized with the goal of improving learning; and in RNN, GRU, and LSTM, hidden size, the number of units, and dropout were used to affect learning or memory and generalization, and the learning rate was carefully chosen for the optimizer. Details regarding the parameters are presented in Table 3.

XSL•FO
RenderX

Table 3. Model hyperparameter tuning results.

Model type	Hyperparameters	
	Raw score assignment	1-hot encoding of raw scores
LR[a] model	• None	• None
RF[b] model	• n_estimatorsc=32 • random_state=2024 • max_depthd=4 • min_samples_splite=5 • min_samples_leaff=1 • max_featuresg="log2"	• n_estimators=5 • random_state=2024 • max_depth=6 • min_samples_split=2 • min_samples_leaf=1 • max_features="log2"
ANN[h] model	• Layers: 2 • Neurones: 32 • Activation: "relui" • Dropout: 0.5 • Optimizer: Adam (lrj=0.001) • Loss: "msek" • Epochs: 50 • Batch size: 10	• Layers: 2 • Neurones: 16 • Activation: "relu" • Dropout: 0.5 • Optimizer: Adam (lr=0.001) • Loss: "mse" • Epochs: 100 • Batch size: 10
RNN[l] model	• Layers: 1 (RNN, 32 units) • Activation: "relu" • Dropout: 0.4 • Optimizer: Adam • Loss: "mse" • Epochs: 100 • Batch size: 10	• Layers: 1 (RNN, 32 units) • Activation: "relu" • Dropout: 0.2 • Optimizer: Adam • Loss: "mse" • Epochs: 100 • Batch size: 50
GRU[m] model	• Layers: 1 (GRU, 64 units) • Activation: "relu" • Optimizer: Adam (lrn=0.01) • Loss: "mse" • Epochs: 100 • Batch size: 50	• Layers: 1 (GRU, 32 units) • Activation: "relu" • Optimizer: Adam (lr=0.001) • Loss: "mse" • Epochs: 200 • Batch size: 10
LSTM[o] model	• Layers: 1 (LSTM, 16 units) • Activation: "relu" • Optimizer: Adam • Loss: "mse" • Epochs: 200 • Batch size: 50	• Layers: 1 (LSTM, 64 units) • Activation: "relu" • Optimizer: Adam • Loss: "mse" • Epochs: 200 • Batch size: 10

[a]LR: linear regression.

[b]RF: random forest.

[c]n_estimators: number of estimators.

[d]random_state: random state.

[e]max_depth: maximum depth.

[f]min_samples_split: minimum samples split.

[g]min_samples_leaf: minimum samples per leaf.

[h]max_features: maximum features.

[i]ANN: artificial neural network.

[j]relu: rectified linear unit.

[k]lr: learning rate.

[l]mse: mean squared error.

[m]RNN: recurrent neural network.

[n]GRU: gated recurrent unit.

[o]LSTM: long short-term memory.

XSL•FO
RenderX

Model Evaluation Metrics

The predictive performance of the ASP-ECD-G model was evaluated by reference to common metrics (ie, root mean square error [RMSE], mean absolute error [MAE], and criterion correlation, r) with the goal of assessing its accuracy and correlation with reference results. Lower RMSE or MAE values indicate higher levels of accuracy, whereas an *r* value close to 1 signifies a strong positive correlation, thus validating the predictive ability of the model.

As part of this study, 10 participants who met the same criteria as those used in study 2 were recruited to assess the test-retest reliability of the ASP-ECD-G tool. The gamified assessment was administered to these participants twice, with a 1-month interval between the 2 administration periods. The sample size of 10 is in line with relevant guidelines for reliability testing in small-scale validation studies, which focus on within-individual consistency over time. The test-retest design focused on evaluating the stability of behavioral responses to identical scenarios. In both administrations, all 34 interactive questions and logical jump paths were retained.

Participants

As part of this study, the Credamo platform (Beijing Yishu Mofa Technology Co, Ltd) was used to digitalize the gamified assessment questionnaire. The questionnaire included demographic variables (ie, gender, age, highest level of education, marital status, employment status, and position), as well as the gamified assessment tool and the ASP questionnaire. The sampling criteria used for this study were as follows: (1) participants aged ≥18 years; (2) participants who had at least 1 job or internship experience; and (3) nonpsychology majors.

Participants who did not meet these criteria were excluded via the platform's custom filtering mechanism. The questionnaire was distributed primarily via online social networks such as WeChat groups and Moments. A total of 291 eligible questionnaires were ultimately collected for this study. After further filtering based on lie detection items, 286 valid questionnaires were obtained.

The descriptive statistics of the sample referenced in study 2 are presented in Table 4. Regarding gender distribution, this study included 166 (58%) male participants and 120 (41.9%) female participants out of the 286 participants. With regard to age distribution, the study included 59.4% (170/286) of the participants aged between 23 and 28 years, followed by 26.6% (76/286) aged between 19 and 22 years. For participants' levels of education, the sample predominantly included individuals with undergraduate degrees (234/286, 81.8%), followed by those with master's degrees (36/286, 12.6%). Regarding marital status, 68.9% (197/286) of the participants were single. With respect to employment status, 62.9% (180/286) of the participants were employed at the time of this study, and 24.1% (69/286) were serving as interns. For participants' positions, 62.6% (179/286) of the participants were junior employees, followed by 21% (60/286) who occupied junior management positions.

XSL•FO
RenderX

Table 4. Demographic characteristics of the study 2 participants (n=286).

Variable and category	Sample, n (%)
Sex	
Male	166 (58)
Female	120 (42)
Intersex	0 (0)
Age (y)	
19-22	76 (26.6)
23-28	170 (59.4)
29-35	36 (12.6)
36-45	4 (1.4)
Highest level of education	
High school or vocational school	2 (0.7)
Associate degree	13 (4.5)
Bachelor's degree	234 (81.8)
Master's degree	36 (12.6)
Doctorate	1 (0.3)
Marital status	
Single	197 (68.9)
Engaged	11 (3.8)
Married, no children	39 (13.6)
Married, with children	30 (10.5)
In a relationship	9 (3.1)
Employment status	
Employed	180 (62.9)
Interning	69 (24.1)
Formerly employed	13 (4.5)
Never employed	21 (7.3)
Other	3 (1)
Position	
Junior employee	179 (62.6)
Junior management	60 (21)
Midlevel management	30 (10.5)
Senior management	12 (4.2)
Other	5 (1.7)
Total	286 (100)

Analysis of Assessment Properties

Study Design

We built on study 2 by incorporating items used to measure pleasure, interest, positive emotions, negative emotions, and immersive experience into the gamified assessment tasks and the ASP questionnaire. These items are scored on a scale from 1 to 9, in which context higher scores indicate stronger experiences. The specific items are detailed in Multimedia Appendix 3.

We used a 2×2 mixed experimental design, in which the assessment format (ie, gamified assessment vs questionnaire assessment) was used as the within-subject variable, and participant incentive (ie, with vs without) was used as the between-subject variable. This study aimed to investigate the impacts of participant incentives on individual performance

XSL•FO
RenderX

across different assessment formats via gamified motivational mechanisms.

After the participants completed the questionnaire, they were asked the following question: "If you were to participate in a company's new employee psychological test, which method would you prefer?" This question aimed to gauge their future willingness to use gamified assessment versus traditional questionnaire assessment.

Participant incentives were introduced as an external motivator to stimulate achievement motivation, in line with previous research that has linked ASP traits to instrumental rationality, that is, the prioritization of strategic self-interest over social norms [26,27]. Individuals who exhibit high levels of manipulativeness or deceitfulness often engage in utility-maximizing behaviors, thereby justifying their self-serving actions as rational responses to their perceived circumstances [26]. Individuals who exhibit stronger tendencies toward an ASP rationalize their unethical behavior as a necessary form of self-preservation, reframing it as a logical choice rather than an antisocial tendency [27].

Rather than directly incentivizing "antisocial tendencies," which carry a strong social stigma, this study framed the incentive based on "rationality," a positively valenced construct that aligns with participants' self-perceived strategic competence. While external incentives can enhance individuals' task focus in challenging situations, they may also lead to deceptive self-presentation [28]. The design of this study was based on the hypothesis that this framing would enhance social desirability effects in questionnaires, in which context participants could adjust their responses consciously. Moreover, participants' gamified assessment scores, which were based on the behavioral choices made in the context of immersive scenarios, were expected to remain unaffected.

The experimental condition involved an increase in the participation payment (from 9-15 RMB [US $1.25-$2.1]) in exchange for "rational results," without defining specific criteria; this approach aimed to mimic the context of real-world strategic self-presentation (eg, job applicants tailoring answers with the goal of appearing competent). Participants in the control group completed the gamified and questionnaire assessments in the usual manner, whereas those in the experimental group were presented with an incentive: "This study hopes that you can be rational. If your results show that you are rational, we will increase your participation payment (from 9 to 15 yuan)."

Participants

The participants were selected based on criteria that were consistent with those used in study 2, and the questionnaires were distributed via the Credamo platform. We used GPower (version 3.1) software to calculate the required sample size to ensure sufficient statistical power for the repeated-measures ANOVA conducted as part of this 2×2 mixed experimental design. Regarding the mixed design featuring 2 between-subject and 2 within-subject assessments, we tested the main effects for both factor types. The analysis, which was based on a medium effect size (f=0.25), α=0.05, power=0.8, and a within-subject correlation assumption of 0.5, indicated minimum sample sizes of 98 for between-subject effects, 34 for within-subject effects, and 68 for interaction effects.

We collected 200 questionnaires across 4 groups with the goal of maximizing participation. A total of 148 valid responses remained after a screening process involving lie detection items. Each group (ie, the group with incentives and the group without incentives) included 74 participants.

Descriptive statistics concerning the sample are presented in Table 5. Of the 148 participants, 94 (63.5%) were female and 54 (36.5%) were male. The participants were predominantly aged between 29 and 35 years (54/148, 36.5%), followed by those aged between 23 and 28 years (45/148, 30.4%). Most participants had obtained bachelor's degrees (98/148, 66.2%), followed by those who had obtained associate degrees (19/148, 12.8%). The majority of participants (95/148, 64.2%) were married with children. Regarding participants' employment status, of the 148 participants, 143 (96.6%) were employed, whereas 5 (3.4%) were serving as interns. The participants' job positions spanned from entry level to senior management, although entry-level employees represented the largest group (59/148, 39.9%).

XSL•FO
RenderX

Table 5. Demographic characteristics of the participants in study 3 (n=148).

Variable and category	Participants, n (%)
Sex	
Male	54 (36.5)
Female	94 (63.5)
Intersex	0 (0)
Age range (y)	
19-22	11 (7.4)
23-28	45 (30.4)
29-35	54 (36.5)
36-45	27 (18.2)
46-50	11 (7.4)
Education	
Junior high school	3 (2)
High school or vocational	7 (4.7)
Associate degree	19 (12.8)
Bachelor's degree	98 (66.2)
Master's degree	21 (14.2)
Marital status	
Single	40 (27)
Engaged	1 (0.7)
Married, no children	9 (6.1)
Married with children	95 (64.2)
In a relationship	3 (2)
Employment	
Employed	143 (96.6)
Intern	5 (3.4)
Position	
Entry-level	59 (39.9)
Junior management	36 (24.3)
Middle management	42 (28.4)
Senior management	11 (7.4)

Ethical Considerations

This study was approved by the ethics committee of Beijing Normal University (approval BNU202503310097), thus ensuring compliance with ethical guidelines. All the participants provided written informed consent. The informed consent form detailed the purpose of the study, the procedures used in this research, and the participants' right to withdraw from the study without penalty; furthermore, it noted that the use of data was limited to the purposes of the study and that the data collected as part of this research would be anonymized for analysis. The participants' privacy was protected via the deidentification of personal data, the secure storage of such data on password-protected computers, and the aggregated reporting of study results. All the participants received a basic participation payment of RMB 9, and those in the experimental group in study 3 received an additional RMB 6 as a reward for "rational outcomes"; these payments were provided via a digital platform and were not linked to trait scores to prevent bias. No identifiable information or images concerning participants are included in the manuscript or the multimedia appendices; thus, no additional consent for personal identification was needed. These measures are consistent with relevant ethical standards for informed consent, confidentiality, and participant well-being as outlined in the institutional and international research guidelines.

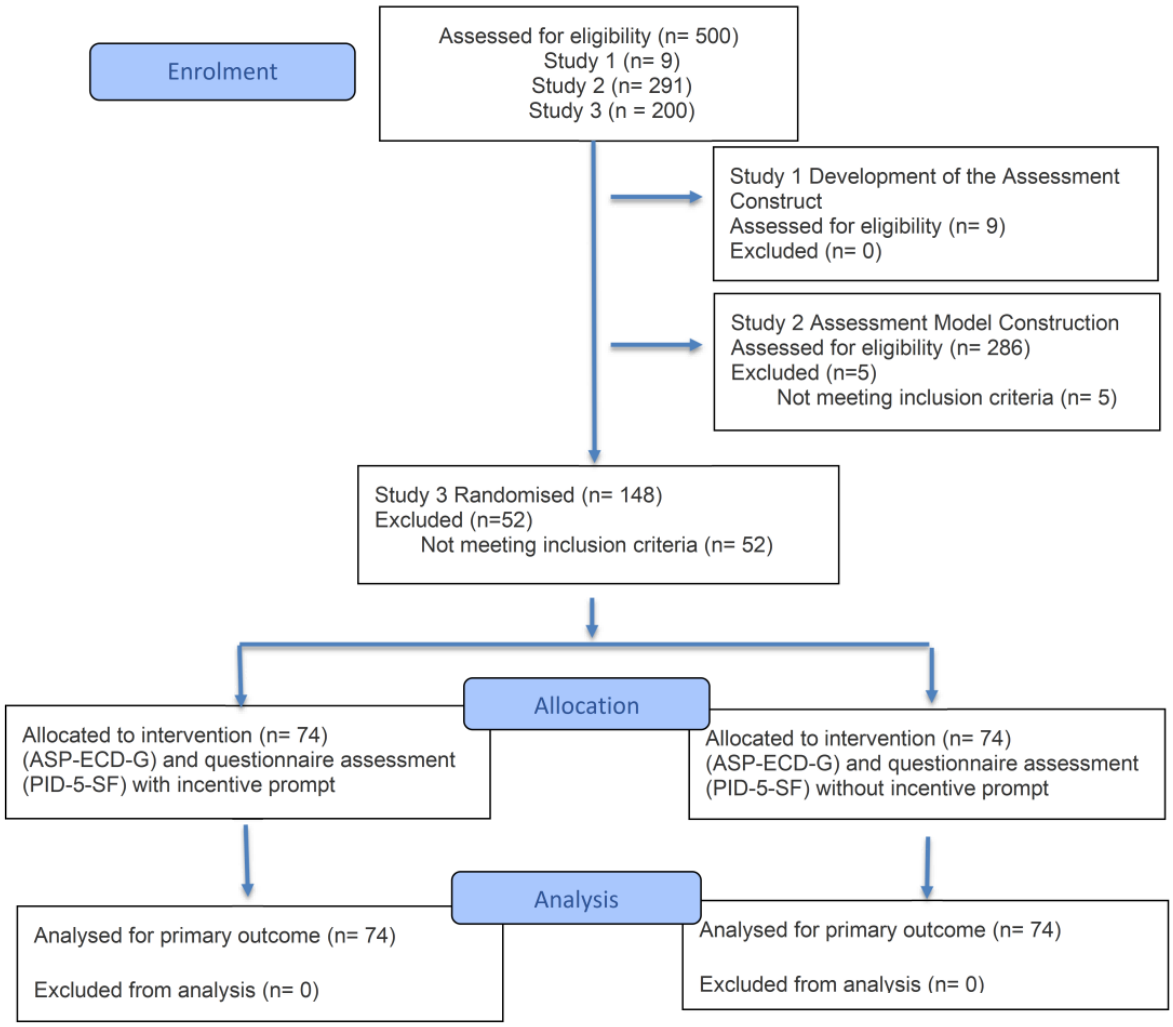
XSL•FO
RenderX

Results

Overview

Study 1 involved semistructured interviews with 9 professionals (1 senior manager, 2 midlevel managers, and 6 frontline employees) with >3 years of experience; each interview took 30-60 minutes. For study 2, the Credamo platform was used to distribute digital questionnaires (demographics, gamified assessment, and ASP) to participants (aged ≥18 years, with work or internship experience, nonpsychology majors); 291 responses were collected, 286 of which were valid after lie detection. Recruitment for study 3 also used Credamo (with the same criteria as study 2); we collected 200 questionnaires, 148 of which were valid (n=74 per group with or without incentives after lie detection). The design process is shown in Figure 3.

Figure 3. Research design flow diagram of three studies.



Development of the Assessment Construct

The ASP-ECD-G is presented as a text-based game provided on a questionnaire platform. The storyline of the ASP-ECD-G simulates a workplace scenario in which players play the role of an employee who has just completed a 1-month probation period and is facing the challenge of securing a permanent position. Players must solve various workplace problems, including conflicts with colleagues, crisis management, ethical dilemmas, and team performance management. Each choice made by the player influences the development of the plot; however, the ultimate outcome involves the player receiving a dismissal notice from human resources.

The design of the ASP-ECD-G incorporates 3 game elements: narrative, immersion, and feedback. The narrative is integrated into the assessment content and is presented in the form of a workplace story in which participants are required to immerse themselves to answer questions. For each question, the game characters provide preset feedback based on the player's choices, thereby driving the plot forward. Players must select the most suitable options from the choices available within the scenario, and they do not have the ability to save their progress or exit the game at a midway point. Incorrect selections can be rectified by returning to the previous page to reanswer the questions. The presentation of the ASP-ECD-G is illustrated in Multimedia Appendix 4.

The ASP-ECD-G, similar to SJTs, is rooted in real-world scenarios and combines questions with options for assessment. However, in comparison with traditional scales and SJTs, the gamified assessment tool developed as part of this study exhibits 3 distinctive features. First, the scenarios included in the ASP-ECD-G follow a narrative sequence, and the options are

XSL•FO
RenderX

characterized by logical transitions, thus violating the independence assumption among the items. Second, because of the logical transitions between questions, different participants may experience varying numbers of scenes during the course of the gamified assessment. Third, each question included in the ASP-ECD-G reflects 1 or more behavioral traits, thereby violating the 1-dimensionality assumption.

Because of these differences in data types, item interdependencies, and dimensional assumptions between the multidimensional nature of the ASP-ECD-G and traditional assessment methods, this study details the findings obtained via multidimensional validation analyses—in which item response theory is used alongside the nominal response model—in Multimedia Appendix 5 for the benefit of readers seeking in-depth insights. This study highlights the fact that these features make the assessment process of the ASP-ECD-G resemble a cohesive story rather than a collection of isolated questions. The ASP-ECD-G aims to replicate real workplace scenarios as closely as possible, and the logical transitions between different options enhance the coherence and immersion of the presented scenario for participants during the assessment process.

Assessment Model Construction

Results of the Expert-Assigned Coding

The expert panel often found it difficult to reach a consensus during the process of scoring game options, thus leading to instability in the scoring results. The task of explaining the sources of differences in scores among participants was challenging, even if scores could be assigned to all paths. The core issue pertained to the task of determining whether different participants who chose the same option in response to the same question should be assigned the same score. Further consideration revealed that treating participants' responses as a path made it difficult to account for the complexity and extensive information involved in this process. This situation led to instability in the scoring results, as the experts faced similar challenges in the process of scoring both question options and different paths within the same scenario.

After a fixed random seed was determined, the performances of different models on the same dataset were subjected to a comprehensive comparison. The evaluation included the RMSE and MAE for the training and testing sets, as well as the correlation (r) between the predicted and reference results on the testing set; this evaluation aimed to assess the performance of each model.

As indicated in Table 6, none of the models exhibited significant overfitting or underfitting following parameter tuning. Among the testing set results, the GRU model exhibited the best performance, as indicated by RMSE and MAE values of 0.380 and 0.313, respectively, and a correlation of 0.676, thus indicating a high level of consistency between the predicted and observed results.

The study further evaluated the performance of each model across different ASP score ranges in the testing set to assess the stability of the predictions. These results are presented in Table 7.

Table 6. Expert-assigned coding model performance for the prediction of antisocial personality scores (n=286; training–test split, 7:3).

Dataset and evaluation metric	LR[a] model	RF[b] model	ANN[c] model	RNN[d] model	GRU[e] model	LSTM[f] model
Training set						
RMSE[g]	0.415	0.432	0.438	0.436	0.363	0.447
MAE[h]	0.332	0.357	0.343	0.353	0.288	0.356
r[i]	0.637	0.694	0.601	0.598	0.762	0.559
T esting set						
RMSE	0.453	0.418	0.462	0.436	0.380	0.433
MAE	0.356	0.336	0.362	0.340	0.313	0.344
r	0.510	0.626	0.529	0.523	0.676	0.567

[a]LR: linear regression.

[b]RF: random forest.

[c]ANN: artificial neural network.

[d]RNN: recurrent neural network.

[e]GRU: gated recurrent unit.

[f]LSTM: long short-term memory.

[g]RMSE: root mean square error.

[h]MAE: mean absolute error.

[i]r represents the correlation between the predicted results and the reference results for the best-performing model.

XSL•FO
RenderX

Table 7. RMSE[a] values of expert-assigned coding models across different ASP[b] score ranges (test set, N=59)[c].

ASP score range	Sample size	LR[d] model	RF[e] model	ANN[f] model	RNN[g] model	GRU[h] model	LSTM[i] model
1-1.5	1	0.594	0.757	0.314	0.760	0.483	0.826
1.5-2	13	0.243	0.325	0.297	0.350	0.318	0.252
2-2.5	23	0.360	0.243	0.387	0.278	0.356	0.281
2.5-3	15	0.451	0.402	0.470	0.353	0.377	0.466
3-3.5	6	0.686	0.682	0.478	0.727	0.465	0.627
3.5-4	1	1.537	1.396	1.380	1.518	.854	1.492

[a]RMSE: root mean square error.

[b]ASP: antisocial personality.

[c]No samples featuring ASP scores of ≥4 were included in the testing set.

[d]LR: linear regression.

[e]RF: random forest.

[f]ANN: artificial neural network.

[g]RNN: recurrent neural network.

[h]GRU: gated recurrent unit.

[i]LSTM: long short-term memory.

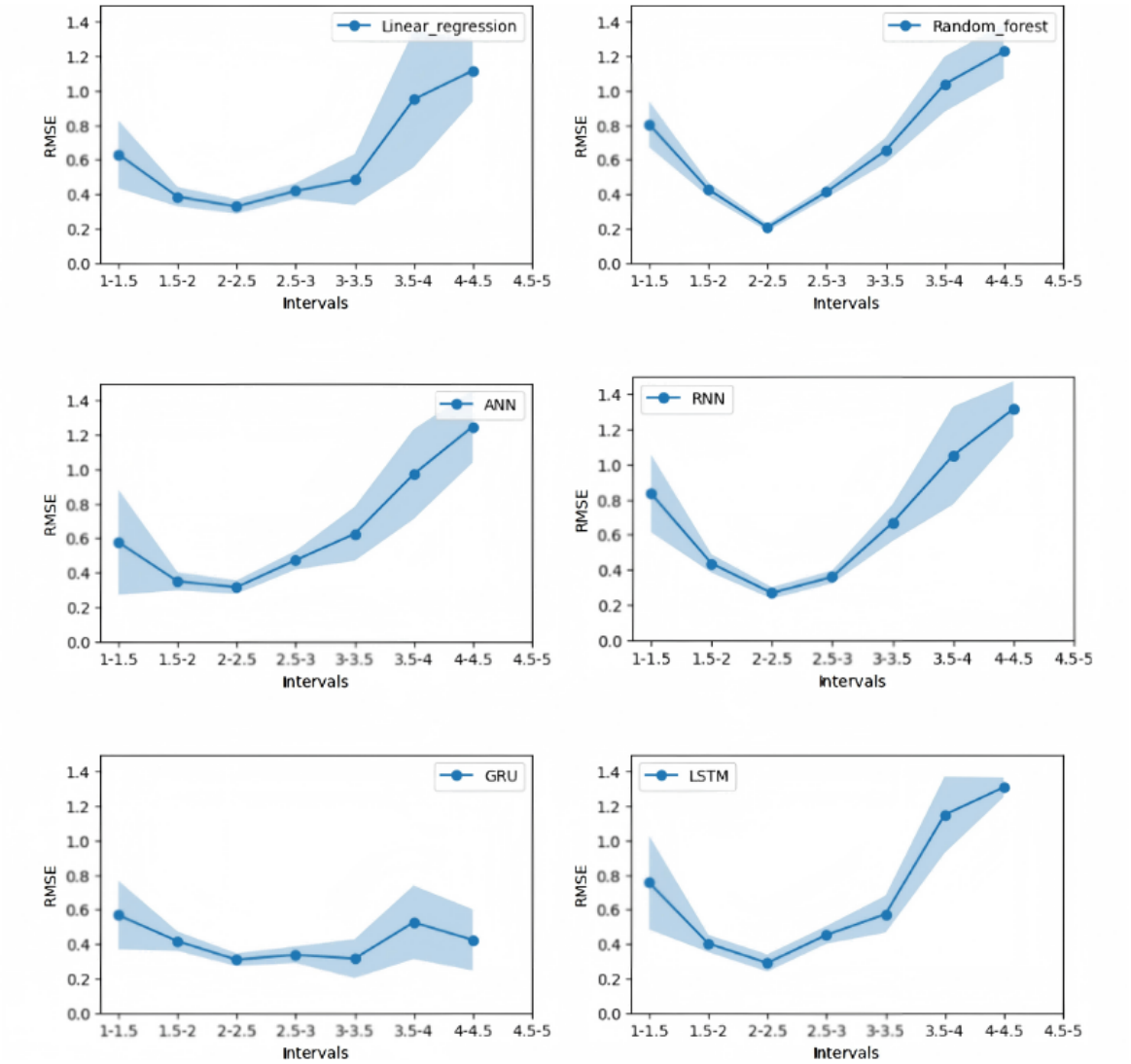
An analysis of model performance across the range of different ASP scores revealed that the LR model struggled to address complex relationships, as indicated by its high RMSE in mid-to-high ranges (eg, 2.5-3: 0.451; 3-3.5: 0.686), although it exhibited a moderate level of performance in the 1.5 to 2 range (0.243). The RF model excelled in the 1.5 to 2.5 range (RMSE 0.243-0.325) but exhibited instability in cases involving extreme scores (3.5-4: 1.396). The ANN model maintained a low RMSE in lower ranges (1-2.5:≤0.387) but showed increased errors in higher ranges (2.5-3: 0.470; 3-3.5: 0.478). The RNN model exhibited midrange consistency (1.5-3: 0.278-0.353) but was characterized by the highest RMSE in the 3.5 to 4 range (1.518). The GRU model exhibited notably balanced performance: the lowest RMSE was observed in the 1.5 to 2 range (0.318), moderate values were observed in the middle range (2-2.5: 0.356; 2.5-3: 0.377), and relatively controlled errors were observed at the extremes (3.5-4: 0.854). While the LSTM model matched the GRU in the 1.5 to 2 range (0.252 vs 0.318), it exhibited the worst performance in high ranges (3.5-4: 1.492), thus highlighting its inferior handling of sparse data. Given its balanced performance and resilience to small samples, the GRU model was optimal for the prediction of ASP scores.

This study used 100 different random seeds to facilitate the random sampling of the test set with the goal of exploring the robustness of each model in greater depth. In these 100 experiments, changes in the RMSE values were observed across different ASP score ranges. The results are presented in Figure 4. The line shown in this figure indicates the average value across 100 iterations, and the light blue area represents the average plus or minus 1 SD. These points also apply to Tables 4 and 5.

These results indicate that, even after 100 iterations, the GRU model consistently outperformed the other models and exhibited higher levels of stability.

XSL•FO
RenderX

Figure 4. Root mean square error (RMSE) variations across 100 random testing set selections for different models (expert-assigned coding). RMSE means are shown with SD bands. ANN: artificial neural network; GRU: gated recurrent unit; LSTM: long short-term memory; RNN: recurrent neural network.



Results of 1-Hot Encoding

We used the same fixed random seed to perform a comprehensive comparison of the predictive performance of different models on the same dataset.

As indicated in Table 8, no significant overfitting or underfitting was observed after the parameters were tuned for all the models. We further evaluated the performance stability of each model across different ASP score ranges in the test set, as detailed in Table 9.

XSL•FO
RenderX

Table 8. Model performance for antisocial personality score prediction using 1-hot encoding (n=286, 34×5–dimensional encoding).

Dataset and metric	LR[a] model	RF[b] model	ANN[c] model	RNN[d] model	GRU[e] model	LSTM[f] model
Training set						
RMSE[g]	0.336	0.384	0.358	0.387	0.303	0.359
MAE[h]	0.269	0.301	0.285	0.299	0.230	0.270
r[i]	0.782	0.744	0.770	0.700	0.789	0.748
Test set						
RMSE	0.638	0.401	0.406	0.420	0.322	0.328
MAE	0.426	0.317	0.323	0.322	0.251	0.233
r	0.492	0.619	0.677	0.581	0.808	0.760

[a]LR: linear regression.

[b]RF: random forest.

[c]ANN: artificial neural network.

[d]RNN: recurrent neural network.

[e]GRU: gated recurrent unit.

[f]LSTM: long short-term memory.

[g]RMSE: root mean square error.

[h]MAE: mean absolute error.

[i]r represents the correlation between the predicted results and the reference results for the best model.

Table 9. RMSE[a] values of 1-hot encoding models across different ASP[b] score ranges (test set, N=59, 5-dimensional vectorization)[c].

ASP score range	Sample size	LR[d] model	RF[e] model	ANN[f] model	RNN[g] model	GRU[h] model	LSTM[i] model
1-1.5	1	0.266	0.825	0.358	0.586	0.452	0.726
1.5-2	13	0.269	0.338	0.235	0.234	0.229	0.428
2-2.5	23	0.833	0.252	0.438	0.350	0.270	0.243
2.5-3	15	0.612	0.381	0.442	0.331	0.400	0.365
3-3.5	6	0.512	0.641	0.445	0.746	0.288	0.503
3.5-4	1	0.571	1.149	0.985	1.278	0.467	1.163

[a]RMSE: root mean square error.

[b]ASP: antisocial personality.

[c]No samples featuring ASP scores of ≥4 were included in the testing set.

[d]LR: linear regression.

[e]RF: random forest.

[f]ANN: artificial neural network.

[g]RNN: recurrent neural network.

[h]GRU: gated recurrent unit.

[i]LSTM: long short-term memory.

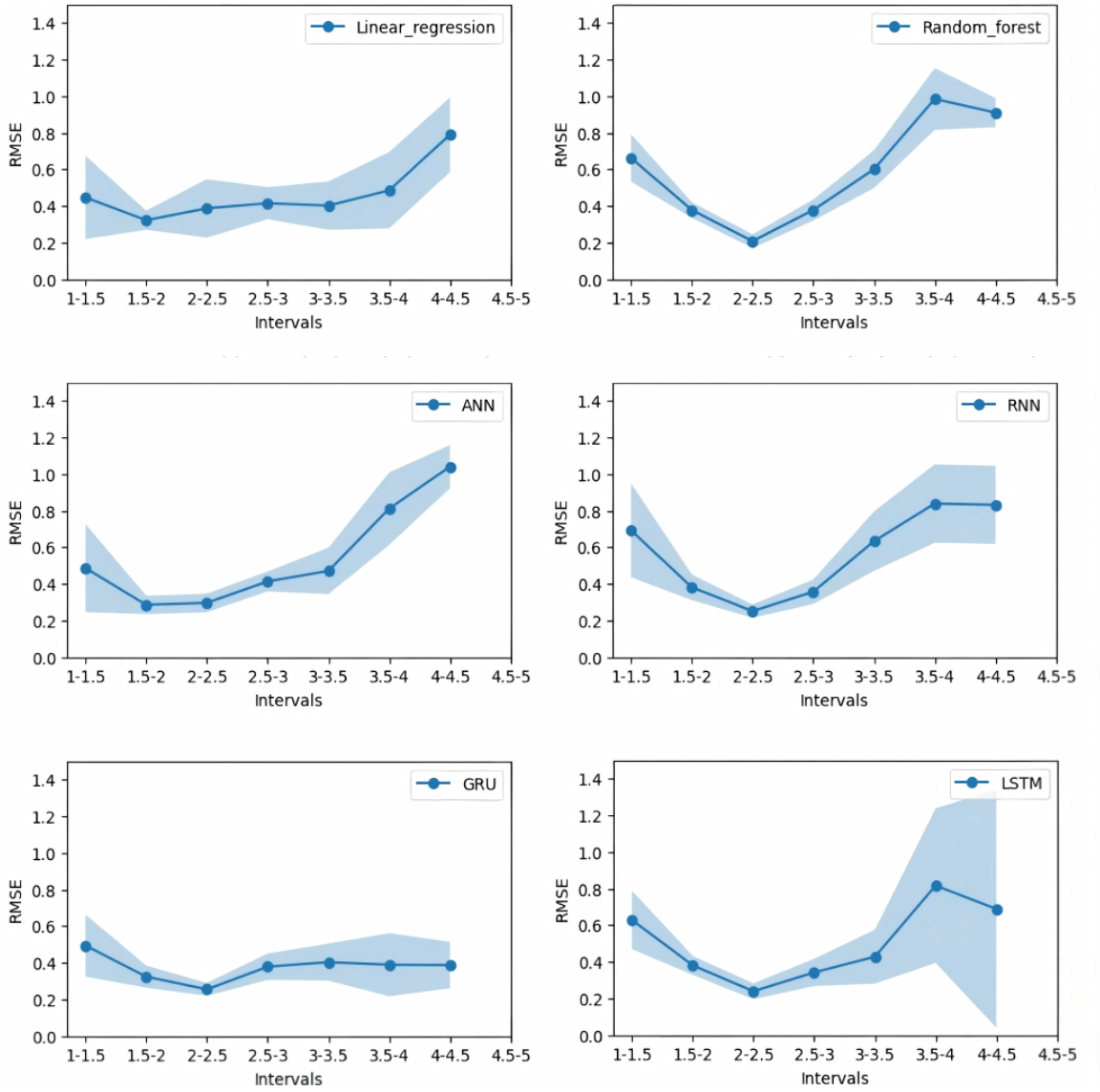
An analysis of the model performance across the range of ASP scores revealed that although the LR model exhibited moderately good performance in the lower range (1-2, RMSE 0.266-0.269), its high RMSE values across the 2 to 3 range indicated that the model struggled in cases involving complex patterns. The RF model exhibited consistent performance in the middle range (1.5-3, RMSE 0.252-0.381) but was characterized by unstable extremes (3.5-4, RMSE 1.149). The ANN model maintained a relatively low RMSE (≤0.445) in the range of 1 to 3.5, suggesting consistent performance; however, in the range of 3.5 to 4, it exhibited an RMSE of 0.985, highlighting the limitations of these rare cases of high scores. The RNN model demonstrated strong performance in the intermediate range (1.5-3, RMSE 0.234-0.350); however, excessive error was observed in the extreme range in this context (3.5-4, RMSE 1.278). The GRU model excelled, exhibiting the lowest RMSE in the 1.5 to 2 range (0.229) alongside consistent performance across most ranges (≤0.467). In contrast, the LSTM model exhibited poor performance in the high range (3.5-4, RMSE 1.163), and its performance in the middle range was inferior to that of the GRU (eg, 2-2.5, RMSE 0.243 compared with 0.288 for the GRU), despite comparable results in the low range (1.5-2, RMSE 0.428 compared with 0.224 for the GRU).

XSL•FO
RenderX

Given its balanced performance and resilience in small samples, the GRU model was optimal for predicting ASP scores.

This study used 100 different random seeds to facilitate random sampling of the test set, with the goal of examining the robustness of each model in greater detail. Across the 100 experiments, the study revealed changes in the RMSE values across various ASP score ranges. The results are presented in Figure 5.

Figure 5. Root mean square error (RMSE) variations across 100 random testing set selections for different models (1-hot encoding). RMSE means are shown with SD bands. ANN: artificial neural network; GRU: gated recurrent unit; LSTM: long short-term memory; RNN: recurrent neural network.



Best Model

Notably, the tuning results varied depending on the data processing method used. When expert scoring was used, models often exhibited underfitting, whereas models that used 1-hot encoding frequently exhibited overfitting. However, both approaches can ultimately be adjusted to achieve satisfactory results through hyperparameter tuning.

A possible reason for this finding is that expert scoring relies on the team's understanding and experience to assign values to each option. The expert scoring method, which involves limited feature dimensions (34×2), may not fully capture all behavioral patterns and dynamic changes, thereby reducing the model's ability to capture comprehensive features and leading to underfitting. In contrast, 1-hot encoding increases the feature dimensions (34×5) and theoretically retains all the information contained in the original data. The higher-dimensional feature representation can capture more details but may also be impacted by noise and outliers, thus leading to overfitting.

Overall, the GRU model not only excelled in predictive robustness but also showed the strongest correlations between the predicted results and the actual observed values. Therefore, after all the indicators were balanced, we decided to use the GRU model as a predictive tool in the subsequent phases of this research.

Overall, the assessment model trained using the GRU model via 1-hot encoding data processing was selected as the final model because its correlation with the ASP questionnaire scores was higher than that of the scores obtained via the expert encoding process (r=0.792>0.742). Detailed results are referred to Table 10.

XSL•FO
RenderX

In addition, as part of this study, models were trained using each of the 7 behavioral traits of ASP as individual prediction labels. The results revealed that models trained on the total score exhibited performance comparable to that of models trained on individual traits, although the former achieved higher levels of precision. Specific details concerning this analysis are provided in Multimedia Appendix 4 for readers seeking further technical insights.

This study assessed the retest reliability of the ASP-ECD-G instrument using the intraclass correlation coefficients (ICCs). The results revealed a moderate to high level of reliability, as indicated by an ICC of 0.776 (95% CI 0.097-0.944, *df*=9; *P*=.02); these findings revealed consistent behavioral responses over a 1-month interval. This statistic exceeded the conventional threshold for acceptable reliability (ICC≥0.70), thus suggesting that measures of antisocial traits stabilized over time. The significant *P* value (*P*=.02<.05) confirmed that the observed retesting consistency was not the result of chance; these results thus indicated the instrument's suitability for longitudinal assessment. Although the sample size was small, the robust ICC provides initial evidence to support the reliability of the gamified assessment; however, validation using a larger longitudinal sample is necessary.

Table 10. Descriptive statistics and correlation results of model predictions and questionnaire assessments.

Variables	Values, mean (SD)	Values, range	ASP[a] questionnaire score	*P* value
ASP questionnaire score	2.303 (0.532)	4.357-1.107	—[b]	—
Expert scoring GRU[c] model	2.383 (0.430)	4.172-1.578	0.742	<.001
1-hot encoding GRU model	2.317 (0.476)	4.264-1.479	0.792	<.001

[a]ASP: antisocial personality.

[b]Not applicable.

[c]GRU: gated recurrent unit.

Analysis of Assessment Properties

Comparison of the Experience of Completing Gamified Assessments and Questionnaire Assessments

The results of the repeated-measures ANOVA are presented in Table 11. Notably, significant differences were evident in the experience dimensions between the gamified assessment results and the questionnaire assessment results (F_{147}=25.522, 47.940, 5.581, 151.109, 14.259; *P*<.05).

A comparison concerning the pleasure dimension revealed that the pleasure associated with the gamified assessment was lower than that associated with the questionnaire assessment (F_{147}=25.522; *P*<.001). This finding indicates that while participants might have experienced some degree of enjoyment during the gamified assessment, this approach may not have fully evoked the level of pleasure expected, possibly because of the conflict, contradictions, and negative features embedded in the gamified assessment scenarios, including the corresponding choices and outcomes (ie, receiving a "dismissal notice" from human resources). This misalignment with participants' expectations could result in an unpleasant experience. However, this unpleasant experience suggests higher levels of engagement on the part of the participants, who were more likely to be influenced by the storyline, thus leading to higher levels of the immersion experience.

XSL•FO
RenderX

Table 11. Comparison of experience ratings between the gamified assessment and the questionnaire assessment (n=148, 2×2 mixed experimental design)[a].

Variable and assessment type	Values, mean (SD)	Mean difference	*F* test (*df*)	*P* value	Partial η^2
Pleasure		–0.750	25.522 (147)	<.001	0.153
ASP-ECD-G[b]	5.432 (2.158)				
ASP[c] questionnaire	6.182 (1.686)				
Interest		0.939	47.940 (147)	<.001	.254
ASP-ECD-G	7.095 (1.643)				
ASP questionnaire	6.155 (1.979)				
Positive emotions		0.392	5.581 (147)	.02	0.038
ASP-ECD-G	6.507 (2.045)				
ASP questionnaire	6.115 (2.062)				
Negative emotions		1.892	151.109 (147)	<.001	0.517
ASP-ECD-G	4.365 (2.330)				
ASP questionnaire	2.473 (1.809)				
Immersion		0.412	14.259 (147)	<.001	0.092
ASP-ECD-G	7.628 (1.225)				
ASP questionnaire	7.216 (1.720)				

[a]The gender, age, highest level of education, marital status, employment status, and position of the participants were included as control variables.

[b]ASP-ECD-G: Antisocial Personality Traits Evidence-Centered Design Gamified assessment tool.

[c]ASP: antisocial personality.

The interest generated by the gamified assessment was greater than that generated by the questionnaire assessment (F_{147}=47.940; *P*<.001), thus indicating that the gamified assessment was relatively successful in providing an interesting and engaging experience. The introduction of gamified elements likely increased the interactivity and the rates of participation in the tasks, thereby increasing the participants' interest and engagement.

The comparison of positive emotions also indicated that the gamified assessment evoked more positive emotions (F_{147}=5.581; *P*=.02). This finding might be because of the reward mechanisms associated with the gamified assessment, which enhanced the participants' sense of achievement and self-efficacy. This positive emotional experience may encourage the participants to engage honestly in the experiment to receive more positive feedback.

The results regarding negative emotions revealed that the scores pertaining to the gamified assessment significantly exceeded those pertaining to the questionnaire assessment (F_{147}=151.109; *P*<.001). This finding is in line with the results concerning pleasure reported previously. The fact that the scores for both positive and negative emotions were higher in the gamified assessment suggests that the overall emotional state of the participants was more intense during the gamified assessment process, indicating higher levels of engagement that can, in turn, enhance immersion and authenticity.

The comparison of immersion experience revealed that the gamified assessment obtained a significantly higher value than the questionnaire assessment (F_{147}=14.259; *P*<.001), thus indicating that the gamified assessment can capture the participants' attention more effectively, thereby making them more focused on the task at hand. A cross-analysis of the gamified and questionnaire assessment scores is presented in Table 12. This study revealed that, among the participants, both assessment methods resulted in a high proportion of the participants who reported high satisfaction scores (ie, simultaneously choosing scores of 7, 8, or 9); furthermore, the immersion experience scores associated with the gamified assessment were consistently ≥4.

XSL•FO
RenderX

Table 12. Cross-tabulation of immersion experience scores between the gamified and questionnaire assessments[a].

Immersion experience score	ASP[b] questionnaire									Total (n=148)
	1 (n=2)	2 (n=1)	3 (n=3)	4 (n=2)	5 (n=16)	6 (n=21)	7 (n=20)	8 (n=46)	9 (n=37)	
ASP-ECD-G										
4	0	0	0	0	0	2	0	0	0	2
5	0	0	0	0	5	2	0	0	0	7
6	1	1	2	1	4	5	3	1	0	18
7	0	0	0	0	5	9	12	5	1	32
8	1	0	1	0	2	1	4	28	10	47
9	0	0	0	1	0	2	1	12	26	42

[a]No participants obtained immersion experience scores of 1 to 3 in the gamified assessment.

[b]ASP: antisocial personality.

[c]ASP-ECD-G: Antisocial Personality Traits Evidence-Centered Design Gamified assessment tool.

Subsequently, 2 participants who rated the immersion experience associated with the gamified assessment higher than that associated with the questionnaire assessment were randomly selected for follow-up interviews. These participants reported that the emphasis of the gamified assessment on interactivity and engagement enabled them to become more deeply absorbed in the experience. In contrast, the questionnaire assessment tended to represent participants' retrospective evaluation of the overall experience, which could entail a slight disconnection from their current state. This feedback further indicates that gamified assessment tools can not only engage participants effectively but also sustain higher levels of immersion over longer periods than is possible through the use of questionnaires.

Examination of Resistance to Faking in the Gamified Assessment Questionnaires

The descriptive statistics pertaining to different game formats according to the presence or absence of participant payment incentives are presented in Table 13.

The results of the ANOVA, which are presented in Table 14, indicate significant differences in questionnaire scores between situations in which participant payment incentives were present and those in which they were absent (F_1=5.740; P=.02). The participants who received payment incentives obtained significantly lower ASP scores than those who did not receive payment incentives. However, no significant differences were observed in the gamified assessment results (F_1=0.138; P=.71).

Two possible explanations may account for this result. First, gamification itself can enhance the intrinsic appeal of the task, thus causing the participants to focus on the task itself rather than on potential external rewards. Second, gamification simulates a real game scenario, in which context the immersive experience that occurs during the response process increases the cost and difficulty of faking, thus making it challenging for the participants to manipulate their responses in line with their subjective goals. Regardless, the high resistance to faking that characterizes gamified assessment plays a crucial role in efforts to reflect the participants' personal traits more authentically, especially in recruitment and selection contexts.

Table 13. Descriptive statistics for different game formats according to the presence or absence of participant payment incentives.

Assessment type and participant payment incentive	Values, mean (SD)	Sample size, n
ASP-ECD-G[a]		
No	2.185 (0.430)	74
Yes	2.156 (0.382)	74
ASP[b] questionnaire		
No	2.201 (0.420)	74
Yes	2.066 (0.373)	74

[a]ASP-ECD-G: Antisocial Personality Traits Evidence-Centered Design Gamified assessment tool.

[b]ASP: antisocial personality.

XSL•FO
RenderX

Table 14. ANOVA results concerning resistance to response manipulation (n=148)[a].

Source and dependent variable	Type III sum of squares	Mean square	*F* test (*df*)	*P* value	Partial η^2
Corrected model					
ASP[b] questionnaire	2.073	0.296	1.915 (7)	.07	0.087
ASP-ECD-G[c]	2.529	0.361	2.333 (7)	.03	0.104
Intercept					
ASP questionnaire	4.090	4.090	26.455 (1)	<.001	0.159
ASP-ECD-G	5.424	5.424	35.023 (1)	<.001	0.200
Gender					
ASP questionnaire	0.450	0.450	2.912 (1)	.09	0.020
ASP-ECD-G	0.175	0.175	1.128 (1)	.29	0.008
Age (y)					
ASP questionnaire	0.006	0.006	0.039 (1)	.84	0.000
ASP-ECD-G	0.021	0.021	0.136 (1)	.71	0.001
Highest education					
ASP questionnaire	0.150	0.150	0.971 (1)	.33	0.007
ASP-ECD-G	0.182	0.182	1.177 (1)	.28	0.008
Marital status					
ASP questionnaire	0.007	0.007	0.048 (1)	.83	0
ASP-ECD-G	1.087	1.087	7.018 (1)	.01	0.048
Employment status					
ASP questionnaire	0.022	0.022	0.142 (1)	.71	0.001
ASP-ECD-G	0.002	0.002	0.015 (1)	.90	0
Position					
ASP questionnaire	0.642	0.642	4.155 (1)	.04	0.029
ASP-ECD-G	0.337	0.337	2.176 (1)	.14	0.015
Participant payment incentive					
ASP questionnaire	0.887	0.887	5.740 (1)	.02	0.039
ASP-ECD-G	0.021	0.021	0.138 (1)	.71	0.001
Error					
ASP questionnaire	21.644 (140)	0.155	__[d]	—	—
ASP-ECD-G	21.680 (140)	0.155	—	—	—
Total					
ASP questionnaire	697.505 (148)	—	—	—	—
ASP-ECD-G	721.585 (148)	—	—	—	—
Corrected total					
ASP questionnaire	23.717 (147)	—	—	—	—
ASP-ECD-G	24.209 (147)	—	—	—	—

[a]Antisocial personality questionnaire: R^2=0.087 (adjusted R^2=0.042); Antisocial Personality Traits Evidence-Centered Design Gamified assessment tool: R^2=0.104 (adjusted R^2=0.060).

[b]ASP: antisocial personality.

[c]ASP-ECD-G: Antisocial Personality Traits Evidence-Centered Design Gamified assessment tool.

[d]Not applicable.

XSL•FO
RenderX

Analysis of Future Assessment Preferences

Despite the higher levels of negative emotions induced by the negative outcomes in the gamified assessment tool, at the end of the survey, approximately 79.1% of the participants indicated that they would prefer to participate in a gamified scenario assessment over a questionnaire assessment in future employee psychological assessments. This finding suggests that gamified assessment tools are inherently more appealing to participants and that they might provide them with more emotional experiences. Such enhanced emotional experience is likely the result of the content of the gamified assessment tool itself.

Discussion

Principal Findings

On the basis of the *DSM-5* criteria, 7 core antisocial traits were operationalized into 24 observable workplace behaviors based on semistructured interviews conducted with 9 industry professionals (Table 1). These behaviors were integrated into a text-based gamified assessment that included 10 subscenarios and 34 interactive questions with a logic rooted in logical jumps (eg, question jumps and progressive paths), thereby ensuring alignment between clinical constructs and simulated workplace decisions (Multimedia Appendix 1). Content validity was confirmed based on an expert review; in this context, 100% correspondence was observed between game behaviors and *DSM-5* traits.

Six machine learning models were compared using PID-5-SF scores as criteria (n=286). The GRU model exhibited optimal performance, achieving a criterion correlation of r=0.850 and a test set prediction error of RMSE 0.273 (Table 6). An analysis conducted across different ASP score ranges revealed the stability of the GRU in the midscore range (1.5-3, RMSE ≤0.377) and resistance to extreme value fluctuations (3.5-4, RMSE 0.854), whereas the LR and LSTM models exhibited greater errors in complex or sparse score intervals (Table 7). A supplementary reliability study conducted with 10 participants yielded a moderate-to-strong ICC (0.776, 95% CI 0.097-0.944, df=9, P=.02), thus highlighting the temporal stability of scores over a 1-month interval (refer to the Results section, Best Model subsection).

A 2×2 mixed experimental design (n=148) revealed significant differences in immersion experience between gamified assessment (mean 7.628) and questionnaire assessment (mean 7.216; F_{147}=14.259; P<.001; Table 11). External incentives (ie, increased participation payments) induced response bias in the questionnaires, such that incentivized participants reported lower ASP scores (mean 2.066 vs 2.201; F_1=5.740; P=.02), whereas gamified assessment scores remained unaffected (F_1=0.138; P=.71; Table 14), thus indicating that the narrative effectively neutralized strategic self-presentation (Table 14). The gamified assessment also elicited stronger negative emotions (mean 4.365 vs 2.473; F_{147}=151.109; P<.001) and higher levels of interest (mean 7.095 vs 6.155; F_{147}=47.940; P<.001) in line with its immersive design. The variance decomposition revealed that immersion experience served as an effective antibias component (F_{147}=14.259; P<.001, η^2=0.092).

The future assessment preference data indicated that 79.1% (117/148) of the participants preferred the gamified assessment over questionnaires, despite the stronger negative emotions induced by the gamified assessment, thus reflecting the acceptability of this approach in organizational settings (Table 12).

This study reports the construction of a replicable ECD framework for gamified psychological assessment, which can be extended beyond the assessment of antisocial traits to encompass broader personality constructs. The framework involves six iterative phases: (1) theoretical and needs analysis, which aligns target constructs (eg, *DSM-5* criteria) with the empirical literature and stakeholder requirements; (2) competency model construction, which involves translating abstract traits into hierarchical, observable behavioral indicators; (3) evidence model development, which entails linking behavioral manifestations with trait dimensions based on expert validation; (4) game ontology design, which focuses on embedding assessment tasks within immersive narratives via dynamic decision paths; (5) technical implementation, which requires integrating interactive interfaces with mechanisms for the capture of real-time data; and (6) model training and validation, which highlights the use of machine learning algorithms to transform behavioral patterns into reliable trait scores that can be validated by reference to established instruments.

Key innovations include the integration of realistic scenarios to reduce response bias; the use of adaptive technology solutions (eg, dynamic jump logic and 1-shot coding) to capture nuanced behaviors; and multidimensional validation through statistical and qualitative methods. This framework, which prioritizes consistency among theoretical constructs, observable behaviors, and technical design, provides a systematic approach to the development of gamification tools that contribute to the broader aim of capturing realistic psychological traits.

Limitations

Although significant progress has been made in the development and validation of the gamified assessment tool, several limitations warrant further investigation.

First, the validation experiments reported in this study relied on a limited sample. Future researchers should investigate more diverse samples, including those from different cultural backgrounds, types of organizations, and employee levels, with the aim of enhancing and validating the generalizability of this tool.

Second, although the tool resists falsification, the underlying psychological mechanisms remain unclear. Future researchers should explore how different gamified elements influence behavior and decision-making and design experiments aimed at elucidating these mechanisms.

While the initial retest analysis indicated acceptable reliability (ICC=0.776), the small sample size limited confidence in longitudinal stability. Future researchers could validate the reliability of this approach based on larger samples and longer time intervals (eg, 3-6 months), thereby enabling them to explore

XSL•FO
RenderX

the validity of the assessment of the ASP-ECD-G in further detail.

Comparison With Previous Work

Unlike traditional tools, such as scales and SJTs, which assume item independence and 1-dimensionality, the ASP-ECD-G uses interconnected items within a simulation, thereby reflecting a multidimensional understanding of antisocial traits. This innovative approach can facilitate a more comprehensive and accurate assessment of personality traits. Recent efforts to gamify personality assessment, such as the framework for integrity testing by Cui et al [29], have emphasized scenario-based design but have lacked systematic integration with clinical diagnostic criteria, such as the *DSM-5*. In contrast, this study explicitly anchors the assessment ontology in *DSM-5* antisocial traits, thereby ensuring content validity and connecting clinical theory with organizational contexts. This ECD approach is superior to ad hoc game development, as it creates a replicable pipeline leading from trait definition to behavioral evidence collection—a step often overlooked by previous gamified tools [30].

Future Work

Regarding technology, this study used the GRU model for data processing. As artificial intelligence and machine learning technologies continue to develop, future researchers should explore advanced prediction algorithms and expand the sample of participants to improve data analysis capabilities and prediction accuracy. In addition, the current questionnaire-based format of this tool is relatively monotonous. The integration of multimodal data (eg, physiological responses or interaction timing) and the development of software that includes situational images and character avatars—with the goal of deepening behavioral insights—are expected to create a more immersive and engaging assessment experience, thereby improving user engagement and ecological validity.

As a foundational step in the process of developing a gamified assessment for ECD, addressing limitations remains critical. Issues that must still be addressed include sample diversification across groups with different cultural backgrounds, occupations, ages, and levels of education, with the aim of determining whether cross-context and cross-domain applications continue to exhibit utility in these contexts. Future work should also explore the temporal boundaries of the validity of relevant measurements by extending the retesting intervals (eg, 6-12 months). Technical improvements such as visual scenario design and adaptive algorithm improvements may further enhance immersion and measurement accuracy. This study highlights the potential of behavioral assessments to complement traditional self-report methods by anchoring gamification tools in systematic frameworks, such as the ECD, and by emphasizing rigorous validation. This approach provides new ways of understanding negative traits in a more nuanced and reliable manner.

Conclusions

The development and validation of the ASP-ECD-G address, in part, the critical limitations of traditional self-report tools used to assess complex antisocial traits by providing a behavior-based, immersive alternative. The ASP-ECD-G exhibits robust content validity (100% alignment with *DSM-5* traits), excellent predictive accuracy (GRU model criterion correlation r=0.792), and moderate to strong test-retest reliability (ICC=0.776; P=.02). It also exhibits resistance to response manipulation, as incentivized participants in this research were associated with altered questionnaire scores but not gamified assessment results (P=.71), thus supporting the stability of this tool over time and identifying it as a reliable instrument for capturing antisocial tendencies in simulated workplace scenarios.

Unlike self-report tools, which are susceptible to social desirability bias, the gamified approach elicits authentic responses through immersive decision-making, as evidenced by significantly higher immersion scores (mean 7.628 vs 7.216; P<.001) and consistent performance under conditions involving incentives. This mechanism, which is supported by flow theory, is correlated with reduced response distortion. These findings highlight the potential of ECD-based gamified assessments to improve the accuracy of personality measurement in high-stakes contexts, such as recruitment, where traditional methods often fail to detect subtle antisocial behaviors.

Ultimately, the ASP-ECD-G offers a theoretically sound and empirically validated framework, thereby contributing to both psychological research on negative personality traits and organizational practices aimed at mitigating the associated risks. Although this tool is not definitive or final, its contribution lies in the fact that it offers a replicable methodology and preliminary evidence for gamified assessment as a credible alternative in the context of trait measurement, thus calling for continued innovation to unlock its full potential for real-world applications.

Acknowledgments

The authors deeply appreciate the partnership and support of the leaders, teachers, staff, parents and caregivers, and students and scholars at Beijing Normal University, Zhuhai.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

XSL•FO
RenderX

Multimedia Appendix 1
Summary of interview results and interview questionnaire.
[DOCX File , 15 KB - games_v13i1e70453_app1.docx]

Multimedia Appendix 2
The presentation of the Antisocial Personality Traits Based on Evidence-Centered Design Gamified assessment tool.
[PNG File , 1070 KB - games_v13i1e70453_app2.png]

Multimedia Appendix 3
Assessment items pertaining to pleasure, fun, positive emotions, negative emotions, and immersion experience, as well as the Personality Inventory for DSM-5–Short Form items used to measure antisocial personality traits.
[DOCX File , 15 KB - games_v13i1e70453_app3.docx]

Multimedia Appendix 4
Model prediction results for different behavioral traits.
[DOCX File , 192 KB - games_v13i1e70453_app4.docx]

Multimedia Appendix 5
Validation of the multidimensional measurement used in gamified assessment tools.
[DOCX File , 18 KB - games_v13i1e70453_app5.docx]

Multimedia Appendix 6
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 12291 KB - games_v13i1e70453_app6.pdf]

References

1. Hogan R, Hogan J. Assessing leadership: a view from the dark side. Int J Sel Assess 2008 Jun 28;9(1-2):40-51. [doi: 10.1111/1468-2389.00162]
2. Fossati A, Barratt ES, Carretta I, Leonardi B, Grazioli F, Maffei C. Predicting borderline and antisocial personality disorder features in nonclinical subjects using measures of impulsivity and aggressiveness. Psychiatry Res 2004 Feb 15;125(2):161-170. [doi: 10.1016/j.psychres.2003.12.001] [Medline: 15006439]
3. Robinson SL, O'Leary-Kelly AM. Monkey see, monkey do: the influence of work groups on the antisocial behavior of employees. Acad Manage J 1998 Dec;41(6):658-672. [doi: 10.2307/256963]
4. Lilienfeld SO. Conceptual problems in the assessment of psychopathy. Clin Psychol Rev 1994 Jan;14(1):17-38. [doi: 10.1016/0272-7358(94)90046-9]
5. Maples JL, Carter NT, Few LR, Crego C, Gore WL, Samuel DB, et al. Testing whether the DSM-5 personality disorder trait model can be measured with a reduced set of items: an item response theory investigation of the Personality Inventory for DSM-5. Psychol Assess 2015 Dec;27(4):1195-1210. [doi: 10.1037/pas0000120] [Medline: 25844534]
6. Baumert A, Schlösser T, Schmitt M. Economic games. Eur J Psychol Assess 2014 Jan 01;30(3):178-192. [doi: 10.1027/1015-5759/a000183]
7. Ployhart RE, Ehrhart MG. Be careful what you ask for: effects of response instructions on the construct validity and reliability of situational judgment tests. Int J Sel Assess 2003 Apr 28;11(1):1-16. [doi: 10.1111/1468-2389.00222]
8. Hare RD. Psychopathy checklist—revised. APA PsycTests. 1991. URL: https://psycnet.apa.org/doiLanding?doi=10.1037%2Ft01167-000 [accessed 2025-07-22]
9. Reyna VF, Farley F. Risk and rationality in adolescent decision making: implications for theory, practice, and public policy. Psychol Sci Public Interest 2006 Sep;7(1):1-44. [doi: 10.1111/j.1529-1006.2006.00026.x] [Medline: 26158695]
10. Hare RD. A research scale for the assessment of psychopathy in criminal populations. Pers Individ Dif 1980;1(2):111-119. [doi: 10.1016/0191-8869(80)90028-8]
11. Coccaro EF, Lee R, McCloskey MS. Relationship between psychopathy, aggression, anger, impulsivity, and intermittent explosive disorder. Aggress Behav 2014;40(6):526-536. [doi: 10.1002/ab.21536] [Medline: 24760575]
12. van Nimwegen C, van Oostendorp H, Modderman J, Bas M. A test case for GameDNA: conceptualizing a serious game to measure personality traits. In: Proceedings of the 16th International Conference on Computer Games. 2011 Presented at: CGAMES 2011; July 27-30, 2011; Louisville, KY. [doi: 10.1109/cgames.2011.6000342]
13. Lievens F, Sackett PR. The validity of interpersonal skills assessment via situational judgment tests for predicting academic success and job performance. J Appl Psychol 2012 Mar;97(2):460-468. [doi: 10.1037/a0025741] [Medline: 21967295]

XSL•FO
RenderX

14. Deterding S, Dixon D, Khaled R, Nacke L. From game design elements to gamefulness: defining "gamification". In: Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments. 2011 Presented at: MindTrek '11; September 28-30, 2011; Tampere, Finland. [doi: 10.1145/2181037.2181040]
15. DiCerbo KE. Game-based assessment of persistence. Educ Technol Soc 2013 Apr;17(1):17-28.
16. Teng CI. Personality differences between online game players and nonplayers in a student sample. Cyberpsychol Behav 2008 Apr;11(2):232-234. [doi: 10.1089/cpb.2007.0064] [Medline: 18422420]
17. Mislevy RJ, Steinberg LS, Almond RG. Rejoinder to commentaries for "on the structure of educational assessments". Meas Interdiscip Res Perspect 2003;1(1):92-101. [doi: 10.1207/s15366359mea0101_04]
18. Zieky MJ. An introduction to the use of evidence-centered design in test development. Psicol Educ 2014;20(2):79-87. [doi: 10.1016/j.pse.2014.11.003]
19. Shute VJ. Stealth assessment in computer-based games to support learning. In: Tobias S, Fletcher JD, editors. Computer Games and Instruction. Charlotte, NC: Information Age Publishing; 2011:503-524.
20. Singh BK, Katiyar M, Gupta S, Ganpatrao NG. A survey on: personality prediction from multimedia through machine learning. In: Proceedings of the 5th International Conference on Computing Methodologies and Communication. 2021 Presented at: ICCMC 2021; April 8-10, 2021; Erode, India. [doi: 10.1109/iccmc51019.2021.9418384]
21. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000 Jan;55(1):68-78. [doi: 10.1037//0003-066x.55.1.68] [Medline: 11392867]
22. Csikszent M. Flow: The Psychology of Optimal Experience. New York, NY: HarperCollins; 1991.
23. Lee SJ, Jeong EJ, Kim DJ, Kong J. The influence of psychological needs and motivation on game cheating: insights from self-determination theory. Front Psychol 2023 Dec 22;14:1278738 [FREE Full text] [doi: 10.3389/fpsyg.2023.1278738] [Medline: 38187409]
24. Liu L, Chen X, Szolnoki A. Coevolutionary dynamics via adaptive feedback in collective-risk social dilemma game. Elife 2023 May 19;12:e82954 [FREE Full text] [doi: 10.7554/eLife.82954] [Medline: 37204305]
25. First MB. Diagnostic and statistical manual of mental disorders, 5th edition, and clinical utility. J Nerv Ment Dis 2013 Sep;201(9):727-729. [doi: 10.1097/NMD.0b013e3182a2168a] [Medline: 23995026]
26. McKinley S, Patrick C, Verona E. Antisocial personality disorder: neurophysiological mechanisms and distinct subtypes. Curr Behav Neurosci Rep 2018 Jan 25;5:72-80. [doi: 10.1007/s40473-018-0142-0]
27. Bandura A. Moral disengagement in the perpetration of inhumanities. Pers Soc Psychol Rev 1999;3(3):193-209. [doi: 10.1207/s15327957pspr0303_3] [Medline: 15661671]
28. Aquino K, Douglas S. Identity threat and antisocial behavior in organizations: the moderating effects of individual differences, aggressive modeling, and hierarchical status. Organ Behav Hum Decis Process 2003 Jan;90(1):195-208. [doi: 10.1016/s0749-5978(02)00517-4]
29. Cui Y, Chu MW, Chen F. Analyzing student process data in game-based assessments with Bayesian knowledge tracing and dynamic Bayesian networks. J Educ Data Min 2019 Jun 24;11(1):80-100. [doi: 10.5281/zenodo.3554751]
30. Wouters P, van Nimwegen C, van Oostendorp H, van der Spek ED. A meta-analysis of the cognitive and motivational effects of serious games. J Educ Psychol 2013;105(2):249-265. [doi: 10.1037/a0031311]

Abbreviations

ANN: artificial neural network
ASP: antisocial personality
ASP-ECD-G: Antisocial Personality Traits Evidence-Centered Design Gamified assessment tool
DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
ECD: evidence-centered design
GRU: gated recurrent unit
ICC: intraclass correlation coefficient
LR: linear regression
LSTM: long short-term memory
MAE: mean absolute error
PID-5-SF: Personality Inventory for DSM-5 Short Form
RF: random forest
RMSE: root mean square error
RNN: recurrent neural network
SJT: situational judgment test

XSL•FO
RenderX

Edited by A Coristine; submitted 22.12.24; peer-reviewed by X Cheng, JH Song; comments to author 11.03.25; revised version received 04.05.25; accepted 02.07.25; published 25.08.25.

Please cite as:
Tang Y, Xu Y, Zhou Q, Bian R
A Gamified Assessment Tool for Antisocial Personality Traits (Antisocial Personality Traits Evidence-Centered Design Gamified): Randomized Controlled Trial
JMIR Serious Games 2025;13:e70453
URL: https://games.jmir.org/2025/1/e70453
doi:10.2196/70453
PMID:

©Yaobin Tang, Yongze Xu, Qunli Zhou, Ran Bian. Originally published in JMIR Serious Games (https://games.jmir.org), 25.08.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

In-Game Need Satisfaction, Frustration, and Gaming Addiction Patterns Across Subgroups of Adolescents Through Structural Equation Modeling: Cross-Sectional and Instrument Validation Study of the Youth Gaming Experience Scales

Amparo Luján-Barrera[1], MSc; Lydia Cervera-Ortiz[1], MSc; Mariano Chóliz[1], PhD

Department of Basic Psychology, Faculty of Psychology, University of Valencia (UV), Valencia, Spain

Corresponding Author:
Amparo Luján-Barrera, MSc
Department of Basic Psychology
Faculty of Psychology
University of Valencia (UV)
Avinguda Blasco Ibáñez, 21
Valencia, 46010
Spain
Phone: 34 689286800
Email: amparo.l.lujan@uv.es

Abstract

Background: Gaming is a prevalent activity during adolescence, a developmental stage characterized by vulnerability to gaming disorder (GD). According to the self-determination theory, gaming environments can satisfy and frustrate basic psychological needs—autonomy, competence, and relatedness—processes linked to GD. However, existing research has primarily focused on adult populations, and validated instruments assessing both in-game need satisfaction (NS) and in-game need frustration (NF) in adolescents are lacking.

Objective: This study aimed to address this gap by validating the Spanish version of the Basic Psychological Need Satisfaction and Frustration Scale for Gaming (BPNSFS-G) in adolescents. We examined its psychometric properties, its relationship with GD and gaming behavior, and differences in NS and NF across subgroups defined by risk factors for GD (sex, the developmental stage, and the gaming modality).

Methods: A total of 1174 adolescents (mean 12.07, SD 1.23; middle adolescents: 815/1174, 69.4%; male: 637/1037, 61.4%; online gamers: 388/511, 76.3%) participated in a school-based GD prevention program and completed a self-report battery, which included the adapted BPNSFS-G, a validated GD scale, and ad hoc items (daily gaming time and weekly frequency). The Spanish adaptation of the BPNSFS-G was developed using the back-translation method to ensure its content validity. A structural equation modeling approach was used to test its structural, construct, convergent, and discriminant validity, as well as measurement invariance. Reliability at both the item and factor levels, along with criterion validity, was also assessed.

Results: Through exploratory and confirmatory factor analyses, the scale structure was confirmed, validating 2 separate measures: a 3-factor second-order model for NS and a unidimensional model for NF, both with adequate internal consistency (NS: Cronbach α=0.75; NF: Cronbach α=0.77). The results also supported its validity, with NF more strongly associated with GD (r=0.49), and NS more closely related to gaming time (r=0.25-0.36). The scales functioned similarly across sex and developmental stage groups, but not across gaming modalities. Notably, NS showed significant differences across all subgroups, especially between boys and girls ($t_{759.31}$=8.28; P<.001; g=0.55) and between online and offline gaming modes ($t_{178.47}$=5.13; P<.001; g=0.58), with no meaningful differences found for NF.

Conclusions: This study provides initial evidence for the validity and reliability of the Spanish version of the BPNSFS-G for adolescents. The resulting Youth Gaming Experience Scales (Youth Satisfying Gaming Experience Scale and Youth Frustrating Gaming Experience Scale) are brief, theoretically grounded, and empirically supported instruments for assessing in-game psychological need satisfaction and frustration. Given NF's potential as a risk factor, these scales have a strong potential to explore self-determination theory–based mechanisms underlying the motivational etiology of GD and to advance both its assessment and prevention in adolescent populations.

XSL•FO
RenderX

(JMIR Serious Games 2025;13:e63612) doi:10.2196/63612

KEYWORDS

self-determination theory; adolescence; in-game need satisfaction; scale validation; basic psychological needs; structural equation modeling; game addiction; psychological needs; Spain

Introduction

Gaming represents one of the most widespread forms of virtual leisure worldwide. Its relevance is reflected in a global gamer population exceeding 3 billion gamers [1]. In Spain, 77% of young people aged 6 to 14 years and 85.1% of adolescents aged 14 to 18 years engage in gaming regularly [2,3]. Although gaming has transcended age barriers and become a pastime across generations [4], it is during childhood that both traditional (nondigital) and digital gaming habits are formed. Both digital and nondigital gaming share a core feature: their capacity to engage players through intrinsic motivation [5]. However, the key distinction between both categories lies in the potential maladaptive use of digital gaming, which can develop into an addictive disorder called gaming disorder (GD). GD is recognized as a persistent gaming behavior that significantly impairs psychosocial functioning in both the *DSM-5-TR* (*Diagnostic and Statistical Manual of Mental Disorders* [Fifth Edition, Text Revision]) and the *ICD-11* (*International Classification of Diseases, 11th Revision*) psychiatric classifications. However, while the *DSM-5-TR* classifies GD as a condition warranting further research, the *ICD-11* recognizes it as a formal disorder [6,7]. Despite these nosological differences, this study adopts the American Psychiatric Association's *DSM-5* (*Diagnostic and Statistical Manual of Mental Disorders* [Fifth Edition]) based framework for GD, which remains widely used in research and clinical settings [8].

GD represents a growing public health concern, affecting approximately 3% of gamers worldwide [9] and 4.6% of young male gamers [10]. In Spain, its prevalence among adolescents aged 14-18 years is 7.1%, rising to 11.3% among male gamers compared to 2.7% among female gamers [3]. This higher prevalence among adolescent male gamers is consistent with established risk factors for GD, such as male sex and adolescence [11], along with gaming-related characteristics like online gaming [12]. Moreover, a growing body of research highlights gaming motivation as a significant factor in the development of GD [13]. Specifically, the self-determination theory (SDT) [14] provides a comprehensive framework for understanding the etiology of GD in youths [15].

SDT posits that gaming facilitates the fulfillment of the 3 fundamental psychological needs—autonomy (volition and personal agency), competence (sense of efficacy), and relatedness (social connection)—both in-game and in real-life [14,16]. Building on SDT, the need density hypothesis [17] proposes that GD may develop when in-game need satisfaction (NS) significantly exceeds real-life need satisfaction [18].

Although widely accepted [19,20], empirical support for this framework is inconsistent. Some studies could not replicate the expected effects of the imbalance between NS and real-life need satisfaction [21], while others found no direct association between NS and GD [18,22]. According to the need density hypothesis, recent research suggests that real-life need frustration may be a stronger predictor of GD than low real-life need satisfaction [22,23], especially when combined with high NS [18,24,25]. However, gaming environments can also frustrate these needs—defined as the active thwarting rather than merely the absence of satisfaction [24,26]—an aspect not yet fully addressed by SDT [27,28].

Originally, SDT posits that NS is linked to positive gaming outcomes, such as enjoyment, intrinsic motivation, vitality, engagement, and well-being [14,16,29-31], while in-game need frustration (NF) is associated with dysfunctional and obsessive use patterns that contribute to addictive behaviors, such as GD [24,32]. Notably, both real-life need frustration and NF have been linked to GD [22,23,25,28]. However, research on NF has received considerably less attention compared to real-life need frustration [25,27,28].

This gap is not unexpected, as SDT's motivational model of gaming has historically centered on the study of NS [16,30], with more limited theoretical development concerning NF [27]. This emphasis on NS is also reflected in the availability of validated instruments designed to measure NS—such as the Player Experience of Need Satisfaction (PENS) [16], the Game Experience Questionnaire [33], and the Ubisoft Perceived Experience Questionnaire [34]—but not in the assessment of NF. To date, the prevailing approach has assessed NF in conjunction with NS, typically through adapted versions of the Basic Psychological Need Satisfaction and Frustration Scale (BPNSFS) [35] for gaming contexts [25,36]. However, most studies have used ad hoc adaptations without evaluating their psychometric properties [37]. This is particularly noteworthy, given that a validated version of the Basic Psychological Need Satisfaction and Frustration Scale for Gaming (BPNSFS-G) specifically designed to assess NS and NF is already available [28].

The current state of the literature reveals several critical gaps that warrant further investigation. First, research on NF and its relationship with GD remains both limited and methodologically inconsistent. Second, most studies have focused on university students or adult samples [25,28,36,37], while only a few studies have examined NS or NF using adolescent samples [20,21]. Finally, the available instruments to assess these constructs have been developed exclusively in English, restricting the generalizability of findings.

In response, this study aimed to translate the BPNSFS-G into Spanish, adapt it for the adolescent population, and validate its psychometric properties in a sample of adolescent gamers. Specifically, the study assessed the scale's structural validity, internal consistency, as well as construct and criterion validity. Once the psychometric adequacy was established, we tested

XSL•FO
RenderX

two hypotheses based on the current state of literature (1) that NF would show a stronger association with GD than NS and (2) that NS would be more closely related to gaming time and frequency than NF. Additionally, we conducted measurement invariance analyses across subgroups defined by known GD risk factors such as the developmental stage (early vs middle adolescence), sex (girls vs boys), and the gaming mode (online vs offline). This approach enabled us to explore the functioning of NS and NF across different adolescent gamer subpopulations. To our knowledge, this is the first study to examine these questions.

Methods

Participants

Participants were recruited through convenience sampling from 27 schools in Valencia, where the "Gamer" GD prevention program [38] was conducted from October 2022 to May 2023. The program lasts 3 weeks with 1 session per week. A pre- and postevaluation was conducted, but this study focuses exclusively on the baseline assessment, using a cross-sectional design. The questionnaires were paper-based and self-administered in groups under the supervision of prevention specialists.

The program targeted students aged 10 to 15 years (fifth year in primary school to second year in secondary school). Participants were classified into early adolescence (10-12 years old) and middle adolescence (13-15 years old) [39]. The inclusion criteria were no missing responses on any questionnaire item, sufficient Spanish proficiency, and identification as gamers on predefined criteria (see Demographic and Gaming Patterns Measures section for details).

Of the 1747 students evaluated, 1174 met the inclusion criteria (mean age 12.07, SD 1.23 years; early adolescents: 359/1174, 30.6% and middle adolescents: 815/1174, 69.4%), with 1037 participants reporting their sex (female: 400/1037, 38.6% and male: 637/1037, 60.6%) and 511 reporting their gaming modality (online gamers: 388/511, 75.9% and offline gamers: 123/511, 24.1%).

Ethical Considerations

The study was approved by the University of Valencia Ethics Committee (approval H1482079199937), in accordance with institutional and national ethical standards, including the Declaration of Helsinki and Spanish regulations on biomedical research and personal data protection. Participation was voluntary and anonymous, and institutional agreement was obtained from the participating schools prior to data collection. Questionnaires used unique anonymized ID codes, with no personally identifiable information collected, and data were stored securely and accessed only by authorized research personnel. No financial compensation was provided.

Procedure

The BPNSFS-G [28] was adapted into Spanish (Multimedia Appendix 1) using the back-translation method [40], following a 6-phase structure [41]. First, 2 independent translations of the original scale were produced in adolescent psychology and gaming motivation research, all experienced in psychometric validation, linguistic adaptation, and proficient in English. Second, these versions were reviewed and merged into a single refined Spanish version. In the third phase, a bilingual English linguist back-translated this version, which was then assessed by the expert panel for its semantic, conceptual, and cultural equivalence with the original scale.

The fourth phase involved adapting the questionnaire to the adolescent population at both linguistic and methodological levels. First, the panel made linguistic modifications to simplify terminology and improve item clarity for adolescents while maintaining the original meaning. Furthermore, the response scale was reduced from 7 to 5 points to enhance reliability within this population [42]. This modification was deemed necessary because simpler formats are more appropriate for the cognitive development–level characteristic of early adolescence [43]. In addition, item wording was adjusted to ensure applicability to both online and offline gaming experiences, referencing gamers' experiences with the games they usually play [16], rather than restricting responses to a specific gaming modality [28].

The fifth phase consisted of a pilot study conducted in 4 schools (mean age 12.26, SD 0.91 years; girls: 92/199, 46.2%; boys: 93/199, 46.7%; nonbinary sex: 6/199, 3%; early adolescents: 49/199, 24.6%; middle adolescents: 150/199, 75.4%). Based on feedback from students, teachers, and prevention specialists, minor linguistic adjustments were made. A second pilot study was conducted in 2 additional schools (mean age 11.79, SD 1.07 years; girls: 25/56, 44.7%; boys: 31/56, 53.4%; early adolescents: 32/56, 55.2%; middle adolescents: 24/56, 41.4%), confirming the scale's comprehensibility. The expert panel approved the final version of the instrument, confirming its content validity for use in the main study.

Measures

Satisfaction and Frustration of Basic Psychological Needs During Gameplay

The BPNSFS-G [28] is a self-administered questionnaire designed to assess NS and NF, each conceptualized as a second-order construct composed of 3 first-order dimensions: autonomy, competence, and relatedness in-game needs.

The original version comprises 15 items rated on a 7-point Likert scale. For this study, the scale was adapted to a 5-point Likert format (0=strongly disagree to 4=strongly agree). In the original questionnaire, each item ended with the phrase "in my current favorite online game." In the adapted version, this was changed to "the games you normally play" as an introductory statement to broaden its applicability beyond online gaming.

The satisfaction subscale consists of 7 items, selected from the PENS [16]. Higher scores indicate that video games are perceived as a source of need fulfillment within the gaming context. The subscale includes 3 items for autonomy (eg, "I feel that the decisions I make are the ones I really want to make"), 2 for competence (eg, "I feel skilled at what I do"), and 2 for relatedness (eg, "I feel close to the people who are important to me in the game").

The frustration subscale consists of 8 items, adapted from the BPNSFS [35]. Higher scores indicate that video games are

XSL•FO
RenderX

perceived as a context where psychological needs are thwarted. The subscale includes 3 items for autonomy (eg, "I feel forced to do many things that I would not do if I had a choice"), 3 for competence (eg, "I feel insecure about my abilities in the game"), and 2 for relatedness (eg, "I feel excluded from the group I want to belong to in the video game").

Gaming Disorder

GD was assessed using an abbreviated version of the Video Game Dependency Test [44], adapted to the DSM-5 and ICD-11 diagnostic criteria for GD [45]. This self-administered, paper-based questionnaire consists of 10 items assessing gaming-related behaviors (eg, "I spend less time doing other activities because video games take up a significant part of my time") and their impact over the past 12 months, rated on a 5-point Likert scale (0=strongly disagree to 4=strongly agree). The total score ranges from 0 to 40, with higher scores indicating greater symptom severity.

This version of the Video Game Dependency Test has been adapted for Spanish-speaking adolescents, demonstrating adequate internal consistency in this study (Cronbach α=0.85) comparable to the original validation (Cronbach α=0.874) [45].

Demographic and Gaming Pattern Measures

This set of ad hoc items was designed to assess participants' sociodemographic characteristics and gaming patterns. Sociodemographic data comprised age, sex, and school grade. Gaming behavior was measured through daily gaming time and weekly gaming frequency, using 3 self-reported items on a 5-point ordinal scale. Daily gaming time was measured separately for weekdays and weekends, with response options ranging from 1=I do not play, 2=less than 1 hour, 3=1-2 hours, 4=2-3 hours, to 5=3 or more hours. Weekly gaming frequency was evaluated with response options ranging from 1=I do not usually play, 2=1-2 days, 3=3-4 days, 4=5-6 days, to 5=every day. Higher scores on time and frequency measures indicate greater gaming engagement. These 3 items were used as inclusion criteria when participants scored ≥3 on both daily gaming time and weekly frequency. Additionally, participants reported their preferred gaming modality, indicating whether they primarily played offline or online.

Statistical Analyses

Data analyses were conducted using SPSS Statistics (version 28; IBM Corp) and RStudio (Posit Software, PBC). Given the study's multivariate design, structural equation modeling was used to assess factorial structure, measurement invariance, and construct validity. The diagonal weighted least squares estimation method was applied due to the ordinal scale and nonnormal distribution of the data [46].

To analyze structural validity, exploratory factor analysis (EFA) was first conducted using principal component analysis with Varimax rotation. Data adequacy was confirmed using the Kaiser-Meyer-Olkin measure (≥0.80) and Bartlett test (P<.001), and factors were then retained based on eigenvalues (>1) and communalities (h^2≥0.30). Second, confirmatory factor analysis (CFA) tested the factor structure derived from the EFA against competing theoretical models, with model fit evaluated using chi-square (P>.001), comparative fit index (CFI; >0.95), root-mean-square error of approximation (RMSEA; <0.06), and standardized root-mean-square residual (SRMR; <0.08) [47]. Third, measurement invariance was assessed across sex (boys vs girls), the developmental stage (early vs middle adolescents), and the gaming modality (online vs offline) using a multilevel approach. This included testing configural, metric, and scalar invariance, with changes in ΔCFI (≤0.01) and ΔRMSEA (≤0.015) as criteria. For ΔSRMR, thresholds were set at ≤0.03 for metric and ≤0.01 for scalar invariance [48].

Following the selection of the final factorial model, descriptive statistics including mean, SD, skewness, and kurtosis were computed to analyze the distribution of scores. Group differences were examined using independent samples 2-tailed t tests (bilateral) with 95% CIs and Hedges g effect sizes (small≥0.2, medium≥0.5, or large≥0.8). Finally, internal consistency was assessed using Cronbach α and composite reliability (CR; ≥0.7). Item reliability was assessed using Cronbach α if an item was deleted (α–item) and the corrected item-total correlation (CITC) was 0.3 [49]. Construct and criterion validity were tested via CFA and Fisher bilateral correlations with gaming behavior indicators (daily gaming time and weekly frequency) and GD. Other indicators of internal validity were assessed using standardized factor loadings (λ >0.4), squared multiple correlations (R^2≥0.5), and average variance extracted (AVE) to evaluate convergent validity (AVE≥0.5; CR≥0.7; and λ >0.5), as well as discriminant validity (AVE>r^2 between factors) [50,51]. AVE and CR were manually calculated using standard formulas [51]:



λ_{ei} represents the standardized factor loading of item i, n is the number of items in the construct, and θ_i refers to the standardized measurement error associated with each item.

Results

Factorial Analyses

The objective of this section was to examine the internal structure and measurement invariance of the scale. The scale was originally designed as a unified instrument to assess both NS and NF. The results of exploratory, confirmatory, and multigroup factor analyses supported a differentiated structure, leading to the validation of 2 independent measures: a 3-factor second-order model for NS and a unidimensional model for NF.

First, the internal structure of the questionnaire was subjected to an EFA. The results of the EFA indicated a 2D structure with 2 first-order factors composed of NS and NF (Kaiser-Meyer-Olkin measure=0.812; χ^2_{105}=3799.0; P<.001). The first factor (NS) obtained an eigenvalue of 2.26 and explained 22.66% of the variance, while the second factor (NF) obtained an eigenvalue of 3.4 and explained 17.73% of the variance. The principal component analysis excluded items with low values in their communalities, including: "I am not clear if

XSL•FO
RenderX

I can do things well in the video game" (competence frustration=0.17) and "I feel that what I choose in video games really expresses how I am" (autonomy satisfaction=0.27).

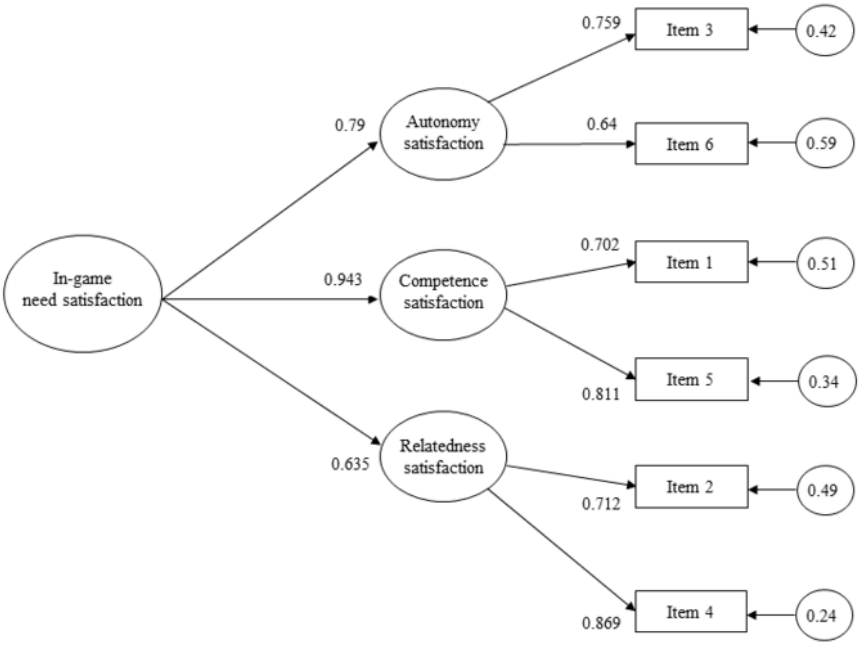
Second, a CFA was performed across the predefined models to identify the best fit for the data. The structure obtained from the EFA was rejected, despite its consistency with the theoretical model (χ^2_{64}=863.9; *P*>.001; CFI=0.82; RMSEA=0.11; 90% CI 0.097-0.12; SRMR=0.09). We then tested 2 factorial structures defined in the literature through competitive CFA. Both models included 6 factors for in-game needs, but model 1 was defined by a second-order structure [28] and model 2 by a first-order structure [35].

Model 2 (χ^2_{75}=617.4; *P*>.001; CFI=0.92; RMSEA=0.072; 90% CI=0.064-0.079; SRMR=0.067) performed better than model 1 (χ^2_{83}=853.1; *P*>.001; CFI=0.89; RMSEA=0.088; 90% CI 0.076-0.09; SRMR=0.084), which improved after eliminating the problematic items identified in EFA (model 1: χ^2_{83}=853.1; *P*>.001; CFI=0.88; RMSEA=0.088; 90% CI 0.076-0.09; SRMR=0.084; and model 2: χ^2_{75}=617.4; *P*>.001; CFI=0.92; RMSEA=0.072; 90% CI 0.064-0.079; SRMR=0.067). However, both models were rejected because of the RMSEA values.

Since the original questionnaire measured NS independently (PENS) [16,28], a second-order model with 3 first-order dimensions for both NS and NF (model 3) was tested. This decision followed the unsatisfactory results of the CFA, where an attempt to declare NS and NF as independent second-order factors resulted in a poor fit (χ^2_{84}=766.1; *P*>.001; CFI=0.88; RMSEA=0.081; 90% CI 0.077-0.088; SRMR=0.09), even after excluding the problematic items (χ^2_{59}=505.3; *P*>.001; CFI=0.91; RMSEA=0.081; 90% CI 0.073-0.09; SRMR=0.084).

The CFA results improved notably when the fit of model 3 was analyzed separately for both scales. For the NS scale, the fit approached ideal values (χ^2_{11}=52.3; *P*>.001; CFI=0.98; RMSEA=0.058; 90% CI 0.041-0.077; SRMR=0.029), which improved further after eliminating the autonomy satisfaction item, as it did not meet the cutoff for factor saturation (λ_e=0.456). Additionally, the 3-factor structure without this item (χ^2_6=12.4; *P*=.05; CFI=0.99; RMSEA=0.052; 90% CI 0.028-0.079; SRMR=0.015) was superior to the unidimensional structure (χ^2_{14}=425.7; *P*>.001; CFI=0.81; RMSEA=0.16; 90% CI 0.15-0.17; SRMR 0.077). Figure 1 shows the final structure of the NS scale.

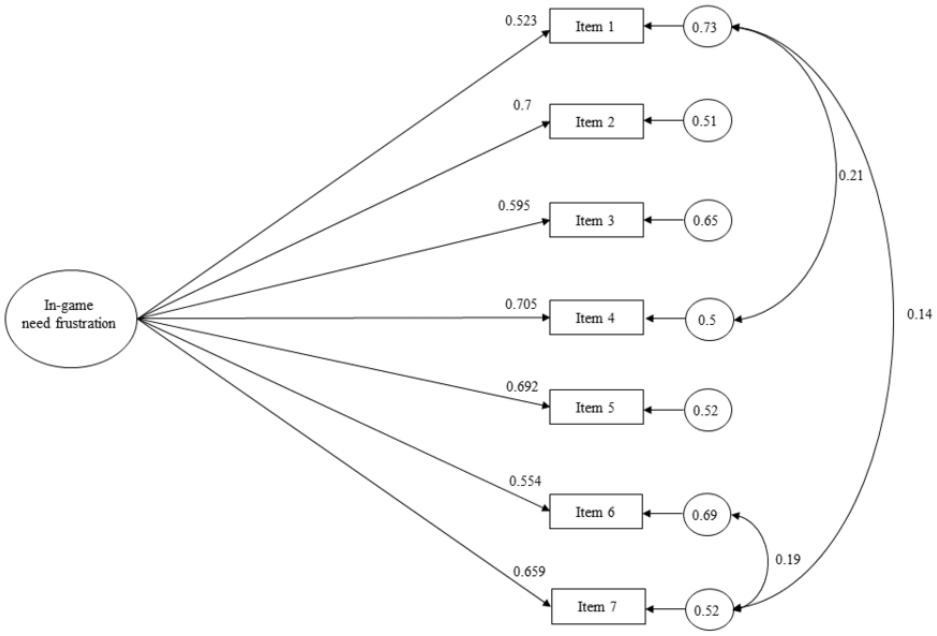
Figure 1. 3D model of in-game need satisfaction.



Conversely, the results of model 3 on the NF scale were quite satisfactory but not enough because of RMSEA (χ^2_{17}=87.4; *P*>.001; CFI=0.957; RMSEA=0.076; 90% CI 0.061-0.092; SRMR=0.033) and a problematic item from the competence frustration factor (λ_e=0.389). Analyses were replicated, removing this item, although the results were lower (χ^2_{11}=67.8; *P*>.001; CFI=0.962; RMSEA=0.087; 90% CI 0.069-0.107; SRMR=0.033). Finally, a unidimensional model was tested (χ^2_{20}=99.1; *P*>.001; CFI=0.952; RMSEA=0.075; 90% CI 0.061-0.089; SRMR=0.036), which did improve by removing this item (χ^2_{14}=82.3; *P*>.001; CFI=0.96; RMSEA=0.08; 90% CI 0.064-0.098; SRMR=0.036), reaching very satisfactory values when relating the error covariances of items 1, 4, 6, and 7 of the scale (χ^2_{11}=32.8; *P*>.001; CFI=0.986; RMSEA=0.053; 90% CI 0.032-0.074; SRMR=0.023). Figure 2 shows the final structure of the NF scale.

XSL•FO
RenderX

Figure 2. Unidimensional model of in-game need frustration.



Finally, CFA supported a 3D second-order structure for NS and a unidimensional structure for NF. After confirming different structures for NS and NF, its measurement invariance was tested. The analyses showed that both scales met the invariance criteria for developmental stage (Table 1) and sex (Table 2) groups, but not for the modality of play (Table 3). Specifically, the NF scale met both levels of invariance for the measurement of girls and boys as well as early and middle adolescents, although the NS scale only reached weak invariance for the developmental stage.

Table 1. Invariance of in-game need satisfaction (NS) and in-game need frustration (NF) based on developmental stage.

Groups	Developmental stage		*P* value	CFI[a] (Δ)	RMSEA[b] (Δ)	SRMR[c] (Δ)
	Early adolescence (358/717), chi-square (*df*)	Middle adolescence (359/717), chi-square (*df*)				
NS						
Configural	5.6 (12)	18.4 (12)	.02	0.994 (—[d])	0.053 (—)	0.023 (—)
Metric	9 (15)	18.9 (15)	.02	0.994 (0)	0.049 (0.004)	0.031 (0.008)
Scalar	9.5 (30)	18.3 (30)	.58	1 (0.006)	0.001 (0.048)	0.024 (0.007)
NF						
Configural	12.9 (22)	43.6 (22)	.001	0.984 (—)	0.066 (—)	0.039 (—)
Metric	17.0 (28)	37.5 (28)	.002	0.988 (0.004)	0.051 (0.015)	0.045 (0.006)
Scalar	26.5 (48)	50.7 (48)	.005	0.986 (0.002)	0.041 (0.01)	0.041 (0.004)

[a]CFI: comparative fit index.

[b]RMSEA: root-mean-square error of approximation.

[c]SRMR: standardized root-mean-square residual.

[d]Not applicable.

XSL•FO
RenderX

Table 2. Invariance of in-game need satisfaction (NS) and in-game need frustration (NF) based on sex.

Groups	Sex		*P* value	CFI[a] (Δ)	RMSEA[b] (Δ)	SRMR[c] (Δ)
	Girls (400/800), chi-square (*df*)	Boys (400/800), chi-square (*df*)				
NS						
Configural	8.3 (12)	18.7 (12)	.008	0.992 (—[d])	0.056 (—)	0.028 (—)
Metric	12.0 (15)	27.4 (15)	.001	0.988 (0.04)	0.064 (0.008)	0.038 (0.01)
Scalar	30.2 (30)	50.5 (30)	.001	0.974 (0.014)	0.065 (0.001)	0.04 (0.004)
NF						
Configural	12.9 (22)	20.7 (22)	.054	0.995 (—)	0.036 (—)	0.028 (—)
Metric	16.6 (28)	22.6 (28)	.078	0.996 (0.001)	0.032 (0.004)	0.034 (0.006)
Scalar	35.7 (48)	38.6 (48)	.009	0.99 (0.006)	0.037 (0.005)	0.034 (0)

[a]CFI: comparative fit index.

[b]RMSEA: root-mean-square error of approximation.

[c]SRMR: standardized root-mean-square residual.

[d]Not applicable.

Table 3. Invariance of in-game need satisfaction (NS) and in-game need frustration (NF) based on the gaming modality.

Groups	Gaming modality		*P* value	CFI[a] (Δ)	RMSEA[b] (Δ)	SRMR[c] (Δ)
	Online (123/246), chi-square (*df*)	Offline (123/246), chi-square (*df*)				
NS						
Configural	5.3 (12)	4.9 (12)	.60	1 (—[d])	0 (—)	0.029 (—)
Metric	9.5 (15)	9.7 (15)	.21	0.994 (0.006)	0.047 (0.047)	0.047 (0.018)
Scalar	15.0 (30)	13.8 (30)	.53	1 (0.006)	0 (0.047)	0.042 (0.005)
NF						
Configural	12.0 (22)	22.8 (22)	.04	0.985 (—)	0.069 (—)	0.052 (—)
Metric	26.1 (28)	28.6 (28)	.002	0.968 (0.017)	0.088 (0.019)	0.075 (0.033)
Scalar	23.2 (48)	34.2 (48)	.17	0.989 (0.021)	0.04 (0.048)	0.06 (0.015)

[a]CFI: comparative fit index.

[b]RMSEA: root mean-square-error of approximation.

[c]SRMR: standardized root-mean-square residual.

[d]Not applicable.

Descriptive Statistics and Differences Between Groups

Overall, NS was higher than NF, which had notably low values, although NS scores showed greater variability (Table 4). Specifically, autonomy (followed by competence) showed the highest satisfaction, while relatedness had the lowest.

XSL•FO
RenderX

Table 4. Description of the scores of the scales of basic need satisfaction and frustration in the gaming context[a].

Scales and items	Mean (SD)	Skewness	Kurtosis
"While playing the video games I normally play ..."			
NS[b]			
1. ... I am confident that I am good at playing.	2.79 (0.98)	–0.49	–0.12
2. ... I feel close to the people who are important to me in the game.	1.86 (1.38)	0.37	–1.21
3. ... I feel like I have a choice and a sense of freedom with the things I do in the game.	2.8 (1.21)	–0.77	–0.37
4. ... I feel close to the people I give to and receive support from in the game.	1.96 (1.38)	–0.07	–1.21
5. ... I feel skilled at what I do.	2.62 (1.20)	–0.64	–0.42
6. ... I feel that the decisions I make are the ones I really want to make.	2.75 (1.28)	–0.8	–0.39
NF[c]			
1. ... most of the things I do I feel I have to do out of obligation.	0.92 (1.12)	10.04	0.17
2. ... I feel excluded from the group I want to belong to in the game.	0.7 (10.0)8	1.53	1.54
3. ... when I make mistakes while playing, I feel like a failure.	0.8 (1.1)	1.35	10.01
4. ... I feel forced to do many things that I would not do if I had a choice.	1.06 (1.21)	0.87	–0.28
5. ... I feel that the people who are important to me in the video game are cold and distant to me during the game.	0.8 (1.11)	1.29	0.78
6. ... I feel insecure about my abilities in the game.	0.84 (10.07)	1.13	0.47
7. ... I feel pressured to do too many things in the game.	0.83 (1.14)	1.26	0.64
Total scores			
NS	14.5 (5.6)	–0.37	–0.14
Autonomy	5.56 (20.09)	–0.68	–0.26
Competence	5.41 (1.89)	–0.52	–0.13
Relatedness	3.82 (2.43)	0.01	–10.01
NF	5.95 (50.04)	0.79	0.01

[a]Originally translated from Spanish (Multimedia Appendix 1).

[b]NS: in-game need satisfaction.

[c]NF: in-game need frustration.

The scoring trend of the different groups was similar to the overall sample scores (Tables 5-7). The highest NS scores were obtained by the online gamers' subgroup, followed by the male participants, and finally by the early adolescents. These same populations also obtained higher scores on the NF, although in this case, the subgroup that presented greater frustration was the early adolescents, followed by the online gamers, and finally the male participants.

Effect sizes indicated that the population variables used to define gamer groups based on needs were relevant for sex and the game modality but not for the developmental stage. However, while the effect size was medium-high in the case of the NS factors, NF had a low or no effect. Therefore, it seems that these personal variables had a greater impact on the experience of NS than on NF.

XSL•FO
RenderX

Table 5. Differences between adolescent groups on the scores of in-game need satisfaction (NS) and in-game need frustration (NF) scales.

	Developmental stage		*t* test (*df*=1172)	*P* value[a]	95% CI	*g*
	Early adolescence (359/1174), mean (SD)	Middle adolescence (815/1174), mean (SD)				
AS[b]	5.57 (2.05)	5.56 (2.12)	0.1	.92	–0.25-0.27	–0.12
CS[c]	5.64 (1.78)	5.31 (1.93)	2.7	.007	0.09-0.56	0.05
RS[d]	4.14 (2.43)	3.67 (2.42)	3.04	.002	0.17-0.77	0.07
NS	17.15 (5.71)	16.23 (5.53)	2.6	.01	0.23-1.62	0.04
NF	6.43 (5.08)	5.75 (5.01)	2.14	.03	0.06-1.31	0.01

[a]The *P* value reflects the 2-tailed hypothesis test.

[b]AS: autonomy satisfaction.

[c]CS: competence satisfaction.

[d]RS: relatedness satisfaction.

Table 6. Differences between sex groups on the scores of in-game need satisfaction (NS) and in-game need frustration (NF) scales.

	Sex		*t* test (*df*)	*P* value[a]	95% CI	*g*
	Girls (400/1037), mean (SD)	Boys (637/1037), mean (SD)				
AS[b]	5.21 (2.25)	5.93 (1.86)	5.37 (730.36)	<.001	0.46-0.99	0.36
CS[c]	4.72 (1.9)	5.86 (1.73)	9.77 (788.67)	<.001	0.91-1.4	0.64
RS[d]	3.24 (2.44)	4.31 (2.36)	7.02 (1035)	<.001	0.77-1.4	0.45
NS	14.9 (5.8)	17.82 (5.04)	8.28 (759.31)	<.001	2.2-3.6	0.55
NF	5.53 (4.93)	5.98 (4.98)	1.46 (1035)	.15	0.17-1.07	0.09

[a]The *P* value reflects the 2-tailed hypothesis test.

[b]AS: autonomy satisfaction.

[c]CS: competence satisfaction.

[d]RS: relatedness satisfaction.

Table 7. Differences between gaming modality groups on the scores of in-game need satisfaction (NS) and in-game need frustration (NF) scales.

	Modality		*t* test (*df*)	*P* value[a]	95% CI	*g*
	Online (388/511), mean (SD)	Offline (123/511), mean (SD)				
AS[b]	6.13 (1.68)	5.79 (1.89)	1.91 (590)	.05	–0.01 to 0.7	0.2
CS[c]	5.94 (1.63)	5.27 (1.8)	3.67 (189.65)	<.001	0.33 to 1.01	0.4
RS[d]	4.82 (2.27)	3.25 (2.55)	6.1 (186.34)	<.001	1.06 to 2.08	0.67
NS	18.81 (4.58)	16 (5.51)	5.126 (178.47)	<.001	1.73 to 3.89	0.58
NF	6.15 (5.02)	5.27 (4.97)	1.7 (509)	.09	–0.14 to 1.9	0.18

[a]The *P* value reflects the 2-tailed hypothesis test.

[b]AS: autonomy satisfaction.

[c]CS: competence satisfaction.

[d]RS: relatedness satisfaction.

Psychometric Evaluation of the Constructs

To ensure the psychometric properties of the scale's factors, we analyzed the main reliability and validity indicators (Table 8) and their relationship with gaming patterns and GD (Table 9). Item correlation analyses revealed higher correlations between items within the same factor than between items of different factors for the NS scale, but not for the NF scale. This supports the multifactorial and single-factor structures, respectively. Both scales showed adequate discriminant validity through the squared correlational values (r^2) between factors (autonomy-competence=0.179; autonomy-relatedness=0.084; competence-relatedness=0.169 [NS]; NS-NF=0.012).

XSL•FO
RenderX

Table 8. Psychometric study of in-game psychological needs factors.

	NS[a]				NF[b]
	AS[c,d]	CS[e,f]	RS[g,h]	TS[i,j]	TS[k]
AS					
r[l] (95% CI)	1	0.47 (0.43 to 0.52)	0.32 (0.27 to 0.37)	0.76 (0.73 to 0.78)	–0.076 (–0.133 to –0.019)
P value	>.99	<.001	<.001	<.001	.009
CS					
r (95% CI)	0.47 (0.43 to 0.52)	1	0.42 (0.37 to 0.47)	0.78 (0.76 to 0.81)	0.061 (0.004 to 0.118)
P value	<.001	>.99	<.001	<.001	.04
RS					
r (95% CI)	0.32 (0.27 to 0.37)	0.42 (0.37 to 0.47)	1	0.78 (0.76 to 0.81)	0.219 (0.164 to 0.273)
P value	<.001	<.001	>.99	<.001	<.001
TS					
r (95% CI)	0.76 (0.73 to 0.78)	0.78 (0.76 to 0.81)	0.78 (0.76 to 0.81)	1	0.098 (0.41 to 0.154)
P value	<.001	<.001	<.001	>.99	<.001

[a]NS: in-game need satisfaction.

[b]NF: in-game need frustration.

[c]AS: autonomy satisfaction.

[d]Cronbach α=0.58; composite reliability=0.58; average variance extracted=0.49; R^2=0.62.

[e]CS: competence satisfaction.

[f]Cronbach α=0.66; composite reliability=0.73; average variance extracted=0.58; R^2=0.89.

[g]RS: relatedness satisfaction.

[h]Cronbach α=0.75; composite reliability=0.68; average variance extracted=0.64; R^2=0.4.

[i]TS: total score.

[j]Cronbach α=0.75; composite reliability=0.82; average variance extracted=0.57.

[k]Cronbach α=0.77; composite reliability=0.83; average variance extracted=0.41.

[l]*r*: Pearson bilateral correlations.

XSL•FO
RenderX

Table 9. Associations between in-game psychological needs and patterns of gaming use and problematic behavior.

	NS[a]				NF[b]
	AS[c]	CS[d]	RS[e]	TS[f]	TS
MGT [g]					
r[h] (95% CI)	0.11 (0.05-0.17)	0.18 (0.13-0.24)	0.15 (0.09-0.24)	0.19 (0.13-0.24)	0.12 (0.06-0.17)
P value	<.001	<.001	<.001	<.001	<.001
WGT [i]					
r (95% CI)	0.25 (0.17-0.28)	0.31 (0.25-0.36)	0.3 (0.25-0.036)	0.36 (0.31-0.41)	0.08 (0.03-0.14)
P value	<.001	<.001	<.001	<.001	.005
WGF [j]					
r (95% CI)	0.19 (0.14-0.25)	0.24 (0.18-0.29)	0.16 (0.1-0.21)	0.25 (0.19-0.3)	0.1 (0.043-0.156)
P value	<.001	<.001	<.001	<.001	<.001
TDV [k]					
r (95% CI)	0.14 (0.08-0.2)	0.3 (0.25-0.35)	0.35 (0.29-0.4)	0.34 (0.29-0.39)	0.49 (0.44-0.53)
P value	<.001	<.001	<.001	<.001	<.001

[a]NS: in-game need satisfaction.
[b]NF: in-game need frustration.
[c]AS: autonomy satisfaction.
[d]CS: competence satisfaction.
[e]RS: relatedness satisfaction.
[f]TS: total score.
[g]MGT: midweek gaming daily time.
[h]*r*: Fisher bilateral bivariate correlation.
[i]WGT: weekend gaming daily time.
[j]WGF: weekly gaming frequency.
[k]TDV: Video Game Dependency Test.

Furthermore, except for the autonomy satisfaction factor, all the factors of the scales presented good convergent validity indices. Specifically, the NS factors obtained higher values compared to the NF factors. It is worth noting that all the constructs obtained good validity results represented by AVE and R^2 (except for NF's AVE and NS-relatedness for R^2).

In addition, internal consistency analysis indicated that the different factors of NS and NF assess reliably and accurately (Table 8). However, the results of the autonomy satisfaction factor were the only results that did not exceed the α and CR cutoff point.

All items of both scales showed high reliability with CITC values above 0.3. In fact, CITC values ranged from 0.413 to 0.56, which showed that the totality of the items possessed excellent representativeness of the construct to which they belong. The R^2 and standardized error values of items 1 (R^2=0.273; ε=0.73), 3 (R^2=0.355; ε=0.65), and 6 (R^2=0.307; ε=0.69) of the NF scale and item 6 of the NS scale (R^2=0.409; ε=0.59) did not show as adequate a measurement as the rest of the items; however, the elimination of any 1 of them would mean a decrease in overall reliability (NS: Cronbach α<0.749 and NF: Cronbach α<0.766).

After testing the psychometric characteristics of the scales, the relationship between NS and NF factors (Table 8), as well as with gaming patterns, was analyzed (Table 9). On the one hand, the NS subdimensions showed strong correlations with the total NS score. Although statistically significant, the correlation between NS and NF was very low. Among the NS dimensions, relatedness satisfaction showed the strongest association with frustration.

On the other hand, the relationship between needs and gaming patterns showed moderate correlations in the case of NS (*r*=0.187-0.359) and low or even null correlations in the case of NF (*r*=0.084-0.115). Weekend gaming time showed the highest correlations, followed by weekly frequency and weekday gaming time with NS and its first-order factors. Specifically, competence exhibited the strongest correlations (*r*=0.182-0.308).

Regarding GD, its strongest correlation was observed with NF, whereas NS also showed a significant but lower association. Notably, the correlation between NS and GD was slightly weaker than the correlation between NS and weekend gaming time, suggesting that satisfaction is more strongly associated with gaming engagement than with problematic use. In contrast, NF showed very low correlations with gaming patterns, reinforcing its specific association with GD rather than general

XSL•FO
RenderX

gaming behavior. Among the satisfied needs, relatedness exhibited the strongest association with GD, followed by competence and autonomy.

Discussion

Principal Findings

This study aimed to adapt and validate the BPNSFS-G into Spanish for adolescents, resulting in the development of the Youth Gaming Experience Scales (in Spanish "Escalas para Jóvenes sobre la Experiencia de Juego" or EJEJ), designed to independently assess NS (Youth Satisfying Gaming Experience Scale [EJEJ-S]) and NF (Youth Frustrating Gaming Experience Scale [EJEJ-F]). The results supported the validity and reliability properties of both scales, showing that NF was more strongly associated with GD, while NS was more closely linked to gaming time. This pattern challenges the need density hypothesis capacity to explain GD by highlighting the central role of NF. Furthermore, for the first time, measurement invariance analyses enabled the empirical exploration of differences in these experiences according to personal and contextual factors widely recognized as GD risk markers. Overall, the EJEJs show both research and applied utility, contributing to the study, evaluation, early detection, and intervention design of GD from an SDT framework.

Factorial analyses compared several models proposed in the literature regarding the internal structure of psychological needs in gaming, including the original BPNSFS model, which treats both NS and NF as either first-order [35] or second-order factors [28]. None of these structures demonstrated satisfactory fit. Instead, a more parsimonious model based on 2 independent scales was confirmed, as suggested by the EFA results. Accordingly, 2 separate scales were defined to assess NS and NF. This model resembles the structure of the PENS scale [16,28] and supports the theoretical distinction between NS and NF as separate constructs, rather than as opposite ends of a single continuum [26,52]. Recent evidence supports this distinction, indicating that both dimensions may coexist within the gaming experience [19,24,28]. While NS and NF are typically negatively correlated in real-life contexts [35], this association does not seem to hold in gaming environments [25,28]. These findings suggest that psychological needs in video games may be simultaneously satisfied and frustrated [19].

The EJEJ-F showed a unidimensional structure, failing to differentiate between the frustration of autonomy, competence, and relatedness needs. This contrasts with the NS dimension in our study, which did show a clear 3-factor structure, in line with previous findings [28,37]. From a methodological perspective, this outcome may be influenced by factors such as the linguistic adaptation, response format, or the developmental characteristics of the adolescent sample [41]. Theoretically, it may also reflect the still-limited conceptual development of NF in gaming research. While recent models have proposed a dimensional structure for NF analogous to NS [27], our findings did not support this configuration. Nonetheless, both scales demonstrated adequate internal consistency, consistent with prior research [28], supporting the reliability of the instrument for assessing the overall construct of NF, as well as the global and specific dimensions of NS.

Regarding validity, the EJEJ-S and EJEJ-F showed distinct patterns of association with external gaming-related variables, supporting the questionnaire's convergent and criterion validity. In line with the hypotheses, NF showed a stronger association with GD, while NS was more closely related to gaming time and frequency. This pattern aligns with SDT, where NF is conceptualized as a form of controlled motivation—driven by rigid and external regulation—linked to compulsive and dysregulated gaming behavior [27,32], while NS is associated with more autonomous and intrinsically regulated gaming, typically reflected in positive engagement indicators such as gaming time [16,29-31]. These findings highlight the EJEJ's sensitivity to differentiate between adaptive and maladaptive gaming.

These findings, along with previous studies questioning the direct link between NS and GD [18,22], raise 2 important implications. First, the need density hypothesis [17], which suggests that GD results from an imbalance between real-life need satisfaction and NS, is called into question. Our results challenge the core assumptions of this hypothesis, highlighting NF—rather than NS—as a stronger and independent indicator of GD [22,23] in both adolescents and adults [25,28].

Second, the weak association observed between NF and gaming time contributes to the growing evidence challenging the use of gaming time as a primary diagnostic criterion for GD [53,54]. While gaming time has traditionally been viewed as a sign of "tolerance" in addiction models [6,55], several studies have questioned its validity as a standalone marker of pathology [12,56,57]. Our findings support this perspective, suggesting that gaming time may reflect normative engagement more than dysregulated use and is likely modulated by motivational factors—such as NS—or by broader regulatory processes [22,23,58,59]. This reinforces recent proposals to integrate motivational variables like NF into diagnostic models to better distinguish between intensive, but healthy, and truly problematic gaming [19,56,60].

In preventive contexts—especially within school-based programs [61]—these indicators can help detect motivational patterns that precede the development of GD. This supports the need for interventions that go beyond limiting gaming time and promote healthier psychological engagement [62], for which SDT has proven to be an effective framework in both clinical and educational contexts [63-65]. Given their sensitivity to problematic gaming patterns, the EJEJs may contribute to early risk detection and to the development of SDT-based interventions.

Supporting this applied potential, the measurement invariance analyses confirmed the factorial equivalence of the EJEJ across sex and the adolescence stage—2 key variables consistently identified as risk factors for GD [11,12]. Previous studies have explored psychological NS and NF across gaming characteristics, such as gaming genres in the BPNSFS-G [28] or level of generality of the need experience (game session, general gaming, and particular session play) in the Basic Needs in Games Scale [37]. However, to our knowledge, no prior

XSL•FO
RenderX

research has examined these experiences in relation to personal variables like the adolescent stage or sex.

These results offer novel exploratory insights into the in-game experience of psychological needs among gamer adolescent subpopulations. Younger adolescents reported higher levels of both NS and NF. This greater experiential intensity may be explained by a stronger emotional involvement with gaming at earlier developmental stages [66], as well as by an increased sensitivity to the psychological experience of NS and NF, as described in SDT for this life stage [24]. Regarding sex, male participants scored higher on NS [67], with no significant differences found in NF. Although male sex is widely recognized as a key risk factor for GD [11,12], these results suggest that NF may represent a shared vulnerability mechanism across sex, which could depend on moderating variables such as gaming motivation (eg, socializing [68]) or coping style [58,59,69]. These findings underscore the need for further research into sex-related differences in the experience of NS and NF [16,22], for which the EJEJ may offer a valuable assessment tool.

In contrast, invariance was not observed across gaming modality (online vs offline), which limits direct comparisons between these groups and suggests differences in how gamers experience their psychological needs depending on the gaming modality [30]. Notably, significant differences were found in several NS dimensions across modalities (except for autonomy), with higher scores among online gamers. This could be especially relevant for the relatedness need [16], given the greater potential of online games for real interpersonal interaction, compared to offline games, which typically involve nonhuman characters. These findings highlight the need to explore how the gaming modality affects the experience and assessment of psychological needs in different gaming contexts. Previous instruments, such as the PENS [16], did not differentiate between online and offline contexts, leading to interpretive ambiguities. More recent tools like the BPNSFS-G [28] have been developed specifically for online gaming. However, the present results suggest that the validity of these assessments may not generalize across modalities, reinforcing the need for modality-specific validation.

Limitations

This study presents several limitations that should be considered in future research. First, the cross-sectional design limits the ability to establish causal relationships. Second, although psychometric quality was supported by multivariate analyses, future studies should implement regression models to examine predictive validity. Third, the use of a convenience sample limits the generalizability of the findings, highlighting the need for future studies to include diverse cultural contexts and settings (eg, different intervention programs). Fourth, longitudinal designs are needed to assess the direction and stability of the associations between in-game needs and GD. Fifth, test-retest reliability studies should be included to evaluate the temporal consistency of the scores, particularly in the context of intervention and prevention. Finally, the lack of invariance suggests the need to develop adapted versions of the instrument for different gaming modalities.

Additionally, the brief format is both a strength—due to its practical application—and a limitation in theoretical research. Future studies should examine the specific dynamics of autonomy, competence, and relatedness in both NS and NF [14,16,27,30], a goal for which the EJEJ-S could be appropriate but not the EJEJ-F due to its unidimensional structure. Furthermore, future research should explore the relationship between in-game and real-life need experiences across adolescent populations [25].

Conclusions

This study provides the first validation of the Spanish version of the BPNSFS-G for adolescent populations, leading to the development of EJEJ-S and EJEJ-F. The psychometric properties of these scales were thoroughly supported, demonstrating their reliability and validity across various adolescent subpopulations, including different GD risk profiles. This robustness supports their applicability to diverse contexts, suggesting their potential utility in both intervention and prevention contexts.

By highlighting NF as a key risk factor for GD, this study reinforces the theoretical value of SDT in understanding the etiology of GD in youths. NF may serve as a predictive indicator of problematic gaming and help differentiate it from normative use, highlighting its potential for early detection and intervention. These findings underscore the relevance of SDT-based approaches for developing prevention programs that promote healthy gaming behaviors.

Although the sample's representativeness limits the generalizability of the findings, the successful adaptation of the questionnaire broadens its applicability to a wider range of populations. Future research should include longitudinal designs, regression models, and test-retest reliability analyses to improve the predictive validity and temporal consistency of the scales. In conclusion, the EJEJs have strong potential to advance both the theoretical understanding and practical application of interventions for GD within the SDT framework.

Acknowledgments

The authors would like to thank the people responsible for managing and running the Gamer program in the Addictions Service (PMD) of the Council of Valencia, Manuel Gómez Tejedor and Rocío Benito Corchete. This work has been done thanks to the cooperation agreement between the Addictions Service of the Council of Valencia and the Gambling and Technological Addictions Research Unit of the Universidad de Valencia in the framework of technology and gambling addiction prevention.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Original version of Youth Gaming Experience Scales.
[DOCX File , 18 KB - games_v13i1e63612_app1.docx]

References

1. Duarte F. How many gamers are there? (New 2025 statistics). Exploding Topics. 2024. URL: https://explodingtopics.com/blog/number-of-gamers [accessed 2025-02-20]
2. Asociación Española de Videojuegos. La industria del videojuego en España en 2021. Anuario. 2021. URL: https://aevi.org.es/videojuego-facturo-1-795-millones-euros-2021-una-base-superior-los-18-millones-usuarios-espana/ [accessed 2023-09-22]
3. Informe sobre: trastornos comportamentales 2022: juego con dinero, uso de videojuegos y uso compulsivo de internet en las encuestas de drogas y otras adicciones en España EDADES y ESTUDES. Observatorio Español de las Drogas y las Adicciones. Ministerio de Sanidad, Delegación del Gobierno para el Plan Nacional sobre Drogas. 2022. URL: https://pnsd.sanidad.gob.es/profesionales/sistemasInformacion/sistemaInformacion/pdf/2022_Informe_Trastornos_Comportamentales.pdf [accessed 2023-12-18]
4. 2020 essential facts about the video game industry. Entertainment Software Association. 2020. URL: https://www.theesa.com/wp-content/uploads/2024/03/Final-Edited-2020-ESA_Essential_facts.pdf [accessed 2023-09-22]
5. Nakamura J, Csikszentmihalyi M. Flow theory and research. In: Oxford Handbook of Positive Psychology. Oxford, MS: Oxford University Press; 2009.
6. Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, Text Revision (DSM-5-TR). Arlington, VA: American Psychiatric Association Publishing; 2022.
7. International Classification of Diseases, 11th Revision. World Health Organization. 2018. URL: https://icd.who.int/browse11/lm/en#/http://id.who.int/icd/entity/1448597234 [accessed 2023-12-16]
8. Anthony WL, Mills DJ, Nower L. Evaluation of the psychometric properties of DSM–5 internet gaming disorder measures: a COSMIN systematic review and meta-analysis. Clin Psychol Sci Pract 2023;30(2):170-185. [doi: 10.1037/cps0000123]
9. Stevens MW, Dorstyn D, Delfabbro PH, King DL. Global prevalence of gaming disorder: a systematic review and meta-analysis. Aust N Z J Psychiatry 2021;55(6):553-568. [doi: 10.1177/0004867420962851] [Medline: 33028074]
10. Fam JY. Prevalence of internet gaming disorder in adolescents: a meta-analysis across three decades. Scand J Psychol 2018;59(5):524-531. [doi: 10.1111/sjop.12459] [Medline: 30004118]
11. Mihara S, Higuchi S. Cross-sectional and longitudinal epidemiological studies of Internet gaming disorder: a systematic review of the literature. Psychiatry Clin Neurosci 2017;71(7):425-444 [FREE Full text] [doi: 10.1111/pcn.12532] [Medline: 28436212]
12. Király O, Koncz P, Griffiths MD, Demetrovics Z. Gaming disorder: a summary of its characteristics and aetiology. Compr Psychiatry 2023;122:152376 [FREE Full text] [doi: 10.1016/j.comppsych.2023.152376] [Medline: 36764098]
13. Bäcklund C, Elbe P, Gavelin HM, Sörman DE, Ljungberg JK. Gaming motivations and gaming disorder symptoms: a systematic review and meta-analysis. J Behav Addict 2022;11(3):667-688 [FREE Full text] [doi: 10.1556/2006.2022.00053] [Medline: 36094861]
14. Ryan RM, Deci EL. Self-Determination Theory: Basic Psychological Needs in Motivation, Development, and Wellness. New York, NY: Guilford Press; 2017:1462528767.
15. Rosendo-Rios V, Trott S, Shukla P. Systematic literature review online gaming addiction among children and young adults: a framework and research agenda. Addict Behav 2022;129:107238. [doi: 10.1016/j.addbeh.2022.107238] [Medline: 35104738]
16. Ryan RM, Rigby CS, Przybylski A. The motivational pull of video games: a self-determination theory approach. Motiv Emot 2006;30(4):344-360. [doi: 10.1007/s11031-006-9051-8]
17. Rigby S, Ryan R. Glued to Games. London: Bloomsbury; 2011:0313362246.
18. Bender PK, Gentile DA. Internet gaming disorder: relations between needs satisfaction in-game and in life in general. Psychol Pop Media 2020;9(2):266-278. [doi: 10.1037/ppm0000227]
19. Ballou N, Deterding S, Iacovides I, Helsby L. Do people use games to compensate for psychological needs during crises? A mixed-methods study of gaming during COVID-19 lockdowns. 2022 Presented at: CHI '22: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems; April 29-May 5, 2022; New Orleans, LA, United States p. 1-15. [doi: https://doi.org/10.1145/3491102.3501858]
20. Bradt L, Vermote B, Zaman B, Vansteenkiste M, Soenens B. Are video games and school conflictual or complementary contexts for affording psychological need fulfillment? Implications for adolescents' problematic gaming and school adjustment. Interact Comput 2024:iwae020 [FREE Full text] [doi: 10.1093/iwc/iwae020]

XSL•FO
RenderX

21. Przybylski AK, Weinstein N. Investigating the motivational and psychosocial dynamics of dysregulated gaming: evidence from a preregistered cohort study. Clin Psychol Sci 2019;7(6):1257-1265. [doi: 10.1177/2167702619859341]
22. Mills DJ, Milyavskaya M, Mettler J, Heath NL, Derevensky JL. How do passion for video games and needs frustration explain time spent gaming? Br J Soc Psychol 2018;57(2):461-481. [doi: 10.1111/bjso.12239] [Medline: 29352471]
23. Mills DJ, Milyavskaya M, Mettler J, Heath NL. Exploring the pull and push underlying problem video game use: a self-determination theory approach. Pers Individ Dif 2018;135:176-181. [doi: 10.1016/j.paid.2018.07.007]
24. Vansteenkiste M, Ryan RM, Soenens B. Basic psychological need theory: advancements, critical themes, and future directions. Motiv Emot 2020;44(1):1-31. [doi: 10.1007/s11031-019-09818-1]
25. Allen JJ, Anderson CA. Satisfaction and frustration of basic psychological needs in the real world and in video games predict internet gaming disorder scores and well-being. Comput Human Behav 2018;84:220-229. [doi: 10.1016/j.chb.2018.02.034]
26. Bartholomew KJ, Ntoumanis N, Ryan RM, Bosch JA, Thøgersen-Ntoumani C. Self-determination theory and diminished functioning: the role of interpersonal control and psychological need thwarting. Pers Soc Psychol Bull 2011;37(11):1459-1473. [doi: 10.1177/0146167211413125] [Medline: 21700794]
27. Ballou N, Deterding S. 'I Just Wanted to Get It Over and Done With': a grounded theory of psychological need frustration in video games. 2023 Presented at: Proceedings of the ACM on Human-Computer Interaction, Volume 7, Issue CHI PLAY; October 4, 2023; New York, NY, United States p. 217-236. [doi: 10.1145/3611028]
28. Kosa M, Uysal A. Need frustration in online video games. Behav Inf Technol 2021;41(11):2415-2426. [doi: 10.1080/0144929x.2021.1928753]
29. Johnson D, Gardner J, Sweetser P. Motivations for videogame play: predictors of time spent playing. Comput Human Behav 2016;63:805-812. [doi: 10.1016/j.chb.2016.06.028]
30. Przybylski AK, Rigby CS, Ryan RM. A motivational model of video game engagement. Rev Gen Psychol 2010;14(2):154-166. [doi: 10.1037/a0019440]
31. Vella K, Johnson D, Hides L. Playing alone, playing with others: differences in player experience and indicators of wellbeing. 2015 Presented at: CHI PLAY '15: Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play; October 5-7, 2015; London, United Kingdom p. 3-12. [doi: https://doi.org/10.1145/2793107.2793118]
32. Holding AC, Verner-Filion J, Lalande D, Schellenberg BJI, Vallerand RJ. The roles of need satisfaction and passion in symptoms of behavioral addiction: the case of video gaming and gambling. Motiv Sci 2021;7(3):345-355. [doi: 10.1037/mot0000241]
33. Johnson D, Gardner MJ, Perry R. Validation of two game experience scales: the Player Experience of Need Satisfaction (PENS) and Game Experience Questionnaire (GEQ). Int J Hum Comput Stud 2018;118:38-46. [doi: 10.1016/j.ijhcs.2018.05.003]
34. Azadvar A, Canossa A. UPEQ: ubisoft perceived experience questionnaire: a self-determination evaluation tool for video games. 2018 Presented at: FDG '18: Proceedings of the 13th International Conference on the Foundations of Digital Games; August 7, 2018; Malmö, Sweden. [doi: https://doi.org/10.1145/3235765.3235780]
35. Chen B, Vansteenkiste M, Beyers W, Boone L, Deci EL, Van der Kaap-Deeder J, et al. Basic psychological need satisfaction, need frustration, and need strength across four cultures. Motiv Emot 2014;39(2):216-236. [doi: 10.1007/s11031-014-9450-1]
36. Chamarro A, Oberst U, Cladellas R, Fuster H. Effect of the frustration of psychological needs on addictive behaviors in mobile videogamers—the mediating role of use expectancies and time spent gaming. Int J Environ Res Public Health 2020;17(17):6429 [FREE Full text] [doi: 10.3390/ijerph17176429] [Medline: 32899364]
37. Ballou N, Denisova A, Ryan R, Rigby CS, Deterding S. The Basic Needs in Games Scale (BANGS): a new tool for investigating positive and negative video game experiences. Int J Hum Comput Stud 2024;188:103289. [doi: 10.1016/j.ijhcs.2024.103289]
38. Luján-Barrera A, Cervera-Ortíz L, Peláez A, Chóliz M. Eficacia del programa escolar gamer: características y predictores de la prevención de la adicción a videojuegos en adolescentes [Effects of the gamer school program: characteristics and predictors of gaming addiction prevention in adolescents]. ResearchGate Preprint posted online on September 2024 [FREE Full text] [doi: 10.13140/RG.2.2.15363.52001]
39. Stages of adolescence. Healthy Children. American Academy of Pediatrics. 2019. URL: https://www.healthychildren.org/english/ages-stages/teen/pages/stages-of-adolescence.aspx [accessed 2024-12-10]
40. Hambleton RK, Lee MK. Methods for translating and adapting tests to increase cross-language validity. In: The Oxford Handbook of Child Psychological Assessment. Oxford: Oxford University Press; 2013:172-181.
41. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. Spine (Phila Pa 1976) 2000;25(24):3186-3191. [doi: 10.1097/00007632-200012150-00014] [Medline: 11124735]
42. Revilla MA, Saris WE, Krosnick JA. Choosing the number of categories in agree–disagree scales. Sociol Methods Res 2013;43(1):73-97. [doi: 10.1177/0049124113509605]
43. Mellor D, Moore KA. The use of Likert scales with children. J Pediatr Psychol 2014;39(3):369-379. [doi: 10.1093/jpepsy/jst079] [Medline: 24163438]
44. Chóliz M, Marco C. Patterns of video game use and dependency in children and adolescents. Ann Psychol 2011;27(2):418-426 [FREE Full text]

XSL•FO
RenderX

45. Herdoiza-Arroyo P. Trastorno por videojuegos en adolescentes iberoamericanos: análisis de las dimensiones de impulsividad [dissertation]. University of Valencia. 2021. URL: https://www.educacion.gob.es/teseo/mostrarRef.do?ref=1944342 [accessed 2024-01-04]
46. Mindrila D. Maximum likelihood (ML) and diagonally weighted least squares (DWLS) estimation procedures: a comparison of estimation bias with ordinal and multivariate non-normal data. Int J Digit Soc 2010;1(1):60-66 [FREE Full text]
47. Kline RB. Promise and pitfalls of structural equation modeling in gifted research. In: Methodologies for Conducting Research on Giftedness. Washington, DC: American Psychological Association; 2010:147-169.
48. Byrne BM. Structural Equation Modeling With AMOS: Basic Concepts, Applications, and Programming. 2nd Edition. London: Routledge; 2013.
49. Wang YS, Wang HY, Shee DY. Measuring e-learning systems success in an organizational context: scale development and validation. Comput Human Behav 2007;23(4):1792-1808. [doi: 10.1016/j.chb.2005.10.006]
50. Cheung GW, Wang C. Current approaches for assessing convergent and discriminant validity with SEM: issues and solutions. Acad Manag Proc 2017;2017(1):12706. [doi: 10.5465/AMBPP.2017.12706abstract]
51. Hair JF. Multivariate Data Analysis. 8th Edition. Andover, Hampshire: Cengage; 2019.
52. Sheldon KM, Hilpert JC. The Balanced Measure of Psychological Needs (BMPN) scale: an alternative domain general measure of need satisfaction. Motiv Emot 2012;36(4):439-451. [doi: 10.1007/s11031-012-9279-4]
53. Király O, Sleczka P, Pontes HM, Urbán R, Griffiths MD, Demetrovics Z. Validation of the Ten-Item Internet Gaming Disorder Test (IGDT-10) and evaluation of the nine DSM-5 Internet Gaming Disorder criteria. Addict Behav 2017;64:253-260. [doi: 10.1016/j.addbeh.2015.11.005] [Medline: 26632194]
54. Przybylski AK, Weinstein N, Murayama K. Internet gaming disorder: investigating the clinical relevance of a new phenomenon. Am J Psychiatry 2017;174(3):230-236 [FREE Full text] [doi: 10.1176/appi.ajp.2016.16020224] [Medline: 27809571]
55. King DL, Herd MC, Delfabbro PH. Motivational components of tolerance in internet gaming disorder. Comput Human Behav 2018;78:133-141. [doi: 10.1016/j.chb.2017.09.023]
56. Charlton JP, Danforth ID. Distinguishing addiction and high engagement in the context of online game playing. Comput Human Behav 2007;23(3):1531-1548. [doi: 10.1016/j.chb.2005.07.002]
57. Griffiths M, Kuss D, Pontes H. A brief overview of internet gaming disorder and its treatment. Aust Clin Psychol 2016:20108-20108 [FREE Full text] [doi: 10.1017/9781108632591.029]
58. Kardefelt-Winther D. Conceptualizing Internet use disorders: addiction or coping process? Psychiatry Clin Neurosci 2017;71(7):459-466 [FREE Full text] [doi: 10.1111/pcn.12413] [Medline: 27278653]
59. Mills DJ, Allen JJ. Self-determination theory, internet gaming disorder, and the mediating role of self-control. Comput Human Behav 2020;105:106209. [doi: 10.1016/j.chb.2019.106209]
60. Kardefelt-Winther D, Heeren A, Schimmenti A, van Rooij A, Maurage P, Carras M, et al. How can we conceptualize behavioural addiction without pathologizing common behaviours? Addiction 2017;112(10):1709-1715 [FREE Full text] [doi: 10.1111/add.13763] [Medline: 28198052]
61. Throuvala MA, Griffiths MD, Rennoldson M, Kuss DJ. School-based prevention for adolescent internet addiction: prevention is the key. A systematic literature review. Curr Neuropharmacol 2019;17(6):507-525 [FREE Full text] [doi: 10.2174/1570159X16666180813153806] [Medline: 30101714]
62. Bonnaire C, Serehen Z, Phan O. Effects of a prevention intervention concerning screens, and video games in middle-school students: influences on beliefs and use. J Behav Addict 2019;8(3):537-553 [FREE Full text] [doi: 10.1556/2006.8.2019.54] [Medline: 31537087]
63. Gillison FB, Rouse P, Standage M, Sebire SJ, Ryan RM. A meta-analysis of techniques to promote motivation for health behaviour change from a self-determination theory perspective. Health Psychol Rev 2019;13(1):110-130. [doi: 10.1080/17437199.2018.1534071] [Medline: 30295176]
64. Li AY, Chau C, Cheng C. Development and validation of a parent-based program for preventing gaming disorder: the game over intervention. Int J Environ Res Public Health 2019;16(11):1984 [FREE Full text] [doi: 10.3390/ijerph16111984] [Medline: 31167457]
65. Sevinçli MC, Eğer Aydoğmuş M. Video games as a part of intervention programs based on self-determination theory. Psikiyatride Güncel Yaklaşımlar 2022;14(2):207-220. [doi: 10.18863/pgy.974968]
66. Granic I, Lobel A, Engels RCME. The benefits of playing video games. Am Psychol 2014;69(1):66-78. [doi: 10.1037/a0034857] [Medline: 24295515]
67. Lafrenière MAK, Vallerand RJ, Donahue EG, Lavigne GL. On the costs and benefits of gaming: the role of passion. Cyberpsychol Behav 2009;12(3):285-290. [doi: 10.1089/cpb.2008.0234] [Medline: 19366320]
68. Demetrovics Z, Urbán R, Nagygyörgy K, Farkas J, Zilahy D, Mervó B, et al. Why do you play? The development of the Motives for Online Gaming Questionnaire (MOGQ). Behav Res Methods 2011;43(3):814-825. [doi: 10.3758/s13428-011-0091-y] [Medline: 21487899]
69. T'ng ST, Ho KH, Pau K. Need frustration, gaming motives, and internet gaming disorder in mobile Multiplayer Online Battle Arena (MOBA) games: through the lens of self-determination theory. Int J Ment Health Addict 2022:1-21 [FREE Full text] [doi: 10.1007/s11469-022-00825-x] [Medline: 35497075]

XSL•FO
RenderX

Abbreviations

AVE: average variance extracted
BPNSFS: Basic Psychological Need Satisfaction Frustration Scale
BPNSFS-G: Basic Psychological Need Satisfaction Frustration Scale for Gaming
CFA: confirmatory factor analysis
CFI: comparative fit index
CITC: corrected item-total correlation
CR: composite reliability
DSM-5: Diagnostic and Statistical Manual of Mental Disorders (Fifth Edition)
DSM-5-TR: Diagnostic and Statistical Manual of Mental Disorders (Fifth Edition, Text Revision)
EFA: exploratory factor analysis
EJEJ: Youth Gaming Experience Scale
EJEJ-F: Youth Frustrating Gaming Experience Scale
EJEJ-S: Youth Satisfying Gaming Experience Scale
GD: gaming disorder
ICD-11: International Classification of Diseases, 11th Revision
NF: in-game need frustration
NS: in-game need satisfaction
PENS: Player Experience of Need Satisfaction
RMSEA: root-mean-square error of approximation
SDT: self-determination theory
SRMR: standardized root-mean-square residual

Edited by S Liu; submitted 25.06.24; peer-reviewed by P Sequeira, Y Feng, JE Teixeira, F Yang; comments to author 03.11.24; revised version received 10.05.25; accepted 20.07.25; published 12.11.25.

Please cite as:
Luján-Barrera A, Cervera-Ortiz L, Chóliz M
In-Game Need Satisfaction, Frustration, and Gaming Addiction Patterns Across Subgroups of Adolescents Through Structural Equation Modeling: Cross-Sectional and Instrument Validation Study of the Youth Gaming Experience Scales
JMIR Serious Games 2025;13:e63612
URL: https://games.jmir.org/2025/1/e63612
doi:10.2196/63612
PMID:41223412

©Amparo Luján-Barrera, Lydia Cervera-Ortiz, Mariano Chóliz. Originally published in JMIR Serious Games (https://games.jmir.org), 12.11.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Prevalence, Sex Differences, and Predictors of Internet Gaming Disorder Among Impoverished Rural Adolescents: Cross-Sectional and Prospective Cohort Study

Chenhan Wang[1*], MA; Yifan Li[2*], MA; Shuhong Lin[2], MA; Qiuping Huang[3], PhD; Yongyan Shi[2], MA; Guangxian Yang[1], PhD; Erjia Huang[4], BA; Xicheng Deng[1], PhD; Jinwen Luo[1], PhD

[1]The Affiliated Children's Hospital of Xiangya School of Medicine (Hunan Children's Hospital), Central South University, No. 86 Ziyuan Road, Yuhua District, Hunan Province, Changsha, China

[2]Department of Psychiatry, The Second Xiangya Hospital of Central South University, Changsha, China

[3]Department of Psychology, School of Humanities and Management, Hunan University of Chinese Medicine, Changsha, China

[4]Clinical Nursing Teaching and Research Section, The Second Xiangya Hospital, Central South University, Changsha, China

[*]these authors contributed equally

Corresponding Author:
Chenhan Wang, MA
The Affiliated Children's Hospital of Xiangya School of Medicine (Hunan Children's Hospital), Central South University, No. 86 Ziyuan Road, Yuhua District, Hunan Province, Changsha, China

Abstract

Background: Internet gaming disorder (IGD) is prevalent globally and linked to significant negative outcomes. Impoverished rural adolescents face unique risks due to limited supervision and unequal digital resources, with limited longitudinal research conducted in this population. Existing studies show sex differences in IGD prevalence, but their manifestations and mechanisms in rural populations remain unclear.

Objective: This is the first large-sample cross-sectional and prospective cohort study targeting impoverished adolescents in rural areas. It aimed to determine the prevalence of IGD among impoverished rural adolescents, identify sex-specific risk and protective factors, and evaluate the longitudinal predictors of IGD.

Methods: In this study, self-administered questionnaires were used to collect demographic characteristics, affective states, impulsivity, gaming time, and scores for IGD. First, the prevalence of IGD at baseline and follow-up, along with sex differences, was calculated. Correlation analysis was conducted to explore variables associated with IGD. Subsequently, multivariate logistic regression analysis was conducted to identify baseline and follow-up predictors of IGD.

Results: The cross-sectional analysis at baseline included 13,931 valid responses (IGD prevalence: n=725, 5.2%; males: 489/7304, 6.7%; females: 236/6627, 3.6%). A 1-year longitudinal follow-up showed IGD prevalence of 5% (692/13,931; males: 511/7304, 7.0%; females: 181/6627, 2.7%; after multiple imputations). Common baseline factors were gaming time (females: odds ratio [OR] 1.11, 95% CI 1.08-1.14, *P*<.001; males: OR 1.11, 95% CI 1.09-1.13, *P*<.001), self-esteem scores (females: OR 0.95, 95% CI 0.92-0.98, *P*=.002; males: OR 0.95, 95% CI 0.92-0.97, *P*<.001), depression scores (females: OR 1.14, 95% CI 1.11-1.16, *P*<.001; males: OR 1.11, 95% CI 1.09-1.13, *P*<.001), and impulsive behavior (females: OR 1.16, 95% CI 1.10-1.22, *P*<.001; males: OR 1.10, 95% CI 1.06-1.14, *P*<.001). Companionship (OR 0.71, 95% CI 0.52-0.97; *P*=.03) was a protective factor for females, while age (OR 1.08, 95% CI 1.02-1.15; *P*=.02) and poor self-regulation (OR 1.07, 95% CI 1.03-1.11; *P*=.001) posed extra risk for males at baseline. Longitudinal predictors were baseline gaming time (females: OR 1.06, 95% CI 1.03-1.09, *P*<.001; males: OR 1.02, 95% CI 1.00-1.05, *P*=.03) and impulsive behavior (females: OR 1.38, 95% CI 1.30-1.46, *P*<.001; males: OR 1.27, 95% CI 1.22-1.31, *P*<.001). Baseline companionship (OR 0.32, 95% CI 0.23-0.43; *P*<.001) was a protective factor for females, while baseline poor self-regulation (OR 1.32, 95% CI 1.27-1.37; *P*<.001) was a predictive factor for males after 1 year.

Conclusions: IGD prevalence was lower in rural than in urban populations and higher in males than in females. Impulsivity, gaming time, and guardian companionship showed sex differences: females relied more on companionship, whereas males were more vulnerable to poor self-regulation. Interventions should address these differences, strengthening family support and psychological adjustment. This study provides novel insights into sex-specific pathways of IGD in rural settings and offers empirical evidence for developing targeted prevention strategies, highlighting its practical significance for public health.

(JMIR Serious Games 2025;13:e83522) doi:10.2196/83522

XSL•FO
RenderX

KEYWORDS

internet gaming disorder; rural adolescents; prevalence; sex differences; longitudinal study; adolescent mental health

Introduction

In 2019, the World Health Organization officially included gaming disorder in the *International Classification of Diseases, 11th Revision*, identifying its core features as a shift in gaming priority, loss of behavioral control, and impairment in social functioning. According to the latest meta-analyses, the prevalence of internet gaming disorder (IGD) among adolescents in East Asia is approximately 6%, with a rate of approximately 6.4% reported in China [1,2].

To date, most research on IGD has focused on urban populations. However, children and adolescents in economically underdeveloped rural areas may face unique challenges such as limited parental supervision resources, lack of guidance on digital device use, and unequal distribution of digital resources between urban and rural areas. These environmental differences may contribute to distinct risk profiles for IGD among rural youth compared to their urban counterparts [3]. Given that IGD has been linked to negative outcomes, including poorer sleep quality, impaired affective regulation, decreased attention in learning, and increased risk of suicidal ideation [4-7], the underlying mechanisms of IGD may be closely associated with reward system dysfunction and impaired cognitive control abilities [8,9]. Thus, systematically assessing the epidemiology and risk mechanisms of IGD in rural areas holds great public health importance. Unfortunately, existing studies targeting rural populations are mostly cross-sectional, which only allow for identification of baseline associations but fall short of revealing temporal dynamics or causal relationships. For example, it remains unclear whether mental health issues lead to IGD, whether IGD triggers subsequent problems, or whether both exhibit a bidirectional relationship among rural adolescents. This limitation greatly restricts the applicability and scalability of current intervention strategies in rural contexts.

Considerable evidence indicates sex differences in the prevalence of IGD, with male individuals typically exhibiting higher rates than female individuals [10,11], but the underlying mechanisms remain insufficiently understood. Historically, higher prevalence in male individuals has been partly attributed to the predominance of male-targeted game designs that emphasize stimulation and control; however, the market is increasingly offering games designed to cater to female players, and the number of male and female gamers is now broadly comparable [12]. This indicates that differences in prevalence cannot be fully explained by game types and motivation design alone. Current research is shifting focus toward potential individual susceptibility factors. For instance, male individuals may exhibit greater vulnerability in impulsivity and reward sensitivity, whereas being female may be more closely associated with difficulties in emotional regulation [13,14]. Crucial questions that remain unanswered include whether the sex disparity observed in previous studies persists in rural samples, whether female students with subthreshold baseline gaming problems are at elevated risk of progression over time, and whether factors such as family support and affective symptoms exert differential effects by sex. These questions warrant targeted longitudinal investigation in rural school-aged populations to clarify sex-specific pathways and inform tailored preventive interventions.

This study aimed to determine the prevalence of IGD among impoverished rural primary and middle school students in China, identify sex-specific risk and protective factors, and evaluate longitudinal predictors of IGD over a 1-year follow-up period. We hypothesized that (1) the prevalence of IGD in this rural school-aged cohort would differ from findings in broader or urban-representative samples; (2) individuals with longer baseline gaming time, more severe depressive state, and higher impulsivity would be more likely to develop IGD subsequently; (3) higher baseline self-esteem and greater guardian or parental involvement would be associated with lower risk of IGD; and (4) predictive factor patterns would differ by sex. The findings of this longitudinal study are expected to provide empirical evidence to inform precise prevention and intervention strategies for IGD in rural schools.

Methods

Study Design and Data Collection Process

This study adopted a cross-sectional and prospective longitudinal cohort design and was conducted and reported in accordance with the Strengthening the Reporting of Observational Studies in Epidemiology statement [15]. The completed Strengthening the Reporting of Observational Studies in Epidemiology checklist is provided as Checklist 1. From February 2023 to July 2024, the baseline survey was conducted in primary and middle schools across 13 impoverished rural counties in Hunan province, as well as in the rural free clinic program of Hunan Children's Hospital. As data collection was bound to the free clinic schedule, no a priori sample size calculation was conducted; instead, all eligible students were enrolled, yielding 18,967 valid questionnaires. Baseline data were obtained via on-site paper questionnaires covering demographics, self-esteem, affective status, impulsivity, gaming time, and IGD.

Follow-up began in January 2024 among students who completed the baseline survey and whose schools arranged a second free clinic. Using the same questionnaire method, follow-up continued until June 2025, with 5254 responses collected. Loss to follow-up mainly reflected schools that had not yet organized a second clinic, resulting in a nonrandom attrition pattern. Identity checks were used to ensure accurate matching of baseline and follow-up data.

Participants

Participants were recruited through their respective schools to respond to on-site surveys by distributing questionnaires collectively in class groups during the free medical consultation event.

The inclusion criteria for this study were age between 6 and 18 years; long-term residence in impoverished rural areas of Hunan

XSL•FO
RenderX

province; regular attendance to rural primary or secondary schools; and voluntary participation of both the student and their primary guardian, with signed informed consent.

Exclusion criteria included the presence of cognitive or language impairments preventing comprehension or completion of the questionnaire, temporary attendance to rural schools for less than 6 months, and refusal to participate or withdrawal during questionnaire completion for personal reasons.

Measurements

A self-developed questionnaire was used to collect demographic information, including age, sex, years of education, only child status, and daily time spent playing internet games (hours per day). The questionnaire also assessed the quality of the relationship between the participants and their primary guardians (rated as very poor, poor, good, or very good), as well as whether the primary guardian was able to provide companionship when needed (hereinafter referred to as *companionship*).

The Chinese version of the Rosenberg Self-Esteem Scale was used to assess overall self-esteem. This scale consists of 10 items and uses a 4-point Likert scale for scoring, with higher scores indicating higher levels of overall self-esteem. The scale's Cronbach α was 0.88 [16]. The Patient Health Questionnaire–9 (PHQ-9) was used to assess depressive mood. This scale includes 9 items and uses a 4-point Likert scale for scoring, with higher total scores reflecting more severe depressive mood. The scale's Cronbach α was 0.90 [17,18]. The Generalized Anxiety Disorder–7 (GAD-7) scale was used to assess anxious mood. This scale comprises 7 items and uses a 4-point Likert scale for scoring, with higher total scores reflecting more severe anxious mood. The scale's Cronbach α was 0.92 [19]. The Brief Barratt Impulsiveness Scale (BBIS) was used to assess impulsivity. This scale includes 8 items and uses a 4-point Likert scale for scoring. The scale covers 2 dimensions: poor self-regulation and impulsive behavior. Higher scores indicate greater impulsivity. The Cronbach α was 0.85 for the total scale and 0.88 and 0.80 for the 2 subscales [20]. The Internet Gaming Disorder Scale–Short Form (IGDS9-SF) was used to assess IGD. The scale contains 9 items and uses a 5-point Likert scale for scoring, with higher scores indicating more severe IGD symptoms. The scale's Cronbach α was 0.91 [21]. In this study, participants scoring >32 on the IGDS9-SF were classified as having IGD.

Quality Control

Before the assessments at each school, 5 professional psychiatrists provided standardized training to the teachers. The training covered the study objectives, interpretation of questionnaire items, informed consent procedures, on-site guidance protocols, and handling of abnormal situations, enabling teachers to assist psychiatrists in guiding students during questionnaire completion.

To ensure data quality, questionnaires were excluded if they met any of the following criteria: (1) careless responding—defined as 3 or more scales having all valid items scored identically or exhibition of systematic repetitive response patterns (eg, 1-2-3-4, 4-3-2-1, or 1-2-1) on any core scale; (2) failed validity check items—the questionnaires included lie detection items requiring the response "most of the time," and deviations from this response were deemed invalid; (3) implausible key information (age outside the range of 6 to 18 years; gaming time outside 0 to 24 hours; missing data for age, sex, years of education, or gaming time; or inconsistency between age and years of education); and (4) more than 10% of items left unanswered.

A total of 6452 invalid questionnaires were excluded. Data entry was conducted using EpiData (version 3.1; EpiData Association). To ensure accuracy, 30% of entered questionnaires were randomly rechecked, maintaining an error rate below 1%.

Handling of Missing Data

As the follow-up survey of this study only targeted students whose schools subsequently arranged a second free medical consultation, there were cases of missing data. To assess the mechanism of missing data, this study adopted the Little missing completely at random (MCAR) test. If the *P* value of this test is less than .05, the null hypothesis that the data are MCAR is rejected. On this basis, to further use all baseline information and reduce potential bias, this study used the *mice* package in the R language (R Foundation for Statistical Computing) to perform multiple imputation for the missing data. All statistical analyses related to the status of IGD during the follow-up period—including IGD prevalence rate, persistence rate, and the longitudinal predictors of IGD—were conducted based on the multiple imputation datasets, whereas the cross-sectional analyses involving only baseline data were completed using the original observed data.

Statistical Analysis

Statistical analyses were conducted using SPSS (version 26.0; IBM Corp). The prevalence of IGD was calculated for the total sample and by sex at baseline and follow-up. The Kolmogorov-Smirnov test was used to assess the normality of the continuous variables. Categorical variables (eg, only child status and relationship with guardian) were analyzed using chi-square tests. Normally distributed continuous variables (eg, age and PHQ-9 scores) were compared using independent-sample *t* tests, whereas nonnormally distributed variables were analyzed using the Mann-Whitney *U* test. Furthermore, the false discovery rate (FDR) multiple-testing (Benjamini-Hochberg) correction method was used to control the false positive rate of differences in variable comparisons between different-sex individuals at baseline, and a postcorrection *P* value of less than .05 was considered statistically significant. The Spearman rank correlation was used to examine associations between IGD indicators and other variables. Multivariate logistic regression was performed to identify factors associated with baseline IGD and predict follow-up IGD. First, baseline IGD status was set as the dependent variable, with variables showing significance in univariate analyses included as independent variables to identify associated factors. Next, follow-up IGD status was set as the dependent variable, with baseline IGD and its independent associated factors included as predictors to explore risk factors. The statistical significance level for correlation and regression was set at $P<.05$ for 2-tailed tests.

XSL•FO
RenderX

Ethical Considerations

This study was conducted in accordance with the Declaration of Helsinki as set forth by the World Medical Association. Ethics approval was obtained from the institutional review board of the Hunan Children's Hospital affiliated with Xiangya School of Medicine, Central South University (KS2025-169). This study was part of a free medical outreach program in impoverished rural areas, so no financial compensation was provided to participants. Before data collection, teachers explained the questionnaire content to participants and their primary guardians. The first section of the questionnaire included informed consent information, and participants or their guardians were free to decline participation or withdraw at any time without penalty. All participants provided written informed consent. The informed consent for the primary data collection complied with the institutional review board's guidelines and permitted secondary data analysis without the need for additional consent.

While the original data contained personal identifiers, all participants' names were hidden during the analysis phase to ensure confidentiality. Data were securely stored in password-protected systems accessible only to authorized researchers. No identification of individual participants or users in any images of the manuscript or supplementary material is possible.

Results

Prevalence of IGD and Sex Differences

A total of 18,967 baseline data entries were collected in this study. After quality control, 13,931 valid questionnaires were obtained for the baseline cross-sectional study; please refer to Figure 1 for the detailed screening process. These data revealed a prevalence of IGD of 5.2% (725/13,931; 95% CI 4.8%-5.6%). The prevalence among male individuals (489/7304, 6.7%; 95% CI 6.1%-7.3%) was significantly higher than that among female individuals (236/6627, 3.6%; 95% CI 3.1%-4%; χ^2_1=69.2; risk difference [RD] 0.31, 95% CI 0.02-0.04; *P*<.001).

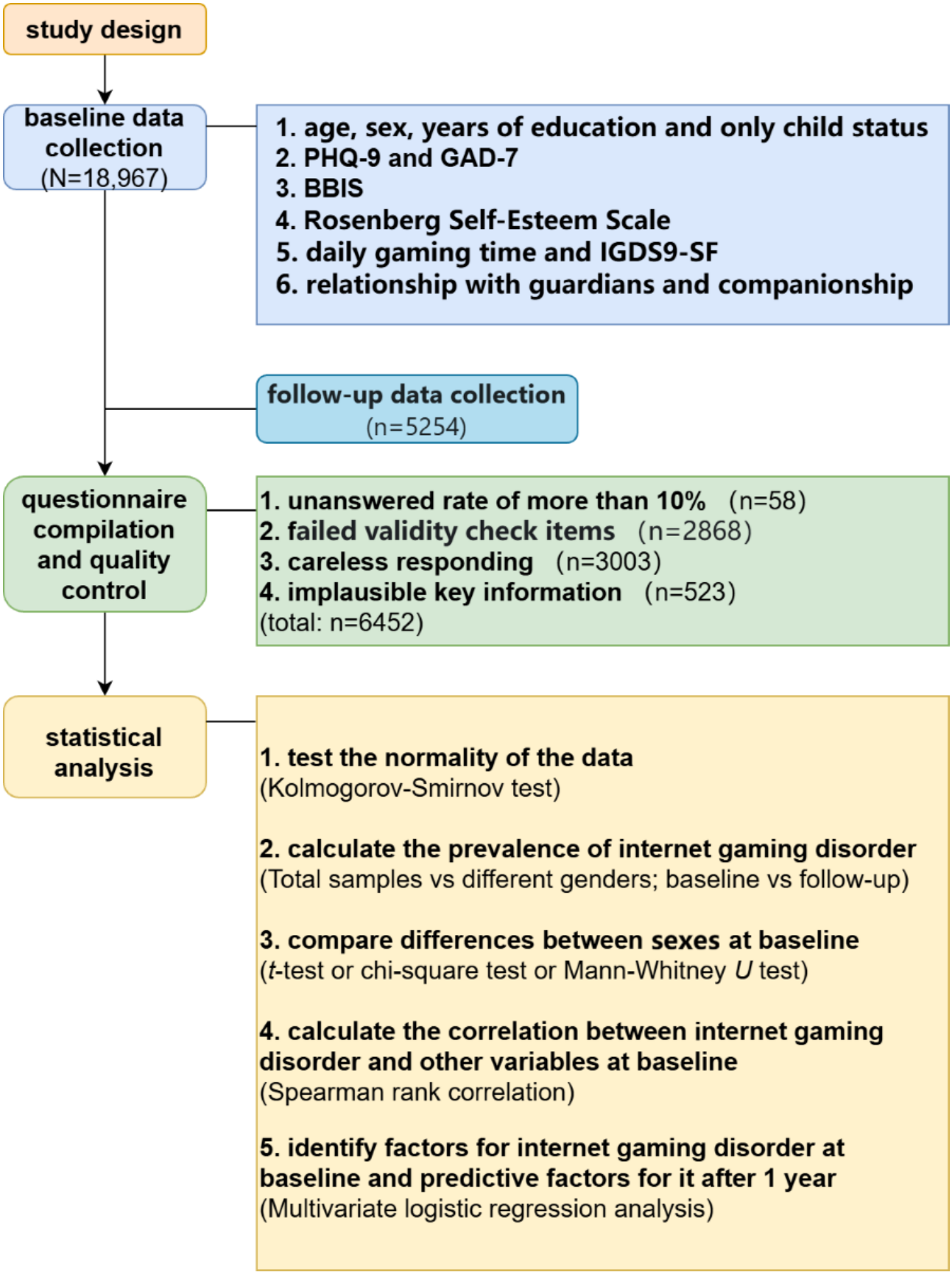
For the 1-year follow-up assessment, 3838 valid questionnaires were obtained (loss to follow-up rate: 10,093/13,931, 72.4%; 95% CI 71.7%-73.2%). The result of the Little MCAR test indicated that the data were not MCAR (χ^2_9=1358.2, *P*<.001). The longitudinal data after multiple imputation showed that the overall prevalence of IGD was 5% (692/13,931; 95% CI 4.6%-5.3%), and the prevalence among male individuals remained significantly higher than that among female individuals (511/7304, 7%; 95% CI 6.4%-7.6% vs 181/6627, 2.7%; 95% CI 2.4%-3.2%, respectively; χ^2_1=133.9; RD 0.04, 95% CI 0.04-0.05; *P*<.001).

Among participants with IGD at baseline, the persistence rate at follow-up of IGD was 9.1% (66/725; 95% CI 7.2%-11.45%). Specifically, the persistence rate was 7.2% (17/236; 95% CI 4.5%-11.2%) in the female subgroup and 10% (49/489; 95% CI 7.7%-13.1%) in the male subgroup. There was no statistically significant difference between the 2 subgroups (χ^2_1=1.5; RD 0.03, 95% CI –0.02 to 0.07; *P*=.22).

As a sensitivity analysis, we also conducted the same analysis on the full-case data, yielding results that were largely consistent with the multiple imputation analysis (see Table S1 in Multimedia Appendix 1 for details).

XSL•FO
RenderX

Figure 1. Flowchart. BBIS: Brief Barratt Impulsiveness Scale; GAD-7: Generalized Anxiety Disorder–7; IGDS9-SF: Internet Gaming Disorder Scale–Short Form; PHQ-9: Patient Health Questionnaire–9.



XSL•FO
RenderX

Sex Differences in Baseline Demographic and Clinical Variables

An analysis of baseline cross-sectional data revealed significant sex differences in several demographic and clinical variables. When grouped by sex, there was a higher proportion of male individuals who were only children than female individuals (χ^2_1=71.8, post-FDR *P*<.001; RD 0.06, 95% CI 0.04-0.07) and a greater percentage who reported "very good" relationships with their primary guardian (*Z*=–4.69; median difference [MD] 0.03, 95% CI 0.02-0.05; post-FDR *P*<.001) and longer daily gaming time (*Z*=–2.83; MD –0.23, 95% CI −0.35 to −0.10; post-FDR *P*=.008). Male individuals also scored higher on the IGDS9-SF (*Z*=–18.79; MD –2.20, 95% CI –2.45 to −1.95; post-FDR *P*<.001) and self-esteem ($t_{13,929}$=–9.89; MD –0.81, 95% CI −0.97 to −0.65; post-FDR *P*<.001). Conversely, male individuals had significantly lower scores on the GAD-7 (*Z*=–13.57; MD 0.83, 95% CI 0.68-0.98; post-FDR *P*<.001), the PHQ-9 (*Z*=–9.27; MD 0.71, 95% CI 0.54-0.89; post-FDR *P*<.001), self-regulation ($t_{13,929}$=3.81; MD 0.18, 95% CI 0.09-0.27; post-FDR *P*<.001), and the total BBIS impulsivity scale ($t_{13,929}$=2.51; MD 0.18, 95% CI 0.04-0.32; post-FDR *P*=.01). For specific baseline variable differences between sexes, please refer to Table 1.

XSL•FO
RenderX

Table . Sex differences in baseline demographic characteristics, family relationship, anxiety, depression, self-esteem, impulsivity, gaming time, and internet gaming disorder scores. The Mann-Whitney *U* test after *z* transformation requires no calculation of df.

Variables	Female individuals (n=6627)	Male individuals (n=7304)	Statistics; Z score, t test, or chi-square (*df*)	Risk difference or median difference (95% CI)	Pre-FDR[a] *P* value	Post-FDR *P* value
Age (y), mean (SD)	12.11 (1.57)	12.12 (1.57)	−0.26 (13,929)[b]	−0.01 (−0.06 to 0.04)	.80	.85
Education (y), mean (SD)	6.33 (1.58)	6.28 (1.58)	1.86 (13,929)[b]	0.05 (−0.003 to 0.10)	.06	.08
Only child, n (%)			71.75 (1)[c]	0.06 (0.04 to 0.07)	<.001	<.001
No	5525 (83.4)	5677 (77.7)				
Yes	1101 (16.6)	1627 (22.3)				
Relationship with primary guardian, n (%)			−4.69 (3)[d]	0.03 (0.02 to 0.05)	<.001	<.001
Very poor	25 (0.4)	43 (0.6)				
Poor	92 (1.4)	93 (1.3)				
Good	1709 (25.8)	1615 (22.1)				
Very good	4801 (72.4)	5553 (76.0)				
Companionship, n (%)			0.49 (1)[c]	0.004 (−0.01 to 0.02)	.48	.54
No	1137 (17.2)	1286 (17.6)				
Yes	5490 (82.8)	6018 (82.4)				
Gaming time (h per d), mean (SD)	2.60 (3.55)	2.83 (3.77)	−2.83 (13,929)[d]	−0.23 (−0.35 to −0.10)	.005	.008
Self-esteem score (range 10-40), mean (SD)	28.15 (4.91)	28.96 (4.75)	−9.89 (13,929)[b]	−0.81 (−0.97 to −0.65)	<.001	<.001
GAD-7[e] score (range 0-21), mean (SD)	4.86 (4.48)	4.04 (4.39)	−13.57 (13,929)[d]	0.83 (0.68 to 0.98)	<.001	<.001
PHQ-9[f] score (range 0-27), mean (SD)	5.37 (5.43)	4.66 (5.16)	−9.27 (13,929)[d]	0.71 (0.54 to 0.89)	<.001	<.001
IGDS9-SF[g] score (range 9-45), mean (SD)	14.51 (6.91)	16.70 (7.97)	−18.79 (13,929)[d]	−2.20 (−2.45 to −1.95)	<.001	<.001
Poor self-regulation score (range 4-16), mean (SD)	9.17 (2.65)	9.00 (2.84)	3.81 (13,929)[b]	0.18 (0.09 to 0.27)	<.001	<.001
Impulsive behavior score (range 4-16), mean (SD)	8.20 (2.51)	8.20 (2.66)	0.03 (13,929)[b]	0.001 (−0.08 to 0.09)	.98	.98
BBIS[h] score (range 8-32), mean (SD)	17.37 (4.13)	17.19 (4.26)	2.51 (13,929)[b]	0.18 (0.04 to 0.32)	.01	.01

[a]FDR: false discovery rate.

[b]*t* test.

[c]Chi-square test.

[d]Z score.

[e]GAD-7: Generalized Anxiety Disorder–7.

[f]PHQ-9: Patient Health Questionnaire–9.

[g]IGDS9-SF: Internet Gaming Disorder Scale–Short Form.

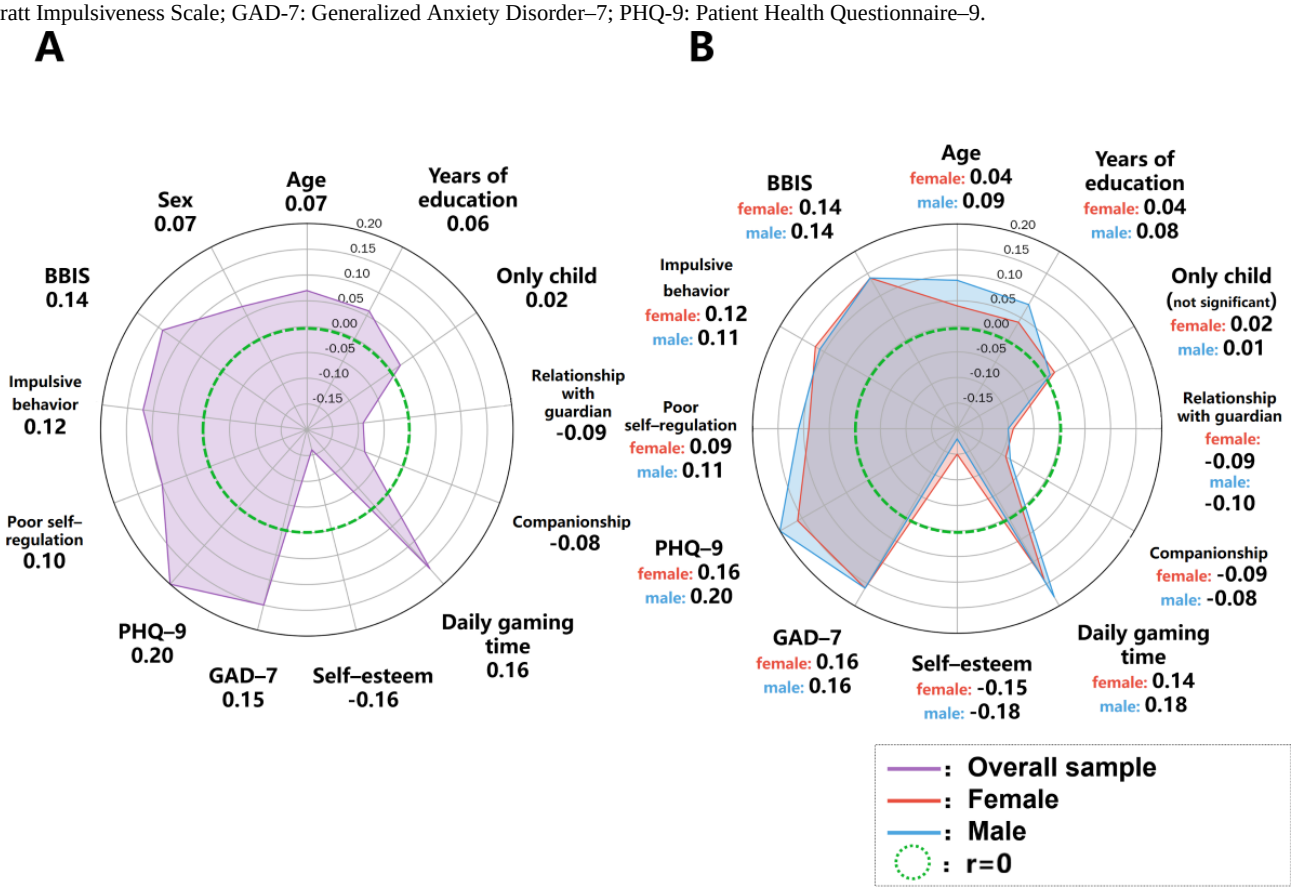
XSL•FO
RenderX

[h]BBIS: Brief Barratt Impulsiveness Scale.

Correlation Analysis of IGD Indicators

On the basis of the baseline cross-sectional data, IGD indicators were significantly correlated with multiple factors (Figure 2), including age (r=0.07, 95% CI 0.05-0.08; P<.001), years of education (r=0.06, 95% CI 0.04-0.08; P<.001), relationship with primary guardian (r=–0.09, 95% CI −0.11 to −0.07; P<.001), companionship (r=–0.08, 95% CI −0.10 to −0.06; P<.001), only child status (r=0.02, 95% CI 0.004-0.04; P=.02), gaming time (r=0.16, 95% CI 0.15-0.18; P<.001), poor self-regulation (r=0.10, 95% CI 0.08-0.12; P<.001), impulsive behavior (r=0.12, 95% CI 0.10-0.13; P<.001), BBIS score (r=0.14, 95% CI 0.12-0.15; P<.001), self-esteem (r=–0.16, 95% CI −0.17 to −0.14; P<.001), GAD-7 score (r=0.15, 95% CI 0.14-0.17; P<.001), and PHQ-9 score (r=0.20, 95% CI 0.18-0.21; P<.001).

Figure 2. Variables associated with internet gaming disorder in (A) the baseline overall sample and (B) male and female individuals. BBIS: Brief Barratt Impulsiveness Scale; GAD-7: Generalized Anxiety Disorder–7; PHQ-9: Patient Health Questionnaire–9.



Subgroup analyses by sex revealed that, for female individuals, there were significant correlations with age (r=0.04, 95% CI 0.02-0.06; P=.002), years of education (r=0.04, 95% CI 0.01-0.06; P=.005), relationship with primary guardian (r=–0.09, 95% CI −0.11 to −0.06; P<.001), companionship (r=–0.09, 95% CI −0.12 to −0.05; P<.001), gaming time (r=0.14, 95% CI 0.11-0.16; P<.001), self-esteem (r=–0.15, 95% CI −0.17 to −0.13; P<.001), GAD-7 score (r=0.16, 95% CI 0.14-0.18; P<.001), PHQ-9 score (r=0.16, 95% CI 0.14-0.17; P<.001), poor self-regulation (r=0.09, 95% CI 0.07-0.12; P<.001), impulsive behavior (r=0.12, 95% CI 0.10-0.15; P<.001), and BBIS score (r=0.14, 95% CI 0.11-0.16; P<.001). For male individuals, significant correlations were found with age (r=0.09, 95% CI 0.07-0.11; P<.001), years of education (r=0.08, 95% CI 0.06-0.10; P<.001), relationship with primary guardian (r=–0.10, 95% CI −0.12 to −0.07; P<.001), companionship (r=–0.08, 95% CI −0.11 to −0.06; P<.001), gaming time (r=0.18, 95% CI 0.16-0.21; P<.001), self-esteem (r=–0.18, 95% CI −0.20 to −0.15; P<.001), GAD-7 score (r=0.16, 95% CI 0.14-0.19; P<.001), PHQ-9 score (r=0.20, 95% CI 0.18-0.23; P<.001), poor self-regulation (r=0.11, 95% CI 0.09-0.13; P<.001), impulsive behavior (r=0.11, 95% CI 0.09-0.14; P<.001), and BBIS score (r=0.14, 95% CI 0.12-0.17; P<.001). Additionally, being male was positively correlated with IGD in the overall sample (r=0.07; P<.001).

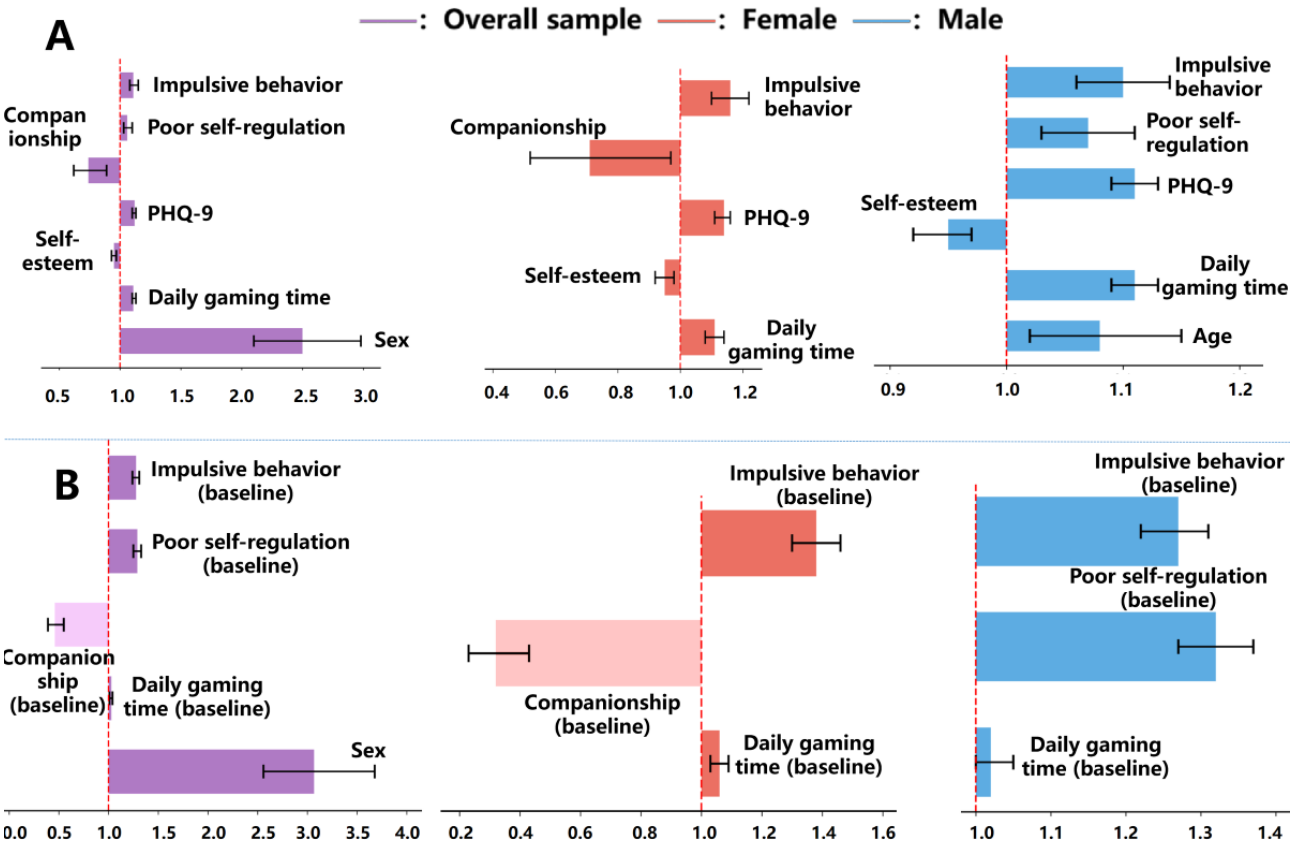
Identification of Factors Associated With IGD

Overview

Multivariate logistic regression models using backward stepwise selection (Wald test) were used, first, to identify factors cross-sectionally associated with IGD at baseline and, second, to longitudinally predict IGD case status at follow-up, as shown in Figure 3.

XSL•FO
RenderX

Figure 3. Odds ratios for (A) identifying baseline factors of internet gaming disorder and (B) predicting follow-up internet gaming disorder. PHQ-9: Patient Health Questionnaire–9.



Baseline IGD Identification Factors (Cross-Sectional Analysis)

In the overall sample, male sex (odds ratio [OR] 2.50, 95% CI 2.10-2.98; *P*<.001), gaming time (OR 1.11, 95% CI 1.10-1.13; *P*<.001), PHQ-9 score (OR 1.12, 95% CI 1.10-1.13; *P*<.001), impulsive behavior (OR 1.11, 95% CI 1.08-1.15; *P*<.001), and self-regulation (OR 1.06, 95% CI 1.03-1.10; *P*<.001) were identified as risk factors. Protective factors included self-esteem score (OR 0.95, 95% CI 0.93-0.97; *P*<.001) and companionship (OR 0.74, 95% CI 0.62-0.89; *P*=.001).

In sex subgroup analyses, common cross-sectional identification factors for IGD included gaming time (female individuals: OR 1.11, 95% CI 1.08-1.14, *P*<.001; male individuals: OR 1.11, 95% CI 1.09-1.13, *P*<.001), self-esteem score (female individuals: OR 0.95, 95% CI 0.92-0.98, *P*=.002; male individuals: OR 0.95, 95% CI 0.92-0.97, *P*<.001), PHQ-9 score (female individuals: OR 1.14, 95% CI 1.11-1.16, *P*<.001; male individuals: OR 1.11, 95% CI 1.09-1.13, *P*<.001), and impulsive behavior (female individuals: OR 1.16, 95% CI 1.10-1.22, *P*<.001; male individuals: OR 1.10, 95% CI 1.06-1.14, *P*<.001). Differences included that companionship was a protective factor for female individuals (OR 0.71, 95% CI 0.52-0.97; *P*=.03), whereas age (OR 1.08, 95% CI 1.02-1.15; *P*=.02) and poor self-regulation (OR 1.07, 95% CI 1.03-1.11; *P*=.001) were additional risk factors for male individuals.

Predictors of IGD at the 1-Year Follow-Up (Longitudinal Analysis)

This study then examined baseline variables as longitudinal predictors of developing IGD at follow-up. In the overall sample, male sex (OR 3.07, 95% CI 2.56-3.68; *P*<.001), baseline gaming time (OR 1.03, 95% CI 1.01-1.04; *P*=.005), baseline poor self-regulation (OR 1.29, 95% CI 1.25-1.33; *P*<.001), and baseline impulsive behavior (OR 1.28, 95% CI 1.24-1.31; *P*<.001) were risk factors. Protective factors included companionship at baseline (OR 0.46, 95% CI 0.39-0.55; *P*<.001).

In sex subgroup analyses, common longitudinal predictors included baseline gaming time (female individuals: OR 1.06, 95% CI 1.03-1.09, *P*<.001; male individuals: OR 1.02, 95% CI 1.00-1.05, *P*=.03) and baseline impulsive behavior (female individuals: OR 1.38, 95% CI 1.30-1.46, *P*<.001; male individuals: OR 1.27, 95% CI 1.22-1.31, *P*<.001). Differences included that companionship at baseline was a protective predictor for female individuals (OR 0.32, 95% CI 0.23-0.43; *P*<.001), whereas poor self-regulation at baseline (OR 1.32, 95% CI 1.27-1.37; *P*<.001) was a predictor for male individuals.

Discussion

Principal Findings

To our knowledge, this study is the first large-scale longitudinal investigation in China focusing on the prevalence, influencing factors, and sex differences regarding IGD among rural adolescents, and it provides new evidence for intervention

XSL•FO
RenderX

strategies targeting IGD among adolescents in impoverished rural areas. This pioneering work fills a critical gap in rural mental health research and has practical implications for designing tailored interventions. Key findings indicate that, in the overall sample, the baseline prevalence of IGD was 5.2% (725/13,931), with a 1-year follow-up prevalence of 5% (692/13,931). Both analyses revealed significantly higher prevalence rates among male individuals than among female individuals. There were common cross-sex risk factors for IGD at both the baseline and follow-up stages: longer gaming time and greater impulsivity behaviors. Sex-specific differences were primarily observed in protective factors and long-term predictive effects—guardian companionship exhibited a significant protective effect for female individuals, whereas poor self-regulation at baseline was a unique long-term predictor for male individuals. Excluding self-esteem and depressive states, these findings align with the 4 hypotheses of this study.

The baseline (725/13,931, 5.2%) and 1-year follow-up (692/13,931, 5%) prevalence rates of IGD among rural students in this study were notably lower than the 10% reported in urban Chinese samples [22]. This discrepancy may be attributed to limited internet infrastructure, fewer gaming devices, and economic constraints restricting frequent gaming in rural areas. Regarding sex differences, male individuals consistently exhibited significantly higher prevalence rates across all stages, aligning with most previous research [23]. This pattern may reflect male individuals' greater preference for competitive, task-oriented games and gaming motivations centered on achievement and self-affirmation [24].

Individual psychological characteristics demonstrated stable effects in identifying or predicting IGD. Depressive symptoms were significant risk factors across sex subgroups at baseline [25], supporting the "emotion escape model" [26], which suggests that individuals with negative emotions such as depression or anxiety may turn to online activities such as gaming as a way to temporarily escape these feelings. This maladaptive coping strategy can contribute to the development of IGD over time. Impulsivity and poor self-regulation were significantly associated with both cross-sectional and longitudinal outcomes, consistent with the interaction of person-affect-cognition-execution model [27]. The interaction of the person-affect-cognition-execution model proposes that addictive behaviors, including IGD, arise from the interaction of personal characteristics (eg, impulsivity), affective responses, cognitive factors, and executive functioning deficits. These factors interact to reduce an individual's ability to resist the immediate rewards offered by gaming, thereby increasing the likelihood of developing a disorder.

Self-esteem served as a cross-sex protective factor at baseline, particularly pronounced in female individuals. Individuals with higher self-esteem may exhibit greater psychological resilience when facing peer pressure or identity challenges, reducing reliance on gaming for self-worth—corroborating the buffering effect of self-esteem in adolescent internet addiction [28]. Thus, psychological interventions targeting affective regulation, self-regulation, and self-evaluation may offer effective pathways for preventing and treating IGD. Notably, after multiple imputation of the entire baseline sample, baseline self-esteem no longer significantly predicted follow-up IGD, suggesting selective bias in the previous complete case analysis. The association patterns between low self-esteem and IGD differed between the follow-up sample and the sample lost to follow-up, and multiple imputation corrected this bias by estimating the IGD status of those lost to follow-up. Evidently, self-esteem's long-term predictive power for IGD is less direct and stable than in cross-sectional studies, potentially mediated or masked by factors such as impulsivity and family companionship.

Gaming time significantly predicted IGD across sex subgroups, consistent with previous studies [29,30]. Excessive gaming not only increases exposure to virtual environments but also diminishes real-life social and emotional interactions, reinforcing immediate reward mechanisms and deepening immersion [27], thereby intensifying dependence. Over time, the weakening of real-world functions may lead individuals to increasingly turn to gaming as an escape from stress, which in turn drives longer gaming sessions, creating a closed loop of addictive behavior. Therefore, targeted interventions to control gaming time should not only reduce opportunities for triggering addictive behavior at its source but also create conditions for individuals to rebuild real-world social connections and restore a sense of purpose, thereby lowering the risk of developing and progressing into IGD.

Guardian companionship emerged as a significant protective factor at both baseline and follow-up, especially for female adolescents. Stable emotional support and supervision reduce adolescents' reliance on virtual worlds, consistent with family functioning research indicating that positive parent-child interactions, emotional support, and behavioral monitoring lower the risk of internet addiction [31,32]. For impoverished rural students, the importance of caregiver companionship is even more pronounced: rural left-behind children commonly experience emotional support deficits [33], and unmet emotional needs may drive compensatory gaming behaviors where immediate feedback and a sense of belonging reinforce use. This finding aligns with attachment theory, which posits that secure attachment reduces dependence on alternative emotional sources. For female adolescents, lack of companionship significantly elevates risk, underscoring the buffering role of emotional support [34].

This study highlights significant sex differences in predictors of IGD. For female adolescents, caregiver companionship was the most salient protective factor, indicating that familial emotional support and presence effectively mitigate the risk of gaming dependence when vulnerability arises. Female individuals with access to emotional support, particularly during adversity, may better regulate negative emotions and reduce the likelihood of gaming addiction [34]. Conversely, poor self-regulation was a significant risk predictor for male adolescents. Due to higher extraversion and lower self-regulation, male individuals are more prone to engaging in high-stimulation, immediate-feedback gaming activities, increasing IGD risk [35]. This pattern aligns with male behavioral tendencies to regulate emotions through external stimulation [36]. The distinct pathways observed between male and female individuals suggest that future research and

XSL•FO
RenderX

intervention strategies should fully consider sex differences, avoiding one-size-fits-all approaches.

Limitations

There were several limitations to this study. First, all data relied on self-report, which may introduce social desirability bias (underreporting gaming time to avoid disapproval) and recall bias (inaccurate memory of past gaming behaviors), affecting data reliability. Second, the limited number of assessment scales restricted the measurement of key factors such as family parenting styles and school interventions, leaving potential associations with IGD unexamined. Third, a single 1-year follow-up failed to capture long-term dynamic changes in IGD and the sustained effects of influencing factors. Fourth, nonrandom sampling (relying on medical programs and government-recommended schools) may limit the generalizability of the results to a broader population of impoverished rural adolescents. Fifth, a second free medical consultation was not arranged in some regions, resulting in the loss of follow-up data. This may limit, to a certain extent, the generalizability of this study's conclusions to adolescents in regions without follow-up. Future research could address these limitations by integrating objective data (eg, mobile phone use logs) to verify self-reports, expanding the scales to include parenting- and school-related variables, increasing follow-up frequency and duration, and adopting random sampling to improve representativeness.

Conclusions

In summary, this pioneering large-scale prospective longitudinal study provides robust evidence on the prevalence and sex-specific predictors of IGD among impoverished rural adolescents in China. The findings reveal that IGD prevalence in rural areas is lower than that in urban regions, with male adolescents exhibiting significantly higher rates than their female counterparts. Impulsivity, gaming time, and caregiver companionship emerged as key influencing factors, demonstrating notable sex differences in both protective and risk profiles. Specifically, female individuals benefited more from the protective role of caregiver companionship, whereas male individuals were more vulnerable to deficits in self-regulation. These findings hold significant practical implications, underscoring the necessity of developing tailored public health strategies that address sex-specific needs, strengthen family support systems, and enhance psychological resilience to effectively prevent and mitigate IGD in these vulnerable communities. Future research should integrate multisource data and conduct long-term follow-ups to deepen understanding of IGD development and refine intervention approaches.

Acknowledgments

The authors sincerely thank each participant who completed the questionnaire and the teachers who assisted their research team in distributing the questionnaires.

Funding

This work was supported by the Hunan Provincial Natural Science Foundation (grant 2023JJ40476) and Hunan Provincial Social Science Achievement Evaluation Committee (grant XSP24YBZ185). The funders did not participate in the research design, data collection, data analysis, interpretation of results, or manuscript writing.

Data Availability

The datasets generated or analyzed during this study are not deposited in public repositories, primarily to protect the privacy of minor participants and comply with the research ethics approval requirements of the affiliated institution, but are available from the corresponding author on reasonable request for academic research purposes.

Authors' Contributions

CW and YL contributed to the study concept and design, data collection and analysis, and manuscript writing and revision. YL, SL, and YS conducted the statistical analysis and data interpretation. QH contributed to study design, data interpretation, and manuscript revision. CW, GY, XD, EH, and JL were involved in data collection and analysis and manuscript revision. QH, GY, JL, and XD supervised the study, obtained funding, and provided critical manuscript revision. All authors had full access to all data and are responsible for the integrity and accuracy of the data and analysis.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Identification factors of internet gaming disorder under baseline conditions, predictive factors of internet gaming disorder under follow-up conditions.
[DOCX File, 34 KB - games_v13i1e83522_app1.docx]

Checklist 1

XSL•FO
RenderX

STROBE checklist.
[DOCX File, 22 KB - games_v13i1e83522_app2.docx]

References

1. Li L, Griffiths MD, Niu Z, Mei S. Fear of missing out (FoMO) and gaming disorder among Chinese university students: impulsivity and game time as mediators. Issues Ment Health Nurs 2020 Dec;41(12):1104-1113. [doi: 10.1080/01612840.2020.1774018] [Medline: 32634040]
2. Liao Z, Chen X, Huang Q, Shen H. Prevalence of gaming disorder in East Asia: a comprehensive meta-analysis. J Behav Addict 2022 Aug 5;11(3):727-738. [doi: 10.1556/2006.2022.00050] [Medline: 35932469]
3. Lenzi M, Elgar FJ, Marino C, et al. Can an equal world reduce problematic social media use? Evidence from the Health Behavior in School-aged Children study in 43 countries. Inf Commun Soc 2022 Oct 26;26(14):2753-2774. [doi: 10.1080/1369118X.2022.2109981]
4. Petrescu TC, Błachnio A, Enea V. Internet gaming disorder in children and adolescents: a systematic review of familial protective and risk factors. Addict Behav 2025 Aug;167:108345. [doi: 10.1016/j.addbeh.2025.108345] [Medline: 40154103]
5. Peng P, Chen Z, Ren S, et al. Internet gaming disorder and suicidal behaviors mediated by sleep disturbance: a large-scale school-based study in 135,174 Chinese middle school students. Eur Child Adolesc Psychiatry 2025 Aug;34(8):2571-2583. [doi: 10.1007/s00787-025-02675-z] [Medline: 40029399]
6. Peng P, Jin J, Chen Z, et al. Impaired sleep quality mediates the relationship between internet gaming disorder and conduct problems among adolescents: a three-wave longitudinal study. Child Adolesc Psychiatry Ment Health 2025 Mar 21;19(1):26. [doi: 10.1186/s13034-025-00889-2] [Medline: 40119352]
7. Peng P, Chen Z, Ren S, et al. Internet gaming disorder predicts the incidence, persistence, and worsening of suicidal ideation: a population-based cohort study of 96,158 Chinese adolescents. J Affect Disord 2025 Jun 15;379:186-193. [doi: 10.1016/j.jad.2025.03.029] [Medline: 40081581]
8. He J, Zhao H, Lei X, et al. How impulsivity and sensation seeking drive IGD tendency in healthy young adults: a resting-state MRI study. J Behav Addict 2025 Aug 16;14(3):1343-1354. [doi: 10.1556/2006.2025.00063] [Medline: 40818056]
9. Yu Y, Li S, Lyu X, Zhang J, Li JB, Lau JT. A cross-lagged panel model of the bidirectional relationships between gaming-specific maladaptive cognitions and internet gaming disorder among adolescents in China. Addict Behav 2025 Oct 8;172:108518. [doi: 10.1016/j.addbeh.2025.108518] [Medline: 41075582]
10. Liao Z, Huang Q, Huang S, et al. Prevalence of internet gaming disorder and its association with personality traits and gaming characteristics among Chinese adolescent gamers. Front Psychiatry 2020 Nov 17;11:598585. [doi: 10.3389/fpsyt.2020.598585] [Medline: 33312143]
11. Yang X, Jiang X, Mo PK, Cai Y, Ma L, Lau JT. Prevalence and interpersonal correlates of internet gaming disorders among Chinese adolescents. Int J Environ Res Public Health 2020 Jan 16;17(2):579. [doi: 10.3390/ijerph17020579] [Medline: 31963197]
12. Aurora data: Honor of Kings Research Report, May 2017 [Report in Chinese]. : Aurora Big Data; 2017 URL: https://www.afenxi.com/44812.html [accessed 2025-08-16]
13. Todorovic L, Bozhar H, de Rooij SR, Bogaerts A, Boyer BE, Larsen H. Bidirectional associations of problematic social media use and problematic gaming with mental health difficulties and strengths in adolescents: sex and social support as potential moderators. J Res Adolesc 2025 Sep;35(3):e70076. [doi: 10.1111/jora.70076] [Medline: 40944323]
14. Tu S, Hu K, Wu H, et al. Effects of cognitive reappraisal on gaming craving and inhibitory control in adolescents with internet gaming disorder. Psychol Addict Behav 2025 Sep;39(6):560-570. [doi: 10.1037/adb0001091] [Medline: 40788716]
15. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. J Clin Epidemiol 2008 Apr;61(4):344-349. [doi: 10.1016/j.jclinepi.2007.11.008] [Medline: 18313558]
16. Shen ZL, Cai TS. Disposal to the 8th item of Rosenberg Self-Esteem Scale (Chinese version). Chin Ment Health J 2008;22(9):661-663 [FREE Full text]
17. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613. [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
18. Zhang YL, Liang W, Chen ZM, et al. Validity and reliability of Patient Health Questionnaire-9 and Patient Health Questionnaire-2 to screen for depression among college students in China. Asia Pac Psychiatry 2013 Dec;5(4):268-275. [doi: 10.1111/appy.12103] [Medline: 24123859]
19. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
20. Steinberg L, Sharp C, Stanford MS, Tharp AT. New tricks for an old measure: the development of the Barratt Impulsiveness Scale-Brief (BIS-Brief). Psychol Assess 2013 Mar;25(1):216-226. [doi: 10.1037/a0030550] [Medline: 23148649]
21. Qin L, Cheng L, Hu M, et al. Clarification of the cut-off score for Nine-Item Internet Gaming Disorder Scale-Short Form (IGDS9-SF) in a Chinese context. Front Psychiatry 2020 May 25;11:470. [doi: 10.3389/fpsyt.2020.00470] [Medline: 32528331]

XSL•FO
RenderX

22. Gou S, Zhang W, Tang Y, Zhang J, He Q. Prevalence of internet gaming disorder among Chinese adolescents: a systematic review and meta-analysis. Asian J Psychiatr 2024 Dec;102:104257. [doi: 10.1016/j.ajp.2024.104257] [Medline: 39366164]
23. Peng P, Chen Z, Ren S, et al. Sex difference in the longitudinal association between depressive symptoms and internet gaming disorder among Chinese adolescents: an explanatory analysis at the aggregate and symptom level. Addict Behav 2025 Sep 17;172:108499. [doi: 10.1016/j.addbeh.2025.108499] [Medline: 40974915]
24. Mihara S, Higuchi S. Cross-sectional and longitudinal epidemiological studies of internet gaming disorder: a systematic review of the literature. Psychiatry Clin Neurosci 2017 Jul;71(7):425-444. [doi: 10.1111/pcn.12532] [Medline: 28436212]
25. Peng P, Chen Z, Ren S, et al. Trajectory of internet gaming disorder among Chinese adolescents: course, predictors, and long-term mental health outcomes. J Behav Addict 2025 Jun 23;14(2):846-860. [doi: 10.1556/2006.2025.00054] [Medline: 40549959]
26. Kardefelt-Winther D. A conceptual and methodological critique of internet addiction research: towards a model of compensatory internet use. Comput Human Behav 2014 Feb;31:351-354. [doi: 10.1016/j.chb.2013.10.059]
27. Brand M, Wegmann E, Stark R, et al. The Interaction of Person-Affect-Cognition-Execution (I-PACE) model for addictive behaviors: update, generalization to addictive behaviors beyond internet-use disorders, and specification of the process character of addictive behaviors. Neurosci Biobehav Rev 2019 Sep;104:1-10. [doi: 10.1016/j.neubiorev.2019.06.032] [Medline: 31247240]
28. Kim HK, Davis KE. Toward a comprehensive theory of problematic internet use: evaluating the role of self-esteem, anxiety, flow, and the self-rated importance of internet activities. Comput Human Behav 2009 Mar;25(2):490-500. [doi: 10.1016/j.chb.2008.11.001]
29. Kuss DJ, Griffiths MD, Karila L, Billieux J. Internet addiction: a systematic review of epidemiological research for the last decade. Curr Pharm Des 2014;20(25):4026-4052. [doi: 10.2174/13816128113199990617] [Medline: 24001297]
30. Lemmens JS, Valkenburg PM, Gentile DA. The internet gaming disorder scale. Psychol Assess 2015 Jun;27(2):567-582. [doi: 10.1037/pas0000062] [Medline: 25558970]
31. Ko CH, Wang PW, Liu TL, Yen CF, Chen CS, Yen JY. Bidirectional associations between family factors and internet addiction among adolescents in a prospective investigation. Psychiatry Clin Neurosci 2015 Apr;69(4):192-200. [doi: 10.1111/pcn.12204] [Medline: 24836367]
32. Shek DT, Zhu X, Ma CM. The influence of parental control and parent-child relational qualities on adolescent internet addiction: a 3-year longitudinal study in Hong Kong. Front Psychol 2018 May 1;9:642. [doi: 10.3389/fpsyg.2018.00642] [Medline: 29765349]
33. Fan F, Su L, Gill MK, Birmaher B. Emotional and behavioral problems of Chinese left-behind children: a preliminary study. Soc Psychiat Epidemiol 2010 Jun;45(6):655-664. [doi: 10.1007/s00127-009-0107-4] [Medline: 19657573]
34. Mikulincer M, Shaver PR. Attachment in Adulthood: Structure, Dynamics, and Change: Guilford Publications; 2007.
35. Anderson EL, Steen E, Stavropoulos V. Internet use and problematic internet use: a systematic review of longitudinal research trends in adolescence and emergent adulthood. Int J Adolesc Youth 2016 Oct 10;22(4):430-454. [doi: 10.1080/02673843.2016.1227716]
36. Roberts BW, Mroczek D. Personality trait change in adulthood. Curr Dir Psychol Sci 2008 Feb 1;17(1):31-35. [doi: 10.1111/j.1467-8721.2008.00543.x] [Medline: 19756219]

Abbreviations

BBIS: Brief Barratt Impulsiveness Scale
FDR: false discovery rate
GAD-7: Generalized Anxiety Disorder–7
IGD: internet gaming disorder
IGDS9-SF: Internet Gaming Disorder Scale–Short Form
MCAR: missing completely at random
MD: median difference
OR: odds ratio
PHQ-9: Patient Health Questionnaire–9
RD: risk difference
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

XSL•FO
RenderX

Edited by S Brini; submitted 04.09.25; peer-reviewed by P Peng, Q Wang; accepted 20.10.25; published 17.11.25.

Please cite as:
Wang C, Li Y, Lin S, Huang Q, Shi Y, Yang G, Huang E, Deng X, Luo J
Prevalence, Sex Differences, and Predictors of Internet Gaming Disorder Among Impoverished Rural Adolescents: Cross-Sectional and Prospective Cohort Study
JMIR Serious Games 2025;13:e83522
URL: https://games.jmir.org/2025/1/e83522
doi:10.2196/83522

© Chenhan Wang, Yifan Li, Shuhong Lin, Qiuping Huang, Yongyan Shi, Guangxian Yang, Erjia Huang, Xicheng Deng, Jinwen Luo. Originally published in JMIR Serious Games (https://games.jmir.org), 17.11.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Reciprocal Relationship Between Self-Control Belief and Gaming Disorder in Children and Adolescents: Longitudinal Survey Study

Shimin Zhu[1,2*], PhD; Di Qi[1*], PhD

[1]Department of Applied Social Sciences, The Hong Kong Polytechnic University, Rm GH348, 11 Yuk Choi Road, Hung Hom, Kowloon, Hong Kong, China (Hong Kong)

[2]Mental Health Research Centre, The Hong Kong Polytechnic University, Hong Kong, China (Hong Kong)

[*]all authors contributed equally

Corresponding Author:
Shimin Zhu, PhD
Department of Applied Social Sciences, The Hong Kong Polytechnic University, Rm GH348, 11 Yuk Choi Road, Hung Hom, Kowloon, Hong Kong, China (Hong Kong)

Abstract

Background: Children and adolescents are often at the crossroads of leisure gaming and excessive gaming. It is essential to identify the modifiable psychosocial factors influencing gaming disorder development. The lay theories of self-control (ie, the beliefs about whether self-control can be improved, also called self-control mindsets) may interplay with self-control and gaming disorder and serve as a promising influential factor for gaming disorder.

Objective: This study aims to answer the research questions “Does believing one’s self-control is unchangeable predict more severe gaming disorder symptoms later?” and “Does the severity of gaming disorder symptoms prospectively predict self-control mindsets?” with a 1-year, 2-wave, school-based longitudinal survey.

Methods: A total of 3264 students (338 in grades 4 - 5 and 2926 in grades 7 - 10) from 15 schools in Hong Kong participated in the classroom surveys. We used cross-lagged panel models to examine the direction of the longitudinal association between self-control mindsets and gaming disorder.

Results: A bidirectional relationship was found between self-control mindsets and gaming disorder symptom severity (the cross-lagged path from mindsets to gaming disorder: regression coefficients [*b*] with 95% CI [0.070, 0.020-0.12o, *P*=.006]; and from gaming disorder to mindsets: *b* with 95% CI [0.11, 0.060-0.160, *P*<.001]). Subgroup analyses of boy and girl participants revealed that more growth mindsets regarding self-control predicted less severe gaming disorder symptoms in girls (*b*=0.12, 95% CI 0.053-0.190, *P*=.001) but not in boys (*b*=0.025, 95% CI –0.050 to 0.100, *P*=.51), while more severe gaming disorder symptoms predicted a more fixed mindset of self-control in both boys (*b*=0.15, 95% CI 0.069-0.230, *P*<.001) and girls (*b*=0.098, 95% CI 0.031-0.170, *P*=.004) after 1 year.

Conclusions: Our findings demonstrated the negative impact of gaming disorder on one’s self-control malleability beliefs and implied that promoting a growth mindset regarding self-control might be a promising strategy for gaming disorder prevention and early intervention, especially for girls.

(JMIR Serious Games 2025;13:e59441) doi:10.2196/59441

KEYWORDS

growth mindset; gaming disorder; cross-lagged panel model; children; self-control; adolescents

Introduction

Internet gaming is popular worldwide, especially among young people. Over 87% of internet users aged 16 - 24 years around the world play video games [1]. However, underdevelopment of cognitive control may lead to excessive gaming among children and adolescents, rendering them susceptible to gaming disorder [2]. The updated prevalence rate of internet gaming disorder (IGD) in adolescents and young adults worldwide was 9.9% in recent meta-analyses [3], and the prevalence of probable IGD among Hong Kong adolescents was reported to be as high as 15.6% [4]. Excessive gaming greatly increased due to the social distancing and quarantine measures during the COVID-19 pandemic [5,6], and this trend might continue in the postpandemic era as the number of video game users keeps increasing worldwide [7]. The negative consequences of excessive gaming include, but are not limited to, lower academic achievement, reduced mental wellness, and comorbid psychiatric disorders [8,9]. Moreover, gaming disorder is difficult to treat with medication as patients usually refuse it [10]. Thus, there is a pressing need to identify the modifiable psychological factors associated with problematic gaming behavior to develop more effective preventive interventions.

XSL•FO
RenderX

Self-control plays a central role in managing both the amount and timing of game playing [11]. Lacking self-control over gaming is a core feature of gaming disorder in both the *DSM-5* (*Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition*) and *ICD-11* (*International Classification of Diseases, Eleventh Edition*) [12,13]. Plentiful literature has observed the negative association between self-control and gaming disorder [14-16]. Apart from the direct effects, research has also found indirect effects of self-control on gaming disorder via maladaptive gaming-related cognitions [16]. However, few studies have investigated the associations between gaming-related behavior and implicit theories. Implicit theories are referred to as mindsets, are lay theories or unconscious beliefs about the changeability of one's attributes. If one has a growth mindset (also known as incremental theories), one believes that one's attributes can be improved and changed; while if one has a fixed mindset (ie, entity theories), one believes that attributes are fixed entities and cannot change. One previous study tested the links between implicit theories of intelligence and serious game learning [17], showing that players with a growth mindset performed better in serious games and demonstrated more self-directed learning during game playing than players with a fixed mindset. However, as mindsets are domain-specific [18], how self-control mindsets interact with gaming behavior and disorder is yet unknown.

Based on the literature, we propose that self-control mindsets may be a promising influential factor for gaming disorder. Previous studies have demonstrated that beliefs about the malleability of human attributes influence self-regulatory processes and outcomes [19,20]. Further meta-analytic studies have revealed that in contrast to fixed mindsets, growth mindsets predicted the 3 core self-regulatory processes of goal setting, goal operating, and goal monitoring, which then predicted goal achievement [21]. A later intervention study indicated that promoting a growth mindset for self-regulation improved self-regulation in both laboratory settings and daily life via decreased effort avoidance and improved persistence [22], corroborating that cultivating a growth mindset might be a promising way to sustain self-control and protect adolescents from excessive gaming and gaming disorder.

In terms of the associations between self-control mindsets and gaming disorder, there are 2 plausible directions. One possibility is that fixed self-control mindsets predict gaming disorder. Adolescents with fixed self-control mindsets may deem themselves as inherently lacking self-control; as a result, it is less likely for them to make continuous efforts to execute self-control to stop their gaming behavior, leading them to play more frequently and for longer periods and increasing the possibility of developing gaming disorder. Besides, addictive behaviors such as gaming disorder may also lead to fixed mindsets about self-control. For instance, teens who are preoccupied with and lack self-control for gaming may be more likely to believe their self-control is fixed and not improvable. Therefore, it is necessary to delineate the reciprocal relationships between self-control mindsets and gaming disorder to contribute to theory development and practical direction. Disentangling the direction of the link between self-control mindsets and gaming disorder can enrich the understanding of developmental trajectories. Furthermore, the 2 alternative prediction models provide different implications for prevention and intervention [23].

Literature has consistently found gaming disorder rates were much higher in boys than in girls [3,24]. Furthermore, previous studies have also demonstrated distinct levels of gaming disorder symptoms among men and women who play games [14], and that girls tend to better understand the reasons for and potential dangers of gaming disorder [25]. Apart from the gender differences in adolescents' gaming behaviors observed, including boys playing more, presenting more disordered gaming and preferring more of competitive genres than girls, gender was also shown to moderate personality's association with gaming disorder, thus gender is highly relevant and gender-dependent differences need more attention when studying gaming-related behaviors [26]. Establishing how the relationships between self-control mindsets and gaming disorder differ across genders is key to tailoring intervention plans for girls and boys.

In this study, we used longitudinal data across 2 academic years from primary and secondary schools to identify the reciprocal associations between beliefs about the malleability of self-control and gaming behavior via a cross-lagged panel model (CLPM). This study aimed to answer the research questions "Does believing one's self-control is unchangeable predict more severe gaming disorder symptoms after one year?" and "Does the severity of gaming disorder symptoms prospectively predict self-control mindsets?" Additionally, we aimed to explore the gender differences in the reciprocal relationships between self-control mindsets and gaming disorder by examining the cross-lagged models among girl and boy participants, respectively.

Methods

Procedure and Participants

This was a 2-wave, longitudinal, and school-based survey study. This study crossed 2 academic years, with surveys administered in June 2021 (T1) and June 2022 (T2). We sent invitations to 23 primary and secondary schools across Hong Kong. School invitations stopped when 4 primary schools and 11 secondary schools agreed to participate. Students from grades 4 - 6 at the primary schools and students from grades 7 - 11 at the secondary schools were invited to participate. The information sheet and parental consent forms detailing research objectives, possible risks, and participation benefits were sent to parents through schools.

The surveys were conducted in classrooms by trained research assistants who provided guidance and answered queries when appropriate. Both English and Chinese versions of the questionnaire were provided. Once completed, the questionnaires were packed, sealed, and returned to the laboratory.

A total of 4286 students (aged 10 - 20 years) from 4 primary schools and 11 secondary schools participated in the T1 survey. Students from grades 6 (n=157) and 11 (n=381) who could not be traced at T2 due to graduation were excluded from analyses.

XSL•FO
RenderX

As this study was focused on investigating students' gaming behaviors, we further excluded those reporting they did not play games at both T1 and T2 (n=484). Ultimately, 3264 participants (1539 boys; 338 students in grades 4 - 5 and 2926 students in grades 7 - 10) were included in the data analyses. The baseline demographic characteristics of the final sample of 3264 students are displayed in Table 1.

Table . Baseline characteristics of participants (N=3264).

Variable	Value
Age (years), mean (SD)	14.02 (1.64)
Gender, n (%)	
Male	1539 (47.2)
Female	1725 (52.8)
Grade, n (%)	
Grade 4 (primary)	180 (5.5)
Grade 5 (primary)	158 (4.8)
Grade 7 (secondary)	998 (30.6)
Grade 8 (secondary)	765 (23.4)
Grade 9 (secondary)	634 (19.4)
Grade 10 (secondary)	529 (16.2)
Ethnicity, n (%)	
Chinese	3200 (98)
Non-Chinese	64 (2)
Family structure, n (%)	
With both parents	2581 (79.1)
With single parent	549 (16.8)
With neither parent	123 (3.8)
Missing	11 (0.3)
Socioeconomic status, n (%)	
Low family affluence	25 (0.8)
Medium family affluence	2670 (81.8)
High family affluence	410 (12.6)
Missing	159 (4.9)

Among the total 3264 participants, 2834 participants (1347 boys; 86.8% of the T1 sample) from the 4 primary schools (n=303) and 11 secondary schools (n=2531) remained in the T2 survey while 430 participants were lost to follow up. Attrition analysis was conducted between participants who participated in both time points (n=2834) and only T1 (n=430), and only significant age differences (*P*=.049) were found. Specifically, those who were older were more likely to drop out at T2. Thus, age was added as an auxiliary variable [27] in our cross-lagged analyses later.

Measures

Gaming disorder symptoms were measured using an adapted version of the Game Addiction Scale (GAS) for Chinese youths consisting of 7 items [28,29]. The GAS was developed, conceptually, based on the diagnostic criteria for IGD in the *DSM-5* [29]. A sample item is "I have thought all day long about playing a game." Respondents needed to indicate the extent of their endorsement of the 7 symptoms on a 6-point Likert scale (1=strongly disagree, 6=strongly agree). Higher mean GAS scores represented more severe gaming disorder symptoms. Both Cronbach α and McDonald ω were 0.85 at T1 and 0.86 at T2, showing good reliability.

Mindset of self-control or self-control mindset (MS) was measured using an adapted version of the 4-item Entity Beliefs Subscale of the Implicit Theories of Intelligence Scale [30]. The sentences referred to the descriptions in the original scale but included the word "self-control" instead of "intelligence," such as "People have a certain amount of self-control, and there isn't much they can do to change it." The 4 items were rated on a 6-point Likert scale (1=strongly disagree, 6=strongly agree). The mean score indicated respondents' self-control mindset (ie, the extent to which they believe their self-control is fixed and unchangeable). The higher the mean score, the more fixed the respondents' self-control mindset was, while a lower mean score indicated more of a growth self-control mindset (ie, belief in changeability of self-control). The reliability of this scale was

XSL•FO
RenderX

also adequately high, as both Cronbach α and McDonald ω were 0.85 at T1 and 0.84 at T2.

Self-control was measured by 9 items from the Brief Self-Control Scale [31,32], which assesses individual differences in people's traits of self-control. Based on the confirmatory factor analysis results of Unger et al [32] for the Chinese version of Tangney's Self-Control Scale, the 13 items in the Brief Self-Control measure belonged to 4 factors (5 items on general capacity for self-discipline, 2 items on deliberate or nonimpulsive action, 2 items on healthy habits, and 4 items on work ethics). We maintained 1 item that showed the largest loading on the factor of general capacity for self-discipline ("I have a hard time breaking bad habits") and kept all 8 items on other factors. The items were also measured on a 6-point Likert scale (1=strongly disagree, 6=strongly agree). Some items were reverse-scored, and a higher mean score on all 9 items represented better self-control. This scale's Cronbach α was 0.80 for T1 and 0.81 for T2, and the McDonald ω was 0.81 for T1 and 0.82 for T2, indicating good reliability as well.

Gaming time (GT) was measured by a single item that asked participants to report their average time (ie, none, half an hour, 1 hour, 2 hours, 3 hours, 4 hours, 5 hours, and more than 5 hours) spent playing video games each day.

Sociodemographic information including gender, age, grade, ethnicity, family structure, and socioeconomic status was collected. Family structure was measured as whether participants lived with both, one, or neither of their parents. The specific details of measuring socioeconomic status and the questionnaires used are reported in Multimedia Appendix 1.

Attention-checking questions were added to the T2 survey to screen out participants not taking the questionnaire seriously. Three items were distributed throughout the questionnaire [33] and instructed respondents to choose one specific option such as choosing "strongly agree" among the 5 choices from 1=strongly disagree to 5=strongly agree.

Statistical Analysis

All analyses in this study were conducted in SPSS (version 26; IBM Corp) and R (R Foundation), with the CLPM analyses conducted using the *lavaan* package in R. Descriptive statistical analysis was conducted for each variable based on the available data. Independent samples *t* tests (2-tailed) were performed to compare gender differences in the severity of gaming disorder symptoms (GAS) and self-control mindset at baseline. After this, Pearson correlation analyses were used to measure the correlations between all variables. Next, CLPM analysis using structural equation modeling was performed to examine the directions of relationships between GAS and self-control mindset. Measurement invariance across the 2 waves was first tested for all variables except GT (single-item assessment), as a prerequisite for the CLPM analyses. To determine which level of measurement invariance was upheld among the configural, metric, scalar, and residual models, a decrease of the comparative fit index (CFI) of no more than 0.01 [34] was regarded as the more parsimonious model being satisfied. Residual invariance was turned out to be upheld for all variables including GAS, self-control mindset, and self-control. Thus, the scale scores of these constructs can be compared across time.

The CLPM analyses with self-control mindset and GAS as latent variables were then conducted, with scale items serving as indicators for each latent variable (self-control mindset and GAS), which helps partially out measurement errors and test all pathways of research interest at the same time [34]. The maximum likelihood estimation with robust SEs was used for the structural equation modeling [35]. Model fit was evaluated using the following criteria [34]: the CFI (acceptable >0.90 and good >0.95), the root mean square error of approximation (RMSEA; acceptable <0.08 and good <0.05), and the standardized root mean square residual (SRMR; acceptable <0.08 and good <0.05).

As we added attention-checking questions in the T2 assessment to ensure data quality, those not answering all 3 attention-checking questions correctly (n=645) were seen as failing attention-checking, and their T2 data were treated as missing, as the same as those lost to follow up at T2 (n=430). When conducting the CLPM analyses, missing data was handled using full information maximum likelihood procedures [36] with age as an auxiliary variable [27].

To test whether the cross-lagged effects were different between boys and girls, subgroup CLPM analyses with age as the auxiliary variable were conducted for girl and boy participants, respectively. We also conducted CLPM analyses to investigate the directions of the relationships between self-control mindset and GT and between self-control mindset and self-control following the same procedures set out above.

Ethical Considerations

Ethical approval was obtained from the Human Subjects Ethics Sub-Committee of the Hong Kong Polytechnic University (reference: HSEARS20210414004-01). Parental and respondents' written consent was obtained before this study. Students were assured that their participation was voluntary, that they could withdraw at any time, and that their responses would not be accessed by teachers or parents. All personally identifiable data were stored on a secure server and are to be destroyed 3 years after the conclusion of this study. We made sure participants' personal data were not identifiable in our research report. We did not apply any images in this paper or supplementary material that could identify the individual participants. Each participant received a piece of stationery (worth about US $5) after completion of each survey as compensation.

Results

The mean GAS score of the sample at baseline was 2.85 with a SD of 1.00, slightly below the midpoint of the scale (midpoint=3.5). The mean baseline GAS score for boys was 3.01 (SD 1) while for girls it was 2.70 (SD 0.97), signifying significantly lower gaming disorder symptom severity in girls versus boys (P<.001). Regarding self-control mindset, the mean score at baseline was 3.26 (SD 1.01), which indicates the mindsets among students overall were average, with neither overly fixed nor growth mindsets prevailing. Furthermore, the

XSL•FO
RenderX

mean baseline self-control mindset score for boys was 3.22 (SD 1.06) and for girls was 3.30 (SD 0.96), indicating a more fixed mindset at baseline in girls than in boys (P=.02).

The descriptive statistics and correlations between the variables across the 2 waves are presented in Table 2. All the variables were significantly correlated in the expected directions. The CLPM for GAS and self-control mindset fit the data (χ^2_{223}=1822.0, P<.001; CFI =0.94; RMSEA=0.047, 90% CI 0.045 to 0.049; SRMR=0.071). The estimates in the CLPM are summarized in Table 3. The results indicated GAS and self-control mindset were significantly related in both T1 and T2. Both stability paths were also statistically significant, and the 2 cross-lagged paths were significant as well. Self-control mindset predicted GAS over time, and GAS also prospectively predicted self-control mindset. For the CLPM analyses between self-control mindset and GT (Table 4; fit measures: χ^2_{37}=225.3, P<.001; CFI=0.98; RMSEA=0.039, 90% CI 0.035 to 0.045; SRMR=0.054), and between self-control mindset and self-control (Table 5; fit measures: χ^2_{317}=2740.7, P<.001; CFI=0.91; RMSEA=0.048, 90% CI 0.047 to 0.050; SRMR=0.081), it was found that GT and self-control predicted self-control mindset after 1 year, but self-control mindset did not predict them 1 year later.

Table . Descriptive statistics and correlation analysis of the variables. All correlations were significant at the P<.001 level.

	Mean (SD)	Self-control[a] T1[b]	MS[c] T1	GAS[d] T1	GT[e] T1	Self-control T2[f]	MS T2	GAS T2	GT T2
Self-control T1	3.47 (0.81)	—[g]	−0.47	−0.52	−0.14	0.68	−0.34	−0.39	−0.087
MS T1	3.26 (1.01)	−0.47	—	0.33	0.10	−0.33	0.44	0.28	0.088
GAS T1	2.85 (1)	−0.52	0.33	—	0.43	−0.41	0.26	0.62	0.32
GT T1	1.95 (1.77)	−0.14	0.10	0.43	—	−0.11	0.12	0.31	0.58
Self-control T2	3.32 (0.79)	0.68	−0.33	−0.41	−0.11	—	−0.45	−0.52	−0.11
MS T2	3.27 (0.98)	−0.34	0.44	0.26	0.12	−0.45	—	0.35	0.13
GAS T2	2.85 (0.98)	−0.39	0.28	0.62	0.31	−0.52	0.35	—	0.41
GT T2	1.82 (1.66)	−0.087	0.088	0.32	0.58	−0.11	0.13	0.41	—

[a]Range: 1-6; a higher value represents higher self-control.

[b]T1: time 1.

[c]MS: mindset regarding self-control (range: 1-6; a higher value represents a more fixed mindset).

[d]GAS: Game Addiction Scale (range: 1-6; a higher value represents more severe gaming disorder symptoms).

[e]GT: gaming time (range: 0-6; 0=none, 0.5=half hour, 1=1 hour, 2=2 hours, 3=3 hours, 4=4 hours, 5=5 hours, and 6=more than 5 hours; a higher value represents more average time spent playing video games each day).

[f]T2: time 2.

[g]Not applicable.

XSL•FO
RenderX

Table . Summary of the CLPM[a] results between self-control mindsets and gaming disorder among the whole sample (N=3264).

	Covariance or coefficient (95% CI)	*P* value	Correlation or standardized coefficient
Concurrent paths [b]			
MS[c] T1[d] and GAS[e] T1	0.36 (0.32-0.4)	<.001	0.36
MS T2[f] and GAS T2	0.38 (0.31-0.44)	<.001	0.38
Autoregressive paths [g]			
MS T1 to MS T2	0.46 (0.4-0.51)	<.001	0.41
GAS T1 to GAS T2	0.62 (0.57-0.67)	<.001	0.52
Cross-lagged paths [g]			
MS T1 to GAS T2	0.070 (0.02-0.12)	.006	0.059
GAS T1 to MS T2	0.11 (0.06-0.16)	<.001	0.099

[a]CLPM: cross-lagged panel model.

[b]The values represent the covariance or correlation coefficients.

[c]MS: mindset regarding self-control.

[d]T1: time 1.

[e]GAS: Game Addiction Scale.

[f]T2: time 2.

[g]The values represent the regression coefficients or standardized regression coefficients.

Table . Summary of the CLPM[a] results between self-control mindsets and gaming time in the whole sample (N=3264).

	Covariance or coefficient (95% CI)	*P* value	Correlation or standardized coefficient
Concurrent paths [b]			
MS[c] T1[d] and GT[e] T1	0.2 (0.14 to 0.27)	<.001	0.12
MS T2[f] and GT T2	0.12 (0.048 to 0.2)	.001	0.091
Autoregressive paths [g]			
MS T1 to MS T2	0.5 (0.45 to 0.55)	<.001	0.44
GT T1 to GT T2	0.56 (0.52 to 0.6)	<.001	0.59
Cross-lagged paths [g]			
MS T1 to GT T2	0.059 (−0.005 to 0.12)	.07	0.035
GT T1 to MS T2	0.021 (0.002 to 0.04)	.03	0.033

[a]CLPM: cross-lagged panel model.

[b]The values represent the covariance or correlation coefficients.

[c]MS: mindset regarding self-control.

[d]T1: time 1.

[e]GT: gaming time.

[f]T2: time 2.

[g]The values represent the regression coefficients or standardized regression coefficients.

XSL•FO
RenderX

Table . Summary of the CLPM[a] results between self-control and self-control mindsets among the whole sample (N=3264).

	Covariance or coefficient (95% CI)	*P* value	Correlation or standardized coefficient
Concurrent paths [b]			
MS[c] T1[d] and self-control T1	−0.52 (−0.55 to −0.48)	<.001	−0.52
MS T2[e] and self-control T2	−0.54 (−0.59 to −0.48)	<.001	−0.54
Autoregressive paths [f]			
MS T1 to MS T2	0.42 (0.36 to 0.48)	<.001	0.38
Self-control T1 to self-control T2	0.66 (0.61 to 0.71)	<.001	0.55
Cross-lagged paths [f]			
MS T1 to self-control T2	−0.013 (−0.063 to 0.038)	.63	−0.01
Self-control T1 to MS T2	−0.15 (−0.21 to −0.089)	<.001	−0.13

[a]CLPM: cross-lagged panel model.

[b]The values represent the covariance or correlation coefficients.

[c]MS: mindset regarding self-control.

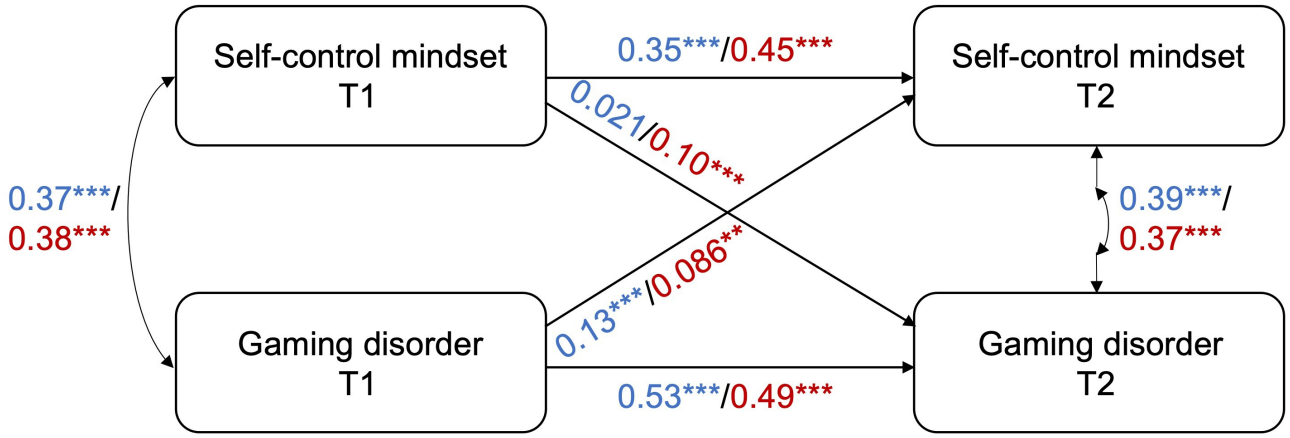
[d]T1: time 1.

[e]T2: time 2.

[f]The values represent the regression coefficients or standardized regression coefficients.

Subgroup CLPM analyses for girls and boys all had adequate fit as well (Table S1 in Multimedia Appendix 1). The CLPM results between self-control mindset and gaming disorder indicated that both the cross-lagged paths were significant in girl students, however, for boys, only the cross-lagged path from GAS in T1 to self-control mindset in T2 was significant, while the opposite was not statistically significant (Figure 1 and Tables S2 and S3 in Multimedia Appendix 1), signifying distinct patterns for the 2 genders. The CLPM results between self-control mindset and GT showed that only the cross-lagged path from GT at T1 to self-control mindset at T2 in girls was significant (Figure 2 and Tables S4 and S5 in Multimedia Appendix 1). For the cross-lagged effects between self-control mindset and self-control, self-control at T1 predicted self-control mindset at T2 in both boys and girls, while self-control mindset at T1 did not predict self-control at T2 (Figure 3 and Tables S6 and S7 in Multimedia Appendix 1).

Figure 1. CLPM results between self-control mindset and gaming disorder for boys and girls separately. The values along the lines are the standardized coefficients of the relevant paths for boys (in blue) and girls (in red). **P*≤.05, ***P*≤.01, ****P*≤.001. Exact *P* values are reported in Tables S2 and S3 in Multimedia Appendix 1. CLPM: cross-lagged panel model; T: time.



XSL•FO
RenderX

Figure 2. CLPM results between self-control mindset and gaming time for boys and girls separately. The values along the lines are the standardized coefficients of the relevant paths for boys (in blue) and girls (in red). *P≤.05, **P≤.01, ***P≤.001. Exact P values are reported in Tables S4 and S5 in Multimedia Appendix 1. CLPM: cross-lagged panel model; T: time.

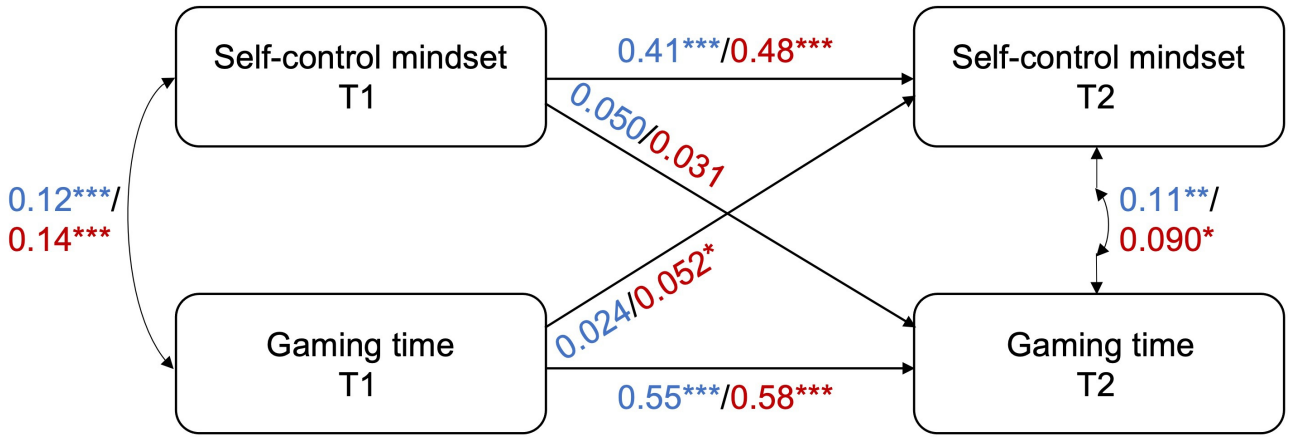
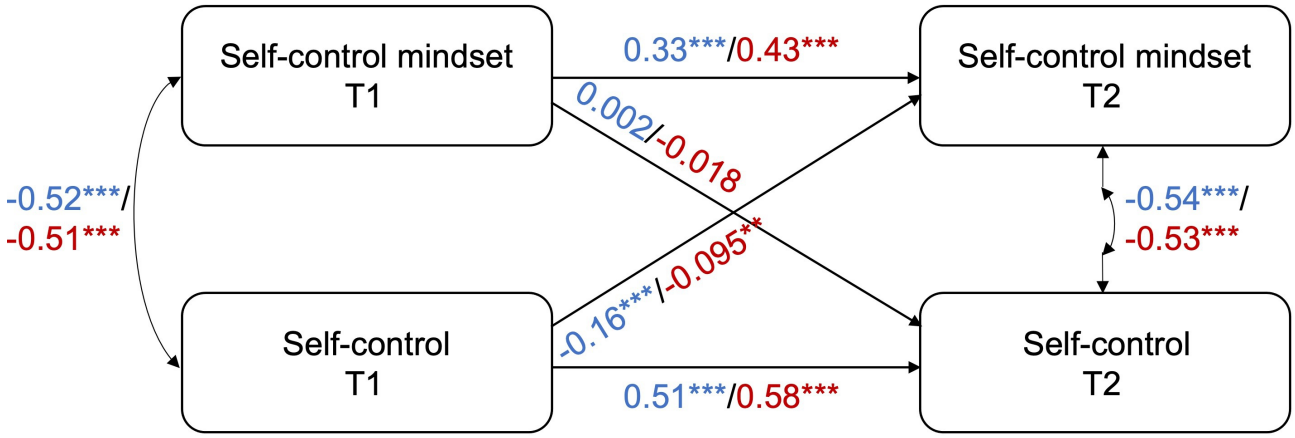


Figure 3. CLPM results between self-control mindset and self-control for boys and girls separately. The values along the lines are the standardized coefficients of the relevant paths for boys (in blue) and girls (in red). *P≤.05, **P≤.01, ***P≤.001. Exact P values are reported in Tables S6 and S7 in Multimedia Appendix 1. CLPM: cross-lagged panel model, T: time.



Discussion

Principal Findings

Gaming disorder is a growing global concern. It is essential to identify the modifiable psychosocial factors influencing gaming disorder to develop suitable prevention strategies. The reciprocal associations between self-control mindsets and gaming disorder were examined with the current longitudinal study among children and adolescents in Hong Kong. The initial evidence of the directions of the aforementioned longitudinal associations helps identify self-control mindsets as a potential key psychological factor for gaming disorder prevention. The findings revealed substantive concurrent associations between fixed mindsets regarding self-control and severe gaming disorder symptoms at both assessment points. Moreover, the 1-year longitudinal association between fixed mindsets regarding self-control and gaming disorder has uncovered the long-term impact of self-control mindsets on gaming disorder and that gaming disorder could diminish one's belief regarding the changeability of self-control.

The separate examinations among the 2 genders further unveiled the gender differences. The significant longitudinal reciprocal relationship only exists among girls. However, among boys, the gaming disorder symptom severity prospectively predicts fixed mindsets regarding self-control, but not reversely, which means boys' gaming disorder symptoms predict their belief of whether they can improve their self-control.

To our knowledge, this is the first study on the reciprocal relationships between implicit theories and gaming disorder over time, and it provides important insights into this field of research. It is also a relatively large-scale study that is school-based and contains representative samples, adding to the applicability and generalizability of the findings.

Comparison to Prior Work

Our study addresses a research gap by clarifying the directional links between gaming disorder and implicit theories of self-control. This study's results revealed the bidirectional relationships between them. Regarding the prediction of self-control mindsets to gaming disorder symptoms after 1 year, our findings were consistent with the results of previous mindset intervention studies. For instance, a brief growth mindset intervention helped to reduce the incidence of clinically significant depressive symptoms during a 9-month follow-up, while depressive symptoms increased among adolescents with a fixed mindset of personality in the control condition [37]. In another study, a brief growth mindset of intelligence intervention was shown to improve secondary school students' academic

XSL•FO
RenderX

achievement and increase their enrollment in advanced courses [38]. This emerging body of research exhibits the effects of growth mindsets on mental health and performance-related outcomes. On the other hand, the findings regarding the alternative pathway of gaming disorder symptoms in predicting later self-control mindsets accorded with past findings that the baseline psychopathology of youth predicted increased fixed mindsets over time [23,30]. Synthesizing the CLPM results of this study, apart from gaming disorder symptoms, GT (especially in girls) and self-control levels also predicted later self-control mindsets, which denotes that mindsets are affected by various factors. The findings of this study demonstrated that experiences like addictive behaviors shape youths' fixed mindsets. Like the findings supporting growth mindsets of intelligence are both antecedents and outcomes of greater academic achievement in youths [39], the bidirectional links between gaming disorder symptoms and implicit theories of self-control identified in our study underscore the reciprocal influences of these factors.

The association between self-control mindsets and gaming behaviors may be explained by the extant literature on how implicit theories impact self-regulation process [21,22]. First, it is postulated that growth mindsets regarding self-control could lead to reduced gaming disorder symptoms through self-reinforcing cycles of motivation and dedicated efforts to try new strategies [38]. Though self-control has been shown to have no causal relationships with gaming disorder by our recent study [40], individual factors encompassing self-control dimensions and gaming motivation explained substantive variance in gaming disorder, much more than gaming-related factors including game genre and platforms [41]. Second, the predictive path of self-control mindsets to gaming disorder symptoms was evident in girls but not in boys in the subgroup analyses, signifying that girls' mindsets shall be more related to gaming behavior regulation. This finding aligned with past research findings regarding gender differences in gaming motives [14,42]. As boys score higher on all gaming motives except escape [14], strong gaming motives might undermine the possible effects of a growth mindset regarding self-control in preventing gaming disorder among boys. The gender differences found in this study also highlight the importance of gender in gaming [26] and more studies are needed to clarify the distinct gaming patterns among the 2 genders.

Implications

The findings of this study have significant practical implications. First, a self-control mindset is a modifiable factor that can be used for developing gaming disorder prevention and early intervention. Given the strong concurrent and longitudinal association between self-control mindsets and gaming disorder revealed in this study, interventions that instill the belief that their self-control can become stronger may be potent to help prevent gaming disorder. Second, childhood and adolescence are ideal developmental stages for developing healthy self-regulatory digital product use. A growth mindset regarding self-control helps strengthen youths' self-efficacy in self-control, especially for girls, and it is promisingly helpful to prevent students from developing gaming disorder in the future. Last but not least, interventions addressing self-control mindsets could be used as motivational interviewing and could be added to existing intervention programs or therapies for gaming disorder to boost patients' confidence and motivation to regain control over their gaming behaviors. A top-up session addressing self-control mindsets will be likely to amplify the efficacy of the original interventions.

Limitations and Future Directions

There are limitations to this study. First, causality cannot be inferred from this study. Self-control mindsets and gaming disorder could still be causally unrelated, and their association could be accounted for by other variables. A recent study revealed a unidirectional negative relationship from self-control to gaming disorder via traditional CLPM analysis but no prospective relationship in a random intercept CLPM analysis [40]. Longitudinal studies containing more waves of data with applying more rigorous analysis methods and studies with experimental designs are needed to inform us better of the causality between self-control mindsets and gaming disorder. Second, though this study's main aim was not the underlying mediators in the relationship between mindsets of self-control and gaming disorder but the directional relationship between them, longitudinal mediation models need to be tested in future research to figure out the relationships between mindsets, self-control, and gaming disorder. Third, the measures including participants' self-control mindsets and gaming disorder symptoms were self-reported; as such, there may be subjective biases influencing the objectivity of this study's findings, although we used attention-checking questions to ensure the quality of the survey. Moreover, GT cannot be perfectly assessed by asking participants to indicate their average GT each day using a scale with several discrete durations; rather, screen monitor tools should be used to record exact GT in future research. Fourth, this study's sample was just Hong Kong children and adolescents and thus the findings might not be able to generalize to a wider population. Future research should validate our study findings with larger and more representative samples.

Conclusion

Internet gaming plays an important role in leisure and educational activities in the digital age. This study is a pioneering investigation into the reciprocal relationships between implicit theories and gaming disorder. This study's results highlight that gaming disorder was associated with a more fixed mindset about self-control after one year in both genders, while only girls' fixed mindsets regarding self-control significantly led to subsequent gaming disorder symptoms after 1 year. The findings lay the groundwork for future research and practices for cultivating growth mindsets to prevent adolescents from developing gaming disorder.

XSL•FO
RenderX

Acknowledgments

We would like to thank the participants and schools for their participation. We would also like to acknowledge the contribution of Xiaomin Li, PhD, as a consultant for handling missing data. This work was supported by General Research Fund projects funded by the Hong Kong Research Grants Council (grants 15601621 and 15608322).

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

SZ handled the study concept and design, methodology and interpretation of data, funding acquisition, study supervision, paper writing, and reviewing and revision. DQ worked on the methodology, data analysis and interpretation of data, and paper writing and revision.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary tables.
[DOCX File, 49 KB - games_v13i1e59441_app1.docx]

References

1. Share of internet users worldwide who play video games on any device as of 3rd quarter 2022, by age group and gender. Statista. 2023. URL: https://www.statista.com/statistics/326420/console-gamers-gender [accessed 2024-12-24]
2. Sugaya N, Shirasaka T, Takahashi K, Kanda H. Bio-psychosocial factors of children and adolescents with internet gaming disorder: a systematic review. Biopsychosoc Med 2019;13(1):3. [doi: 10.1186/s13030-019-0144-5] [Medline: 30809270]
3. Gao YX, Wang JY, Dong GH. The prevalence and possible risk factors of internet gaming disorder among adolescents and young adults: systematic reviews and meta-analyses. J Psychiatr Res 2022 Oct;154:35-43. [doi: 10.1016/j.jpsychires.2022.06.049] [Medline: 35926424]
4. Wang CW, Chan CLW, Mak KK, Ho SY, Wong PWC, Ho RTH. Prevalence and correlates of video and internet gaming addiction among Hong Kong adolescents: a pilot study. Sci World J 2014;2014(874648):874648. [doi: 10.1155/2014/874648] [Medline: 25032242]
5. King DL, Delfabbro PH, Billieux J, Potenza MN. Problematic online gaming and the COVID-19 pandemic. J Behav Addict 2020 Jun;9(2):184-186. [doi: 10.1556/2006.2020.00016] [Medline: 32352927]
6. Zhu S, Zhuang Y, Lee P, Li JCM, Wong PWC. Leisure and problem gaming behaviors among children and adolescents during school closures caused by COVID-19 in Hong Kong: quantitative cross-sectional survey study. JMIR Serious Games 2021 May 7;9(2):e26808. [doi: 10.2196/26808] [Medline: 33960954]
7. Number of video game users worldwide from 2017 to 2027 (in billions). Statista. 2024. URL: https://www.statista.com/statistics/1344693/video-games-users-in-the-world-forecast/ [accessed 2024-12-24]
8. van Rooij AJ, Kuss DJ, Griffiths MD, Shorter GW, Schoenmakers MT, van de Mheen D. The (co-)occurrence of problematic video gaming, substance use, and psychosocial problems in adolescents. J Behav Addict 2014 Sep;3(3):157-165. [doi: 10.1556/JBA.3.2014.013] [Medline: 25317339]
9. Wu AMS, Chen JH, Tong KK, Yu S, Lau JTF. Prevalence and associated factors of Internet gaming disorder among community dwelling adults in Macao, China. J Behav Addict 2018 Mar 1;7(1):62-69. [doi: 10.1556/2006.7.2018.12] [Medline: 29463097]
10. Gao XJ, Sun JJ, Xiang M. Positive psychological intervention for anxiety, depression and coping in subjects addicted to online games. World J Clin Cases 2021 May 16;9(14):3287-3293. [doi: 10.12998/wjcc.v9.i14.3287] [Medline: 34002137]
11. Seay AF, Kraut RE. Project massive: self-regulation and problematic use of online gaming. Presented at: CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems; Apr 28 to May 3, 2007; San Jose, California, USA. [doi: 10.1145/1240624.1240749]
12. Diagnostic and Statistical Manual of Mental Disorders, 5th edition: American Psychiatric Association; 2013.
13. 6C51 Gaming disorder. ICD-11 for Mortality and Morbidity Statistics. 2018. URL: https://icd.who.int/dev11/l-m/en#/http%3a%2f%2fid.who.int%2ficd%2fentity%2f1448597234 [accessed 2024-12-24]
14. Cudo A, Wojtasiński M, Tużnik P, Fudali-Czyż A, Griffiths MD. The relationship between depressive symptoms, loneliness, self-control, and gaming disorder among polish male and female gamers: the indirect effects of gaming motives. Int J Environ Res Public Health 2022 Aug 22;19(16):10438. [doi: 10.3390/ijerph191610438] [Medline: 36012072]
15. Mills DJ, Allen JJ. Self-determination theory, internet gaming disorder, and the mediating role of self-control. Comput Human Behav 2020 Apr;105:106209. [doi: 10.1016/j.chb.2019.106209]

XSL•FO
RenderX

16. Yu Y, Mo PKH, Zhang J, Li J, Lau JTF. Impulsivity, self-control, interpersonal influences, and maladaptive cognitions as factors of internet gaming disorder among adolescents in China: cross-sectional mediation study. J Med Internet Res 2021 Oct 27;23(10):e26810. [doi: 10.2196/26810] [Medline: 34704960]
17. Lee YH, Heeter C, Magerko B, Medler B. Gaming mindsets: implicit theories in serious game learning. Cyberpsychol Behav Soc Netw 2012 Apr;15(4):190-194. [doi: 10.1089/cyber.2011.0328] [Medline: 22165916]
18. Zhu S, Zhuang Y, Cheung SH. Domain specificity or generality: assessing the Chinese implicit theories scale of six fundamental psychological attributes. Front Psychol 2020;11:142. [doi: 10.3389/fpsyg.2020.00142] [Medline: 32116943]
19. Dweck CS. Can personality be changed? The role of beliefs in personality and change. Curr Dir Psychol Sci 2008 Dec;17(6):391-394. [doi: 10.1111/j.1467-8721.2008.00612.x]
20. Molden DC, Dweck CS. Finding “meaning” in psychology: a lay theories approach to self-regulation, social perception, and social development. Am Psychol 2006 Apr;61(3):192-203. [doi: 10.1037/0003-066X.61.3.192] [Medline: 16594836]
21. Burnette JL, O’Boyle EH, VanEpps EM, Pollack JM, Finkel EJ. Mind-sets matter: a meta-analytic review of implicit theories and self-regulation. Psychol Bull 2013 May;139(3):655-701. [doi: 10.1037/a0029531] [Medline: 22866678]
22. Mrazek AJ, Ihm ED, Molden DC, Mrazek MD, Zedelius CM, Schooler JW. Expanding minds: growth mindsets of self-regulation and the influences on effort and perseverance. J Exp Soc Psychol 2018 Nov;79:164-180. [doi: 10.1016/j.jesp.2018.07.003]
23. Schleider JL, Weisz JR. Implicit theories relate to youth psychopathology, but how? a longitudinal test of two predictive models. Child Psychiatry Hum Dev 2016 Aug;47(4):603-617. [doi: 10.1007/s10578-015-0595-2] [Medline: 26443503]
24. Stevens MW, Dorstyn D, Delfabbro PH, King DL. Global prevalence of gaming disorder: a systematic review and meta-analysis. Aust N Z J Psychiatry 2021 Jun;55(6):553-568. [doi: 10.1177/0004867420962851] [Medline: 33028074]
25. Bonnaire C, Serehen Z, Phan O. Effects of a prevention intervention concerning screens, and video games in middle-school students: influences on beliefs and use. J Behav Addict 2019 Sep 1;8(3):537-553. [doi: 10.1556/2006.8.2019.54] [Medline: 31537087]
26. López-Fernández FJ, Mezquita L, Griffiths MD, Ortet G, Ibáñez MI. The role of personality on disordered gaming and game genre preferences in adolescence: gender differences and person-environment transactions. Adicciones 2021 Jul 1;33(3):263-272. [doi: 10.20882/adicciones.1370] [Medline: 32100046]
27. Graham JW. Adding missing-data-relevant variables to FIML-based structural equation models. Struct Equ Modeling 2003 Jan;10(1):80-100. [doi: 10.1207/S15328007SEM1001_4]
28. Hsiao YH. The relationship between the mobile game using behaviors and positive mental health among taiwanese adolescents [Master’s thesis]. : National Taiwan Normal University; 2016. [doi: 10.6345/NTNU202201890]
29. Khazaal Y, Chatton A, Rothen S, et al. Psychometric properties of the 7-item game addiction scale among French and German speaking adults. BMC Psychiatry 2016 May 10;16:132. [doi: 10.1186/s12888-016-0836-3] [Medline: 27160387]
30. Dweck CS. Self-Theories: Their Role in Motivation, Personality, and Development: Psychology Press; 1999.
31. Tangney JP, Baumeister RF, Boone AL. High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. J Pers 2004 Apr;72(2):271-324. [doi: 10.1111/j.0022-3506.2004.00263.x] [Medline: 15016066]
32. Unger A, Bi C, Xiao YY, Ybarra O. The revising of the Tangney Self-Control Scale for Chinese students. Psych J 2016 Jun;5(2):101-116. [doi: 10.1002/pchj.128] [Medline: 27144924]
33. Schroder HS, Dawood S, Yalch MM, Donnellan MB, Moser JS. The role of implicit theories in mental health symptoms, emotion regulation, and hypothetical treatment choices in college students. Cognit Ther Res 2015 Apr;39(2):120-139. [doi: 10.1007/s10608-014-9652-6] [Medline: 35474696]
34. Kline RB. Principles and Practice of Structural Equation Modeling: Guilford Publications; 2015.
35. Marshall SK, Faaborg-Andersen P, Tilton-Weaver LC, Stattin H. Peer sexual harassment and deliberate self-injury: longitudinal cross-lag investigations in Canada and Sweden. J Adolesc Health 2013 Dec;53(6):717-722. [doi: 10.1016/j.jadohealth.2013.06.009] [Medline: 23890776]
36. Acock AC. Working with missing values. J Marriage Fam 2005 Nov;67(4):1012-1028. [doi: 10.1111/j.1741-3737.2005.00191.x]
37. Miu AS, Yeager DS. Preventing symptoms of depression by teaching adolescents that people can change: effects of a brief incremental theory of personality intervention at 9-month follow-up. Clin Psychol Sci 2015;3(5):726-743. [doi: 10.1177/2167702614548317]
38. Yeager DS, Hanselman P, Walton GM, et al. A national experiment reveals where a growth mindset improves achievement. Nat New Biol 2019 Sep;573(7774):364-369. [doi: 10.1038/s41586-019-1466-y] [Medline: 31391586]
39. Jones BD, Wilkins JLM, Long MH, Wang F. Testing a motivational model of achievement: how students’ mathematical beliefs and interests are related to their achievement. Eur J Psychol Educ 2012 Mar;27(1):1-20. [doi: 10.1007/s10212-011-0062-9]
40. Qi D, Li X, Zhu S. Unpacking the myth in the associations between self-control and gaming disorder in children and adolescents: a comparison between traditional and random intercept cross-lagged panel model analyses. Int J Ment Health Addict ;2024:1-15. [doi: 10.1007/s11469-024-01294-0]

XSL•FO
RenderX

41. Cudo A, Kopiś-Posiej N, Griffiths MD. The role of self-control dimensions, game motivation, game genre, and game platforms in gaming disorder: cross-sectional and longitudinal findings. Psychol Res Behav Manag 2023;16:4749-4777. [doi: 10.2147/PRBM.S435125] [Medline: 38024654]
42. Wilhelm C. Gender role orientation and gaming behavior revisited: examining mediated and moderated effects. Inf Comm Soc 2018 Feb;21(2):224-240. [doi: 10.1080/1369118X.2016.1271902]

Abbreviations

CFI: comparative fit index
CLPM: cross-lagged panel model
***DSM-5*:** *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition*
GAS: Game Addiction Scale
GT: gaming time
***ICD-11*:** *International Classification of Diseases, Eleventh Edition*
IGD: internet gaming disorder
RMSEA: root mean square error of approximation
SRMR: standardized root mean square residual

Edited by A Coristine, T Leung; submitted 12.04.24; peer-reviewed by FJ Lopez-Fernandez, MB Ellur, S Ni, S Hewage; revised version received 30.10.24; accepted 01.11.24; published 15.01.25.

Please cite as:
Zhu S, Qi D
Reciprocal Relationship Between Self-Control Belief and Gaming Disorder in Children and Adolescents: Longitudinal Survey Study
JMIR Serious Games 2025;13:e59441
URL: https://games.jmir.org/2025/1/e59441
doi:10.2196/59441

© Shimin Zhu, Di Qi. Originally published in JMIR Serious Games (https://games.jmir.org), 15.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Effects of Progressive Aerobic Training on Executive-Reward Network Connectivity and Symptoms of Internet Gaming Disorder: Randomized Controlled Trial

Shaoyu Cui[1*], MPSYD; Xuefeng Xu[2*], PsyD; Xin Luo[1*], MPSYD; Meiting Wei[1], PsyD; Guang-Heng Dong[1], PsyD

[1]Department of Psychology, Yunnan Normal University, 1st Lianda Street, Chenggong District, Kunming, Yunnan Province, China

[2]Faculty of Education, Yunnan Normal University, 1st Lianda Street, Chenggong District, Kunming, Yunnan Province, China

[*]these authors contributed equally

Corresponding Author:
Guang-Heng Dong, PsyD
Department of Psychology, Yunnan Normal University, 1st Lianda Street, Chenggong District, Kunming, Yunnan Province, China

Abstract

Background: Internet gaming disorder (IGD) causes neurocognitive deficits and brain dysfunction. Progressive aerobic training (PAT) seems more practical. However, its effect on IGD and the underlying neural mechanism remains unclear.

Objective: This preregistered, randomized controlled, single-blind study examined the efficacy of a novel non-pharmacological intervention by elucidating the neurocognitive mechanisms in IGD.

Methods: A total of 72 participants with IGD (meeting the DSM-5 criteria for IGD, Internet Addiction Test [IAT] score >50, and no comorbid conditions) were recruited and randomly assigned to a PAT group (received 12 weeks of PAT with the intensity dynamically adjusted based on real-time performance) or a free training (FT) group (completed sessions of equal duration but without a progressive structure, allowing free choice of exercise mode and intensity). No intervention-related adverse events were reported. Sixty-four participants completed the experiment (PAT: 33; FT: 31), including pretreatment and posttreatment functional magnetic resonance imaging scans and 20 PAT sessions in a month. Regional homogeneity and degree centrality are calculated; the overlapping brain regions were used as seed points for functional connectivity (FC) analysis. The correlation between FC and behavioral data and neurotransmitters was also evaluated.

Results: The PAT group demonstrated a significantly greater reduction in IAT scores compared to the FT group (t_{32}=4.333, P_{bonf}<0.001, Cohen d=0.754, 95% CI 0.362-1.137), accompanied by a specific reduction in game craving within the PAT group (t_{32}=2.278, P_{bonf}=0.03, Cohen d=0.397, 95% CI 0.045-0.851]). FC analysis revealed that PAT significantly enhanced FC within the executive control network (ECN), increasing connectivity between the right medial superior frontal gyrus (R-mSFG) and key regions, including the left postcentral gyrus ($F_{1, 62}$=7.95, P=.006), bilateral superior parietal gyrus (right: $F_{1, 62}$=5.68, P=.02; left: $F_{1, 62}$=8.85, P=.004), and left inferior frontal gyrus ($F_{1, 62}$=11.37, P=.001). PAT also strengthened ECN-reward network (RN) integration, enhancing FC between the R-mSFG and bilateral insula (right: $F_{1, 62}$=11.41, P=.001; left: $F_{1, 62}$=7.94, P=.006) and left substantia nigra ($F_{1, 62}$=10.60, P=.002). These neural changes were behaviorally relevant, as pretest game craving positively correlated with post-intervention FC strength between the R-mSFG and left postcentral gyrus (r=0.36, P=.04, 95% CI 0.02 - 0.63) and right precentral gyrus (r=0.40, P=.02, 95% CI 0.07 - 0.65). Furthermore, the FC changes were significantly associated with cannabinoid (CB1) (P=.003) and metabotropic glutamate (mGluR5) receptor distributions (P=.005).

Conclusions: This study demonstrates the efficacy of a progressive and adaptive PAT intervention in reducing IGD severity, moving beyond static protocols by dynamically tailoring intensity to individual performance. The therapeutic effect may be mediated by modulating the functional connectivity between the ECN and the RN, potentially enhancing top-down control. These results, supported by the correlation between FC, behavior, and neurotransmitter systems, indicate that PAT represents a promising non-pharmacological intervention approach worthy of further investigation.

Trial Registration: Chinese clinical trial registry; ChiCTR2400090834; https://www.chictr.org.cn/hvshowprojectEN.html?id=261866

(JMIR Serious Games 2025;13:e83597) doi:10.2196/83597

XSL•FO
RenderX

KEYWORDS

functional connectivity; internet gaming disorder; progressive aerobic training; randomized controlled trial; neural mechanism; brain dysfunction

Introduction

Internet gaming disorder (IGD) has emerged as a severe mental health issue that needs to be addressed [1,2]. Both the *DSM-5* (*The Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition*) and ICD-11 (*The International Classification of Diseases, 11th Revision*) define this disorder as a repetitive or continuous gaming behavior pattern that lasts for at least 12 months, unless severe symptoms occur [3-5]. IGD can lead to severe negative consequences for players, including neurocognitive deficits, impaired executive functioning, and emotional dysregulation [6,7], which in turn generate physical, psychological, and behavioral effects on players.

Neuroimaging studies have identified significant alterations in functional brain networks associated with IGD [8], including impaired executive control over gaming cravings and strong gaming cravings in reward processing-related brain regions [9,10]. Multiple studies have observed relatively decreased functional connectivity (FC) between executive-control–related and reward-related regions [11,12]. One study found that the IGD group showed decreased FC between the right superior frontal gyrus (SFG) and right posterior insula compared to controls [13]. Another found that the IGD group showed significantly decreased FC values between the right middle frontal gyrus (as well as dorsolateral prefrontal cortex; DLPFC) and the right anterior cingulate cortex [11,14]. Moreover, reduced FC strength revealed weakened connectivity in executive-control–related regions, including SFG-DLPFC [15] and DLPFC-postcentral gyrus [16].

Traditional interventions for IGD have focused mainly on psychological, pharmacological, and exercise-based therapies [17]. These treatment strategies need to perform by professional trained psychiatrist and usually expensive and hard to perform. Comparing with these treatment strategies, the exercise-based interventions may be more practical [18] as it is widely acceptable and easily performed in daily life. Exercise intervention, such as the aerobic exercise, is known to increase brain-derived neurotrophic factor, and the increased brain-derived neurotrophic factor enhances cognitive function by modifying neurotransmission, stimulating neurogenesis, and promoting synaptic plasticity in the brain [19]. Aerobic exercise has been proven to restore and regulate nerve cells, enhancing their adaptability to external variations, and optimizing cognitive abilities, executive functions, and inhibitory control, while fostering positive effects and mitigating negative impacts [20-22]. In previous studies on aerobic training for IGD, Liu et al [23] found through systematic review and meta-analysis that mind–body exercises like Tai Chi and Qigong can significantly improve executive function. In a study of training on brain function in methamphetamine addicts, aerobic training temporarily altered the activation of neural networks in the reward system of drug-dependent individuals, effectively inhibiting their drug craving [24]. Rensburg et al [25] explored the effects of training on the brain during withdrawal and suggested that individuals exercising 15 hours after quitting smoking show significant reductions in smoking craving and significant activation of the brain reward system. Meanwhile, physical training normalizes dopaminergic and glutamatergic transmissions [26]. Cannabinoids are thought to play a crucial role in mediating the mood-regulating effects of exercise [27], while mGluR5 is pivotal for the central reward processing in addicted individuals [28]. These evidence supports the hypothesis that aerobic training mitigates IGD by restoring the functional balance between the reward and control networks, a process potentially mediated by its synergistic effects on dopaminergic, glutamatergic, and endocannabinoid systems.

Although extensive research underscores the therapeutic potential of exercise, the neurofunctional effects of fixed, progressive aerobic training (PAT) regimens in individuals with IGD remain uncharted in the published literature. A study on morphine-addicted rats shows that compared to the fixed aerobic protocol, progressive aerobic exercise significantly outperformed in reversing morphine-induced impairments in memory, memory consolidation, and locomotor activity [29]. One study suggests that 24-week PAT could significantly improve the anxiety symptoms, psychological pressure and sleep state of male drug addicts [30]. However, the specific impact of PAT on the neural mechanism with IGD remains unclear.

Both regional homogeneity (ReHo) and degree centrality (DC) are two main resting-state functional magnetic resonance imaging metrics in a way of voxel-wise whole-brain analysis [31,32]. ReHo is a voxel-based measure of brain activity that estimates the degree of synchronization between the time series of a given voxel and its nearest neighbors [33,34]. The validity of this measure is based on the premise that intrinsic brain activity primarily reflects the coordinated activation of voxel clusters rather than isolated single voxels. Consequently, higher ReHo values are thought to represent greater synchronization of local field potentials associated with neuronal activity in the human brain [35]; DC considers Pearson correlation coefficient as metric for FC estimation between each pair of voxels, assigning to each voxel the global number of functional connections between it and all other voxels across the brain [36]. In its more diffuse implementations, it involves the use of a threshold over which a pair of voxels is considered connected [37,38]. The combination of ReHo and DC captures both local and remote neural synchronization [39], demonstrating superior performance over traditional methods for selecting regions of interest (ROIs) in resting-state functional magnetic resonance imaging (fMRI) [40,41].

While IGD research has not employed PAT intervention, dance as a sport enhances frontal-striatal circuit connectivity in individuals with IGD [23]. Research on heroin addiction further suggests that physical exercise serves as an effective non-pharmacological intervention, capable of improving patients' executive control function [42]. Therefore, this study evaluated the efficacy and neural mechanisms of PAT on IGD

XSL•FO
RenderX

through FC and correlation analysis. We hypothesized that PAT would reduce the severity of game craving and addiction severity in IGD by altering the FC of brain regions involved in executive control network (ECN) and reward network (RN).

Methods

Trial Design and Involvement

No patient or public involvement was incorporated into the design, conduct, reporting, or dissemination of this research. This was a prospective, randomized, controlled, single-blind trial with two parallel groups. The research was conducted at the Department of Psychology of Yunnan Normal University in Kunming, China, from November 1, 2024, to December 1, 2024. The study design and reporting adhered to the CONSORT 2025 guidelines, specifically following the CONSORT checklist for reporting information in randomized trials (Checklist 1) [43]. The trial proceeded to its planned completion without early termination.

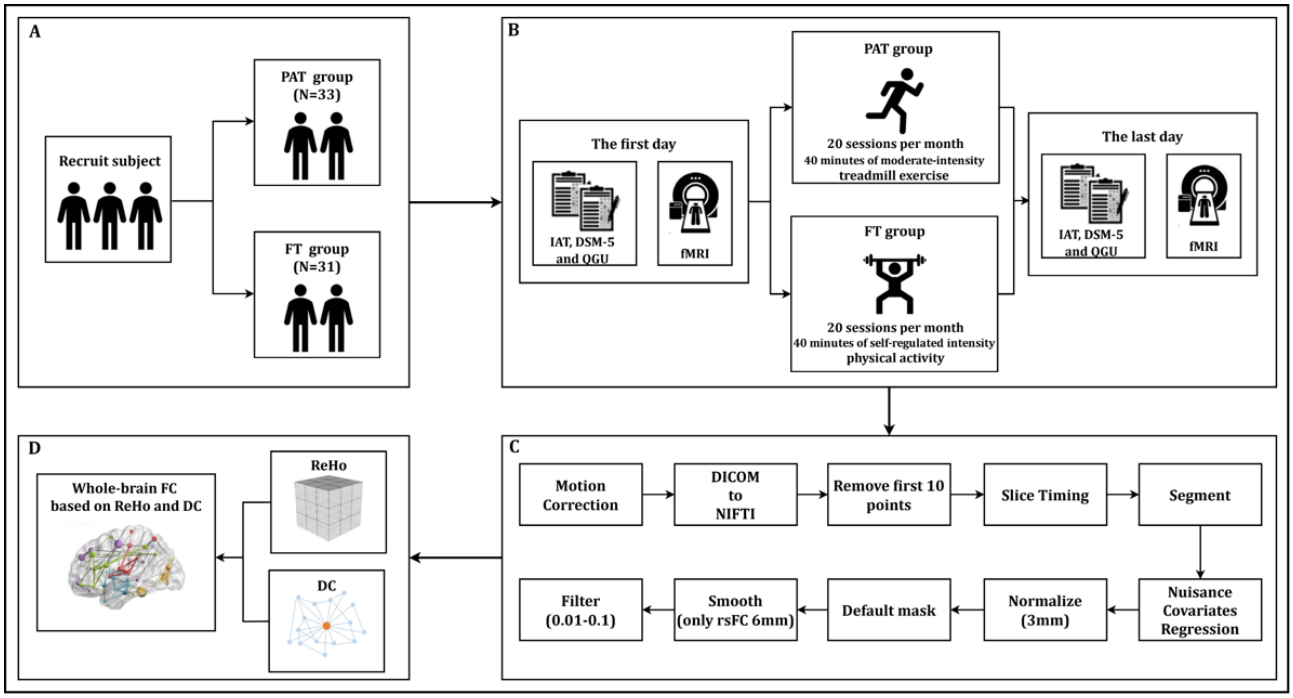
Ethical Considerations

The study protocol was reviewed and approved by the Human Investigation Committee (Institutional Review Board) at Yunnan Normal University (Approval No: yunnuethic2024-044; registered prospectively on 2024-10-14). No changes were made to the trial protocol after its commencement. Written informed consent was obtained from all participants prior to their enrollment in the study. To ensure participant privacy and confidentiality, all collected data were immediately deidentified following collection. Participants were assigned anonymous codes, and all analyses were performed using these coded identifiers. No personally identifiable information is included in this manuscript or any supplementary materials. All participants who completed the entire study protocol, including pretest and posttest assessments and all training sessions, received financial compensation of 400 RMB (approximately US $56.29).

Participants

Figure 1 shows the process and sequence of the study. This study initially recruited 72 participants with IGD. Eight participants dropped out during the intervention training, resulting in a final cohort of 64 participants (PAT: n=33; FT: n=31). No missing data were encountered (no additional prespecified or post-hoc analyses were performed).

Figure 1. Experimental timeline and procedures for the randomized controlled trial of progressive aerobic training versus free training in individuals with internet gaming disorder. (**A**) Subject; (**B**) treatment processing; (**C**) data preprocessing; (**D**) data analysis. PAT: progressive aerobic training; FT: free training; IAT: internet addiction test; DSM: *Diagnostic and Statistical Manual of Mental Disorders*; QGU: questionnaire for gaming urges; fMRI: functional magnetic resonance imaging; ReHo: regional homogeneity; DC: degree centrality; FC: functional connectivity.



We recruit patients with IGD from schools through advertisements to obtain treatment opportunities for improving mental health. Researchers require interested participants to participate in the internet addiction test (IAT) online [44], which has been proven to effectively diagnose internet addiction. Participants with scores greater than 50 are required to undergo an offline professional interview with a psychiatrist. Select eligible research subjects using *DSM-5* criteria [45], and Mini International Neuropsychiatric Interview (MINI) [46].

The criteria for inclusion in participants are as follows: (1) IAT score >50; (2) *DSM-5* score >5; (3) active engagement in online gaming exceeding 1 year; (4) BMI within normal to overweight range (15≤BMI≤28); (5) no cognitive impairment and no other psychiatric disorders confirmed by MINI; (6) no previous treatment for IGD; (7) no experience of regular use of any psychotropic drugs. The exclusion criteria for participants were as follows: (1) majors related to sports; (2) *DSM-5* score <5; (3) the body contains metals. All participants who met the inclusion criteria were randomly assigned to either the PAT

XSL•FO
RenderX

group or the FT group. Detailed demographic data for the two groups are shown in Table 1.

Table . Demographic and clinical characteristics of participants with internet gaming disorder at baseline.

Characteristic	PAT[a] group, n=33	FT[b] group, n=31	*t* test[c]*(df)*	*P* value
Sex, n (%)			—[d]	—[d]
Male	11 (33.33)	13 (41.93)		
Female	22 (66.66)	18 (58.07)		
Age, mean (SD)	19.97 (1.36)	19.87 (1.61)	0.266 (62)	.79
IAT[e] score, mean (SD)	62.88 (12.196)	62.94 (13.878)	−0.017 (62)	.99
DSM-5[f] score, mean (SD)	5.24 (2.092)	5.19 (2.088)	0.093 (62)	.93
QGU[g] score, mean (SD)	34.48 (16.333)	30.13 (17.957)	1.016 (62)	.31

[a]PAT: progressive aerobic training.

[b]FT: free training.

[c]The *t* tests reported in the table are all two-tailed independent samples tests.

[d]not applicable.

[e]IAT: internet addiction test.

[f]DSM-5: Diagnostic and Statistical Manual of Mental Disorders-5.

[g]QGU: Questionnaire for Gaming Urges.

Intervention Processes

In this study, the American College of Sports Medicine guidelines were used to develop an aerobic training program with an aerobic intensity of 46% in the first phase and 55% in the second phase, as well as a training duration of 40 minutes per session for a total of 20 sessions in 1 month [47-49]. The PAT group trained only on the treadmill at the speed calculated from the step test [50]. The specific training process and related contents are shown in Multimedia Appendix 1. The interventions were delivered by research assistants who were trained in the specific protocols for both the PAT and FT groups to ensure consistency and safety.

Adverse Events

As adverse events were not systematically assessed, none were reported during the trial.

Sample Size

A comparison between study completers and dropouts showed no significant differences in key demographic or clinical measures at baseline (Multimedia Appendix 1), indicating that attrition bias is unlikely to have substantially influenced the primary findings. This trial involved no interim analyses or stopping guidelines.

Using G-power, setting the effect size at f=0.25 and α error probability at 0.05, we estimated the required sample size at 66. The current sample is a bit lower than the required 66.

Randomization to the PAT or FT Group

Participants were randomly allocated to either the PAT group or the FT group using simple randomisation with a 1:1 allocation ratio. The random allocation sequence was generated by an independent researcher not involved in participant recruitment or intervention. The sequence was concealed using sequentially numbered, opaque, sealed envelopes. The principal investigator enrolled participants, and a research assistant then assigned them to groups by opening the next envelope in the sequence.

Blinding

The participants were indeed blinded to the existence of the other experimental group. Both groups were in the same training room but did not know the existence of the other group (the PAT group trained from 14:00-19:00; the FT group trained from 20:00-22:00). Research assistants conducting assessments were aware of group allocations due to the practical constraints of implementing the intervention protocols.

Behavioral Measurements

The current analysis is a per-protocol analysis; only subjects who finished the pretest, 20-time training, and posttest were included. Before the first training and after the entire training period, all participants' addiction levels were measured based on the IAT, *DSM-5* criteria, and craving scores. Craving was measured with the Questionnaire for Gaming Urges, adapted from the Tiffany Questionnaire for Smoking Urges [51]. Meanwhile, before the first training session, after 10 interventions and all 20 interventions, we conducted a 3-minute step test [50] on all participants to obtain the maximum oxygen uptake for dynamically adjusting the intensity of PAT.

Magnetic Resonance Imaging Data Acquisition

The magnetic resonance imaging data were acquired using a Siemens Trio 3T scanner. The specific parameters used were as follows: repetition time=2000 ms, 54 interleaved slices, echo time=30 ms, thickness=3.0 mm, flip angle=90°, field of view=220 mm × 220 mm, and matrix=64×64. Head motions were minimized by filling the empty space around the subjects' heads with sponge and fixing their lower jaws with tape.

XSL•FO
RenderX

Data Preprocessing

The preprocessing of the rs-fMRI data was completed according to Statistical Parametric Mapping principles [52] using SPM12 [53], and data processing and analysis were performed using Data Processing & Analysis of Brain Imaging (DPABI) [54,55]. The preprocessing of the functional rs-fMRI data included the following steps: (1) the quality of the raw rs-fMRI data for all the participants was examined; (2) the raw data were transformed from the Digital Imaging and Communication in Medicine (DICOM) format to the Neuroimaging Informatics Technology Initiative (NIFTI) format; (3) the initial 10 volumes were discarded to maintain magnetic field stabilization; (4) slice timing was performed; (5) realignment was performed (excluding subjects with maximum head motion >2.5 mm or rotation >2.5°); (6) regressing nuisance covariates were included, including the signal noise of white matter, cerebrospinal fluid, and Friston-24 head motion parameters [56]; (7) the data were normalized using the standard procedure within the DPABI pipeline, which registers the images to the MNI-152 template; (8) the data were filtered by a bandpass filter with a frequency range of 0.01 to 0.1 Hz before the calculation of ReHo and DC; and (9) the data were smoothed using a Gaussian kernel with a full width at half maximum of 6 mm.

ROI Selection

We first computed ReHo and DC values for both the PAT and FT groups using pretest and posttest measurements. Repeated-measure ANOVAs were then conducted for two groups using the DPABI Viewer toolbox, with a statistical threshold set at $P<.01$ and a cluster size threshold of >100 voxels (corrected for the false discovery rate).

Statistical Analysis

The prespecified primary outcome was the change in the Questionnaire for Gaming Urges score from baseline to post-intervention. Secondary outcomes included other behavioral measures and functional connectivity measures (including their correlation with normative neurotransmitter maps).

We conducted whole-brain FC analysis using the ROI obtained by overlapping DC and ReHo. Then, significant brain region data were extracted with the DPABI toolbox for repeated measures ANOVA analysis in JASP (Version 0.19.3) [57,58], applying a Bonferroni correction ($\alpha=.05$, corrected for multiple comparisons). The correlation analysis of behavioral data was also completed using JASP. Subsequently, we performed a correlation analysis between FC results and neurotransmitter levels using JuSpace 1.5 (10,000 permutations, $P<.05$, with correction for the false discovery rate) [59]. The analysis demonstrates a spatial correspondence between our FC findings and the normative neurotransmitter maps, rather than a correlation with measured individual neurochemical levels.

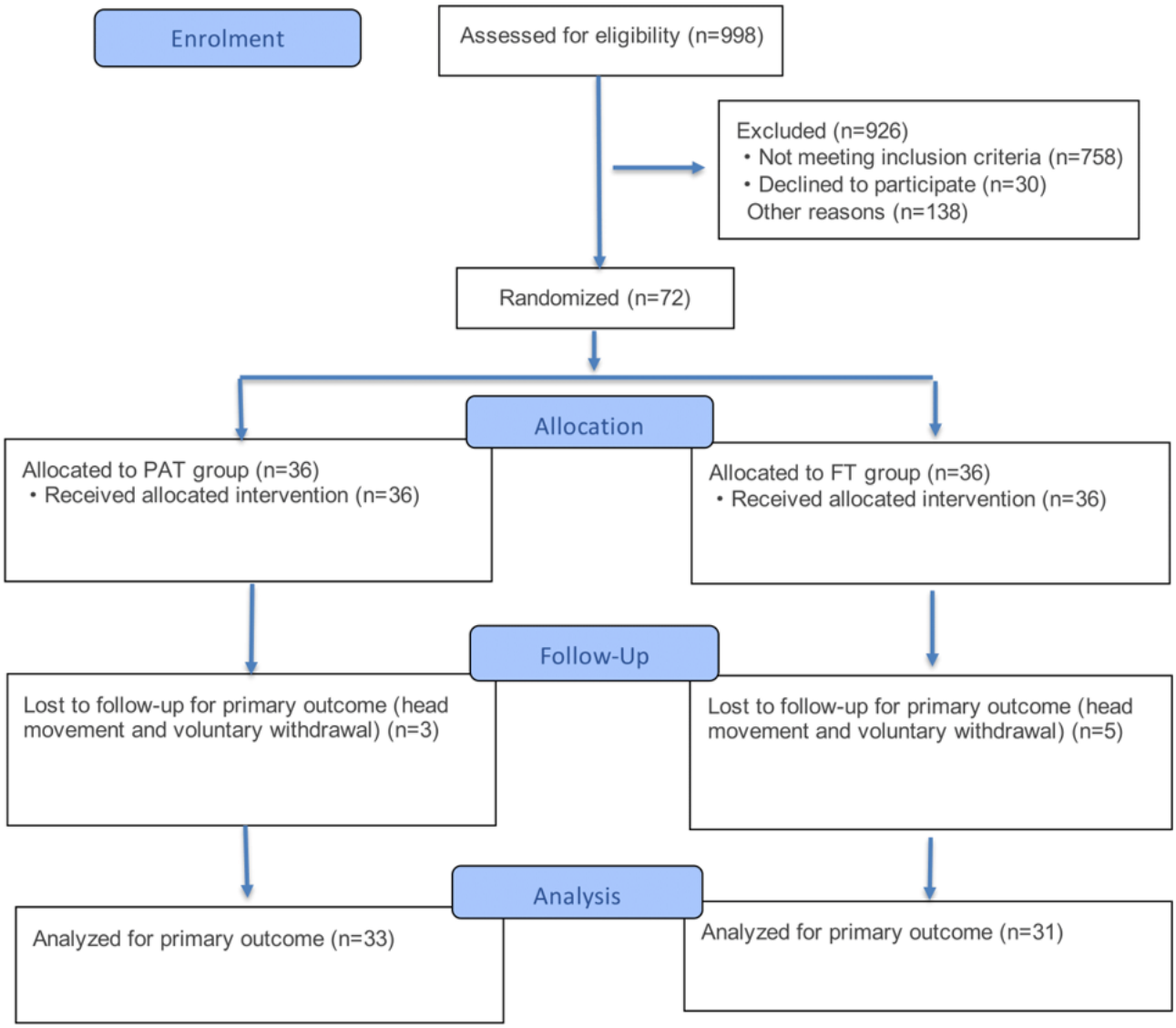
Results

Overview

The following analysis was conducted with 64 participants who completed the interventions without receiving other therapies, as shown in Figure 2. Consistent with the methods, no intervention-related adverse events were observed, and no additional prespecified or post-hoc analyses were performed.

XSL•FO
RenderX

Figure 2. CONSORT flow diagram of participant screening, randomization, and analysis in a trial of progressive aerobic training (PAT) versus free training (FT) for internet gaming disorder.

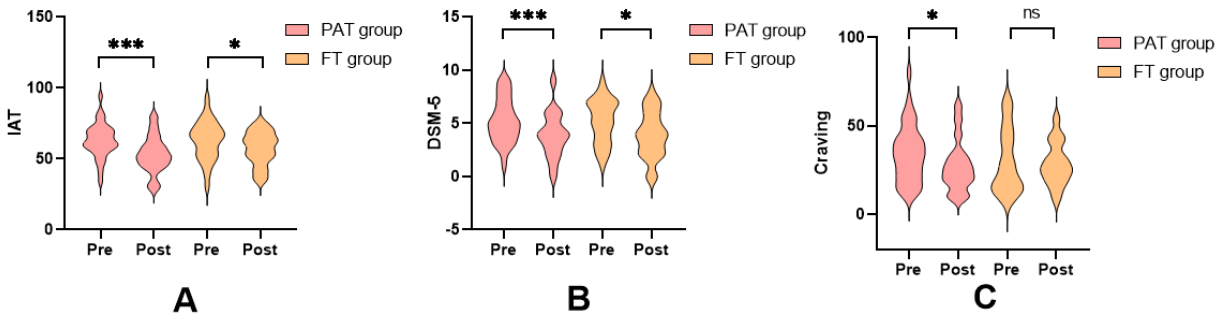


Behavioral Measures

We performed repeated measures ANOVA on game craving in the PAT and FT groups and found that the interaction was not significant. Then, paired-samples *t* tests were conducted to compare pretest and posttest gaming cravings in the PAT group and the FT group. The results revealed a significant reduction in gaming cravings in the PAT group (t_{32}=2.278, P_{bonf}=0.03, Cohen *d*=0.397, 95% CI 0.045-0.851), whereas no significant change was observed in the FT group (t_{30}=0.481, P_{bonf}=1, Cohen *d*=0.086, 95% CI −0.306 to 0.502). Compared to the FT group (t_{30}=2.277, P_{bonf}=0.03, Cohen *d*=0.409, 95% CI 0.039-0.773), the PAT group showed a significantly greater reduction in the IAT score (t_{32}=4.333, P_{bonf}<0.001, Cohen *d*=0.754, 95% CI 0.362-1.137). This finding was consistent with *DSM-5* scores (Figure 3).

XSL•FO
RenderX

Figure 3. Behavioral changes in internet gaming disorder following progressive aerobic training (PAT) versus free training (FT). DSM, Diagnostic and Statistical Manual of Mental Disorders, version 5; Pre, pretest before training; Post, posttest after training. * *P*<.05, ** *P*<.01, *** *P*<.001, ns: no significance.



Functional Connectivity Results

FC Results Within the ECN

Repeated-measures ANOVA showed a significant interaction between the changes in FC between the right medial superior frontal gyrus (R-mSFG) and left postcentral gyrus ($F_{1, 62}$=7.953, *P*=.006), bilateral superior parietal gyrus (SPG; right: $F_{1, 62}$=5.675, *P*=.02; left: $F_{1, 62}$=8.853, *P*=.004), left inferior frontal gyrus ($F_{1, 62}$=11.373, *P*=.001), right precentral gyrus ($F_{1, 62}$=7.325, *P*=.009), right middle cingulate gyrus ($F_{1, 62}$=6.659, *P*=.012) in both groups before and after intervention.

Subsequent post-hoc analysis demonstrated that the PAT group exhibited significantly enhanced FC between the R-mSFG and left postcentral gyrus (t_{62}=−2.726, P_{bonf}=0.008, Cohen *d*=−0.633, 95% CI −1.268 to 0.002), bilateral SPG (right: t_{62}=−2.417, P_{bonf}=0.019, Cohen *d*=−0.525, 95% CI −1.126 to 0.076; left: t_{62}=−2.861, P_{bonf}=0.006, Cohen *d*=−0.693, 95% CI −1.358 to −0.027), left inferior frontal gyrus (t_{62}=−3.254, P_{bonf}=0.002, Cohen *d*=−0.642, 95% CI −1.202 to −0.082), right precentral gyrus (t_{62}=−2.26, P_{bonf}=0.027, Cohen *d*=−0.517, 95% CI −1.141 to 0.107), and right middle cingulate gyrus (t_{62}=−2.42, P_{bonf}=0.018, Cohen *d*=−0.567, 95% CI −1.221 to 0.087). No such effects were observed in the FT group.

FC Results Between the ECN and the RN

Repeated-measures ANOVA showed a significant interaction between the changes in FC between the R-mSFG and bilateral insula (right: $F_{1, 62}$=11.408, *P*=.001; left: $F_{1, 62}$=7.94, *P*=0.006) and left substantia nigra ($F_{1, 62}$=10.604, *P*=.002) in both groups before and after intervention.

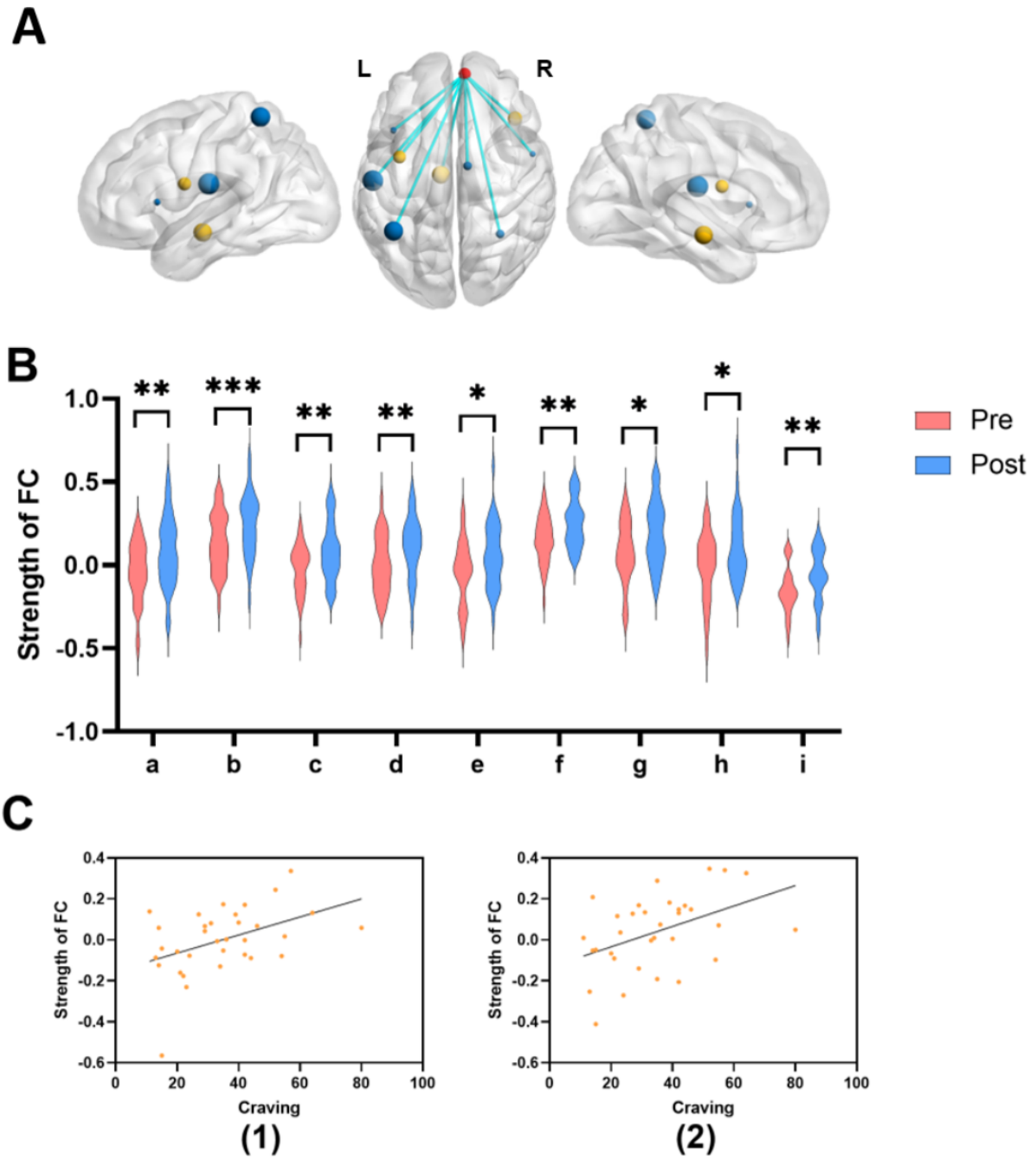
Subsequent post-hoc analysis confirmed that the PAT group showed significantly enhanced FC between the R-mSFG and bilateral insula (right: t_{62}=−3.372, P_{bonf}=0.001, Cohen *d*=−0.59, 95% CI −1.089 to −0.092; left: t_{62}=−2.761, P_{bonf}=0.008, Cohen *d*=−0.611, 95% CI −1.233 to 0.01) and left substantia nigra (t_{62}=−3.157, P_{bonf}=0.002, Cohen *d*=−0.778, 95% CI −1.476 to −0.080). However, no significant changes were detected in the FT group.

Correlation Analyses Between FC Results and Behavioral Data

The ReHo analysis revealed significant activation in the R-mSFG, while the DC analysis identified significant regions in the bilateral medial superior frontal gyrus and the left precuneus. Subsequently, we performed an overlap analysis between the ReHo and DC results to identify common brain regions. At *P*<.01, the R-mSFG emerged as the only significant overlapping region (Figure 4). Based on these findings, we selected the R-mSFG as the ROI for subsequent whole-brain FC analysis. In the PAT group, a significant positive correlation (*r*=0.359, *P*=.04, 95% CI 0.018-0.625) between pretest game cravings and FC strength between the R-mSFG and left postcentral gyrus was found. The same result was found for the R-mSFG to the right precentral gyrus (*r*=0.399, *P*=.021, 95% CI 0.065-0.653; Figure 4).

XSL•FO
RenderX

Figure 4. Functional connectivity (FC) alterations and their association with game craving following progressive aerobic training. (**A**) Visualization of the region of interest and functional connectivity. The size of the ball corresponds to the t value obtained from the simple effect. The red ball indicates the right medial superior frontal gyrus (R-mSFG); the blue ball indicates the executive control network including the left postcentral, bilateral superior frontal gyrus (SFG), left inferior frontal gyrus (IFG), right precentral gyrus, and right middle cingulate gyrus (MCG); and the yellow ball indicates the reward network including the bilateral insula and left substantia nigra. (**B**) Whole brain–based functional connectivity (FC) results. The x-axis shows the different functional connections: a, the R-mSFG to the left postcentral gyrus; b, the R-mSFG to the right insula; c, the R-mSFG to the left superior parietal gyrus (SPG); d, the R-mSFG to the left IFG; e, the R-mSFG to the right precentral gyrus; f, the R-mSFG to the left substantia nigra; g, the R-mSFG to the right MCG; h, the R-mSFG to the right SPG; i, the R-mSFG to the left insula. Pre, pretest before the training; Post, posttest after the training. * P<.05, ** P<.01, *** P<.001. (**C**) Correlation analysis between FC results and levels of game craving. (1) Correlation results between game craving and FC strength between the R-mSFG and left postcentral gyrus in the progressive aerobic training group. (2) Correlation results between game craving and FC strength between the R-mSFG and the right precentral gyrus.



Correlation Analyses Between FC Results and Neurotransmitters

In the pretest of the PAT group, compared to the posttest results, significant correlations were observed between the spatial distribution of CB1 receptors (P_{bonf}=.003), mGluR5_rosaneto receptors (P_{bonf}=.005), and mGluR5_dubois receptors (P_{bonf}=.005) based on FC changes in the R-mSFG (Multimedia Appendix 2).

XSL•FO
RenderX

Discussion

Principal Findings

In this study, we evaluated the efficacy and neural mechanisms of the PAT intervention for IGD. The results indicate that PAT was associated with reduced severity of addiction and craving levels, alongside alterations in the FC within the ECN and between the ECN and the RN.

PAT Reduces Addiction Severity and Gaming Cravings in IGD

Following the PAT intervention, participants in the PAT group exhibited markedly lower gaming cravings and decreased addiction severity relative to the AT group. Although PAT has not been used in previous literature for the treatment of IGD, previous studies on alcohol and nicotine dependence have shown that PAT training can reduce cravings for addictive substances and the severity of addiction [60,61]. Therefore, this study found decreased game craving and addiction severity after PAT, indicating that PAT is an effective intervention for patients with IGD, significantly mitigating the symptoms of IGD.

PAT Strengthened FC Within the ECN

This study found enhanced FC between the prefrontal cortex (mSFG and IFG), SPG, middle cingulate gyrus, and postcentral and precentral gyrus in IGD after PAT. The prefrontal cortex and postparietal cortex are implicated in a wide range of cognitive functions, particularly cognitive control [62,63]. The SPG has a crucial role in visual perception, spatial cognition, and attention [64,65]. Resting-state fMRI has found that compared to healthy controls, patients with IGD showed decreased FC in the ECN (including several parts of the frontoparietal cortex) [66,67], indicating impaired executive functioning. The findings of this study demonstrate that PAT contributed to enhancing the functional integrity of ECN implicated in IGD. The precentral gyrus and postcentral gyrus are the key regions of sensorimotor networks associated with integrating sensorimotor information and coordinating physical movement [68,69]. Although the precentral gyrus is not part of the ECN, executive control processes frequently require integrating sensory information transmitted from the precentral gyrus to achieve efficient behavioral regulation [70]. Therefore, this study found that the FC within the ECN was significantly enhanced following PAT, indicating that PAT may help normalize sensorimotor dysfunction of IGD, thereby contributing to the restoration of executive control and emotional regulation capabilities.

PAT Strengthened FC Between the ECN and the RN

The present study found enhanced FC between the mSFG, bilateral insula, and substantia nigra. The insula plays a pivotal role in the RN, engaging in reward processing and the regulation of motivated behaviors by integrating sensory and emotional information. It influences decision-making and is closely associated with mental health [71,72]. In the context of addiction, insula dysfunction may contribute to heightened cravings and impulsive behaviors, disrupting the balance between impulse control and reward processing [73]. The substantia nigra, a core component of the basal ganglia, plays a pivotal role in reward-related processing [74]. Meanwhile, the superior frontal gyrus, a key region within the ECN and a default mode network seed region, is critical for top-down cognitive control, including decision-making, selective attention, inhibitory control, and emotion regulation [75,76]. Consistent with prior research, individuals with IGD exhibit reduced FC between the ECN and RN compared to healthy controls [77-79]. Therefore, this study revealed that the FC between the ECN and RN was significantly enhanced following PAT, indicating that PAT may weaken gaming cue reactivity and strengthen top-down regulation of reward processing in patients with IGD.

Association Between PAT and Molecular Systems in IGD

Few studies have correlated the neuroimaging changes of IGD with the activity of microscopic neurotransmitters [80,81]. This study found an association between the FC and activity maps of CB1 and mGluR5 in the PAT group. CB1 cannabinoid receptors are located presynaptically in multiple brain regions, including reward-related structures like the caudate-putamen, substantia nigra, and globus pallidus [82,83]. It is well-established that aerobic exercise is a potent trigger for the release of endogenous cannabinoids, such as anandamide, in the brain [84]. The primary central receptor for these endogenous cannabinoids is the CB1 receptor. Therefore, our finding that functional connectivity changes correlate with the spatial distribution of CB1 receptors provides a direct neurobiological substrate for the effects of PAT. We posit that the PAT in our study enhanced the signaling of endogenous cannabinoids, which in turn modulated neural circuitry in CB1-rich brain regions that are critically involved in IGD pathology.

Furthermore, the significant correlation with the mGluR5 map is highly informative. The mGluR5 is highly expressed in key addiction-related regions, including the nucleus accumbens, cerebral cortex, hippocampal formation, and amygdala [28]. Both CB1 and mGluR5 are considered to exhibit functional interactions and are known to form heteromeric complexes that regulate synaptic plasticity [85,86]. This provides a compelling rationale for our concurrent findings regarding both receptors. We thus hypothesize that the therapeutic action of PAT is mediated not by a single system, but by the coordinated modulation of CB1 and mGluR5 signaling, which promotes adaptive synaptic plasticity to normalize the dysregulated neural circuitry underlying IGD.

Limitations

This study has several limitations. First, while our FC analysis revealed altered connectivity between key brain regions, it did not elucidate the directionality of these changes. Future studies could employ effective connectivity analysis to determine the causal interactions underlying these FC modifications. Second, the FT group exhibited a slight therapeutic effect on IGD symptoms, suggesting a potential placebo effect. To address this limitation, future research could incorporate multiple control groups to better isolate the specific effects of PAT. Third, our sample was recruited solely from schools, which may limit the generalizability of the findings to older individuals with IGD. Fourth, it should be noted that potential confounding factors,

XSL•FO
RenderX

such as general fitness improvement, social interaction during training, and participant expectancy effects, may have contributed to the observed outcomes. Finally, the modest sample size in this study may limit the generalizability of the findings. Future research with a larger, multicenter cohort, incorporating long-term follow-ups, would enhance the statistical power and robustness of the findings.

Conclusions

This study provides evidence for the beneficial effects of PAT on IGD symptoms. A key innovation of this trial is its implementation of a progressive aerobic intervention, which dynamically adjusted intensity based on the real-time performance of individuals with IGD. This approach moves beyond conventional fixed-intensity protocols to offer a more personalized and potentially more effective therapeutic approach. At the cognitive-neural level, our findings suggest that PAT may enhance executive control function in patients with IGD concurrently with alterations in its regulatory influence over the reward system. These neural improvements were substantiated by significant correlations between neurofunctional markers and behavioral measures. The results offer insights into the neural mechanisms associated with PAT's positive outcomes in IGD and provide a foundation for future clinical intervention research.

Funding

The current research was supported by Yunnan Provincial Department of Science and Technology Basic Research Project (2,02,405AC350075; 2,02,501AS070042) and the Innovation Team Program in Philosophy and Social Science of Yunnan Province (Research on psychological adaptation and development of China's ethnic minority students in border areas).

Data Availability

The data are stored at our lab-based network attachment system and will be made available on request.

Authors' Contributions

Conceptualization: XX (lead), GHD (supporting)
Data curation: SC (lead), XX (supporting)
Formal analysis: SC
Funding acquisition: GHD
Investigation: SC (lead), XX (supporting), MW (supporting), XL (supporting)
Methodology: XX (lead), MW (supporting), XL (supporting)
Project administration: GHD
Resources: XX (lead), MW (supporting), XL (supporting)
Software: XX (lead), MW (supporting), XL (supporting)
Supervision: GHD
Validation: SC
Visualization: SC
Writing – original draft: SC
Writing – review & editing: SC (lead), GHD (supporting)

Conflicts of Interest

None declared.

Multimedia Appendix 1
The specific training process and related contents follow progressive aerobic training versus free training.
[DOCX File, 23 KB - games_v13i1e83597_app1.docx]

Multimedia Appendix 2
Correlations between functional connectivity and neurotransmitter systems in the progressive aerobic training (PAT) group for internet gaming disorder. 5-HT, 5-hydroxytryptamine; CB1, cannabinoid 1 receptor; mGluR5, metabotropic glutamate receptor 5. ** P<.01.
[PNG File, 71 KB - games_v13i1e83597_app2.png]

Checklist 1
CONSORT-EHEALTH checklist (V1.6.1).
[DOCX File, 38 KB - games_v13i1e83597_app3.docx]

References

XSL•FO
RenderX

1. Mental health. Science Direct. URL: https://www.sciencedirect.com/topics/psychology/mental-health [accessed 2025-11-20]
2. Gorowska M, Tokarska K, Zhou X, Gola MK, Li Y. Novel approaches for treating internet gaming disorder: a review of technology-based interventions. Compr Psychiatry 2022 May;115:152312. [doi: 10.1016/j.comppsych.2022.152312] [Medline: 35405548]
3. Behavior (neuroscience). Science Direct. URL: https://www.sciencedirect.com/topics/neuroscience/behavior-neuroscience [accessed 2025-11-20]
4. Diagnostic and Statistical Manual of Mental Disorders: DSM-5: American Psychiatric Association, DSM-5 Task Force; 2013. URL: https://doi.org/10.1176/appi.books.9780890425596 [doi: 10.1176/appi.books.9780890425596]
5. Jo YS, Bhang SY, Choi JS, Lee HK, Lee SY, Kweon YS. Clinical characteristics of diagnosis for internet gaming disorder: comparison of DSM-5 IGD and ICD-11 GD diagnosis. J Clin Med 2019 Jun 28;8(7):945. [doi: 10.3390/jcm8070945] [Medline: 31261841]
6. Gao YX, Wang JY, Dong GH. The prevalence and possible risk factors of internet gaming disorder among adolescents and young adults: systematic reviews and meta-analyses. J Psychiatr Res 2022 Oct;154:35-43. [doi: 10.1016/j.jpsychires.2022.06.049] [Medline: 35926424]
7. Wang M, Zheng H, Zhou W, et al. Disrupted dynamic network reconfiguration of the executive and reward networks in internet gaming disorder. Psychol Med 2023 Sep;53(12):5478-5487. [doi: 10.1017/S0033291722002665] [Medline: 36004801]
8. Dong GH, Dai J, Potenza MN. Ten years of research on the treatments of internet gaming disorder: a scoping review and directions for future research. JBA 2024;13(1):51-65. [doi: 10.1556/2006.2023.00071]
9. Wang M, Dong H, Zheng H, Du X, Dong GH. Inhibitory neuromodulation of the putamen to the prefrontal cortex in Internet gaming disorder: how addiction impairs executive control. J Behav Addict 2020 Jun;9(2):312-324. [doi: 10.1556/2006.2020.00029] [Medline: 32663381]
10. Dong G, Zheng H, Liu X, Wang Y, Du X, Potenza MN. Gender-related differences in cue-elicited cravings in Internet gaming disorder: The effects of deprivation. J Behav Addict 2018 Dec 1;7(4):953-964. [doi: 10.1556/2006.7.2018.118] [Medline: 30556781]
11. Dong H, Wang M, Zhang J, Hu Y, Potenza MN, Dong GH. Reduced frontostriatal functional connectivity and associations with severity of internet gaming disorder. Addict Biol 2021 Jul;26(4):e12985. [doi: 10.1111/adb.12985] [Medline: 33236526]
12. Kim JY, Chun JW, Park CH, et al. The correlation between the frontostriatal network and impulsivity in internet gaming disorder. Sci Rep 2019;9(1):1191. [doi: 10.1038/s41598-018-37702-4]
13. Zhang Y, Mei W, Zhang JX, Wu Q, Zhang W. Decreased functional connectivity of insula-based network in young adults with internet gaming disorder. Exp Brain Res 2016 Sep;234(9):2553-2560. [doi: 10.1007/s00221-016-4659-8] [Medline: 27119360]
14. Gyri. Science Direct. URL: https://www.sciencedirect.com/topics/neuroscience/gyri [accessed 2025-11-21]
15. Dong G, Wang Z, Wang Y, Du X, Potenza MN. Gender-related functional connectivity and craving during gaming and immediate abstinence during a mandatory break: implications for development and progression of internet gaming disorder. Prog Neuropsychopharmacol Biol Psychiatry 2019 Jan 10;88:1-10. [doi: 10.1016/j.pnpbp.2018.04.009] [Medline: 29684536]
16. Dong H, Wang M, Zheng H, Zhang J, Dong GH. The functional connectivity between the prefrontal cortex and supplementary motor area moderates the relationship between internet gaming disorder and loneliness. Prog Neuropsychopharmacol Biol Psychiatry 2021 Jun 8;108:110154. [doi: 10.1016/j.pnpbp.2020.110154] [Medline: 33137406]
17. Zajac K, Ginley MK, Chang R. Treatments of internet gaming disorder: a systematic review of the evidence. Expert Rev Neurother 2020 Jan;20(1):85-93. [doi: 10.1080/14737175.2020.1671824] [Medline: 31544539]
18. Li S, Wu Q, Tang C, Chen Z, Liu L. Exercise-based interventions for internet addiction: neurobiological and neuropsychological evidence. Front Psychol 2020;11:1296. [doi: 10.3389/fpsyg.2020.01296] [Medline: 32670157]
19. Müller P, Duderstadt Y, Lessmann V, Müller NG. Lactate and BDNF: key mediators of exercise induced neuroplasticity? J Clin Med 2020 Apr 15;9(4):1136. [doi: 10.3390/jcm9041136] [Medline: 32326586]
20. Kim YH, Jeong JG, Kim TY. Effect of percutaneous electrical nerve stimulation on the spinal neuron excitability in STZ-induced diabetic rats [Article in Korean]. J Korean Acad Clin Electrophysiol 2004;2(2):13-23 [FREE Full text]
21. Rahman N, Mihalkovic A, Geary O, Haffey R, Hamilton J, Thanos PK. Chronic aerobic exercise: autoradiographic assessment of GABA(a) and mu-opioid receptor binding in adult rats. Pharmacol Biochem Behav 2020 Sep;196:172980. [doi: 10.1016/j.pbb.2020.172980] [Medline: 32593790]
22. Xue Y, Yang Y, Huang T. Effects of chronic exercise interventions on executive function among children and adolescents: a systematic review with meta-analysis. Br J Sports Med 2019 Nov;53(22):1397-1404. [doi: 10.1136/bjsports-2018-099825] [Medline: 30737201]
23. Jia S, Du Z, Chu D, et al. A network meta-analysis of mind-body exercise interventions for internet addiction symptoms in young adults. Front Public Health 2025;13:1565372. [doi: 10.3389/fpubh.2025.1565372] [Medline: 40606075]
24. Wang H, Chen Y, Li X, Wang J, Zhou Y, Zhou C. Moderate-intensity aerobic exercise restores appetite and prefrontal brain activity to images of food among persons dependent on methamphetamine: a functional near-infrared spectroscopy study. Front Hum Neurosci 2019;13:400. [doi: 10.3389/fnhum.2019.00400] [Medline: 31798434]

XSL•FO
RenderX

25. Janse Van Rensburg K, Taylor A, Hodgson T, Benattayallah A. Acute exercise modulates cigarette cravings and brain activation in response to smoking-related images: an fMRI study. Psychopharmacology (Berl) 2009 Apr;203(3):589-598. [doi: 10.1007/s00213-008-1405-3] [Medline: 19015835]
26. Lynch WJ, Peterson AB, Sanchez V, Abel J, Smith MA. Exercise as a novel treatment for drug addiction: a neurobiological and stage-dependent hypothesis. Neurosci Biobehav Rev 2013 Sep;37(8):1622-1644. [doi: 10.1016/j.neubiorev.2013.06.011] [Medline: 23806439]
27. Matei D, Trofin D, Iordan DA, et al. The endocannabinoid system and physical exercise. Int J Mol Sci 2023 Jan 19;24(3):1989. [doi: 10.3390/ijms24031989] [Medline: 36768332]
28. Brown RM, Mustafa S, Ayoub MA, Dodd PR, Pfleger KDG, Lawrence AJ. mGlu5 receptor functional interactions and addiction. Front Pharmacol 2012;3:84. [doi: 10.3389/fphar.2012.00084] [Medline: 22586398]
29. Ashtari N, Radahmadi M, Alaei H. Preventive effects of fixed and progressive forced exercises on memory and brain electrical activity in morphine-addicted rats. An Acad Bras Cienc 2022;94(4):e20200596. [doi: 10.1590/0001-3765202220200596] [Medline: 35830068]
30. Gui Z, Luo L, Yuan D. Effects of 24 weeks progressive aerobic training on physical and mental health of male drug addicts with different depression levels: —data analysis based on smart big data acquisition system for drug treatment. Presented at: 2021 International Conference on Information Technology and Contemporary Sports (TCS); Jan 15-17, 2021; Guangzhou, China p. 253-258. [doi: 10.1109/TCS52929.2021.00059]
31. Somandepalli K, Kelly C, Reiss PT, et al. Short-term test-retest reliability of resting state fMRI metrics in children with and without attention-deficit/hyperactivity disorder. Dev Cogn Neurosci 2015 Oct;15:83-93. [doi: 10.1016/j.dcn.2015.08.003] [Medline: 26365788]
32. Zang YF, Zuo XN, Milham M, Hallett M. Toward a meta-analytic synthesis of the resting-state fMRI literature for clinical populations. Biomed Res Int 2015;2015:435265. [doi: 10.1155/2015/435265] [Medline: 26171391]
33. Synchronism. Science Direct. URL: https://www.sciencedirect.com/topics/earth-and-planetary-sciences/synchronism [accessed 2025-11-21]
34. Zang Y, Jiang T, Lu Y, He Y, Tian L. Regional homogeneity approach to fMRI data analysis. Neuroimage 2004 May;22(1):394-400. [doi: 10.1016/j.neuroimage.2003.12.030] [Medline: 15110032]
35. Lv Y, Margulies DS, Villringer A, Zang YF. Effects of finger tapping frequency on regional homogeneity of sensorimotor cortex. PLoS ONE 2013;8(5):e64115. [doi: 10.1371/journal.pone.0064115]
36. Correlation coefficient. Science Direct. URL: https://www.sciencedirect.com/topics/earth-and-planetary-sciences/correlation-coefficient [accessed 2025-11-21]
37. Aiello M, Salvatore E, Cachia A, et al. Relationship between simultaneously acquired resting-state regional cerebral glucose metabolism and functional MRI: a PET/MR hybrid scanner study. Neuroimage 2015 Jun;113:111-121. [doi: 10.1016/j.neuroimage.2015.03.017] [Medline: 25791784]
38. Cao Q, Zang Y, Sun L, et al. Abnormal neural activity in children with attention deficit hyperactivity disorder: a resting-state functional magnetic resonance imaging study. Neuroreport 2006 Jul 17;17(10):1033-1036. [doi: 10.1097/01.wnr.0000224769.92454.5d] [Medline: 16791098]
39. Luo FF, Xu H, Zhang M, Wang Y. Abnormal regional spontaneous brain activity and its indirect effect on spasm ratings in patients with hemifacial spasm. Front Neurosci 2020;14:601088. [doi: 10.3389/fnins.2020.601088] [Medline: 33362459]
40. Xu X, Ma X, Ni H, et al. Mindfulness meditation training reduces gaming cravings by reshaping the functional connectivity between the default mode network and executive control-related brain regions. Biol Psychiatry Cogn Neurosci Neuroimaging 2024 Aug;9(8):827-836. [doi: 10.1016/j.bpsc.2024.04.011] [Medline: 38692393]
41. Zhu CZ, Zang YF, Liang M, et al. Discriminative analysis of brain function at resting-state for attention-deficit/hyperactivity disorder. Med Image Comput Comput Assist Interv 2005;8(Pt 2):468-475. [doi: 10.1007/11566489_58] [Medline: 16685993]
42. Fang CL, Zheng GH, Fang CL, Zheng GH. Impact of physical exercise on the executive function and mental health of heroin addicts. Exerc Sci 2024 Nov 1;33(4):477-493. [doi: 10.15857/ksep.2024.00402]
43. Hopewell S, Chan AW, Collins GS, et al. CONSORT 2025 statement: updated guideline for reporting randomised trials. BMJ 2025 Apr 14;389:e081123. [doi: 10.1136/bmj-2024-081123] [Medline: 40228833]
44. Young K. Internet Addiction Test (IAT): Stoelting; 2016. URL: https://stoeltingco.com/Psychological/media/IAT_web_sample.pdf [accessed 2025-11-20]
45. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, 5th edition 2013. URL: https://psychiatryonline.org/doi/book/10.1176/appi.books.9780890425596 [doi: 10.1176/appi.books.9780890425596]
46. Lecrubier Y, Sheehan D, Weiller E, et al. The MINI International Neuropsychiatric Interview (MINI). A short diagnostic structured interview: reliability and validity according to the CIDI. Eur psychiatr 1997;12(5):224-231. [doi: 10.1016/S0924-9338(97)83296-8]
47. Li G, Wang Z, Hao Y, et al. Consensus statement of Chinese experts on exercise prescription (2023). Sports Med Health Sci 2024 Jun;6(2):200-203. [doi: 10.1016/j.smhs.2024.02.003] [Medline: 38708321]
48. Liguori G. ACSM's Guidelines for Exercise Testing and Prescription, 11th edition 2023. URL: https://shop.lww.com/ACSM-s-Guidelines-for-Exercise-Testing-and-Prescription/p/9781975219208?srsltid=AfmBOopMq-c63QBP5gqfSx66zvXrXHCOxCnb1sILLpyxyXwlWpnxKHMi [accessed 2025-11-20]

XSL•FO
RenderX

49. Piercy KL, Troiano RP, Ballard RM, et al. The Physical Activity Guidelines for Americans. JAMA 2018 Nov 20;320(19):2020-2028. [doi: 10.1001/jama.2018.14854] [Medline: 30418471]
50. McArdle WD, Katch FI, Pechar GS, Jacobson L, Ruck S. Reliability and interrelationships between maximal oxygen intake, physical work capacity and step-test scores in college women. Med Sci Sports 1972;4(4):182-186. [doi: 10.1249/00005768-197200440-00019] [Medline: 4648576]
51. Tiffany ST, Drobes DJ. The development and initial validation of a questionnaire on smoking urges. Br J Addict 1991 Nov;86(11):1467-1476. [doi: 10.1111/j.1360-0443.1991.tb01732.x] [Medline: 1777741]
52. Statistical parametric mapping. Science Direct. URL: https://www.sciencedirect.com/topics/neuroscience/statistical-parametric-mapping [accessed 2025-11-21]
53. Statistical parametric mapping. URL: http://www.fil.ion.ucl.ac.uk/spm [accessed 2025-11-21]
54. Yan CG, Wang XD, Zuo XN, Zang YF. DPABI: data processing & analysis for (resting-state) brain imaging. Neuroinformatics 2016 Jul;14(3):339-351. [doi: 10.1007/s12021-016-9299-4] [Medline: 27075850]
55. DPABI: a toolbox for data processing & analysis for brain imaging. The R-fMRI Network. URL: https://rfmri.org/DPABI [accessed 2025-11-21]
56. Friston KJ, Williams S, Howard R, Frackowiak RSJ, Turner R. Movement-related effects in fMRI time-series. Magn Reson Med 1996 Mar;35(3):346-355. [doi: 10.1002/mrm.1910350312] [Medline: 8699946]
57. Love J, Selker R, Marsman M, et al. JASP: graphical statistical software for common statistical designs. J Stat Soft 2019;88(2):1-17. [doi: 10.18637/jss.v088.i02]
58. JASP: a fresh way to do statistics. URL: https://jasp-stats.org/ [accessed 2025-11-21]
59. Dukart J, Holiga S, Rullmann M, et al. JuSpace: a tool for spatial correlation analyses of magnetic resonance imaging data with nuclear imaging derived neurotransmitter maps. Hum Brain Mapp 2021 Feb 15;42(3):555-566. [doi: 10.1002/hbm.25244] [Medline: 33079453]
60. Abrantes AM, Farris SG, Minami H, Strong DR, Riebe D, Brown RA. Acute effects of aerobic exercise on affect and smoking craving in the weeks before and after a cessation attempt. Nicotine Tob Res 2018 Apr 2;20(5):575-582. [doi: 10.1093/ntr/ntx104] [Medline: 28505303]
61. Sari S, Bilberg R, Søgaard Nielsen A, Roessler KK. The effect of exercise as adjunctive treatment on quality of life for individuals with alcohol use disorders: a randomized controlled trial. BMC Public Health 2019 Jun 11;19(1):727. [doi: 10.1186/s12889-019-7083-8] [Medline: 31185955]
62. Nee DE. Integrative frontal-parietal dynamics supporting cognitive control. Elife 2021 Mar 2;10:e57244. [doi: 10.7554/eLife.57244] [Medline: 33650966]
63. Thomas Yeo BT, Krienen FM, Sepulcre J, et al. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. J Neurophysiol 2011 Sep;106(3):1125-1165. [doi: 10.1152/jn.00338.2011]
64. Katsumi Y, Putcha D, Eckbo R, et al. Anterior dorsal attention network tau drives visual attention deficits in posterior cortical atrophy. Brain (Bacau) 2023 Jan 5;146(1):295-306. [doi: 10.1093/brain/awac245] [Medline: 36237170]
65. Chen S, Yan J, Lock M, et al. Alterations of gray matter asymmetry in internet gaming disorder. Sci Rep 2024;14(1):28282. [doi: 10.1038/s41598-024-79659-7]
66. Dong G, Lin X, Potenza MN. Decreased functional connectivity in an executive control network is related to impaired executive function in internet gaming disorder. Prog Neuropsychopharmacol Biol Psychiatry 2015 Mar 3;57:76-85. [doi: 10.1016/j.pnpbp.2014.10.012] [Medline: 25445475]
67. Dong G, Lin X, Hu Y, Xie C, Du X. Imbalanced functional link between executive control network and reward network explain the online-game seeking behaviors in internet gaming disorder. Sci Rep 2015 Mar 17;5(1):9197. [doi: 10.1038/srep09197] [Medline: 25779894]
68. Graziano MSA, Taylor CSR, Moore T. Complex movements evoked by microstimulation of precentral cortex. Neuron 2002 May 30;34(5):841-851. [doi: 10.1016/s0896-6273(02)00698-0] [Medline: 12062029]
69. Desmurget M, Richard N, Harquel S, et al. Neural representations of ethologically relevant hand/mouth synergies in the human precentral gyrus. Proc Natl Acad Sci U S A 2014 Apr 15;111(15):5718-5722. [doi: 10.1073/pnas.1321909111] [Medline: 24706796]
70. Xu X, Yang B, Zheng Y, Ni H, Dong GH. Addicted brains on brief abstinence: similarities and differences in functional connectivity patterns between internet gaming disorder and tobacco use disorder. Prog Neuropsychopharmacol Biol Psychiatry 2025 Mar 20;137:111286. [doi: 10.1016/j.pnpbp.2025.111286] [Medline: 39971120]
71. Haber SN, Knutson B. The reward circuit: linking primate anatomy and human imaging. Neuropsychopharmacology 2010 Jan;35(1):4-26. [doi: 10.1038/npp.2009.129] [Medline: 19812543]
72. Ko M, Chi SH, Lee JH, Suh SI, Lee MS. Altered functional connectivity of the nucleus accumbens and amygdala in cyber addiction: a resting state functional magnetic resonance imaging study. Clin Psychopharmacol Neurosci 2023 May 30;21(2):304-312. [doi: 10.9758/cpn.2023.21.2.304] [Medline: 37119223]
73. Wei L, Zhang S, Turel O, Bechara A, He Q. A tripartite neurocognitive model of internet gaming disorder. Front Psychiatry 2017;8:285. [doi: 10.3389/fpsyt.2017.00285] [Medline: 29312016]
74. Raiha S, Yang G, Wang L, et al. Altered reward processing system in internet gaming disorder. Front Psychiatry 2020;11:599141. [doi: 10.3389/fpsyt.2020.599141] [Medline: 33343426]

XSL•FO
RenderX

75. Hopfinger JB, Buonocore MH, Mangun GR. The neural mechanisms of top-down attentional control. Nat Neurosci 2000 Mar;3(3):284-291. [doi: 10.1038/72999] [Medline: 10700262]
76. Corbetta M, Shulman GL. Control of goal-directed and stimulus-driven attention in the brain. Nat Rev Neurosci 2002 Mar;3(3):201-215. [doi: 10.1038/nrn755] [Medline: 11994752]
77. Han X, Wang Y, Jiang W, et al. Resting-state activity of prefrontal-striatal circuits in internet gaming disorder: changes with cognitive behavior therapy and predictors of treatment response. Front Psychiatry 2018;9:341. [doi: 10.3389/fpsyt.2018.00341] [Medline: 30123144]
78. Yuan K, Yu D, Cai C, et al. Frontostriatal circuits, resting state functional connectivity and cognitive control in internet gaming disorder. Addict Biol 2017 May;22(3):813-822. [doi: 10.1111/adb.12348] [Medline: 26769234]
79. Chen CY, Yen JY, Wang PW, Liu GC, Yen CF, Ko CH. Altered functional connectivity of the insula and nucleus accumbens in internet gaming disorder: a resting state fMRI study. Eur Addict Res 2016;22(4):192-200. [doi: 10.1159/000440716] [Medline: 26863028]
80. Ma L, Mei B, Zhang M, et al. Integrative gray matter volume and molecular analyses of altered intrinsic neural timescale in internet gaming disorder. Prog Neuropsychopharmacol Biol Psychiatry 2025 Mar 20;137:111296. [doi: 10.1016/j.pnpbp.2025.111296] [Medline: 39988256]
81. Ma L, Li X, Mei B, et al. Individualized gray matter morphological abnormalities unveil two neuroanatomical internet gaming disorder subtypes. J Psychiatr Res 2025 Nov;191:180-188. [doi: 10.1016/j.jpsychires.2025.09.064] [Medline: 41027320]
82. Hirvonen J, Zanotti-Fregonara P, Gorelick DA, et al. Decreased cannabinoid CB1 receptors in male tobacco smokers examined with positron emission tomography. Biol Psychiatry Cogn Neurosci Neuroimaging 2018 Nov;84(10):715-721. [doi: 10.1016/j.biopsych.2018.07.009]
83. Wenger T, Moldrich G, Furst S. Neuromorphological background of cannabis addiction. Brain Res Bull 2003 Jul 15;61(2):125-128. [doi: 10.1016/s0361-9230(03)00081-9] [Medline: 12831997]
84. Desai S, Borg B, Cuttler C, et al. A Systematic Review and Meta-Analysis on the Effects of Exercise on the Endocannabinoid System. Cannabis Cannabinoid Res 2022 Aug;7(4):388-408. [doi: 10.1089/can.2021.0113] [Medline: 34870469]
85. Batista EML, Doria JG, Ferreira-Vieira TH, et al. Orchestrated activation of mGluR5 and CB1 promotes neuroprotection. Mol Brain 2016 Aug 20;9(1):80. [doi: 10.1186/s13041-016-0259-6] [Medline: 27543109]
86. Stachowicz K. Interactions between metabotropic glutamate and CB1 receptors: implications for mood, cognition, and synaptic signaling based on data from mGluR and CB1R-targeting drugs. Pharmacol Rep 2024 Dec;76(6):1286-1296. [doi: 10.1007/s43440-024-00612-6] [Medline: 38941064]

Abbreviations

DC: degree centrality
DICOM: Digital Imaging and Communication in Medicine
DLFPC: dorsolateral prefrontal cortex
DPABI: Data Processing & Analysis of Brain Imaging
DSM-5: The Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
ECN: executive control network
FC: functional connectivity
fMRI: functional magnetic resonance imaging
FT: free training
ICD-11: The International Classification of Diseases, 11th Revision
MINI: Mini International Neuropsychiatric Interview
PAT: progressive aerobic training
R-mSFG: right medial superior frontal gyrus
ReHo: regional homogeneity
RN: reward network
ROI: region of interest
SFG: superior frontal gyrus
SPG: superior parietal gyrus

XSL•FO
RenderX

Edited by S Brini; submitted 05.09.25; peer-reviewed by C Ahmadu, SS Kollala; revised version received 13.11.25; accepted 14.11.25; published 28.11.25.

Please cite as:
Cui S, Xu X, Luo X, Wei M, Dong GH
Effects of Progressive Aerobic Training on Executive-Reward Network Connectivity and Symptoms of Internet Gaming Disorder: Randomized Controlled Trial
JMIR Serious Games 2025;13:e83597
URL: https://games.jmir.org/2025/1/e83597
doi:10.2196/83597

© Shaoyu Cui, Xuefeng Xu, Xin Luo, Meiting Wei, Guang-Heng Dong. Originally published in JMIR Serious Games (https://games.jmir.org), 28.11.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Correction: A Novel Approach Using Serious Game Data to Predict the WISC-V Processing Speed Index in Children With Attention-Deficit/Hyperactivity Disorder: Machine Learning Study

Jun-Su Kim[1*], MS; Yoo Joo Jeong[1,2,3*], BS; Seung-Jae Kim[1], MS; Su Jin Jun[1], BS; Jin-Yeop Park[1], BS; Hyang-Sook Hoe[1,2,3], PhD; Jeong-Heon Song[1], PhD

[1]AI-based Neurodevelopmental Diseases Digital Therapeutics Group, Korea Brain Research Institute (KBRI), 61, Cheomdan-ro, Daegu, Republic of Korea

[2]Neurodegenerative Disease Group, Korea Brain Research Institute (KBRI), Daegu, Republic of Korea

[3]Department of Brain and Cognitive Sciences, Gyeongbuk Institute of Science & Technology, Daegu, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Jeong-Heon Song, PhD
AI-based Neurodevelopmental Diseases Digital Therapeutics Group, Korea Brain Research Institute (KBRI), 61, Cheomdan-ro, Daegu, Republic of Korea

Related Article:

Correction of: https://games.jmir.org/2025/1/e73408

(JMIR Serious Games 2025;13:e88500) doi:10.2196/88500

In "A Novel Approach Using Serious Game Data to Predict the WISC-V Processing Speed Index in Children With Attention-Deficit/Hyperactivity Disorder: Machine Learning Study" [1], the authors noted one error.

In the *Acknowledgments* section, the following information has been added:

JHS and HSH contributed equally as co-corresponding authors on this work, and the latter can be reached at sookhoe72@kbri.re.kr.

The correction will appear in the online version of the paper on the JMIR Publications website, together with the publication of this correction notice. Because this correction was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Reference

1. Kim JS, Jeong YJ, Kim SJ, et al. A novel approach using serious game data to predict the WISC-V processing speed index in children with attention-deficit/hyperactivity disorder: machine learning study. JMIR Serious Games 2025 Oct 14;13:e73408. [doi: 10.2196/73408] [Medline: 41086393]

Submitted 26.11.25; this is a non–peer-reviewed article;accepted 05.12.25; published 22.12.25.

Please cite as:
Kim JS, Jeong YJ, Kim SJ, Jun SJ, Park JY, Hoe HS, Song JH
Correction: A Novel Approach Using Serious Game Data to Predict the WISC-V Processing Speed Index in Children With Attention-Deficit/Hyperactivity Disorder: Machine Learning Study
JMIR Serious Games 2025;13:e88500
URL: https://games.jmir.org/2025/1/e88500
doi:10.2196/88500

© Jun-Su Kim, Yoo Joo Jeong, Seung-Jae Kim, Su Jin Jun, Jin-Yeop Park, Hyang-Sook Hoe, Jeong-Heon Song. Originally published in JMIR Serious Games (https://games.jmir.org), 22.12.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly

XSL•FO
RenderX

cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Correction: Exergame (ExerG)-Based Physical-Cognitive Training for Rehabilitation in Adults With Motor and Balance Impairments: Usability Study

Silvia Herren[1*], MSc; Barbara Seebacher[2,3,4*], PhD; Sarah Mildner[2], MSc; Yanick Riederer[5], MSc; Ulrike Pachmann[6], MBA, MSc; Nija Sonja Böckler[7], MA; Stephan Niedecken[5], Dipl; Sabrina Alicia Sgandurra[8], MA; Leo Bonati[1,9,10], MD; Isabella Hotz[2], MSc; Alexandra Schättin[5], PhD; Roman Jurt[7], BA; Christian Brenneis[4,11], MD; Katharina Lenfert[2], MA; Frank Behrendt[1,12], PhD; Stefan Schmidlin[7], MA, MSc; Lennart Nacke[13], PhD; Corina Schuster-Amft[1,12,14*], PhD; Anna Lisa Martin-Niedecken[5,7*], PhD

[1]Research Department, Reha Rheinfelden, Rheinfelden, Switzerland
[10]Department of Clinical Research, University of Basel, Basel, Switzerland
[11]Department of Neurology, Clinic for Rehabilitation Muenster, Muenster, Austria
[12]School of Engineering and Computer Science, Bern University of Applied Sciences, Burgdorf, Switzerland
[13]HCI Games Group, Stratford School of Interaction Design and Business, University of Waterloo, Waterloo, ON, Canada
[14]Department for Sport, Exercise and Health, University of Basel, Basel, Switzerland
[2]Department of Rehabilitation Science, Clinic for Rehabilitation Muenster, Gröben 700, Muenster, Austria
[3]Clinical Department of Neurology, Medical University of Innsbruck, Innsbruck, Austria
[4]Karl Landsteiner Institute for Interdisciplinary Rehabilitation Research, Muenster, Austria
[5]Sphery Ltd, Zurich, Switzerland
[6]VASCage – Centre on Clinical Stroke Research, Innsbruck, Austria
[7]Department of Design, Institute for Design Research, Zurich University of the Arts, Zurich, Switzerland
[8]English Language and Literature, Faculty of Arts, University of Waterloo, Waterloo, ON, Canada
[9]Stroke Center and Department of Neurology, University Hospital Basel, Basel, Switzerland
[*]these authors contributed equally

Corresponding Author:
Barbara Seebacher, PhD
Department of Rehabilitation Science, Clinic for Rehabilitation Muenster, Gröben 700, Muenster, Austria

Related Article:

Correction of: https://games.jmir.org/2025/1/e66515

(JMIR Serious Games 2025;13:e73405) doi:10.2196/73405

In "Exergame (ExerG)-Based Physical-Cognitive Training for Rehabilitation in Adults With Motor and Balance Impairments: Usability Study" (JMIR Serious Games 2025;13:e66515) the authors noted one error.

The files attached to the manuscript as Multimedia Appendices 4 and 5 were mixed. To correct this, the PDF that was originally attached as:

> *Multimedia Appendix 4: Further illustration of outcomes in both user groups.*

Has been changed to:

> *Multimedia Appendix 4: Case report forms of primary and secondary end users.*

Furthermore, the PDF that was originally attached to the article as:

> *Multimedia Appendix 5: Case report forms of primary and secondary end users.*

Has been changed to:

> *Multimedia Appendix 5: Further illustration of outcomes in both user groups.*

These files have been attached to this correction notice, with their appropriate captions, as Multimedia Appendix 1 and Multimedia Appendix 2.

The correction will appear in the online version of the paper on the JMIR Publications website, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

XSL•FO
RenderX

Multimedia Appendix 1
Case report forms of primary and secondary end users.
[PDF File, 2277 KB - games_v13i1e73405_app1.pdf]

Multimedia Appendix 2
Further illustration of outcomes in both user groups.
[PDF File, 234 KB - games_v13i1e73405_app2.pdf]

Submitted 04.03.25; this is a non–peer-reviewed article;accepted 04.03.25; published 04.04.25.

Please cite as:
Herren S, Seebacher B, Mildner S, Riederer Y, Pachmann U, Böckler NS, Niedecken S, Alicia Sgandurra S, Bonati L, Hotz I, Schättin A, Jurt R, Brenneis C, Lenfert K, Behrendt F, Schmidlin S, Nacke L, Schuster-Amft C, Martin-Niedecken AL
Correction: Exergame (ExerG)-Based Physical-Cognitive Training for Rehabilitation in Adults With Motor and Balance Impairments: Usability Study
JMIR Serious Games 2025;13:e73405
URL: https://games.jmir.org/2025/1/e73405
doi:10.2196/73405

© Silvia Herren, Barbara Seebacher, Sarah Mildner, Yanick Riederer, Ulrike Pachmann, Nija Sonja Böckler, Stephan Niedecken, Sabrina Alicia Sgandurra, Leo Bonati, Isabella Hotz, Alexandra Schättin, Roman Jurt, Christian Brenneis, Katharina Lenfert, Frank Behrendt, Stefan Schmidlin, Lennart Nacke, Corina Schuster-Amft, Anna Lisa Martin-Niedecken. Originally published in JMIR Serious Games (https://games.jmir.org), 4.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Correction: A Mixed Reality–Based Telesupervised Ultrasound Education Platform on 5G Network Compared to Direct Supervision: Prospective Randomized Pilot Trial

Minha Kim[1,2*], MD; Meong Hi Son[1,3,4*], MD; Suhyeon Moon[5], MS; Won Chul Cha[1,3,4], MD, PhD; Ik Joon Jo[1], MD, PhD; Hee Yoon[1*], MD, PhD

[1]Department of Emergency Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, 81, Irwon-ro, Gangnam-gu, Seoul, Republic of Korea

[2]Department of Medical Sciences, Graduate School of Kangwon National University, Chuncheon, Republic of Korea

[3]Department of Digital Health, Samsung Advanced Institute for Health Science & Technology, Sungkyunkwan University, Seoul, Republic of Korea

[4]Data Innovation Center, Samsung Medical Center, Seoul, Republic of Korea

[5]Division of Biostatistics, Department of Academic Research, Kangbuk Samsung Hospital, Seoul, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Hee Yoon, MD, PhD
Department of Emergency Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, 81, Irwon-ro, Gangnam-gu, Seoul, Republic of Korea

Related Article:

Correction of: https://games.jmir.org/2025/1/e63448

(JMIR Serious Games 2025;13:e77586) doi:10.2196/77586

In "A Mixed Reality–Based Telesupervised Ultrasound Education Platform on 5G Network Compared to Direct Supervision: Prospective Randomized Pilot Trial" (JMIR Serious Games 2025;13:e63448) the authors noted one error.

Affiliation 3 has been changed from:

> *Department of Health Sciences and Technology, Samsung Advanced Institute for Health Science & Technology, Sungkyunkwan University, Seoul, Republic of Korea*

To appear as follows:

> *Department of Digital Health, Samsung Advanced Institute for Health Science & Technology, Sungkyunkwan University, Seoul, Republic of Korea*

The correction will appear in the online version of the paper on the JMIR Publications website, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 15.05.25; this is a non–peer-reviewed article;accepted 15.05.25; published 12.06.25.

Please cite as:
Kim M, Hi Son M, Moon S, Cha WC, Jo IJ, Yoon H
Correction: A Mixed Reality–Based Telesupervised Ultrasound Education Platform on 5G Network Compared to Direct Supervision: Prospective Randomized Pilot Trial
JMIR Serious Games 2025;13:e77586
URL: https://games.jmir.org/2025/1/e77586
doi:10.2196/77586

© Meong Hi Son, Suhyeon Moon, Won Chul Cha, Ik Joon Jo, Hee Yoon. Originally published in JMIR Serious Games (https://games.jmir.org), 12.6.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic

XSL•FO
RenderX

information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Corrigenda and Addenda

Correction: Effects and Acceptability of a 1-Week Home-Based Virtual Reality Training for Supporting the Management of Stress and Anxiety: Randomized Pilot Trial

Federica Pallavicini[1], PhD; Eleonora Orena[2], MA; Lisa Arnoldi[1,3], MA; Federica Achille[3], MA; Stefano Stefanini[3], MA; Maddalena Cassa[2], MA; Alessandro Pepe[1], PhD; Guido Veronese[1], PhD; Luca Bernardelli[4], MA; Francesca Sforza[4], MA; Sara Fascendini[3], MA; Carlo Alberto Defanti[3], MA; Marco Gemma[2], MA; Massimo Clerici[5], PhD; Giuseppe Riva[6,7], PhD; Fabrizia Mantovani[1], PhD

[1]Department of Human Sciences for Education "Riccardo Massa", University of Milano-Bicocca, Milan, Italy

[2]IRCCS Neurological Institute Carlo Besta, Milan, Italy

[3]European Biomedical Research Foundation, Gazzaniga, Italy

[4]Become-Hub, Milan, Italy

[5]Department of Medicine and Surgery, University of Milano-Bicocca, Milan, Italy

[6]Humane Technology Lab, Department of Psychology, Università Cattolica del Sacro Cuore, Milan, Italy

[7]Applied Technology for Neuro-Psychology Lab, IRCCS Istituto Auxologico Italiano, Milan, Italy

Corresponding Author:
Federica Pallavicini, PhD
Department of Human Sciences for Education "Riccardo Massa"
University of Milano-Bicocca
Piazza dell'Ateneo Nuovo 1, 20126
Milan, 20100
Italy
Phone: 39 02644484944
Email: federica.pallavicini@unimib.it

Related Article:

Correction of: https://games.jmir.org/2025/1/e50326

(JMIR Serious Games 2025;13:e77791) doi:10.2196/77791

In "Effects and Acceptability of a 1-Week Home-Based Virtual Reality Training for Supporting the Management of Stress and Anxiety: Randomized Pilot Trial" (JMIR Serious Games 2025;13:e50326) the authors noted one error.

In the original article, the affiliation of Giuseppe Riva was incorrectly listed as:

> *"Applied Technology for Neuro-Psychology Laboratory, IRCCS Italian Auxological Institute, Milan, Italy."*

This has been replaced by:

> *"Applied Technology for Neuro-Psychology Lab, IRCCS Istituto Auxologico Italiano, Milan, Italy."*

The correction will appear in the online version of the paper on the JMIR Publications website, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

XSL•FO
RenderX

Submitted 20.05.25; this is a non–peer-reviewed article;accepted 23.05.25; published 10.11.25.

Please cite as:
Pallavicini F, Orena E, Arnoldi L, Achille F, Stefanini S, Cassa M, Pepe A, Veronese G, Bernardelli L, Sforza F, Fascendini S, Defanti CA, Gemma M, Clerici M, Riva G, Mantovani F
Correction: Effects and Acceptability of a 1-Week Home-Based Virtual Reality Training for Supporting the Management of Stress and Anxiety: Randomized Pilot Trial
JMIR Serious Games 2025;13:e77791
URL: https://games.jmir.org/2025/1/e77791
doi:10.2196/77791
PMID:41212613

©Federica Pallavicini, Eleonora Orena, Lisa Arnoldi, Federica Achille, Stefano Stefanini, Maddalena Cassa, Alessandro Pepe, Guido Veronese, Luca Bernardelli, Francesca Sforza, Sara Fascendini, Carlo Alberto Defanti, Marco Gemma, Massimo Clerici, Giuseppe Riva, Fabrizia Mantovani. Originally published in JMIR Serious Games (https://games.jmir.org), 10.11.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Corrigenda and Addenda

Correction: Serious Game for the Nursing Assessment of Home-Dwelling Older Adults: Development and Validation Study

Erica Busca[1,2*], PhD; Silvia Caristia[3*], PhD; Sara Palmira Bidone[4], MSc; Alessia Bolamperti[2,5], MSc; Sara Campagna[6], PhD; Arianna Cattaneo[1], MSc; Rosaria Lea[2,5], MSc; Doriana Montani[2], MSc; Antonio Scalogna[5], MSc; Francesca Piovesan[7], PhD; Erika Bassi[1,2], PhD; Alberto Dal Molin[1,2], PhD

[1]Department of Translational Medicine, University of Piemonte Orientale, Novara, Italy

[2]Azienda Ospedaliero Universitaria Maggiore della Carità, Novara, Italy

[3]Department for Sustainable Development and Ecological Transition, University of Piemonte Orientale, Vercelli, Italy

[4]Health and Socio-Health Surveillance Commission, Azienda Sanitaria Locale Alessandria, Alessandria, Italy

[5]Interdepartmental Centre for Innovative Didactics and Simulation in Medicine and Health Professions, University of Piemonte Orientale, Novara, Italy

[6]Department of Public Health and Pediatrics, University of Torino, Torino, Italy

[7]BIP (BraIn Plasticity and Behaviour Changes) Research Group, Department of Psychology, University of Turin, Turin, Italy

[*]these authors contributed equally

Corresponding Author:
Silvia Caristia, PhD
Department for Sustainable Development and Ecological Transition
University of Piemonte Orientale
Piazza Sant' Eusebio 5
Vercelli, 13100
Italy
Phone: 39 3287472630
Email: silvia.caristia@uniupo.it

Related Article:

Correction of: https://games.jmir.org/2024/1/e52644

(JMIR Serious Games 2025;13:e81884) doi:10.2196/81884

In “Serious Game for the Nursing Assessment of Home-Dwelling Older Adults: Development and Validation Study” [1] the authors made one addition.

Francesca Piovesan has been added as a co-author on the manuscript. FR made a substantial intellectual contribution to the manuscript, specifically during the writing and critical revision phases. Her involvement was acknowledged and agreed upon by the research team, and was inadvertently omitted.

The authorship list has been changed as follows:

> *Erica Busca, PhD; Silvia Caristia, PhD; Sara Palmira Bidone, MSc; Alessia Bolamperti, MSc; Sara Campagna, PhD; Arianna Cattaneo, MSc; Rosaria Lea, MSc; Doriana Montani, MSc; Antonio Scalogna, MSc; Francesca Piovesan, PhD; Erika Bassi, PhD; Alberto Dal Molin, PhD*

And the following affiliation has been added, with subsequent affiliations renumbered accordingly:

> *BIP (BraIn Plasticity and Behaviour Changes) Research Group, Department of Psychology, University of Turin, Turin, Italy*

The correction will appear in the online version of the paper on the JMIR Publications website, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Reference

1. Busca E, Caristia S, Bidone SP, Bolamperti A, Campagna S, Cattaneo A, et al. Serious game for the nursing assessment of home-dwelling older adults: development and validation study. JMIR Serious Games 2024 Nov 26;12:e52644 [FREE Full text] [doi: 10.2196/52644] [Medline: 39591607]

XSL•FO
RenderX

Edited by P Kharbanda; submitted 05.08.25; this is a non–peer-reviewed article;accepted 12.08.25; published 17.12.25.

Please cite as:
Busca E, Caristia S, Bidone SP, Bolamperti A, Campagna S, Cattaneo A, Lea R, Montani D, Scalogna A, Piovesan F, Bassi E, Dal Molin A
Correction: Serious Game for the Nursing Assessment of Home-Dwelling Older Adults: Development and Validation Study
JMIR Serious Games 2025;13:e81884
URL: https://games.jmir.org/2025/1/e81884
doi:10.2196/81884
PMID:

©Erica Busca, Silvia Caristia, Sara Palmira Bidone, Alessia Bolamperti, Sara Campagna, Arianna Cattaneo, Rosaria Lea, Doriana Montani, Antonio Scalogna, Francesca Piovesan, Erika Bassi, Alberto Dal Molin. Originally published in JMIR Serious Games (https://games.jmir.org), 17.12.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Correction: FITLIGHT Training and Its Influence on Visual-Motor Reactions and Dribbling Speed in Female Basketball Players: Prospective Evaluation Study

Ahmad K Hassan[1,2], PhD

[1]Department of Physical Education, College of Education, King Faisal University, P.O. Box 400, Al-Ahsa, Saudi Arabia

[2]Department of Team Sports and Racket Games, Faculty of Physical Education, Minia University, Minia, Egypt

Corresponding Author:
Ahmad K Hassan, PhD
Department of Physical Education, College of Education, King Faisal University, P.O. Box 400, Al-Ahsa, Saudi Arabia

Related Article:

Correction of: https://games.jmir.org/2025/1/e70519

Abstract

(JMIR Serious Games 2025;13:e82614) doi:10.2196/82614

In "FITLIGHT Training and Its Influence on Visual-Motor Reactions and Dribbling Speed in Female Basketball Players: Prospective Evaluation Study" (JMIR Serious Games 2025;13:e70519), the author noted two corrections.

The corresponding email address was changed to "amohhamed@kfu.edu.sa."

In addition, the following affiliation was added: "Department of Team Sports and Racket Games, Faculty of Physical Education, Minia University, Minya, Egypt."

The correction will appear in the online version of the paper on the JMIR Publications website, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 18.08.25; this is a non–peer-reviewed article;accepted 18.08.25; published 12.09.25.

Please cite as:
Hassan AK
Correction: FITLIGHT Training and Its Influence on Visual-Motor Reactions and Dribbling Speed in Female Basketball Players: Prospective Evaluation Study
JMIR Serious Games 2025;13:e82614
URL: https://games.jmir.org/2025/1/e82614
doi:10.2196/82614

© Ahmad K Hassan. Originally published in JMIR Serious Games (https://games.jmir.org), 12.9.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Author's Reply: Refining Open-World Game and Nostalgic Film Interventions for Broader and More Reliable Therapeutic Impact

Andreas Benedikt Eisingerich[1]; Annisa Arigayota[1]; Barbara Duffek[2]; Congcong Hou[3]

[1]Imperial College Business School, Imperial College London, South Kensington Campus, London, United Kingdom

[2]Robinson College of Business, Georgia State University, Atlanta, GA, United States

[3]Kyushu Sangyo University, Fukuoka, Japan

Corresponding Author:
Andreas Benedikt Eisingerich
Imperial College Business School, Imperial College London, South Kensington Campus, London, United Kingdom

Related Articles:

https://games.jmir.org/2025/1/e76522

https://games.jmir.org/2025/1/e82157

(JMIR Serious Games 2025;13:e83239) doi:10.2196/83239

KEYWORDS

open-world games; nostalgia; lab experimental study; postgraduate students; My Neighbor Totoro; Kiki's Delivery Service; Nintendo

We greatly welcome and thank Sun and Meng for their thoughtful comments in their letter to the editor [1] in response to our research [2]. We are grateful for the insightful points shared on how work on creative media, such as video games and films, may help challenge stereotypes about gaming and engaging with storytelling to potentially open doors accessible and innovative tools for well-being. Open-world games may have the potential to enhance young people's mental well-being [3], whilst social media and other forms of internet usage may lead to stress and anxiety [4,5]. We respond to the kind and insightful points shared [2] to further help bridge creative media and mental health science.

First, we concur regarding the stated importance to examine the observed effects in our study [1], using participants who may not already be predisposed to enjoy the stimuli such as graduate students. As we note in our article, we recruited participants by telling them that they had a chance to take part in a university study on daily activities and well-being. Thus, the danger of people who already like watching Studio Ghibli films and enjoy playing The Legend of Zelda: Breath of the Wild to self-select to take part in the study is low. We also controlled for study participants' familiarity with and enjoyment of the creative art content. That said, we fully agree that that future research should study the impact of creative media on well-being using a larger, broader and more varied sample in terms of different study participant demographics (age, cultural background, etc).

Second, we concur that due to the intervention's brevity (ie, 30 min of video gameplay and roughly 7 minutes of film viewing), only a snapshot of the possible psychological effect may be captured and may constitute a conservative test. This is an excellent point and we encourage future research to explore the impact of long-term intervention as well as varied intervention formats (eg, multi-session, spaced over several days and weeks and possibly even years) on long-term well-being gains. Additional work is needed to help examine the long-term effects of playing open-world games such as The Legend of Zelda: Breath of the Wild and watching Studio Ghibli classics such as My Neighbor Totor and Kiki's Delivery Service on mental well-being. In addition, we invite future research to explore the potential effects of games and films other than the ones invested in our study [1] to determine the extent various art forms may lead to benefits in calmness, exploration, enhanced meaning, and sense of purpose in life as well as overall life happiness.

Third, study participants were randomly assigned to the study's experimental conditions. We agree that the study of how mood, recent stressors, sleep quality (or lack thereof), social interactions, the consumption of one's favorite beverage (eg, coffee, Red Bull, etc) may interact with the intervention's effects is richly deserving.

In summary, we thank Meng and Sun for their thoughtful and excellent comments on how future research may help bridge gaming and storytelling with mental health science.

Authors' Contributions

All authors contributed equally.

XSL•FO
RenderX

Conflicts of Interest

None declared.

References

1. Meng K, Sun C. Refining open-world games and nostalgic film interventions for broader and more reliable therapeutic impact. JMIR Serious Games 2025 Sep 30:e82157. [doi: 10.2196/82157]
2. Arigayota A, Duffek B, Hou C, Eisingerich AB. Effects of the legend of Zelda: breath of the wild and studio Ghibli films on young people's sense of exploration, calm, mastery and skill, purpose and meaning, and overall happiness in life: exploratory randomized controlled study. JMIR Serious Games 2025 Aug 1;13:e76522. [doi: 10.2196/76522] [Medline: 40750097]
3. Anto A, Basu A, Selim R, Foscht T, Eisingerich AB. Open-world games' affordance of cognitive escapism, relaxation, and mental well-being among postgraduate students: mixed methods study. J Med Internet Res 2024 Dec 17;26:e63760. [doi: 10.2196/63760] [Medline: 39689301]
4. Anto A, Asif RO, Basu A, et al. Exploring the impact of social media on anxiety among university students in the United Kingdom: qualitative study. JMIR Form Res 2023 Jun 16;7:e43037. [doi: 10.2196/43037] [Medline: 37327030]
5. Selim R, Basu A, Anto A, Foscht T, Eisingerich AB. Effects of large language model-based offerings on the well-being of students: qualitative study. JMIR Form Res 2024 Dec 27;8:e64081. [doi: 10.2196/64081] [Medline: 39729617]

Edited by A Coristine; submitted 29.08.25; this is a non–peer-reviewed article;accepted 05.09.25; published 30.09.25.

Please cite as:
Eisingerich AB, Arigayota A, Duffek B, Hou C
Author's Reply: Refining Open-World Game and Nostalgic Film Interventions for Broader and More Reliable Therapeutic Impact
JMIR Serious Games 2025;13:e83239
URL: https://games.jmir.org/2025/1/e83239
doi:10.2196/83239

© Andreas Benedikt Eisingerich, Annisa Arigayota, Barbara Duffek, Congcong Hou. Originally published in JMIR Serious Games (https://games.jmir.org), 30.9.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Refining Open-World Game and Nostalgic Film Interventions for Broader and More Reliable Therapeutic Impact

Chenlei Sun[1], PhD; Ke Meng[2], PhD

[1]Department of Music, Yeungnam University, 280 Daehak-Ro, Gyeongsan, Republic of Korea

[2]Department of Global Convergence, Kangwon National University, Chuncheon, Republic of Korea

Corresponding Author:
Chenlei Sun, PhD
Department of Music, Yeungnam University, 280 Daehak-Ro, Gyeongsan, Republic of Korea

Related Articles:

https://games.jmir.org/2025/1/e76522

Comment in: https://games.jmir.org/2025/1/e83239

(JMIR Serious Games 2025;13:e82157) doi:10.2196/82157

KEYWORDS

open-world games; nostalgia; lab experimental study; postgraduate students; My Neighbor Totoro; Kiki's Delivery Service; Nintendo

We read with great interest the recent study by Arigayota et al [1] that explored the interplay between immersive gameplay, nostalgic film experiences, and psychological well-being. As art therapists, we find this approach inspiring, as it bridges creative media and emotional health. However, in considering how these findings might translate to real-world interventions, several methodological aspects deserve discussion.

First, the recruitment strategy—inviting graduate students to volunteer—may inadvertently favor individuals already inclined toward gaming, Studio Ghibli films, or similar forms of media engagement. Such self-selection could produce a sample with higher baseline openness to immersive and nostalgic experiences, potentially amplifying the observed effects. In therapeutic practice, participants often have varied or even limited familiarity with certain media. Future research could broaden recruitment to include individuals with differing media backgrounds and use screening tools to capture baseline engagement preferences. This would help clarify whether the observed benefits generalize beyond those already predisposed to enjoy these stimuli.

Second, the intervention's brevity—30 minutes of gameplay and approximately 7 minutes of film viewing—captures only a snapshot of the possible psychological impact. While short sessions are pragmatic in experimental settings, art therapy often relies on sustained engagement to foster deeper emotional processing and behavioral change [2,3]. Extending the intervention into a multi-session format, spaced over days or weeks, could reveal whether benefits in calmness, exploration, and meaning consolidate into more enduring well-being gains. Moreover, varying session length could help identify the minimal "dose" necessary to elicit therapeutic outcomes, offering valuable guidance for applied settings.

Third, while the study accounted for familiarity and liking of the media, other potential confounding variables—such as the participants' mood before the session, recent stressors, sleep quality, or social interactions—were not fully controlled. These factors can significantly influence immediate emotional responses and may have interacted with the intervention's effects. In a therapeutic context, we routinely monitor and adapt to such variables, as they shape how individuals respond to creative activities. Incorporating pre-session mood assessments or controlling for recent life events could help disentangle the intervention's unique contributions from the influence of transient emotional states.

In summary, this study offers a valuable foundation for integrating immersive and nostalgic media into well-being interventions. Addressing self-selection bias, exploring longer or repeated exposures, and more comprehensively controlling for confounding factors could strengthen future research and enhance its applicability in therapeutic settings. By refining these aspects, the promising combination of open-world gaming and nostalgic film could evolve into a versatile, evidence-based tool for supporting emotional health across diverse populations.

XSL•FO
RenderX

Authors' Contributions

KM, writing–original draft, writing–review and editing, conceptualization; CS, writing–original draft, writing–review and editing, investigation.

Conflicts of Interest

None declared.

References

1. Arigayota A, Duffek B, Hou C, Eisingerich AB. Effects of the legend of Zelda: breath of the wild and studio Ghibli films on young people's sense of exploration, calm, mastery and skill, purpose and meaning, and overall happiness in life: exploratory randomized controlled study. JMIR Serious Games 2025 Aug 1;13:e76522. [doi: 10.2196/76522] [Medline: 40750097]
2. Czamanski-Cohen J, Weihs KL. The bodymind model: a platform for studying the mechanisms of change induced by art therapy. Arts Psychother 2016 Sep;51:63-71. [doi: 10.1016/j.aip.2016.08.006] [Medline: 27777492]
3. Quinn P. Art therapy's engagement of brain networks for enduring recovery from addiction. Front Psychiatry 2024;15:1458063. [doi: 10.3389/fpsyt.2024.1458063] [Medline: 39834572]

Edited by A Coristine; submitted 10.08.25; this is a non–peer-reviewed article;accepted 04.09.25; published 01.10.25.

Please cite as:
Sun C, Meng K
Refining Open-World Game and Nostalgic Film Interventions for Broader and More Reliable Therapeutic Impact
JMIR Serious Games 2025;13:e82157
URL: https://games.jmir.org/2025/1/e82157
doi:10.2196/82157

© Chenlei Sun, Ke Meng. Originally published in JMIR Serious Games (https://games.jmir.org), 1.10.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Authors' Reply: Is the Pinball Machine a Blind Spot in Serious Games Research?

Luis Carlos Rodríguez Timaná[1], MSc; Javier Ferney Castillo García[2], MSc, PhD; Teodiano Bastos Filho[3], MSc, PhD; Alvaro Alexander Ocampo González[4], MSc, PhD; Nazly Rocio Hincapié Monsalve[1], ING; Nicolas Jacobo Valencia Jimenez[1], MSc, PhD

[1]Faculty of Engineering, Universidad Santiago de Cali, Calle 5 # 62-00 - Pampalinda, Santiago de Cali, Colombia

[2]Faculty of Engineering, Universidad Autónoma de Occidente, Cali, Colombia

[3]Faculty of Engineering, Universidade Federal do Espírito Santo, Espiritu Santo, Brazil

[4]Faculty of Health, Universidad Santiago de Cali, Santiago de Cali, Colombia

Corresponding Author:
Luis Carlos Rodríguez Timaná, MSc
Faculty of Engineering, Universidad Santiago de Cali, Calle 5 # 62-00 - Pampalinda, Santiago de Cali, Colombia

Related Articles:

Comment in: https://games.jmir.org/2024/1/e59053

Comment in: https://games.jmir.org/2025/1/e72354

(JMIR Serious Games 2025;13:e73034) doi:10.2196/73034

KEYWORDS

serious games; research; interventions; arcade technology; digital game paradigm; pinball gaming; arcade gaming; executive functions; neurodiversity; cognitive training; therapeutic interventions

We appreciate the insightful comments and reflections from the author of the letter [1] regarding our study on the impact of serious games on executive functions and their application in neurodiverse populations [2]. The suggestion to consider pinball machines as a tool within the serious games paradigm presents an interesting avenue for further exploration.

At the time of our study, our focus was primarily on conventional and emerging digital technologies, such as virtual reality, mobile devices, and sensor-based interactions. However, we acknowledge that pinball, with its unique combination of physical and digital interactions, may offer valuable cognitive and therapeutic benefits, particularly in the context of executive function training. The references provided in the letter [1] highlight historical and recent research supporting its potential applications in various populations, reinforcing the idea that this arcade technology could play a role in future serious game developments.

Given the evidence presented on pinball's ability to engage attention, impulse control, cognitive flexibility, and problem-solving skills, we recognize its potential as a tool to enhance executive function training. Future work in this area could explore the adaptation of pinball mechanics within digital serious games or investigate its direct application as a therapeutic tool in controlled settings.

Additionally, we acknowledge that the development of assistive technologies for neurodiverse populations often encounters blind spots, where certain tools or approaches are overlooked. Our intention with the published paper is to provide a road map for researchers, highlighting that there remains substantial work to be done in this area. By identifying these gaps, we aim to offer a starting point for ongoing and future investigations.

Several studies have underscored the challenges and opportunities in designing technologies for neurodiverse users. For instance, Frauenberger et al [3] discuss the importance of involving neurodiverse children in the technology design process to ensure that their unique needs are met. Similarly, Benton and Johnson [4] highlight lessons from neurodiverse communities, emphasizing the necessity of tailored technological interventions. These perspectives reinforce the need for comprehensive research and development efforts to address the diverse requirements of neurodiverse populations.

We thank the author of the letter [1] for broadening the discussion on serious game technologies. Their insights open the door to new interdisciplinary research possibilities that could further enrich this field.

XSL•FO
RenderX

Conflicts of Interest

None declared.

References

1. Eckardt JP. Is the pinball machine a blind spot in serious games research? JMIR Serious Games 2025;13:e72354. [doi: 10.2196/72354]
2. Rodríguez Timaná LC, Castillo García JF, Bastos Filho T, Ocampo González AA, Hincapié Monsalve NR, Valencia Jimenez NJ. Use of serious games in interventions of executive functions in neurodiverse children: systematic review. JMIR Serious Games 2024 Dec 18;12:e59053. [doi: 10.2196/59053] [Medline: 39693133]
3. Frauenberger C, Good J, Keay-Bright W. Designing technology for children with special needs: bridging perspectives through participatory design. CoDesign 2011 Mar;7(1):1-28. [doi: 10.1080/15710882.2011.587013]
4. Benton L, Johnson H. Widening participation in technology design: a review of the involvement of children with special educational needs and disabilities. Int J Child Comput Interact 2015;3-4:23-40. [doi: 10.1016/j.ijcci.2015.07.001]

Edited by D Majekodunmi; submitted 23.02.25; this is a non–peer-reviewed article;accepted 11.03.25; published 02.04.25.

Please cite as:
Rodríguez Timaná LC, Castillo García JF, Bastos Filho T, Ocampo González AA, Hincapié Monsalve NR, Valencia Jimenez NJ
Authors' Reply: Is the Pinball Machine a Blind Spot in Serious Games Research?
JMIR Serious Games 2025;13:e73034
URL: https://games.jmir.org/2025/1/e73034
doi:10.2196/73034

© Luis Carlos Rodríguez Timaná, Javier Ferney Castillo García, Teodiano Bastos Filho, Alvaro Alexander Ocampo González, Nazly Rocio Hincapié Monsalve, Nicolas Jacobo Valencia Jimenez. Originally published in JMIR Serious Games (https://games.jmir.org), 2.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Is the Pinball Machine a Blind Spot in Serious Games Research?

Jens Peter Eckardt, BSW, MSc

Research Unit/Knowledge Center (Videnscenter), Bedre Psykiatri, Gammeltorv 14. 2 sal, Copenhagen, Denmark

Corresponding Author:
Jens Peter Eckardt, BSW, MSc
Research Unit/Knowledge Center (Videnscenter), Bedre Psykiatri, Gammeltorv 14. 2 sal, Copenhagen, Denmark

Related Articles:

https://games.jmir.org/2024/1/e59053

Comment in: https://games.jmir.org/2025/1/e73034

(JMIR Serious Games 2025;13:e72354) doi:10.2196/72354

KEYWORDS

serious games; research; interventions; arcade technology; digital game paradigm; pinball gaming; arcade gaming; executive functions; neurodiversity; cognitive training; therapeutic interventions

In their insightful and comprehensive study, Rodríguez Timaná et al [1] illuminate the impact and effectiveness of serious games on executive functions and their integration into educational and therapeutic settings for neurodiverse populations aged 3 - 22 years, using various technologies. However, by focusing on conventional and emerging technologies like virtual reality, smartphones, tablets, PCs, and cameras or sensors, one may overlook long-established technology such as the pinball machine. Its enduring presence in the gaming industry, with a rich history of gameplay designs, mechanics, electronics, and both kinetic and digital formats, raises questions. Could pinball's format, rooted in both physical and digital realms, also bridge the gap between traditional and modern approaches to serious gaming? Could it offer a more tangible, interactive experience as a promising therapeutic tool (or as an adjunct) compared to conventional serious games?

For decades, numerous studies of varying quality and methodologies have examined the use of pinball machines as an intervention for individuals with conditions such as autism, severe disabilities, intellectual disabilities, visual impairments, and Parkinson disease. In 1961, the actions of two autistic children were recorded in an experiment involving devices like a pinball machine, aiming to demonstrate how reinforcement worked and to control their behavior [2]. In 1981, a study concluded that adolescents with severe and profound intellectual disabilities could effectively learn and generalize the use of an electronic pinball machine as a leisure skill [3]. In 2021, Luszeck [4] demonstrated the positive benefits of pinball for individuals with Parkinson disease, showing improvements in balance, fine motor skills, and quality of life. In 2022, Johnson et al [5] explored the mental health benefits of pinball for men (including participants with attention-deficit/hyperactivity disorder), revealing improved well-being both directly and through social connections.

Pinball is a game that engages multiple cognitive processes, strengthening executive functions. Players must maintain attention on the ball and targets. The ball moves quickly, requiring focus to react in real time, thus enhancing sustained attention. Pinball involves quick decision-making under pressure. Players control impulses and emotions, making strategic decisions about when and how to launch the ball or use the flippers, thus improving self-regulation. Players also need to remember the board layout, previous actions, and strategies to maximize points, tapping into working memory and planning skills. Pinball requires players to adjust strategies in response to unpredictable events (eg, the ball bouncing off targets), exercising cognitive flexibility. Players anticipate ball movement and plan for upcoming moves, supporting problem-solving skills. Frequent switching of focus between parts of the machine (flippers, bumpers, targets, and video) helps develop task-switching abilities.

Why is pinball gaming and technology underexplored in serious game research, including theoretical exploration, despite its ongoing development? Is it less adaptable or perhaps too costly, compared to other technologies? While pinball is considered entertainment, the cognitive demands it places on players may provide an effective way to exercise executive function skills. To fully harness its potential as a serious game, researchers must expand their focus, paving the way for arcade applications such as the pinball machine within the digital-centric paradigm.

Conflicts of Interest

None declared.

XSL•FO
RenderX

References

1. Rodríguez Timaná LC, Castillo García JF, Bastos Filho T, Ocampo González AA, Hincapié Monsalve NR, Valencia Jimenez NJ. Use of serious games in interventions of executive functions in neurodiverse children: systematic review. JMIR Serious Games 2024 Dec 18;12:e59053. [doi: 10.2196/59053] [Medline: 39693133]
2. Ferster CB, DeMyer MK. The development of performances in autistic children in an automatically controlled environment. J Chronic Dis 1961 Apr;13:312-345. [doi: 10.1016/0021-9681(61)90059-5] [Medline: 13699188]
3. Hill JW, Wehman P, Horst G. Toward generalization of appropriate leisure and social behavior in severely handicapped youth: pinball machine use. Res Pract Persons Severe Disabl 1981 Dec;6(4):38-44. [doi: 10.1177/154079698200600406]
4. Luszeck T. Effects of pinball on fine motor skills and balance on Parkinson's disease: a case report. Arch Phys Med Rehabil 2021 Apr;102(4):E10. [doi: 10.1016/j.apmr.2021.01.033]
5. Johnson D, Wiley K, Passmore C, Horton EM, Altizer R, Mandryk RL. Shoulder-to-shoulder: how pinball supports men's wellbeing. Proc ACM Hum-Comput Interact 2022 Oct 31;6(CHI PLAY):1-25. [doi: 10.1145/3549514]

Edited by D Majekodunmi; submitted 08.02.25; this is a non–peer-reviewed article;accepted 03.03.25; published 02.04.25.

Please cite as:
Eckardt JP
Is the Pinball Machine a Blind Spot in Serious Games Research?
JMIR Serious Games 2025;13:e72354
URL: https://games.jmir.org/2025/1/e72354
doi:10.2196/72354

© Jens Peter Eckardt. Originally published in JMIR Serious Games (https://games.jmir.org), 2.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Gamification Integration in Technological Devices for Motor Rehabilitation in Parkinson Disease: Scoping Review

Pere Bosch-Barceló[1,2,3], MSc; Oriol Martínez-Navarro[1,2,4], PhD; Maria Masbernat-Almenara[1,2,4], PhD; Carlos Tersa-Miralles[1,2,4], PhD; Anni Pakarinen[3], PhD; Helena Fernández-Lago[1,2,4], PhD

[1]Department of Nursing and Physiotherapy, Universitat de Lleida, C/ Montserrat Roig, nº2, Lleida, Spain

[2]Health Care Research Group (GReCS), Lleida Institute for Biomedical Research Dr Pifarré Foundation (IRBLleida), Lleida, Spain

[3]Department of Nursing Science, University of Turku, Turku, Finland

[4]Consolidated Research Group: Society, Health, Education and Culture (GESEC), Department of Nursing and Physiotherapy, Universitat de Lleida, Lleida, Spain

Corresponding Author:
Oriol Martínez-Navarro, PhD
Department of Nursing and Physiotherapy, Universitat de Lleida, C/ Montserrat Roig, nº2, Lleida, Spain

Abstract

Background: Parkinson disease (PD) is a rapidly growing neurological condition worldwide. While physiotherapy and exercise are effective interventions, the addition of motivational aspects that improve adherence could be beneficial for people with PD. Incorporating technological devices into motor rehabilitation, coupled with gamification elements, could enhance the relevance of rehabilitation and alleviate motor symptoms.

Objective: The aim of this scoping review was to identify and classify the technological devices that integrate gamification elements used in motor rehabilitation in PD, and to describe the justification behind the use of these devices and elements in this context.

Methods: We conducted a scoping review following the framework proposed by Joanna Briggs Institute, along with the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) guidelines. Major health science databases (MEDLINE, EMBASE, Scopus, Cochrane, Web of Science, PsycINFO, and Epistemonikos) were systematically searched. Relevant studies were included if they used technological interventions with gamification elements for motor symptom rehabilitation in PD. Gamification elements were extracted and categorized based on established frameworks, and content analysis was used to review the justifications for the use of technologies integrating gamification.

Results: A total of 4451 studies were retrieved from the search. After the abstract and full-text screening, 81 studies were eligible for data extraction. The analysis identified 453 gamification elements across studies, with development and accomplishment being the most prominent core drive. Progress/feedback was the most frequently used element (79/81, 98% of studies), followed by points (70/81, 86%) and levels/progression (66/81, 81%). Other notable elements included badges, leaderboards, and customization, while several core drives, like ownership and possession, lacked reported elements. Most interventions were delivered through commercial video game consoles (33/81, 41%), followed by computer-based systems (32/81, 40%). Tablet-based applications and integrated rehabilitation platforms were used in 11% (9/81) and 10% (8/81) of the studies, respectively. The expected roles of technology were clear, but intentional use of gamification was scarce.

Conclusions: This scoping review highlights the widespread adoption of technologies integrating gamification elements for motor symptom rehabilitation in individuals with PD. However, it also underscores a critical gap in understanding and justifying gamification mechanics. The current landscape relies heavily on commercial video games and emphasizes performance-based experiences, lacking theoretical grounding.

Trial Registration: OSF Registries tx3d9; https://osf.io/tx3d9

(JMIR Serious Games 2025;13:e69433) doi:10.2196/69433

KEYWORDS

gamification; motor rehabilitation; Parkinson disease; rehabilitation; physiotherapy

Introduction

Parkinson disease (PD) is the second most prevalent neurodegenerative disease in the world, as well as the fastest-growing one [1,2]. Over the last 20 years, the global impact of PD in terms of deaths and disability has increased by more than twofold [3]. As of 2021, nearly 12 million people

XSL•FO
RenderX

around the world were living with PD, and this number is expected to increase by approximately 50% by 2035 [4].

PD is mainly characterized by the presence of several cardinal motor symptoms, mainly stiffness, bradykinesia, tremor, and postural instability [5]. The progression of these symptoms heavily impacts the ability of people with PD to manage daily activities [6]. Pharmacological treatment options for PD motor symptoms are mainly comprised of dopamine-based options: levodopa, dopamine agonists, and monoamine oxidase-B inhibitors [7]. However, these pharmacological choices come with their own array of issues, for example, Levodopa is known to eventually cause undesirable and involuntary movements known as dyskinesias in over half of patients after 6 years [8]. Due to these secondary effects, other treatment options are needed for people with PD. Physical therapy and exercise training have proven capable of inducing long-term improvements in motor symptoms and physical function among individuals with PD [9]. These interventions may also enhance the efficacy of pharmacological treatments, and help delay disease progression and age-related decline, with certain modalities additionally improving quality of life, thus becoming a crucial component in the management of people with PD [9,10].

In the last decades, a wide array of technological devices has been incorporated into the rehabilitation of neurological conditions to assist training. In this context, options such as virtual reality (VR) [11,12], robotics [13], smartphone apps [14], telerehabilitation options [15], and videogames and exergames [16] have been studied and shown positive effects on people with PD. This massive implementation of technological solutions often follows the implication that, in trying to affect and improve the users' experience and behavior, mechanics reminiscent of games are used. This concept of "the use of game design elements in non-game contexts" is widely known as gamification [17]. In the context of rehabilitation, the integration of gamification elements into technological devices is designed to create a more engaging and challenging experience for users. This approach is anticipated to enhance adherence and motivation among participants [18]. The combination of gamification and VR could even add benefits for older people in physical and cognitive domains while allowing complex training scenarios such as dual task training to be performed [19,20].

To effectively incorporate gamification elements across various intervention designs, it is essential to thoroughly understand gamification principles and their interaction with human motivation. Frameworks play a pivotal role in organizing available gamification elements into coherent categories, thereby guiding developers in selecting the most appropriate elements to achieve the desired user experience. Common frameworks include the Octalysis Framework by Kai [21], which identifies 8 core drives of human motivation, such as epic meaning and social influence; and the Marczewski 52 Gamification Mechanics and Elements [22], which classify elements based on the type of player they best cater to. However, the decision of which gamification elements to apply in interventions is often not thoroughly considered, in a context where treatment options should cater to the needs of end users [23]. Recent studies point to most VR rehabilitation options for PD being administered through commercial videogames, which are designed for entertainment in healthy individuals, far away from the context of rehabilitation for people with PD [24-26].

In PD, where apathy and depression are often present, long-term adherence to exercise and rehabilitation is concerning [27,28]. Previous research has shown that the implementation of VR does not warrant a higher adherence to treatment [11]. Strategies to promote motivation to stay involved in the rehabilitation process are key to long-term success in a population with a tendency to drop out [29]. Gamification, by integrating elements like feedback, rewards, and goal setting, may help sustain engagement and counteract the apathy and depression common in PD, thereby supporting long-term adherence to rehabilitation. However, current literature lacks information on the implementation of gamification in technological devices aimed at rehabilitating motor conditions in PD. Understanding which elements are currently used and the types of technological solutions they are integrated into is essential for developing further interventions that address the specific needs of individuals with PD. Therefore, the aim of this scoping review is to identify and classify the gamification elements and technologies used in motor rehabilitation in PD and to describe the justification behind the use of gamification and technology in this context. Specifically, this review focuses on the classification of gamification mechanics available, the types of technological devices used, and the motor symptoms these interventions target. Additionally, it explores the justifications used by each study for the inclusion of gamification elements and technological devices for motor rehabilitation in PD.

Methods

Registration and Framework

This Scoping Review is registered at the Open Science Framework database (OSF.IO/tx3d9). This review followed the Joanna Briggs Institute methodology for scoping reviews [30] and also adhered to the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) to ensure quality, transparency and reproducibility [31].

Objective

The objective of this study is to explore the use of technological devices integrating gamification in the rehabilitation of motor symptoms for individuals with PD. Specifically, it aims to identify the gamification elements and technological devices used in this context, classify the gamification elements based on the core motivational drives they target and the devices according to the motor symptoms they address, and describe the rationale behind their application in enhancing motor symptom rehabilitation for people with PD.

Eligibility Criteria

All peer-reviewed articles and peer-reviewed conference papers published in English or Spanish were included. Gray literature sources were excluded to maintain methodological consistency, ensuring all included studies underwent formal peer evaluation and were fully accessible. To guide the eligibility criteria, the

XSL•FO
RenderX

Population, Concept, and Context framework was used, in which population, concept, and context are defined to guide study selection [30]. Population refers to people with PD, with no restrictions applied regarding specific subpopulations, gender, disease stage, or geographical location. The Concept covered intervention studies that described motor rehabilitation through technological solutions incorporating clearly identifiable gamification elements. Studies that used commercial video games originally developed for entertainment were also included. While Deterding's [17] definition of gamification refers to the use of game elements in nongame contexts, implying the exclusion of full video games aimed at entertainment, the therapeutic application of these games in rehabilitation effectively shifts their context. As a result, built-in motivational features such as points, progression systems, and feedback mechanics can be considered relevant gamification elements when used for rehabilitation in PD. In situations where not enough details were given regarding the intervention to chart its gamification elements, the studies were excluded. No restrictions regarding Context or year of publication were established. Articles that met the selection criteria were incorporated without assessing their quality, as the objective of scoping reviews is to chart the pertinent works in a particular area and pinpoint areas lacking research [30].

Information Sources and Search Strategy

The search strategy was developed and conducted by a librarian on November 23, 2023, across 7 electronic databases: MEDLINE, EMBASE, Scopus, Cochrane, Web of Science, PsycINFO, and Epistemonikos. A combination of controlled vocabulary and free-text keywords was used to ensure retrieval of all potentially relevant studies. The search focused on terms related to PD, gamification, rehabilitation, motor symptoms, and various technological rehabilitation strategies, such as VR, videogames, and exergames. Boolean operators were applied to refine the search, and search strings were adapted to match the indexing structures of each database. The search strategy used for each database is detailed in Multimedia Appendix 1. No systematic or scoping reviews on the specified topic were identified in the Joanna Briggs Institute Systematic Review Register, PROSPERO, or biomedical and nursing databases.

Study Selection and Data Extraction

All articles extracted from databases were imported to Covidence [32] and checked for duplicates by using its duplicate detection system. Given the large volume of references obtained, the initial title and abstract screening were independently conducted by 3 reviewers (PBB, OMN, and MMA) rather than 2, as initially proposed in the registered protocol. Any disagreements during this screening stage were resolved through consensus discussions involving all 3 reviewers. In the next step, full-text screening was independently performed by the same 3 reviewers, and inconsistencies were again resolved by consensus. Data extraction of all included articles was conducted by PBB.

A total of 4451 articles were retrieved from all databases. After duplicates were removed, both by Covidence and manually, 48.48% (2158/4451) of all articles underwent title and abstract reviewing. After this first title and abstract screening, 87.62% (1891/2158) of articles were excluded. A total of 267 articles were left for full-text screening, 69.66% (186/267) of which were excluded.

Synthesis of Results

General study information such as location, sample size, age, disease duration, gender, year, and study design were collected. Specific data pertinent to the review's objectives was also collected.

In order to identify and name the gamification elements found in each study, the Marczewski 52 Gamification Mechanics [22] nomenclature was used due to its clear terminology and ease of use. All gamification elements identified were then classified using the Kai Octalysis Framework [21] due to its focus on motivation and end user interaction with the system. The Octalysis framework delineates human motivation into 8 different core drives: epic meaning and calling, development and accomplishment, empowerment of creativity and feedback, ownership and possession, social influence and relatedness, scarcity and impatience, unpredictability and curiosity, and loss and avoidance. All gamification elements identified using Marczewski's framework [22] were classified with the core drive they most effectively addressed.

Technological devices integrating gamification were classified based on the type of platform they were used with and also based on which symptoms they intended to address.

Finally, a qualitative content analysis using an inductive approach was conducted to identify the justifications used by each study for the inclusion of gamification elements and technological devices for motor rehabilitation in PD [33]. Data from all included studies were coded and organized into categories, which were then clustered into broader themes with the support of a hierarchical tree diagram. Definitions were developed for each code, category, and subcategory. Two reviewers (PBB and HFL) carried out the analysis independently using Excel, and then discussed how to group codes, align interpretations, and resolve discrepancies. Final themes were reviewed and refined through team-based triangulation to ensure consistency and credibility.

Results

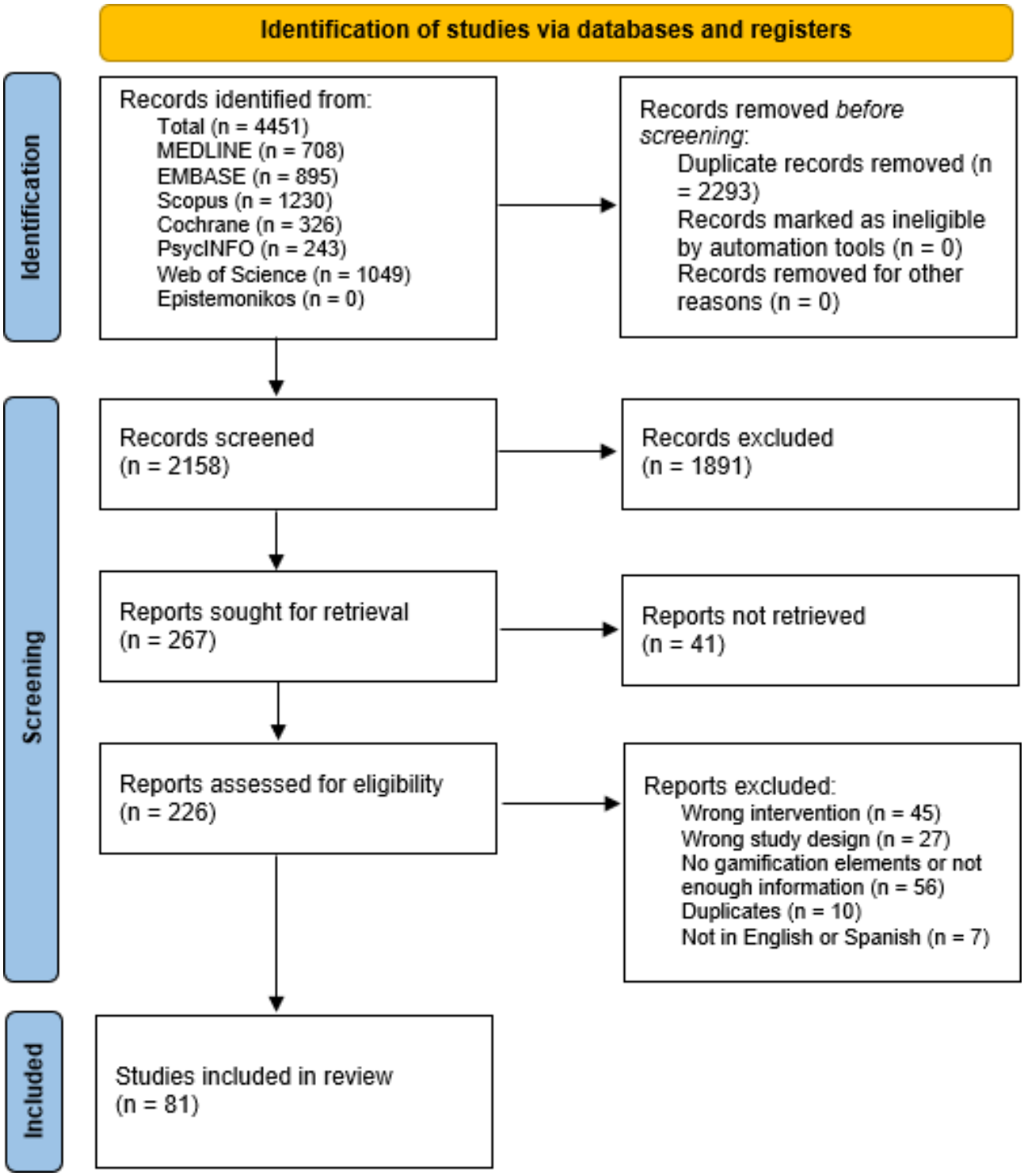
General Information

Altogether 81 studies were included in the final review (Multimedia Appendix 2) [34-114]. The PRISMA flow diagram (Figure 1) [115] details the steps followed during the screening process and data on number of excluded studies as well as reasons for exclusion in the full-text step. Of the 81 articles included, over 50% (41/81) were published in the year 2019 or later. Italy and Brazil were the 2 countries where most research was produced, with 18.5% (14/81) and 17.3% (13/81), respectively. The most frequent study design consisted of randomized controlled trials, which accounted for 49% (40/81) of all studies, followed by quasi-experimental studies, which accounted for 46% (37/81) of all studies. Sample sizes from the studies included ranged from 2 to 302 participants, with an average sample size of 32.2 participants. The mean age for participants was 66.64 (SD 5.05) years, with a mean of 2.29

XSL•FO
RenderX

(SD 0.57) on the Hoehn and Yahr [116] scale for disease severity and a mean disease duration of 7.35 (SD 2.70) years. A total of 39.39% of all participants were female. Most studies (62/81, 77%) implemented their interventions in an in-person format, while 23% (19/81) used telerehabilitation. Additionally, 1% (1/81) of studies used a hybrid approach, combining both in-person and remote intervention formats.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram.



Identification and Classification of Gamification Elements and Technological Devices

Gamification Elements

Detailed data on the frequency and distribution of the identified gamification elements and their association with the Octalysis Framework's 8 Core Drives of motivation can be found in Table 1.

XSL•FO
RenderX

Table . Distribution of use of gamification elements and classification according to the Octalysis Framework.

Octalysis' core drive and gamification elements	Times used (out of 81 studies)	Percentage of total studies where used	Percentage of total used elements
Epic meaning and calling			
Quests [59,60]	2	2	<1
Development and accomplishment			
Progress/Feedback [34-86,88-110,112-114]	79	98	17
Points [34-40,42-44,47,50-52,54-58,62-71,73-107,109-114]	70	86	15
Levels/Progression [34-39,41-47,50-52,54-58,62-68,70-89,91-97,99-103,106-114]	66	81	15
Badges [34-52,54-56,58,62-66,68-71,73-77,79-86,89-92,96-97,99-101,114]	30	37	7
Leaderboards [34-52,54-56,58,62-66,68-71,73-77,79-89,91-97,99-101,114]	29	36	6
Challenges [34-52,54-56,58,62-66,68-71,73-77,79-82,84-97,99-101,114]	26	32	6
Empowerment of creativity and feedback			
Customization [34-52,54-56,58,62-66,68-71,73-77,79-86,89-92,96-97,99-101,114]	32	40	7
Ownership and possession			
—[a]	—	—	—
Social influence and relatedness			
Competition [34-52,54-56,58,62-66,68-71,73-77,79-86,89-92,96-97,99-101,114]	29	36	6
Knowledge share [59,60]	2	2	<1
Guilds/Teams [59,60]	2	2	<1
Scarcity and impatience			
Unlockables [50]	1	1	<1
Unpredictability and curiosity			
—	—	—	—
Loss and avoidance			
Consequences [34-52,54-56,58,62-66,68-71,73-77,79-86,89-92,96-97,99-101,114]	30	37	7
Time Pressure [34-52,54-56,58,62-66,68-71,73-77,79-86,89-92,96-97,99-101,114]	37	46	8

[a]Not available.

A total of 453 gamification elements were identified across all included studies, categorized into 14 distinct types based on the Marczewski classification of gamification mechanics [22]. For the epic meaning and calling core drive, only quests were identified as a relevant gamification element, appearing 2 times across studies. This element was used in 2.47% (2/81) of the studies, accounting for 0.44% (2/4453) of the total gamification elements used. The development and accomplishment core drive was the most prominent category, containing a variety of gamification elements and showing extensive use across the studies. Progress/feedback mechanics were the most frequently used element, appearing in 79 studies (97.53% of studies) and representing 17.44% (80/453) of the total gamification elements used. Other commonly used elements included points, which appeared in 86.42% (70/81) of studies and accounted for 15.45% (70/453) of total elements; levels/progression, which appeared in 81.48% (66/81) of studies and represented 14.57% (66/453) of total elements; badges, used in 37.04% (30/81) of studies, accounting for 6.62% (30/453) of total elements; leaderboards, used in 35.80% (29/81) of studies, making up 6.40% (29/453) of total elements; and challenges, which appeared in 32.09% (26/81) of studies and represented 5.72% (26/453) of total elements. The empowerment of creativity and feedback core drive included only one gamification element, customization,

XSL•FO
RenderX

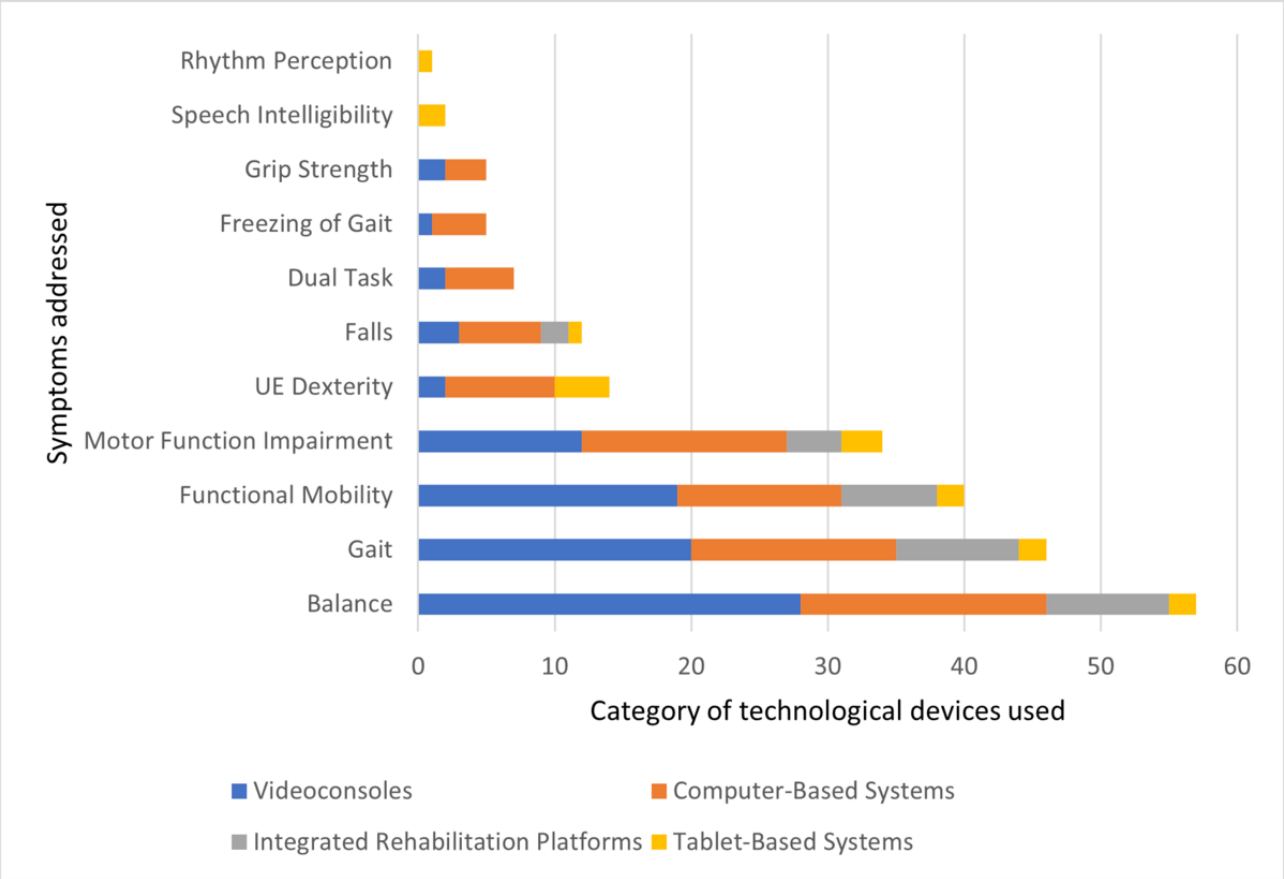
which was used in 32 studies (39.51% of studies) and accounted for 7.06% (32/453) of total elements. Neither the ownership and possession core drive nor the unpredictability and curiosity drive showed any specific gamification elements reported across the studies. Under the social influence and relatedness core drive, elements like competition, knowledge share, and guilds/teams were identified. Competition was used in 35.80% (29/81) of studies, contributing 6.40% (29/81) to the total elements, while both knowledge share and guilds/teams appeared only in 2.47% (2/81) of studies, accounting for a mere 0.44% (2/453) of total elements each. The scarcity and impatience drive included only one element, unlockables, which was used in just one study, appearing in 1.23% (1/81) of studies and making up only 0.22% (1/453) of the total gamification elements. Last, the loss and avoidance core drive included the elements "consequences" and "time pressure". Consequences were used in 37.04% (30/81) of studies, contributing 6.62% (30/453) to the total elements, while time pressure appeared in 45.68% (37/81) of studies, making up 8.14% (37/453) of the total elements used.

Technological Devices Integrating Gamification

Overview

The technological devices integrating gamification identified across 81 studies were categorized into 4 groups based on the platforms they were used with: video game consoles, computer-based systems, tablet-based systems, and integrated rehabilitation platforms. Detailed information on each category and the specific motor symptoms in PD they address can be found in Figure 2.

Figure 2. Symptoms addressed by each type of technological device used in rehabilitation. UE: upper extremity.



Video Game Consoles

Video game consoles like the Nintendo Wii and Microsoft Xbox, originally designed for entertainment, were frequently repurposed for rehabilitation. The Nintendo Wii was the most used device, appearing in 27% (22/81) of all studies, almost exclusively in combination with the Wii Balance Board (20/22, 91% of all Nintendo Wii studies). Most studies with this video game console addressed balance (18/22, 82% of studies), gait (12/22, 55%), functional mobility (11/22, 50%), and motor function impairment (9/22, 41%). The Microsoft Xbox, used with the Kinect motion sensor in 14% (11/81) of studies, was similarly focused on balance (10/11, 91% of studies), gait (8/11, 73%), functional mobility (7/11, 73%), and motor function impairment (3/11, 28%). None of the video game consoles were created with rehabilitation as their primary goal. A total of 85% (28/33) of studies using video game consoles delivered their interventions in an in-person format, while 13% (4/33) used telerehabilitation and 3% (1/33) combined both formats.

Computer-Based Systems

Computer-based systems were used in 40% (32/81) of studies, often in combination with motion-sensing devices like Kinect (12/32, 37.5% of computer-based systems), Leap Motion Controller (7/32, 22%), and balance boards or force plates (7/32, 22%). Kinect was applied mainly to improve balance (8/12, 67%), gait (7/12, 58%), and motor function impairment (5/12, 42%), while balance boards and force plates focused on balance

XSL•FO
RenderX

(7/7, 100%), gait (4/7, 57%), and functional mobility (4/7, 57%), and the Leap Motion Controller targeted upper extremity dexterity (7/7, 100%), motor function impairment (4/7, 57%), and grip strength (3/7, 43%). The HTC Vive Pro VR device was used (2/32, 6%), enhancing immersion for balance (2/2, 100%), gait (2/2, 100%), and functional mobility (2/2, 100%) training. Additionally, some studies combined treadmills with Kinect (3/32, 9%), mainly to focus on falls (3/3, 100%). A total of 91% (29/32) of all computer-based systems were created with rehabilitation as their primary goal, and 78% (25/32) of studies using computer-based systems delivered their interventions in an in-person format, while 22% (7/32) used telerehabilitation.

Tablet-Based Systems

Tablet-based systems appeared in 11% (9/81) of all studies, either as standalone devices (6/9, 67% of all tablet-based systems) or combined with motion-sensing devices such as the Leap Motion Controller (1/9, 11%) or handheld tools like GripAble (1/9, 11%). These tablet-based systems mainly targeted upper extremity dexterity (4/9, 44%) and motor function impairment (3/9, 33%). A total of 100% (9/9) of tablet-based systems were created for rehabilitation as their primary goal and 22% (2/2) of tablet-based systems delivered their interventions in an in-person format, while 78% (7/9) used telerehabilitation.

Integrated Rehabilitation Platforms

Integrated rehabilitation platforms appeared in 10% (8/81) of all studies, testing systems explicitly designed for comprehensive rehabilitation. These platforms were used to rehabilitate balance (8/8, 100%), gait (8/8, 100%), functional mobility (5/8, 63%), and motor function impairment (2/8, 25%). All integrated rehabilitation platforms were created for rehabilitation as their primary goal. A total of 100% of interventions using integrated rehabilitation platforms were delivered in an in-person format.

Justification Behind the Use of Technological Devices Integrating Gamification

Overview

In order to determine the justifications used by each study for the inclusion of gamification elements and technological devices for motor rehabilitation in PD, four key categories were identified: (1) gamification as a benefit for therapeutic interventions, (2) evidence of effectiveness of technology in rehabilitation for PD, (3) evidence of effectiveness of technology in rehabilitation across populations, and (4) specific properties of technological solutions in rehabilitation.

Gamification as a Benefit for Therapeutic Interventions

Gamification was described as a means to support adaptive and individualized rehabilitation [45,47]. Systems were used to monitor user performance, adjust task difficulty based on scoring, and provide real-time feedback throughout the intervention. Some interventions also allowed for the customization of the experience to better match user needs and progression.

Evidence of Effectiveness of Technology in Rehabilitation for PD

Improvement of Motor Symptoms in PD

Technological devices were frequently described as effective [34,44,47,51,57,61,62,69-71,74-76,79,83,85-87,91,93,96,102,104-106,112-114], promising [38, 48, 52, 64, 67, 68, 79, 81, 88, 102, 111], or feasible [61,72,88,114] tools for addressing motor symptoms in individuals with PD. Studies emphasized their role in improving key functions such as balance, gait, and postural control, while others framed them as having strong potential for rehabilitation, even when not yet fully established through outcomes. Articles also pointed to their feasibility as motor rehabilitation strategies, highlighting aspects like affordability and accessibility to support their use in both clinical and home-based settings.

Improvement of Nonmotor Symptoms in PD

Technological devices were described as having a supportive role in addressing cognitive function and emotional well-being in people with PD. Studies reported improvements in areas such as cognition, executive functioning, self-esteem, pain, and mental health [43,48,61,69,71,75,81,87,92,96,97,105,106,109]. Other aspects such as the potential of these technologies to create controlled, structured environments that can enhance the delivery of cognitive therapies and provide emotional support were also highlighted [64,81,100,105].

Positive Influence on Brain Activity Networks in PD

Studies discussed the role of technology in supporting underlying neurological processes relevant to rehabilitation [37,40,43,55,58,64,65,69-71,83,94,98,102]. VR and exergames were noted for their potential to promote neuroplasticity, motor learning, and motor-cognitive integration. These interventions were often associated with repetitive, goal-directed tasks designed to engage and optimize brain networks involved in movement and coordination.

Evidence of Effectiveness of Technology in Rehabilitation Across Populations

General Evidence for Physiotherapy

Studies highlighted that technology could offer rehabilitation outcomes comparable to, and in some cases exceeding, those of traditional physiotherapy [39,47,71,77,84,101]. Articles specifically noted the effectiveness of VR and exergames in enhancing therapy for both older adults and individuals with PD, supporting their use as a valid approach across different populations.

Evidence for Non-PD Populations

Technology was also supported through evidence drawn from broader populations, including stroke patients and older adults [53,54,74,75,77,91,102,112,113]. In these contexts, VR and video game–based interventions were associated with improvements in key rehabilitation outcomes such as balance, mobility, and fall reduction. These findings were used to justify the applicability of similar technological approaches in the context of PD rehabilitation.

XSL•FO
RenderX

Improvement of Clinical Parameters

Some studies pointed to the broader clinical value of technology in rehabilitation, highlighting its role in improving quality of life, physical function, daily activity performance, and pain management [48,51,52,74,86,97]. These outcomes were observed across a range of populations, including children, adults, and individuals with PD.

Specific Properties of Technological Solutions in Rehabilitation

Enhances Motivation Through an Engaging Experience

Technology was described as a tool to make rehabilitation more engaging and enjoyable [36,39,42,43,45,48,50,53,58,61,64,66,67,69-71,83,85,87-89,91,94,96-100,102,103,105,107-109,111,113,114]. By offering interactive and rewarding experiences, these interventions were seen as helpful in encouraging active participation and supporting long-term adherence to therapy programs.

Provides Quality Biofeedback

Real-time feedback was noted as a valuable feature in many interventions [37,38,41-45,51,54,55,58,63,65,66,73-75,77,80,81,83,84,91,92,94,98,100,102,103,106,107,109-112,114]. These feedback mechanisms were associated with enhanced motor learning, improved performance, and greater awareness during rehabilitation tasks.

Customization of Training Parameters

The ability to tailor exercises to individual needs emerged as a key strength of technological solutions [39-41,45,48,51,53,57,63,65,70,72,73,76,84,91,94,97,100,101,105,107,111]. Customizable settings enabled repeated practice in varied contexts, supporting motor skill development and encouraging problem-solving during rehabilitation.

Enriched or Ecological Environment

Some technologies were highlighted for their capacity to simulate real-world environments, helping users practice tasks in settings that reflect daily life [48,54,55,60,63,72,74,76,78,81,84,92,97,98,101,109,112,113]. These ecologically valid scenarios were viewed as beneficial for promoting skill transfer beyond the therapeutic setting.

Supports Dual-Task Integration

Studies mentioned that technological interventions incorporated features that combined physical and cognitive tasks, making it possible to train complex dual-task scenarios [62,66,76,81,83,92,93,95,97,100,106,111]. This was particularly relevant for engaging users both mentally and physically within a single rehabilitation activity.

Enables Telerehabilitation

Technology was presented as a practical solution for home-based rehabilitation [36,58,61,83,97,100,103,106]. Devices described as user-friendly and affordable were positioned as a means of extending therapy beyond clinical settings while maintaining effectiveness.

Accessible and Cost-Effective

Portability and low cost were mentioned as noted advantages, especially in tools designed for home use [36,61,74,79,82,87,97,104]. These features lead to an easier integration into daily routines, which may contribute to better consistency and accessibility in rehabilitation.

Safe Intervention

Studies emphasized the safety of technological interventions, particularly in controlled environments [50,89,98,112,113]. This was considered relevant for tasks that carry a higher risk of falls or injury, offering a secure space for functional improvement.

Discussion

Principal Findings

This scoping review aimed to identify and classify the technological devices integrating gamification used in motor rehabilitation for PD, as well as to describe the rationale behind their application. This review is particularly relevant as it addresses a growing interest in leveraging innovative, technology-driven approaches to manage PD's motor symptoms, a field where gamification has been suggested to enhance patient engagement and therapeutic outcomes [20]. Despite this potential, the integration of gamification into PD rehabilitation interventions remains underexplored, with a limited understanding of how these elements are being applied in practice or their theoretical justifications.

Findings reveal that while technological devices integrating rehabilitation are widely used, the range of gamification mechanics is narrow, with a strong focus on the development and accomplishment core drive. This result aligns with statements by Kai [21], according to which the core drive of development and accomplishment is the easiest to design and the one drawing the most focus from developers. Commonly used elements like feedback mechanisms, points, and progression systems dominate interventions, reflecting an emphasis on monitoring and improving performance. This aligns closely with the most targeted motor symptoms—balance, gait, and functional mobility—which significantly benefit from feedback-driven adjustments offered by motion-sensing technologies, aiding postural control in individuals with PD [117]. Gamification mechanics such as time pressure and consequences from the loss and avoidance core drive are used less frequently with the goal of instilling urgency and accountability, improving speed and efficiency. Elements from this core drive should be carefully considered in designing future interventions, since different motor phenotypes of PD appear to respond differently to reward or punishment-based systems [118]. Similarly, competition mechanics foster motivation and social interaction, aiding long-term engagement. However, more intrinsic motivational pathways, such as those tied to unpredictability and curiosity or ownership and possession, remain underused. Notably absent are complex gamification features like narratives or exploratory elements, highlighting a gap in current approaches.

XSL•FO
RenderX

Out of all the studies included in this review, only 2 [45,47] explicitly refer to the concept or use of gamification when describing their interventions, and none used a theoretical framework to justify their use. Additionally, 37 out of the 81 included studies used technological devices originally designed for entertainment in healthy populations rather than as rehabilitation tools tailored to individuals with health conditions. These findings underline clear research gaps in applying gamification to its full potential in the rehabilitation of people with PD, as well as highlighting an overreliance on video game consoles and other devices designed for entertainment purposes (ie, Nintendo Wii and Microsoft Kinect) rather than rehabilitation-specific technologies. The high presence of interventions based on commercial video consoles and video games for rehabilitation presents an intriguing landscape within this context. On the one hand, leveraging refined, off-the-shelf commercial software offers distinct advantages, eliminating the need for costly development efforts and providing products inherently designed for entertainment, which may enhance participant engagement. However, it is essential to recognize that these interventions lack tailored considerations for populations affected by neurodegenerative conditions, such as PD. Furthermore, the feedback mechanisms embedded in these software options are geared toward game performance rather than exercise and motor learning, potentially rendering them less relevant for the rehabilitation of individuals with PD. As a result, gamification elements, digital environments, and gameplay patterns may not align optimally with the requirements and objectives of this specific population [119]. User-centered design would be an essential part of development in this context, incorporating the needs and preferences of individuals with PD into the interventions [120]. Notably, people with PD exhibit distinct differences from the younger populations typically targeted by commercial video games, which could lead to unintended outcomes when using them to apply rehabilitation interventions [121].

The delivery format of interventions varied by technology type. Video game consoles and computer-based systems were primarily used in in-person settings (85% [28/33] and 78% [25/32], respectively), while tablet-based systems were predominantly used in telerehabilitation (7/9, 78%). Despite most devices being designed for at-home use, researchers tended to implement interventions in supervised settings, likely to ensure safety, adherence, and standardized assessments. This preference aligns with findings suggesting that motion-sensing and immersive rehabilitation technologies often require clinical oversight for effective implementation, particularly in neurodegenerative conditions such as PD [122].

This review identified 4 distinct categories through content analysis of the justifications provided by studies for using technological devices integrating gamification for rehabilitation. Collectively, these categories underscore expectations of such solutions as being safe, motivational, and engaging tools for addressing diverse conditions. They are bolstered by extensive evidence of potential and proven effectiveness and offer innovative training opportunities, including telerehabilitation, dual-task integration, and advanced biofeedback. Despite these well-defined expectations, articles give minimal or no attention to the gamification elements that significantly contribute to their interventions. Current evidence of gaming interventions is unclear, with no studies isolating the effects of gamification used in interventions on people with PD [123]. These findings suggest that while technological solutions show great promise for advancing rehabilitation for individuals with PD, greater emphasis on thoughtfully integrating gamification elements could enhance outcomes by providing more tailored and effective interventions for this population [24].

Implications for Future Research and Clinical Practice

Many of the interventions in this review relied on commercial entertainment video games, which are not inherently designed for rehabilitation. Their built-in feedback systems and progression mechanics are primarily meant to enhance gameplay rather than support motor learning or therapeutic outcomes for individuals with PD. However, these games remain widely used because they are engaging, cost-effective, and easily accessible. Rather than focusing solely on developing new rehabilitation-specific games, a practical approach could involve evaluating existing commercial games to determine their suitability for therapy. Establishing clear assessment criteria, such as a standardized checklist, could help clinicians identify which games align best with rehabilitation goals and patient needs.

Future research should also explore how well these technologies adapt to the specific challenges faced by individuals with PD, particularly in terms of usability and accessibility. Additionally, a deeper understanding of how different gamification elements influence engagement and adherence, especially when grounded in motivational theories, could help refine rehabilitation strategies. Identifying the most effective elements for sustaining participation and improving outcomes would allow clinicians to make better-informed choices and develop interventions that truly meet the needs of people with PD. To advance this field, research should also prioritize a structured, evidence-based approach to gamification, ensuring that technological solutions are designed with both therapeutic effectiveness and user experience in mind. Integrating User-Centered Design principles with established motivational frameworks could help create interventions that are both engaging and clinically relevant, thus optimizing their impact in neurorehabilitation settings.

Limitations

This scoping review adhered to the PRISMA-ScR [31] recommendations to ensure methodological rigor. However, several limitations warrant acknowledgment:

Due to the substantial volume of retrieved studies, an initial screening was conducted using only titles and abstracts. This step involved 3 different reviewers who reached a consensus in cases of disagreement. Despite this method, there remains a risk of inadvertently excluding relevant papers.

Another limitation consists of the review's scope being bound by the search date at which it was performed. Consequently, any studies published after this date were not considered. There is also a limitation in that gray literature, which may contain valuable insights, was not included in the search. As a result, some relevant research might have been overlooked. Articles

XSL•FO
RenderX

published in languages other than English and Spanish were also not included, which may also omit relevant information from this review.

This review did not systematically analyze the usability, accessibility, or adaptability of the interventions, as these factors were not part of the predefined data extraction criteria. While these aspects are important for the success of rehabilitation technologies, more research is needed to understand how well gamified interventions address the unique motor and cognitive challenges of individuals with PD.

Finally, there is also a limitation in how the selection of gamification elements for this study was performed. While efforts were made to identify gamification elements in all included studies based on manuscript information and web-based sources, certain custom-made software options and commercial games were inaccessible or had limited available information. Consequently, some gamification elements may not have been fully accounted for.

Conclusions

This scoping review sheds light on the widespread adoption of technologies integrating gamification elements for motor symptom rehabilitation in individuals with PD. However, it also reveals a notable gap in understanding and acknowledging gamification mechanics among researchers and developers. The current landscape of gamification elements within the reviewed interventions predominantly emphasizes performance-based experiences, focusing on straightforward elements like scoring and feedback. Expectations of technological devices integrating gamification applied in rehabilitation contexts are clear, but there is a notable absence of justification for the specific inclusion of gamification through the application of relevant theories or frameworks, indicating a lack of standardization or rationale for the design of these interventions. Further studies in this field should focus on the importance of user-centered design for the inclusion of gamification mechanics that could modify the experience of users, so as to optimize the development efforts to optimally reach the needs of people with PD. There is also a clear need for the development of studies that report on the effects of gamification itself, which would allow developers to precisely deploy elements according to the expected role of these mechanics within their interventions.

Acknowledgments

This study has been funded by Instituto de Salud Carlos III through the project “PI20/00403” (cofunded by European Regional Development Fund/European Social Fund “A way to make Europe”/“Investing in your future”).

Data Availability

All data generated and analyzed during this study will be made available upon reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy for each database.
[DOCX File, 25 KB - games_v13i1e69433_app1.docx]

Multimedia Appendix 2
Details of all included studies.
[XLSX File, 37 KB - games_v13i1e69433_app2.xlsx]

Checklist 1
PRISMA-ScR checklist.
[PDF File, 112 KB - games_v13i1e69433_app3.pdf]

References

1. Poewe W, Seppi K, Tanner CM, et al. Parkinson disease. Nat Rev Dis Primers 2017 Mar 23;3(1–21):17013. [doi: 10.1038/nrdp.2017.13] [Medline: 28332488]
2. Feigin VL, Abajobir AA, Abate KH, et al. Global, regional, and national burden of neurological disorders during 1990–2015: a systematic analysis for the global burden of disease study 2015. Lancet Neurol 2017 Nov;16(11):877-897. [doi: 10.1016/S1474-4422(17)30299-5]
3. Deuschl G, Beghi E, Fazekas F, et al. The burden of neurological diseases in Europe: an analysis for the global burden of disease study 2017. Lancet Public Health 2020 Oct;5(10):e551-e567. [doi: 10.1016/S2468-2667(20)30190-0] [Medline: 33007212]

XSL•FO
RenderX

4. Luo Y, Qiao L, Li M, Wen X, Zhang W, Li X. Global, regional, national epidemiology and trends of Parkinson's disease from 1990 to 2021: findings from the global burden of disease study 2021. Front Aging Neurosci 2024;16:1498756. [doi: 10.3389/fnagi.2024.1498756] [Medline: 39868382]
5. Leite Silva ABR, Gonçalves de Oliveira RW, Diógenes GP, et al. Premotor, nonmotor and motor symptoms of Parkinson's disease: a new clinical state of the art. Ageing Res Rev 2023 Feb;84:101834. [doi: 10.1016/j.arr.2022.101834] [Medline: 36581178]
6. Shulman LM, Gruber-Baldini AL, Anderson KE, et al. The evolution of disability in Parkinson disease. Mov Disord 2008 Apr 30;23(6):790-796. [doi: 10.1002/mds.21879] [Medline: 18361474]
7. Armstrong MJ, Okun MS. Diagnosis and treatment of Parkinson disease. JAMA 2020 Feb 11;323(6):548. [doi: 10.1001/jama.2019.22360]
8. Manson A, Stirpe P, Schrag A. Levodopa-induced-dyskinesias clinical features, incidence, risk factors, management and impact on quality of life. J Parkinsons Dis 2012;2(3):189-198. [doi: 10.3233/JPD-2012-120103] [Medline: 23938226]
9. Tomlinson CL, Patel S, Meek C, et al. Physiotherapy versus placebo or no intervention in Parkinson's disease. Cochrane Database Syst Rev 2013 Sep 10;2013(9):CD002817. [doi: 10.1002/14651858.CD002817.pub4] [Medline: 24018704]
10. Mak MK, Wong-Yu IS, Shen X, Chung CL. Long-term effects of exercise and physical therapy in people with Parkinson disease. Nat Rev Neurol 2017 Nov;13(11):689-703. [doi: 10.1038/nrneurol.2017.128] [Medline: 29027544]
11. Dockx K, Bekkers EM, Van den Bergh V, et al. Virtual reality for rehabilitation in Parkinson's disease. Cochrane Database Syst Rev 2016 Dec 21;12(12):CD010760. [doi: 10.1002/14651858.CD010760.pub2] [Medline: 28000926]
12. Truijen S, Abdullahi A, Bijsterbosch D, et al. Effect of home-based virtual reality training and telerehabilitation on balance in individuals with Parkinson disease, multiple sclerosis, and stroke: a systematic review and meta-analysis. Neurol Sci 2022 May;43(5):2995-3006. [doi: 10.1007/s10072-021-05855-2] [Medline: 35175439]
13. Picelli A, Capecci M, Filippetti M, et al. Effects of robot-assisted gait training on postural instability in Parkinson's disease: a systematic review. Eur J Phys Rehabil Med 2021 Jun;57(3):472-477. [doi: 10.23736/S1973-9087.21.06939-2] [Medline: 33826278]
14. Putzolu M, Manzini V, Gambaro M, et al. Home-based exercise training by using a smartphone app in patients with Parkinson's disease: a feasibility study. Front Neurol 2023;14:1205386. [doi: 10.3389/fneur.2023.1205386] [Medline: 37448748]
15. Vellata C, Belli S, Balsamo F, Giordano A, Colombo R, Maggioni G. Effectiveness of telerehabilitation on motor impairments, non-motor symptoms and compliance in patients with Parkinson's disease: a systematic review. Front Neurol 2021;12:627999. [doi: 10.3389/fneur.2021.627999] [Medline: 34512495]
16. Chuang CS, Chen YW, Zeng BY, et al. Effects of modern technology (exergame and virtual reality) assisted rehabilitation vs conventional rehabilitation in patients with Parkinson's disease: a network meta-analysis of randomised controlled trials. Physiotherapy 2022 Dec;117:35-42. [doi: 10.1016/j.physio.2022.07.001] [Medline: 36242929]
17. Deterding S, Khaled R, Nacke LE, Dixon D. Gamification: toward a definition. Presented at: CHI 2011 Gamification Workshop Proceedings; May 7-12, 2011; Vancouver, BC, Canada URL: http://gamification-research.org/wp-content/uploads/2011/04/02-Deterding-Khaled-Nacke-Dixon.pdf [accessed 2025-06-30]
18. Alfieri FM, da Silva Dias C, de Oliveira NC, Battistella LR. Gamification in musculoskeletal rehabilitation. Curr Rev Musculoskelet Med 2022 Dec;15(6):629-636. [doi: 10.1007/s12178-022-09797-w] [Medline: 36301514]
19. Adlakha S, Chhabra D, Shukla P. Effectiveness of gamification for the rehabilitation of neurodegenerative disorders. Chaos Solit Fractals 2020 Nov;140:110192. [doi: 10.1016/j.chaos.2020.110192]
20. De Freitas TB, Leite PHW, Doná F, Pompeu JE, Swarowsky A, Torriani-Pasin C. The effects of dual task gait and balance training in Parkinson's disease: a systematic review. Physiother Theory Pract 2020 Oct;36(10):1088-1096. [doi: 10.1080/09593985.2018.1551455] [Medline: 30501424]
21. Kai CY. Actionable Gamification: Beyond Points, Badges, and Leaderboards: Packt Publishing Ltd; 2019. URL: https://yukaichou.com/gamification-book/ [accessed 2025-06-30]
22. Marczewski A. The periodic table of gamification elements - gamified UK - gamification expert. : gamified.uk; 2017 URL: https://www.gamified.uk/2017/04/03/periodic-table-gamification-elements/ [accessed 2025-06-26]
23. Sevcenko K, Lindgren I. The effects of virtual reality training in stroke and Parkinson's disease rehabilitation: a systematic review and a perspective on usability. Eur Rev Aging Phys Act 2022 Jan 25;19(1):4. [doi: 10.1186/s11556-022-00283-3] [Medline: 35078401]
24. Rodríguez-Mansilla J, Bedmar-Vargas C, Garrido-Ardila EM, et al. Effects of virtual reality in the rehabilitation of Parkinson's disease: a systematic review. J Clin Med 2023 Jul 26;12(15):10419374. [doi: 10.3390/jcm12154896] [Medline: 37568298]
25. Zhang J, Luximon Y, Pang MYC, Wang H. Effectiveness of exergaming-based interventions for mobility and balance performance in older adults with Parkinson's disease: systematic review and meta-analysis of randomised controlled trials. Age Ageing 2022 Aug 2;51(8):afac175. [doi: 10.1093/ageing/afac175] [Medline: 35930726]
26. Cano Porras D, Siemonsma P, Inzelberg R, Zeilig G, Plotnik M. Advantages of virtual reality in the rehabilitation of balance and gait: systematic review. Neurology (ECronicon) 2018 May 29;90(22):1017-1025. [doi: 10.1212/WNL.0000000000005603] [Medline: 29720544]

XSL•FO
RenderX

27. den Brok M, van Dalen JW, van Gool WA, Moll van Charante EP, de Bie RMA, Richard E. Apathy in Parkinson's disease: a systematic review and meta-analysis. Mov Disord 2015 May;30(6):759-769. [doi: 10.1002/mds.26208] [Medline: 25787145]
28. Aarsland D, Larsen JP, Lim NG, et al. Range of neuropsychiatric disturbances in patients with Parkinson's disease. J Neurol Neurosurg Psychiatry 1999 Oct;67(4):492-496. [doi: 10.1136/jnnp.67.4.492] [Medline: 10486397]
29. van Nimwegen M, Speelman AD, Overeem S, et al. Promotion of physical activity and fitness in sedentary patients with Parkinson's disease: randomised controlled trial. BMJ 2013 Mar 1;346(7898):f576. [doi: 10.1136/bmj.f576] [Medline: 23457213]
30. Edoardo A, Zachary M, Joanna Briggs Institute. JBI manual for evidence synthesis. : Joanna Briggs Institute; 2020 URL: https://tinyurl.com/47tcykxd [accessed 2025-06-26]
31. Tricco AC, Lillie E, Zarin W, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 2;169(7):467-473. [doi: 10.7326/M18-0850] [Medline: 30178033]
32. Covidence systematic review software. 2023. URL: www.covidence.org [accessed 2025-06-26]
33. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
34. Albiol-Pérez S, Gil-Gómez JA, Muñoz-Tomás MT, Gil-Gómez H, Vial-Escolano R, Lozano-Quilis JA. The effect of balance training on postural control in patients with Parkinson's disease using a virtual rehabilitation system. Methods Inf Med 2017 Mar 23;56(2):138-144. [doi: 10.3414/ME16-02-0004] [Medline: 28244545]
35. Allen NE, Song J, Paul SS, et al. An interactive videogame for arm and hand exercise in people with Parkinson's disease: a randomized controlled trial. Parkinsonism Relat Disord 2017 Aug;41:66-72. [doi: 10.1016/j.parkreldis.2017.05.011] [Medline: 28528804]
36. Alves MLM, Mesquita BS, Morais WS, Leal JC, Satler CE, Dos Santos Mendes FA. Nintendo WiiTM versus Xbox KinectTM for assisting people with Parkinson's disease. Percept Mot Skills 2018 Jun;125(3):546-565. [doi: 10.1177/0031512518769204] [Medline: 29665760]
37. Anwar N, Akram S, Ilyas A, Khalid K, Munir M, Khizer Hayar M. Comparison of virtual reality and conventional balance training to improve balance and walking in Parkinson's disease patients. Med Forum Mon 2021;32(7):129-133.
38. Bacha JMR, Cunha MCC, Freitas TB, et al. Effects of virtual rehabilitation on postural control of individuals with Parkinson disease. Motricidade 2021;17(3):220-227.
39. Barth M, Möbius R, Themann P, et al. Functional improvement of patients with Parkinson syndromes using a rehabilitation training software. Front Neurol 2023;14:1210926. [doi: 10.3389/fneur.2023.1210926] [Medline: 37645604]
40. Bekkers EMJ, Mirelman A, Alcock L, et al. Do patients with Parkinson's disease with freezing of gait respond differently than those without to treadmill training augmented by virtual reality? Neurorehabil Neural Repair 2020 May;34(5):440-449. [doi: 10.1177/1545968320912756] [Medline: 32202203]
41. Bevilacqua R, Benadduci M, Renato Riccardi G, et al. SI-ROBOTICS system: a preliminary study on usability of a rehabilitation program in patients with parkinson's disease. Presented at: CEUR Workshop Proceedings; Dec 16, 2022; Florence, Italy URL: https://ceur-ws.org/Vol-3323/paper3.pdf [accessed 2025-06-30]
42. Brachman A, Marszałek W, Kamieniarz A, Michalska J, Pawłowski M, Juras G. Biomechanical measures of balance after balance-based exergaming training dedicated for patients with Parkinson's disease. Gait Posture 2021 Jun;87:170-176. [doi: 10.1016/j.gaitpost.2021.04.036] [Medline: 33940308]
43. Calabrò RS, Naro A, Cimino V, et al. Improving motor performance in Parkinson's disease: a preliminary study on the promising use of the computer assisted virtual reality environment (CAREN). Neurol Sci 2020 Apr;41(4):933-941. [doi: 10.1007/s10072-019-04194-7] [Medline: 31858331]
44. Carpinella I, Cattaneo D, Bonora G, et al. Wearable sensor-based biofeedback training for balance and gait in Parkinson disease: a pilot randomized controlled trial. Arch Phys Med Rehabil 2017 Apr;98(4):622-630. [doi: 10.1016/j.apmr.2016.11.003] [Medline: 27965005]
45. Chua LK, Chung YC, Bellard D, et al. Gamified dual-task training for individuals with Parkinson disease: an exploratory study on feasibility, safety, and efficacy. Int J Environ Res Public Health 2021 Nov 25;18(23):12384. [doi: 10.3390/ijerph182312384] [Medline: 34886114]
46. Cikajlo I, Zajc D, Dolinšek I. Precise hand movement telerehabilitation with virtual cubes for patients with parkinson's disease. Presented at: ACM International Conference Proceeding Series; Oct 13-14, 2016; Lisbon, Portugal p. 17-20. [doi: 10.1145/3051488.305149]
47. Cikajlo I, Hukić A, Dolinšek I, et al. Can telerehabilitation games lead to functional improvement of upper extremities in individuals with Parkinson's disease? Int J Rehabil Res 2018 Sep;41(3):230-238. [doi: 10.1097/MRR.0000000000000291] [Medline: 29757774]
48. Cikajlo I, Peterlin Potisk K. Advantages of using 3D virtual reality based training in persons with Parkinson's disease: a parallel study. J Neuroeng Rehabil 2019 Oct 17;16(1):119. [doi: 10.1186/s12984-019-0601-1] [Medline: 31623622]
49. Cikajlo I, Hukić A, Zajc D. Exergaming as part of the telerehabilitation can be adequate to the outpatient training: preliminary findings of a non-randomized pilot study in Parkinson's disease. Front Neurol 2021;12:625225. [doi: 10.3389/fneur.2021.625225] [Medline: 33815252]

XSL•FO
RenderX

50. Dauvergne C, Bégel V, Gény C, Puyjarinet F, Laffont I, Dalla Bella S. Home-based training of rhythmic skills with a serious game in Parkinson's disease: usability and acceptability. Ann Phys Rehabil Med 2018 Nov;61(6):380-385. [doi: 10.1016/j.rehab.2018.08.002] [Medline: 30193992]
51. de Melo GEL, Kleiner AFR, Lopes JBP, et al. Effect of virtual reality training on walking distance and physical fitness in individuals with Parkinson's disease. NeuroRehabilitation 2018;42(4):473-480. [doi: 10.3233/NRE-172355] [Medline: 29660956]
52. de Melo Cerqueira TM, de Moura JA, de Lira JO, Leal JC, D'Amelio M, do Santos Mendes FA. Cognitive and motor effects of kinect-based games training in people with and without Parkinson disease: a preliminary study. Physiother Res Int 2020 Jan;25(1):e1807. [doi: 10.1002/pri.1807] [Medline: 31468656]
53. Ellis T, Latham NK, DeAngelis TR, Thomas CA, Saint-Hilaire M, Bickmore TW. Feasibility of a virtual exercise coach to promote walking in community-dwelling persons with Parkinson disease. Am J Phys Med Rehabil 2013;92(6):472-485. [doi: 10.1097/PHM.0b013e31828cd466]
54. Esculier JF, Vaudrin J, Bériault P, Gagnon K, Tremblay LE. Home-based balance training programme using Wii Fit with balance board for Parkinsons's disease: a pilot study. J Rehabil Med 2012 Feb;44(2):144-150. [doi: 10.2340/16501977-0922] [Medline: 22266676]
55. Fernández-González P, Carratalá-Tejada M, Monge-Pereira E, et al. Leap motion controlled video game-based therapy for upper limb rehabilitation in patients with Parkinson's disease: a feasibility study. J Neuroeng Rehabil 2019 Nov 6;16(1):133. [doi: 10.1186/s12984-019-0593-x] [Medline: 31694653]
56. Ferraz DD, Trippo KV, Duarte GP, Neto MG, Bernardes Santos KO, Filho JO. The effects of functional training, bicycle exercise, and exergaming on walking capacity of elderly patients with Parkinson disease: a pilot randomized controlled single-blinded trial. Arch Phys Med Rehabil 2018 May;99(5):826-833. [doi: 10.1016/j.apmr.2017.12.014] [Medline: 29337023]
57. Fundarò C, Maestri R, Ferriero G, Chimento P, Taveggia G, Casale R. Self-selected speed gait training in Parkinson's disease: robot-assisted gait training with virtual reality versus gait training on the ground. Eur J Phys Rehabil Med 2019 Aug;55(4):456-462. [doi: 10.23736/S1973-9087.18.05368-6] [Medline: 30370751]
58. Gandolfi M, Geroin C, Dimitrova E, et al. Virtual reality telerehabilitation for postural instability in Parkinson's disease: a multicenter, single-blind, randomized, controlled trial. Biomed Res Int 2017;2017:7962826. [doi: 10.1155/2017/7962826] [Medline: 29333454]
59. Ganzeboom M, Bakker M, Beijer L, Rietveld T, Strik H. Speech training for neurological patients using a serious game. Brit J Educational Tech 2018 Jul;49(4):761-774. [doi: 10.1111/bjet.12640]
60. Ganzeboom M, Bakker M, Beijer L, Strik H, Rietveld T. A serious game for speech training in dysarthric speakers with Parkinson's disease: exploring therapeutic efficacy and patient satisfaction. Int J Lang Commun Disord 2022 Jul;57(4):808-821. [doi: 10.1111/1460-6984.12722] [Medline: 35338751]
61. Goffredo M, Baglio F, DE Icco R, et al. Efficacy of non-immersive virtual reality-based telerehabilitation on postural stability in Parkinson's disease: a multicenter randomized controlled trial. Eur J Phys Rehabil Med 2023 Dec;59(6):689-696. [doi: 10.23736/S1973-9087.23.07954-6] [Medline: 37847247]
62. Gonçalves GB, Leite MAA, Orsini M, Pereira JS. Effects of using the Nintendo Wii Fit plus platform in the sensorimotor training of gait disorders in Parkinson's disease. Neurol Int 2014 Jan 17;6(1):5048. [doi: 10.4081/ni.2014.5048] [Medline: 24744845]
63. Gulcan K, Guclu-Gunduz A, Yasar E, Ar U, Sucullu Karadag Y, Saygili F. The effects of augmented and virtual reality gait training on balance and gait in patients with Parkinson's disease. Acta Neurol Belg 2023 Oct;123(5):1917-1925. [doi: 10.1007/s13760-022-02147-0] [Medline: 36443623]
64. Hajebrahimi F, Velioglu HA, Bayraktaroglu Z, Helvaci Yilmaz N, Hanoglu L. Clinical evaluation and resting state fMRI analysis of virtual reality based training in Parkinson's disease through a randomized controlled trial. Sci Rep 2022 May 16;12(1):8024. [doi: 10.1038/s41598-022-12061-3] [Medline: 35577874]
65. Hashemi Y, Taghizadeh G, Azad A, Behzadipour S. The effects of supervised and non-supervised upper limb virtual reality exercises on upper limb sensory-motor functions in patients with idiopathic Parkinson's disease. Hum Mov Sci 2022 Oct;85:102977. [doi: 10.1016/j.humov.2022.102977] [Medline: 35932518]
66. Herz NB, Mehta SH, Sethi KD, Jackson P, Hall P, Morgan JC. Nintendo Wii rehabilitation ("Wii-hab") provides benefits in Parkinson's disease. Parkinsonism Relat Disord 2013 Nov;19(11):1039-1042. [doi: 10.1016/j.parkreldis.2013.07.014] [Medline: 23968649]
67. Holmes JD, Gu ML, Johnson AM, Jenkins ME. The effects of a home-based virtual reality rehabilitation program on balance among individuals with Parkinson's disease. Phys Occup Ther Geriatr 2013 Sep;31(3):241-253. [doi: 10.3109/02703181.2013.814743]
68. Imbimbo I, Coraci D, Santilli C, et al. Parkinson's disease and virtual reality rehabilitation: cognitive reserve influences the walking and balance outcome. Neurol Sci 2021 Nov;42(11):4615-4621. [doi: 10.1007/s10072-021-05123-3] [Medline: 33661481]

XSL•FO
RenderX

69. Jäggi S, Wachter A, Adcock M, et al. Feasibility and effects of cognitive-motor exergames on fall risk factors in typical and atypical Parkinson's inpatients: a randomized controlled pilot study. Eur J Med Res 2023 Jan 16;28(1):30. [doi: 10.1186/s40001-022-00963-x] [Medline: 36647177]
70. Kashif M, Ahmad A, Bandpei MAM, Syed HA, Raza A, Sana V. A randomized controlled trial of motor imagery combined with virtual reality techniques in patients with Parkinson's disease. J Pers Med 2022 Mar 12;12(3):450. [doi: 10.3390/jpm12030450] [Medline: 35330450]
71. Kashif M, Ahmad A, Bandpei MAM, Gilani SA, Hanif A, Iram H. Combined effects of virtual reality techniques and motor imagery on balance, motor function and activities of daily living in patients with Parkinson's disease: a randomized controlled trial. BMC Geriatr 2022 Apr 30;22(1):381. [doi: 10.1186/s12877-022-03035-1] [Medline: 35488213]
72. Lau J, Regis C, Burke C, Kaleda MJ, McKenna R, Muratori LM. Immersive technology for cognitive-motor training in Parkinson's disease. Front Hum Neurosci 2022;16:863930. [doi: 10.3389/fnhum.2022.863930] [Medline: 35615742]
73. Lee NY, Lee DK, Song HS. Effect of virtual reality dance exercise on the balance, activities of daily living, and depressive disorder status of Parkinson's disease patients. J Phys Ther Sci 2015 Jan;27(1):145-147. [doi: 10.1589/jpts.27.145] [Medline: 25642060]
74. Liao YY, Yang YR, Cheng SJ, Wu YR, Fuh JL, Wang RY. Virtual reality-based training to improve obstacle-crossing performance and dynamic balance in patients with Parkinson's disease. Neurorehabil Neural Repair 2015 Aug;29(7):658-667. [doi: 10.1177/1545968314562111] [Medline: 25539782]
75. Liao YY, Yang YR, Wu YR, Wang RY. Virtual reality-based Wii Fit training in improving muscle strength, sensory integration ability, and walking abilities in patients with Parkinson's disease: a randomized control trial. Int J Gerontol 2015 Dec;9(4):190-195. [doi: 10.1016/j.ijge.2014.06.007]
76. Maranesi E, Casoni E, Baldoni R, et al. The effect of non-immersive virtual reality exergames versus traditional physiotherapy in Parkinson's disease older patients: preliminary results from a randomized-controlled trial. Int J Environ Res Public Health 2022 Nov 10;19(22):14818. [doi: 10.3390/ijerph192214818] [Medline: 36429537]
77. Mazur A, Pietraszko W, Krygowska-Wajs A. Visual feedback training using WII Fit improves balance in Parkinson's disease. Folia Med Cracov 2013;LIII(1):65-78.
78. Melo G, Kleiner AFR, Lopes J, et al. P100 - effects of virtual reality training on mobility in individuals with Parkinson's disease. Gait Posture 2018 Sep;65:394-395. [doi: 10.1016/j.gaitpost.2018.06.071]
79. Mhatre PV, Vilares I, Stibb SM, et al. Wii Fit balance board playing improves balance and gait in Parkinson disease. PM R 2013 Sep;5(9):769-777. [doi: 10.1016/j.pmrj.2013.05.019] [Medline: 23770422]
80. Mirelman A, Maidan I, Herman T, Deutsch JE, Giladi N, Hausdorff JM. Virtual reality for gait training: can it induce motor learning to enhance complex walking and reduce fall risk in patients with Parkinson's disease? J Gerontol A Biol Sci Med Sci 2011 Feb;66(2):234-240. [doi: 10.1093/gerona/glq201] [Medline: 21106702]
81. Mirelman A, Rochester L, Maidan I, et al. Addition of a non-immersive virtual reality component to treadmill training to reduce fall risk in older adults (V-TIME): a randomised controlled trial. Lancet 2016 Sep 17;388(10050):1170-1182. [doi: 10.1016/S0140-6736(16)31325-3] [Medline: 27524393]
82. Negrini S, Bissolotti L, Ferraris A, Noro F, Bishop MD, Villafañe JH. Nintendo Wii Fit for balance rehabilitation in patients with Parkinson's disease: a comparative study. J Bodyw Mov Ther 2017 Jan;21(1):117-123. [doi: 10.1016/j.jbmt.2016.06.001] [Medline: 28167167]
83. Nuic D, Vinti M, Karachi C, Foulon P, Van Hamme A, Welter ML. The feasibility and positive effects of a customised videogame rehabilitation programme for freezing of gait and falls in Parkinson's disease patients: a pilot study. J Neuroeng Rehabil 2018 Apr 10;15(1):31. [doi: 10.1186/s12984-018-0375-x] [Medline: 29636105]
84. Nuic D, van de Weijer S, Cherif S, et al. Home-based exergaming to treat gait and balance disorders in patients with Parkinson's disease: a phase II randomized controlled trial. Eur J Neurol 2024 Jan;31(1):e16055. [doi: 10.1111/ene.16055] [Medline: 37691341]
85. Oña ED, Balaguer C, Cano-de la Cuerda R, Collado-Vázquez S, Jardón A. Effectiveness of serious games for leap motion on the functionality of the upper limb in Parkinson's disease: a feasibility study. Comput Intell Neurosci 2018;2018:7148427. [doi: 10.1155/2018/7148427] [Medline: 29849550]
86. Özgönenel L, Çağırıcı S, Çabalar M, Durmuşoğlu G. Use of game console for rehabilitation of Parkinson's disease. Balkan Med J 2016 Jul;33(4):396-400. [doi: 10.5152/balkanmedj.2016.16842] [Medline: 27606134]
87. Palacios-Navarro G, García-Magariño I, Ramos-Lorente P. A kinect-based system for lower limb rehabilitation in Parkinson's disease patients: a pilot study. J Med Syst 2015 Sep;39(9):103. [doi: 10.1007/s10916-015-0289-0] [Medline: 26265237]
88. Pastore-Wapp M, Kaufmann BC, Nyffeler T, Wapp S, Bohlhalter S, Vanbellingen T. Feasibility of a combined intermittent theta-burst stimulation and video game-based dexterity training in Parkinson's disease. J Neuroeng Rehabil 2023 Jan 12;20(1):2. [doi: 10.1186/s12984-023-01123-w] [Medline: 36635679]
89. Pazzaglia C, Imbimbo I, Tranchita E, et al. Comparison of virtual reality rehabilitation and conventional rehabilitation in Parkinson's disease: a randomised controlled trial. Physiotherapy 2020 Mar;106:36-42. [doi: 10.1016/j.physio.2019.12.007] [Medline: 32026844]
90. Pelosin E, Ponte C, Putzolu M, et al. Motor-cognitive treadmill training with virtual reality in Parkinson's disease: the effect of training duration. Front Aging Neurosci 2021;13:753381. [doi: 10.3389/fnagi.2021.753381] [Medline: 35069171]

XSL•FO
RenderX

91. Pompeu JE, Mendes F, Silva KD, et al. Effect of Nintendo WiiTMBased motor and cognitive training on activities of daily living in patients with Parkinson's disease: a randomised clinical trial. Physiotherapy 2012 Sep;98(3):196-204. [doi: 10.1016/j.physio.2012.06.004] [Medline: 22898575]
92. Pompeu JE, Arduini LA, Botelho AR, et al. Feasibility, safety and outcomes of playing Kinect Adventures!. Physiotherapy 2014 Jun;100(2):162-168. [doi: 10.1016/j.physio.2013.10.003] [Medline: 24703891]
93. Pompeu JE, Torriani-Pasin C, Doná F, Ganança FF, da Silva KG, Ferraz HB. Effect of kinect games on postural control of patients with parkinson's disease. : Association for Computing Machinery; 2015 Oct Presented at: REHAB '15; Lisbon Portugal p. 54-57 URL: https://dl.acm.org/doi/proceedings/10.1145/2838944 [doi: 10.1145/2838944.2838958]
94. Pullia M, Ciatto L, Andronaco G, et al. Treadmill training plus semi-immersive virtual reality in Parkinson's disease: results from a pilot study. Brain Sci 2023 Sep 12;13(9):37759913. [doi: 10.3390/brainsci13091312] [Medline: 37759913]
95. Puyjarinet F, Bégel V, Geny C, et al. At-home training with a rhythmic video game for improving orofacial, manual, and gait abilities in Parkinson's disease: a pilot study. Front Neurosci 2022;16:874032. [doi: 10.3389/fnins.2022.874032] [Medline: 35769698]
96. Qayyum S, Hashmi Z, Waqas S, Tariq M, Mughal MW, Tanvir M. Effects of exer–gaming on balance and gait in Parkinson's patients. PJMHS 2022 Dec 30;16(12):213-214 [FREE Full text] [doi: 10.53350/pjmhs20221612213]
97. Ribas CG, Alves da Silva L, Corrêa MR, Teive HG, Valderramas S. Effectiveness of exergaming in improving functional balance, fatigue and quality of life in Parkinson's disease: a pilot randomized controlled trial. Parkinsonism Relat Disord 2017 May;38:13-18. [doi: 10.1016/j.parkreldis.2017.02.006] [Medline: 28190675]
98. Sánchez-Herrera-Baeza P, Cano-de-la-Cuerda R, Oña-Simbaña ED, et al. The impact of a novel immersive virtual reality technology associated with serious games in Parkinson's disease patients on upper limb rehabilitation: a mixed methods intervention study. Sensors (Basel) 2020 Apr 11;20(8):2168. [doi: 10.3390/s20082168] [Medline: 32290517]
99. Santos P, Machado T, Santos L, Ribeiro N, Melo A. Efficacy of the Nintendo Wii combination with conventional exercises in the rehabilitation of individuals with Parkinson's disease: a randomized clinical trial. NeuroRehabilitation 2019;45(2):255-263. [doi: 10.3233/NRE-192771] [Medline: 31498138]
100. Schaeffer E, Busch JH, Roeben B, et al. Effects of exergaming on attentional deficits and dual-tasking in Parkinson's disease. Front Neurol 2019;10(JUN):646. [doi: 10.3389/fneur.2019.00646] [Medline: 31275234]
101. Severiano MIR, Zeigelboim BS, Teive HAG, Santos GJB, Fonseca VR. Effect of virtual reality in Parkinson's disease: a prospective observational study. Arq Neuropsiquiatr 2018 Feb;76(2):78-84. [doi: 10.1590/0004-282X20170195] [Medline: 29489960]
102. Shih MC, Wang RY, Cheng SJ, Yang YR. Effects of a balance-based exergaming intervention using the Kinect sensor on posture stability in individuals with Parkinson's disease: a single-blinded randomized controlled trial. J Neuroeng Rehabil 2016 Aug 27;13(1):78. [doi: 10.1186/s12984-016-0185-y] [Medline: 27568011]
103. Song J, Paul SS, Caetano MJD, et al. Home-based step training using videogame technology in people with Parkinson's disease: a single-blinded randomised controlled trial. Clin Rehabil 2018 Mar;32(3):299-311. [doi: 10.1177/0269215517721593] [Medline: 28745063]
104. Summa S, Basteris A, Betti E, Sanguineti V. Adaptive training with full-body movements to reduce bradykinesia in persons with Parkinson's disease: a pilot study. J Neuroeng Rehabil 2015 Feb 14;12(1):16. [doi: 10.1186/s12984-015-0009-5] [Medline: 25885094]
105. Syed UE, Kamal A. Video game-based and conventional therapies in patients of neurological deficits: an experimental study. Disabil Rehabil Assist Technol 2021 Apr;16(3):332-339. [doi: 10.1080/17483107.2019.1679266] [Medline: 31684776]
106. van den Heuvel MRC, Kwakkel G, Beek PJ, Berendse HW, Daffertshofer A, van Wegen EEH. Effects of augmented visual feedback during balance training in Parkinson's disease: a pilot randomized clinical trial. Parkinsonism Relat Disord 2014 Dec;20(12):1352-1358. [doi: 10.1016/j.parkreldis.2014.09.022] [Medline: 25283070]
107. van Beek JJW, van Wegen EEH, Bohlhalter S, Vanbellingen T. Exergaming-based dexterity training in persons with Parkinson disease: a pilot feasibility study. J Neurol Phys Ther 2019 Jul;43(3):168-174. [doi: 10.1097/NPT.0000000000000278] [Medline: 31136450]
108. van der Kolk NM, de Vries NM, Kessels RPC, et al. Effectiveness of home-based and remotely supervised aerobic exercise in Parkinson's disease: a double-blind, randomised controlled trial. Lancet Neurol 2019 Nov;18(11):998-1008. [doi: 10.1016/S1474-4422(19)30285-6] [Medline: 31521532]
109. Wang YW, Chen CH, Lin YC. Balance rehabilitation system for parkinson's disease patients based on augmented reality. Presented at: 2020 IEEE Eurasia Conference on IOT, Communication and Engineering (ECICE); Oct 23-25, 2020; Yunlin, Taiwan p. 191-194 URL: https://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=9301908 [doi: 10.1109/ECICE50847.2020.9302018]
110. Yang WC, Wang HK, Wu RM, Lo CS, Lin KH. Home-based virtual reality balance training and conventional balance training in Parkinson's disease: a randomized controlled trial. J Formos Med Assoc 2016 Sep;115(9):734-743. [doi: 10.1016/j.jfma.2015.07.012] [Medline: 26279172]
111. Yen CY, Lin KH, Hu MH, Wu RM, Lu TW, Lin CH. Effects of virtual reality-augmented balance training on sensory organization and attentional demand for postural control in people with Parkinson disease: a randomized controlled trial. Phys Ther 2011 Jun;91(6):862-874. [doi: 10.2522/ptj.20100050] [Medline: 21474638]

XSL•FO
RenderX

112. Yuan RY, Chen SC, Peng CW, Lin YN, Chang YT, Lai CH. Effects of interactive video-game-based exercise on balance in older adults with mild-to-moderate Parkinson's disease. J Neuroeng Rehabil 2020 Jul 13;17(1):91. [doi: 10.1186/s12984-020-00725-y] [Medline: 32660512]
113. Yun SJ, Hyun SE, Oh BM, Seo HG. Fully immersive virtual reality exergames with dual-task components for patients with Parkinson's disease: a feasibility study. J Neuroeng Rehabil 2023 Jul 18;20(1):92. [doi: 10.1186/s12984-023-01215-7] [Medline: 37464349]
114. Zeigelboim BS, José MR, Severiano MIR, et al. The use of exergames in the neurorehabilitation of people with Parkinson disease: the impact on daily life. Int Arch Otorhinolaryngol 2021 Jan;25(1):e64-e70. [doi: 10.1055/s-0040-1702973] [Medline: 33542753]
115. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71. [doi: 10.1136/bmj.n71] [Medline: 33782057]
116. Hoehn MM, Yahr MD. Parkinsonism: onset, progression and mortality. Neurology (ECronicon) 1967 May;17(5):427-442. [doi: 10.1212/wnl.17.5.427] [Medline: 6067254]
117. Caudron S, Guerraz M, Eusebio A, Gros JP, Azulay JP, Vaugoyeau M. Evaluation of a visual biofeedback on the postural control in Parkinson's disease. Neurophysiol Clin 2014 Jan;44(1):77-86. [doi: 10.1016/j.neucli.2013.10.134] [Medline: 24502908]
118. van Nuland AJ, Helmich RC, Dirkx MF, et al. Effects of dopamine on reinforcement learning in Parkinson's disease depend on motor phenotype. Brain (Bacau) 2020 Dec 5;143(11):3422-3434. [doi: 10.1093/brain/awaa335] [Medline: 33147621]
119. Steiner B, Elgert L, Saalfeld B, Wolf KH. Gamification in rehabilitation of patients with musculoskeletal diseases of the shoulder: scoping review. JMIR Serious Games 2020 Aug 25;8(3):e19914. [doi: 10.2196/19914] [Medline: 32840488]
120. Chen Y. Exploring design guidelines of using user-centered design in gamification development: a delphi study. Int J Hum Comput Interact 2019 Aug 9;35(13):1170-1181. [doi: 10.1080/10447318.2018.1514823]
121. Gregor P, Newell AF, Zajicek M. Designing for dynamic diversity. Presented at: ASSETS02; Jul 8-10, 2002; Edinburgh, Scotland p. 151-156.
122. Cunha B, Ferreira R, Sousa ASP. Home-based rehabilitation of the shoulder using auxiliary systems and artificial intelligence: an overview. Sensors (Basel) 2023 Aug 11;23(16):7100. [doi: 10.3390/s23167100] [Medline: 37631637]
123. Ernst M, Folkerts AK, Gollan R, et al. Physical exercise for people with Parkinson's disease: a systematic review and network meta-analysis. Cochrane Database Syst Rev 2023 Jan 5;1(1):CD013856. [doi: 10.1002/14651858.CD013856.pub2] [Medline: 36602886]

Abbreviations

PD: Parkinson disease
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews
VR: virtual reality

Edited by A Coristine; submitted 29.11.24; peer-reviewed by A Poliani, E Vallefuoco, H Söbke, RS Gomaa Mahmoud; revised version received 29.04.25; accepted 29.04.25; published 04.07.25.

Please cite as:
Bosch-Barceló P, Martínez-Navarro O, Masbernat-Almenara M, Tersa-Miralles C, Pakarinen A, Fernández-Lago H
Gamification Integration in Technological Devices for Motor Rehabilitation in Parkinson Disease: Scoping Review
JMIR Serious Games 2025;13:e69433
URL: https://games.jmir.org/2025/1/e69433
doi:10.2196/69433

© Pere Bosch-Barceló, Oriol Martínez-Navarro, Maria Masbernat-Almenara, Carlos Tersa-Miralles, Anni Pakarinen, Helena Fernández-Lago. Originally published in JMIR Serious Games (https://games.jmir.org), 4.7.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Home-Based Virtual Reality Training for Enhanced Balance, Strength, and Mobility Among Older Adults With Frailty: Systematic Review and Meta-Analysis

Hammad Alhasan[*], PhD; Elaf Alandijani[*], PT; Lara Bahamdan[*], PT; Ghofran Khudary[*], PT; Yara Aburaya[*], PT; Abdulaziz Awali[*], PhD; Mansour Abdullah Alshehri[*], PhD

Department of Medical Rehabilitation Sciences, Faculty of Applied Medical Sciences, Umm Al-Qura University, Makkah 24382, Saudi Arabia
[*]all authors contributed equally

Corresponding Author:
Hammad Alhasan, PhD
Department of Medical Rehabilitation Sciences, Faculty of Applied Medical Sciences, Umm Al-Qura University, Makkah 24382, Saudi Arabia

Abstract

Background: Frailty is a geriatric syndrome associated with increased risk of falls, hospitalization, and reduced quality of life. Traditional exercises may be unsuitable for older adults with frailty due to mobility issues and accessibility barriers. Virtual reality (VR) offers an engaging, home-based alternative by providing interactive training with real-time feedback. VR interventions have shown potential benefits for improving balance, strength, and mobility.

Objective: This systematic review and meta-analysis aimed to evaluate the effectiveness of VR-based home training programs in improving balance, strength, and mobility among older adults with frailty and prefrailty.

Methods: A systematic search was conducted in PubMed, Scopus, and Web of Science from inception to November 1, 2023, using terms related to older adults, frailty, virtual reality, balance, mobility, and strength. Eligible studies included randomized and nonrandomized trials involving adults with frailty or prefrailty aged ≥65 years who received home-based VR interventions aimed at improving balance, strength, or functional mobility. Comparator groups included no intervention, traditional exercise, or standard care. Studies involving participants with neurological or cognitive disorders were excluded. Study quality was assessed using the Physiotherapy Evidence Database scale. A random-effects meta-analysis was performed to calculate pooled mean differences (MD) and 95% CIs for 3 primary outcomes: Berg Balance Scale, Timed Up and Go, and Chair Stand.

Results: A total of 1063 records were identified, with 1023 screened after duplicate removal. Six studies met the inclusion criteria, involving 407 participants (mean age 75.2, SD 6.4 y), of whom 198 were allocated to VR interventions and 159 to control groups. VR interventions lasted a mean of 13.3 (SD 7.7) weeks, with an average of 39.6 (SD 5.2) sessions lasting 25.3 (SD 5) minutes. Methodological quality was high in 5 studies (mean Physiotherapy Evidence Database score=5.6, SD 1.3). Four studies were included in the meta-analysis. Significant improvements were observed in balance, as measured by the Berg Balance Scale (MD=3.62; 95% CI 2.29 - 4.95; $P<.001$; $I^2=0\%$). No significant effects were found for mobility (Timed Up and Go: MD=−0.37; 95% CI −1.16 to 0.41; $P=.35$; $I^2=0\%$) or strength (Chair Stand: MD=−0.20; 95% CI −1.70 to 1.29; $P=.79$; $I^2=21\%$).

Conclusions: VR-based home exercise interventions show promise in improving balance among older adults with frailty and prefrailty. However, their effects on strength and functional mobility remain unclear. Variability in study designs and outcome measures limits the generalizability of current findings. Further high-quality research is needed to determine optimal VR training protocols and assess long-term adherence and clinical effectiveness.

Trial Registration: PROSPERO (International Prospective Register of Systematic Review) CRD42023478330; https://www.crd.york.ac.uk/PROSPERO/view/CRD42023478330

(JMIR Serious Games 2025;13:e67146) doi:10.2196/67146

KEYWORDS

older adults with frailty; balance; strength; functional mobility; home-based training; virtual reality; exergames; systematic review; older adults; PRISMA

XSL•FO
RenderX

Introduction

Frailty is a major concern for older adults, significantly affecting their well-being and quality of life [1,2]. It is characterized by a significant decline in the performance of various physiological systems and lacks a universal phenotype, signifying its heterogeneity as a geriatric syndrome [3,4]. Instead, it varies among individuals, considering their unique characteristics and circumstances, with a consensus that frailty is characterized by an increased vulnerability to adverse health outcomes [5,6]. Individuals with frailty, who are prone to experiencing functional decline and disability, face a higher risk of falls, hospitalization, and mortality [7,8]. Therefore, falls are a major concern for older adults with frailty, as they can lead to loss of autonomy, injuries, and even death [9,10].

Frailty can be categorized into 3 stages: prefrailty, frailty, and frailty complications. In the prefrailty stage, individuals may experience 1 or 2 symptoms that directly indicate limitations in their physical function or health; with early intervention and appropriate responses, successful management of these challenges is possible [11]. The frailty stage is characterized by hallmark symptoms such as weight loss, exhaustion, low physical activity, slowness, and weakness that lead to limitations in the individual's functioning and worsening of the overall quality of life [12]. The frailty complications stage occurs when an individual's functional independence is significantly impaired with accompanying behavioral patterns that may lead to death [13]. The Fried Frailty Phenotype stands out as a widely used tool for assessing frailty. It assesses physical frailty using 5 criteria: unintentional weight loss, low energy or self-reported exhaustion, reduced grip strength, reduced physical activity, and slowness by slowed walking speed. When 1 or 2 criteria are present, the individual is considered to be in a prefrail state, while the presence of more than 2 criteria indicates frailty [14,15].

Frailty prevalence increases with age, affecting 46% of older adults in the prefrail stage and 15% - 11% in the frail stage [16]. Socioeconomic factors, nutritional status, and ethnic background also play significant roles in frailty prevalence. Longitudinal studies on frailty progression are limited, but some indicate that frailty status can improve, remain stable, or worsen over time [17,18]. These statistics underscore the widespread impact of frailty among older adults, highlighting the need for targeted interventions.

Given the above, society is faced with the challenge of finding effective rehabilitation solutions to promote healthy aging [19,20]. Traditional exercises are often not preferred by older adults due to factors such as lack of motivation, perceived physical limitations, and the repetitive and monotonous nature of the exercises [21-23]. Virtual reality (VR) technology presents a promising alternative that could effectively address these challenges as it provides practical and easy-to-use solutions [24-28]. In this review, VR refers to interactive, digital systems that simulate task-oriented environments to encourage physical rehabilitation, which includes platforms such as sensor-enabled gaming platforms, exergames, and nonimmersive VR [29,30]. This definition incorporates motion-tracking systems that do not require the use of head-mounted displays but require real-time feedback and user engagement through physical movement, which are commonly used in VR-based rehabilitation interventions [31].

In comparison with traditional exercises, VR offers numerous advantages, such as structured guidance, real-time feedback, and adaptable difficulty, enabling users to engage safely within their abilities, especially for those at risk of falling [32]. Additionally, the interactive elements of VR have been shown to increase motivation, adherence, and cognitive engagement [26,33-35]. It should also be emphasized that VR-based exercises can significantly improve motor and cognitive functions [29,32,36,37]. A recent randomized controlled trial (RCT) compared the effect of VR training to Otago exercises [38]. Balance was used as an outcome, and the results indicated that the VR group showed significant improvements compared to the Otago exercise group. However, the study used a pre-post intervention design without a control group, which makes it challenging to attribute improvements solely to the VR intervention. Additionally, the study's findings may not apply to older adults with frailty as they recruited community-dwelling older adults. Similarly, a trial compared the effect of traditional versus VR treadmill on mobility and cognition among individuals with frailty [39]. Both modalities yield positive effects, but there is a preference for VR over traditional treadmill exercises due to the added benefit of cognitive improvement. A recent systematic review found that supervised VR training in rehabilitation settings can improve balance and reduce fall risk among older adults with frailty [40,41].

While traditional, center-based training has demonstrated the aforementioned benefits, older adults with frailty often face challenges to participating in traditional center-based exercise due to mobility limitations, fear of falling, and transportation challenges [42,43]. The COVID-19 pandemic further highlighted the importance of remote care models for this population [44,45]. These limitations emphasize the need for other approaches that are both safe and accessible. VR-based training may be specifically suitable for older adults with frailty, who frequently face challenges to engaging in traditional exercise due to impaired mobility or fear of falling [36,46]. Hence, home-based VR training is increasingly being studied as a feasible method to enhance physical activity among individuals with frailty. This prompts the need to investigate whether similar findings can be achieved through the use of VR at home, taking into consideration its ease of access, affordability, and other advantages. Therefore, the objective of the current systematic review was to examine the effectiveness of VR as home-based training on balance, strength, and mobility outcomes among older adults with frailty or prefrailty.

Methods

Protocols and Registration

The current systematic review was registered on PROSPERO with the following registration (CRD42023478330 [47]).

XSL•FO
RenderX

Data Sources

The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) guidelines were followed for this review. The PRISMA item checklist can be found in Checklist 1. The search time frame was from inception to November 1, 2023. The goal was to identify recent studies on the effects of VR training for enhanced balance, strength, and mobility at home among older adults with frailty and prefrailty. Two authors independently performed searches in the following databases: Scopus, Web of Science, and PubMed.

Search Strategies

The search terms were specific to each database. The following is an example of the search terms used in Scopus: risk of fall OR balance OR strength OR function AND frail OR prefrail AND older adult AND virtual reality OR video games OR mobile game. A detailed overview of the search terms and strategies used is provided in Multimedia Appendix 1.

Selection Criteria

The study comprised all English-language papers, including those that used a single-group design in which a VR as a home-based exercise intervention was compared with no intervention or other interventions for enhanced strength, balance, and mobility among older adults with frailty or prefrailty. The PICOS (Population, Intervention, Comparison, Outcome, and Study Design) framework for the current review was as follows:

- Of population (P), older adults with frailty or prefrailty aged 65 years or more.
- Of intervention (I) VR, a home-based exercise that is used to improve balance, strength, and mobility.
- Of comparison (C) no intervention, traditional exercises, or standard care.
- Of outcomes (O) balance, strength, and mobility measured using validated outcome measures.
- Of study design (S), RCT and non-RCT.

Table 1 summarizes the inclusion and exclusion criteria structured according to the PICOS framework.

Table . Summary of the inclusion and exclusion criteria.

Category	Inclusion criteria	Exclusion criteria
Population	Older adults with frailty or prefrailty (aged 65+ years)	Adults without frailty and younger populations (<64 years)
Intervention	VR[a], video games, or mobile game-based interventions targeting balance, strength, or function	Interventions not involving VR or gaming, or not targeting balance, strength, or function
Comparison	Any (eg, standard care, other exercise modalities, or no intervention)	None required
Outcome	Outcomes related to balance, strength, and function	Studies not reporting functional outcomes related to balance, strength, mobility, or fall risk
Study design	Randomized controlled trials or clinical trials	Observational studies, reviews, or case reports
Language	English	Non-English publications
Publication type	Peer-reviewed full-text papers	Abstracts, dissertations, or protocols

[a]VR: virtual reality.

Participants

The included studies comprised male or female older adults with a mean age of 65 years or older, described as older adults with frailty, older adults with prefrailty, aged, geriatric, or older adults living in the community, independently, in retirement centers or nursing homes. Studies that included participants with specific medical conditions, such as stroke, Parkinson disease, or cognitive impairment, were excluded. Table 2 summarizes the frailty status of participants across the included studies.

XSL•FO
RenderX

Table . Frailty status of participants in included studies.

Study	Frailty status (N)	Frailty status description
[48]	Frail (61)	Participants were explicitly described as frail, with reduced mobility.
[49]	Prefrail (202)	Participants were described as having a moderate fall risk.
[50]	Prefrail (59)	Participants showed fall risk indicators and used assistive devices.
[51]	Prefrail (30)	Participants were sedentary, nonexercising older adults with some support needs.
[52]	Frail (18)	All participants had a history of falls and lived in a nursing home.
[53]	Prefrail (37)	All participants were explicitly classified as pre-frail.

Interventions

The included studies explored various interventions to improve balance, mobility, and physical function; these included video-guided exercises with resistance bands, VR balance training guided by a physiotherapist, or sensor-based exergames exercised at home. Each intervention was compared to traditional exercise programs or participants' usual activities to assess their effectiveness.

Outcome Measures

The included studies assessed the effectiveness of various VR interventions for older adults by measuring changes in physical function and mobility. Outcomes such as the knee extension strength test and Sit-to-Stand test were used to assess lower extremity strength, the Berg Balance Scale (BBS) test was used to assess balance, and the Timed Up and Go (TUG) test was used to assess functional mobility.

Quality Assessment

The Physiotherapy Evidence Database (PEDro) assessment tool was used to assess the methodological quality of the included studies [54]. The total PEDro score reflected the quality of the study as follows: a total score of ≥6 indicated high quality, 4 - 5 represented fair quality, and ≤3 indicated poor quality [55].

Data Extraction

The reviewers independently assessed the trials for eligibility by reviewing the titles and abstracts. If a paper title or abstract was deemed relevant, the full text was retrieved for evaluation against the inclusion and exclusion criteria. Any disagreement between the authors was resolved by the lead author. A data extraction form was created, and the data were extracted by the independent reviewers.

Data Analysis

A random-effects meta-analysis was performed using Review Manager (version 5.4; Cochrane) software. Three primary outcomes were included: BBS, TUG, and Chair Stand (CS) tests. The purpose was to identify the mean difference (MD) in balance, risk of falls, and strength between VR groups and conventional intervention or control groups, and also to determine the overall treatment effect size. Heterogeneity was assessed with the I^2 index, which has 4 classification levels: unimportant heterogeneity (0% - 40%), moderate heterogeneity (30% - 60%), substantial heterogeneity (50% - 90%), and considerable heterogeneity (75% - 100%) [56]. Effect sizes were calculated for all studies using Cohen *d*, with <0.2 indicating a "trivial" effect size, with 0.2 indicating a "small" effect size, 0.5 indicating a "medium" effect size, and 0.8 indicating a "large" effect size [57].

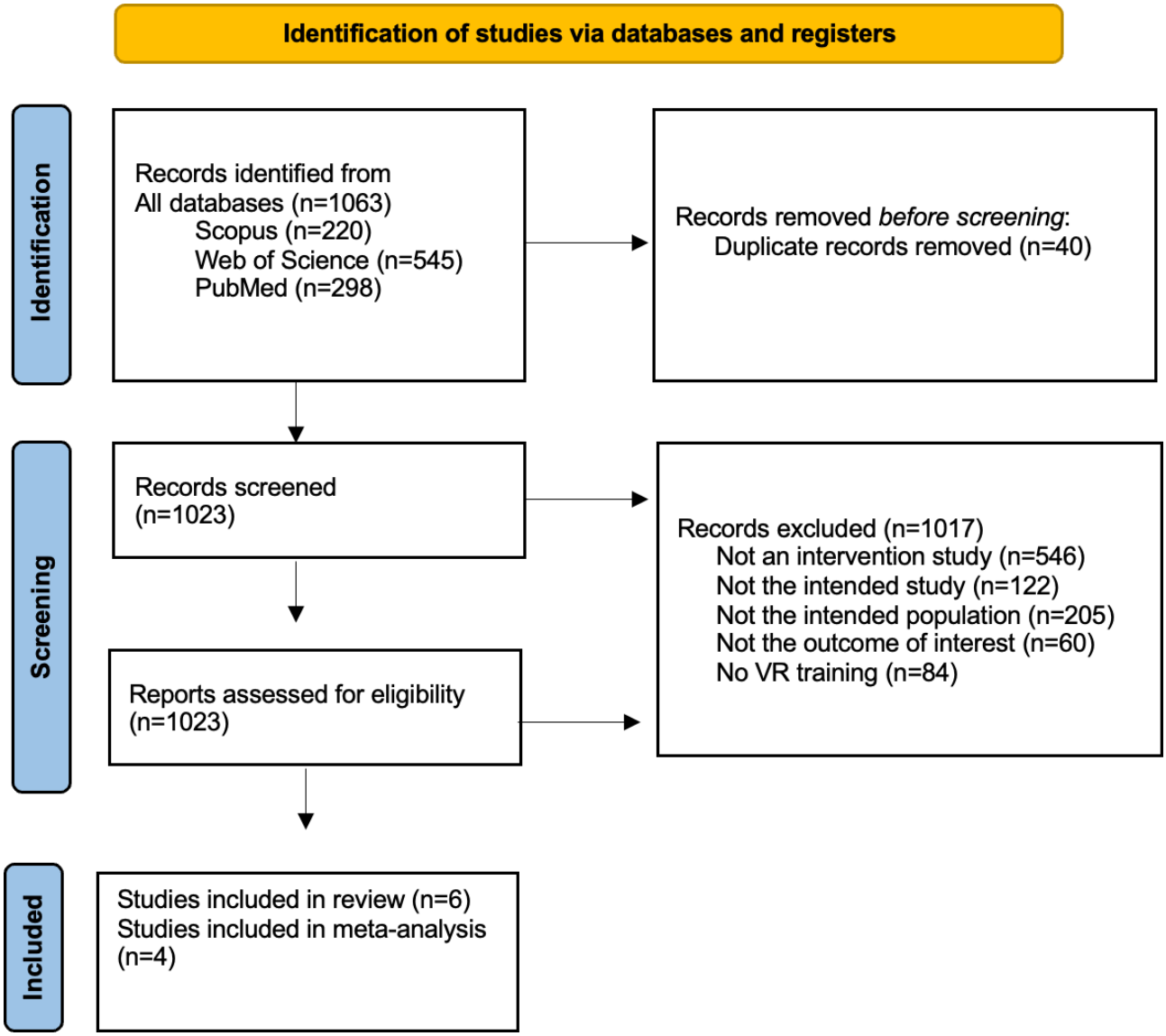
Results

Overview

A total of 1063 papers were identified as relevant; 40 were duplicates. The remaining 1023 papers were screened. After the initial screening, 1017 papers were excluded based on the titles and abstracts. The final review included 6 papers. The selection process for this systematic review is presented in the flow diagram in Figure 1.

XSL•FO
RenderX

Figure 1. The results of the literature search conducted on November 1, 2023. VR: virtual reality.



Methodological Quality

The mean PEDro score was 5.6 (SD 1.3) with 5 studies [48,50-53] graded as high quality and 1 [49] as poor quality. Table 3 presents the results of the quality assessment of the included studies.

XSL•FO
RenderX

Table . Physiotherapy evidence database scale assessment for included studies.

Study	1[a]	2[b]	3[c]	4[d]	5[e]	6[f]	7[g]	8[h]	9[i]	10[j]	11[k]	Total
[48]	✓[l]	✓	N[m]	✓	N	N	N	✓	✓	✓	✓	6/10
[49]	✓	N	N	N	N	N	N	N	✓	✓	✓	3/10
[50]	✓	✓	N	N	N	N	✓	✓	✓	✓	✓	6/10
[51]	✓	✓	N	✓	N	N	N	✓	✓	✓	✓	6/10
[52]	✓	✓	N	✓	N	N	✓	✓	N	✓	✓	6/10
[53]	✓	✓	✓	✓	N	N	✓	✓	N	✓	✓	7/10
Total	6/6	5/6	1/6	4/6	0/6	0/6	3/6	5/6	4/6	6/6	6/6	

[a]1: Eligibility criteria.
[b]2: Random allocation.
[c]3: Concealed allocation.
[d]4: Baseline comparability.
[e]5: Blind people.
[f]6: Blind therapists.
[g]7: Blind assessors.
[h]8: Adequate follow-up.
[i]9: Intention-to-treat analysis.
[j]10: Between-group comparisons.
[k]11: Point estimates and variability.
[l]✓: yes.
[m]N: no.

Characteristics of Included Studies

Participants and Study Designs

Four of the included studies [48,49,51,52] were RCTs and the remainder were experimental non-RCTs [50,53]. The total number of participants recruited studies was 407, with a mean age of 75.2 (SD 6.4) years, while 357 participants were included in the analysis. Of those analyzed, the VR groups comprised 198 participants, and the control groups comprised 159 participants. The number of individuals in the VR groups ranged from 7 to 63, with a mean of 33 (SD 21.1). In the control groups, the number of individuals ranged from 11 to 61, with a mean of 31.8 (SD 20.1). All studies were conducted at home or in nursing homes and in public home care. Characteristics of included studies are summarized in Table 4.

XSL•FO
RenderX

Table . Characteristics of included studies.

Study	Design	Total number of participants	Intervention	Training type for the intervention group	Total sessions, weeks, duration	Number of samples analyzed	Outcome measures	Effect size
[48]	RCT[a]	61	IG[b]: video + resistance band; CG[c]: standard care	Individual home-based exercise using video and booklet	5 months, 3 sessions/week, 26 min/session	IG=25; CG=28	Chair Stand	d=0.32
[49]	RCT	37	IG: step pad + CSRT[d]; CG: usual activities	Home-based interactive step game	8 weeks, 2 - 3 sessions/week, 15 - 20 min/session	IG=15; CG=17	TUG[e], Chair Stand, proprioception	TUG: d=0.14; CS: d=0.29
[50]	Pilot study	148	IG1: Microsoft-Kinect exergames; IG2: SMT[f]; CG: usual activities	Unsupervised home programs (WEBB[g], Otago, SMT)	16 weeks, 3 sessions/week	IG1=24; IG2=39; CG=61	TUG, Chair Stand, proprioception	TUG: d=0.24; CS: d=0.16
[51]	RCT	100	IG: Kinect video games; CG: balance, stretching, or strength	Play Kinect games supervised by nurse	6 weeks, 5 sessions/week, 30 min/session	IG=48; CG=42	TUG, BBS[h]	BBS: d=1.10; TUG: d=0.05
[52]	RCT	21	IG: balance training with BTS NIRVANA[i] VR[j]; CG: conventional balance exercises	VR-based balance exercises supervised by a PT[k] in a nursing home	6 weeks, 3 sessions/week, ~35 - 45 min/session	IG=7; CG=11	TUG, BBS	TUG: *P*=.01; BBS: d=0.71
[53]	Pilot study	40	IG: Otago-based exercise; CG: none	Lower-body strength and balance exercises	6 months (3 supervised + 3 unsupervised)	40	TUG, Chair Stand	TUG: *P*=.006; CS: *P*=.03

[a]RCT: randomized control trial.

[b]IG: intervention group.

[c]CG: control group.

[d]CSRT: choice stepping reaction time.

[e]TUG: Timed Up and Go.

[f]SMT: Step Mat Training.

[g]WEBB: Weight-Bearing Exercise for Better Balance.

[h]BBS: Berg Balance Scale.

[i]BTS NIRVANA: innovative therapeutic systems aiding the rehabilitation process of patients affected by neuro-motor disease by multisensorial stimulation.

[j]VR: virtual reality.

[k]PT: physical therapist.

Number and Duration of Intervention

The duration of VR sessions ranged from 6 to 24 weeks, with a mean of 13.3 (SD 7.7) weeks. The total number of sessions ranged from 16 to 120, with a mean of 39.6 (SD 5.2) sessions. The length of the sessions was 10 to 50 minutes, with a mean of 25.3 (SD 5) minutes. Descriptions of the exergame interventions used in the included studies can be found in Table 5 and Multimedia Appendix 2.

XSL•FO
RenderX

Table . Descriptions of VR[a] or exergame interventions used in included studies.

Study	VR tool	Description
[48]	Television-based video exercise program	Participants followed a prerecorded exercise video featuring strength and balance activities based on the Otago Exercise Program.
[49]	Dance Dance Revolution style step pad game	Participants used a step pad linked to a screen to play rhythm-based games, requiring directional steps and incorporating cognitive challenges to enhance executive function.
[50]	Microsoft Kinect-based exergame	Participants engaged in movement-based games that require stepping, shifting, and reaching, using motion-sensor technology.
[51]	Microsoft Kinect-based exergame	Participants used a motion-sensing camera system to perform stepping, squatting, and weight-shifting, guided by visual cues on a screen.
[52]	Nintendo Wii Fit system	Participants performed exergames using the Wii Balance Board, performing static and dynamic activities with real-time visual feedback.
[53]	Tablet-based video program with wearable motion sensors	Participants used a tablet app that delivered exercise videos inspired by the Otago program, and a necklace-worn motion sensor was used for activity monitoring and feedback.

[a]VR: virtual reality.

Qualitative Analysis

BBS

Two studies [51,52] used the BBS as a key outcome measure. Significant improvements were observed in VR groups. Both studies assessed BBS postintervention within 6 weeks.

TUG

Five studies [49-53] used the TUG as a key outcome measure. Significant improvements were found in the VR group in 3 studies [50-52], where 2 studies [51,52] assessed TUG postintervention within 6 weeks and 1 within 6 months of follow-up [50]. The remaining 2 studies [49,53] did not show significant improvements for all groups, where 1 study [49] assessed TUG postintervention within 8 weeks and the other one [53] within 16 weeks.

Strength

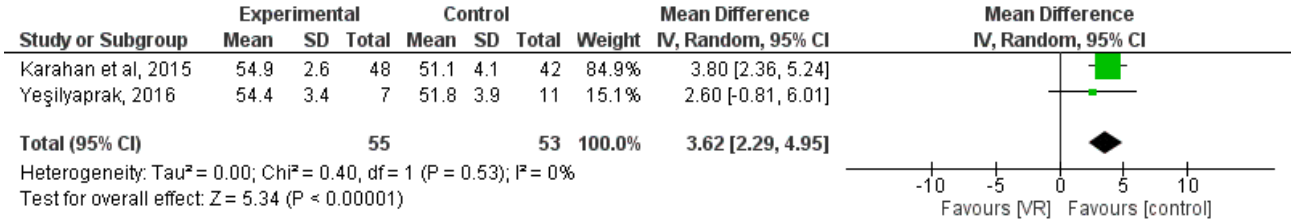
Two studies [49,50] used knee extension strength as a key outcome. Both studies did not show significant improvements in any group. Four studies assessed CS performance as a key outcome measure [48-50,53]. All studies did not show significant improvements for all groups.

Meta-Analysis

Overview

Only four studies [49-52] were included in the meta-analysis. The study by Vestergaard et al [48] did not use standardized outcomes compatible for pooling, and the study byGeraedts et al [53] was a single-arm study without a control group, preventing calculation of comparative effect sizes. Therefore, both were excluded from the meta-analysis [43-46]. Figures 2-4 show the overall treatment effect size and the results of each study on the BBS, TUG, and CS.

Figure 2. Forest plot for the mean difference of the effect of VR compared with conventional interventions and control on the BBS; lower BBS mean score indicates higher risk of falling [51,52]. BBS: Berg Balance Scale; VR: virtual reality.



XSL•FO
RenderX

Figure 3. Forest plot for the mean difference of the effect of VR compared with conventional interventions and control on the time (in seconds) of the TUG; lower TUG mean score indicates better mobility performance [49-52]. TUG: Timed Up and Go; VR: virtual reality.

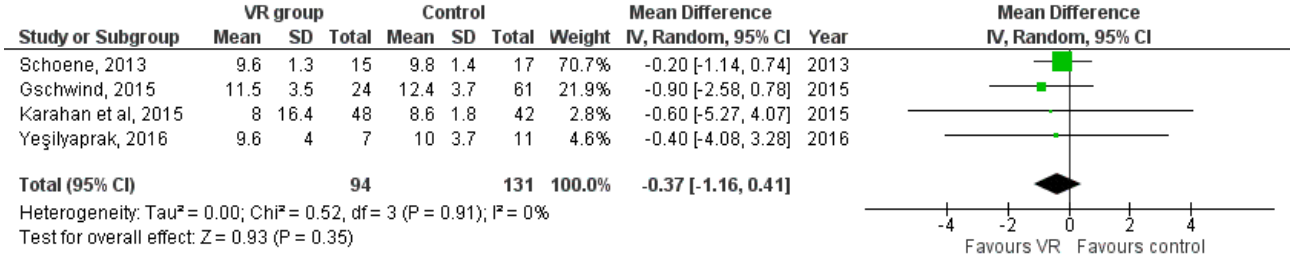
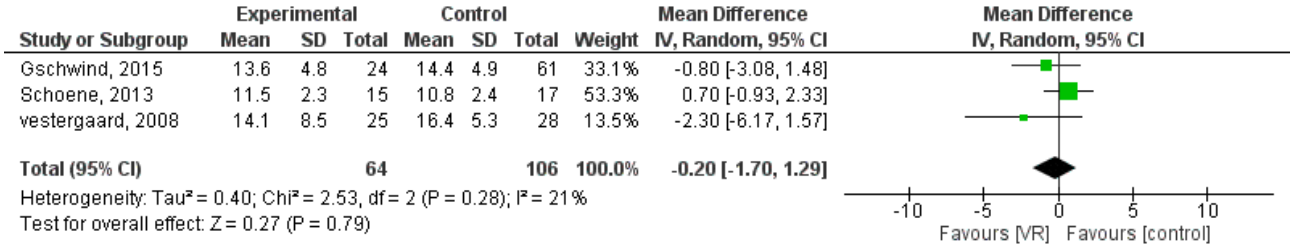


Figure 4. Forest plot for the mean difference of the effect of VR compared with conventional interventions and control on the Chair Stand test; lower mean score indicates better mobility performance [48-50]. VR: virtual reality.



BBS

Two [51,52] studies with 55 participants were eligible for inclusion in this meta-analysis. A forest plot revealed that VR yielded better results than conventional interventions and no intervention (control) in terms of improvements in postural control (MD=3.62; 95% CI 2.29 to 4.95; P<.001; I^2=0%). For conventional interventions and the control, the results of the meta-analysis showed a lower mean score on the BBS, indicating a higher risk of falling (Figure 2).

TUG

Four studies [49-52] with 94 participants were eligible for inclusion in this meta-analysis. The forest plot showed no significant differences between the interactive video games and other interventions or the control on TUG (MD=−0.37; 95% CI −1.16 to 0.41; P=.35; I^2=0%; Figure 3).

Strength

Three studies [48-50] with 64 participants were eligible for inclusion in this meta-analysis. The forest plot showed no significant differences between interactive video games and other interventions or the control on CS (MD=−0.20; 95% CI −1.70 to 1.29; P=.79; I^2=21%; Figure 4).

Discussion

Principal Findings

This current systematic review aimed to identify studies that have investigated the use of VR training at home to improve balance, strength, and mobility among older adults with frailty and prefrailty. Although few studies have focused on this aim, the findings showed that VR was effective at improving balance but not strength and mobility. One potential reason for this could be due to limited experience with independent technology use among older adults and therapists, a challenge that has been validated in previous studies [41,58]. Nevertheless, the studies included in this review are examples of how technological advancements can reshape health care delivery and demonstrate that certain VR interventions can be used safely with older adults, as no study reported any safety hazards.

Comparison to Prior Work

The findings from this review showed that home-based VR intervention produces great variability in effectiveness on different outcomes. On balance, both [51,52] found a significant effect on balance in the VR group when using the BBS. This finding aligns with a recent review [59], which also found a significant improvement in BBS scores following VR intervention in older adults with balance impairments. Similarly, another study [60] found that VR intervention significantly improved BBS scores in older adults residing in nursing homes compared with those living in communities. The populations studied in these studies were similar, indicating that these findings may be generalizable.

For strength outcomes, the CS test did not show significant improvements in 4 of the included studies. In the study by Vestergaard et al [48], the lack of improvement may be due to insufficient progression in the exercises included in the VR training group. Without adequate progression or intensity, participants may not have experienced meaningful gains in strength or functional mobility. However, this remains a possible explanation rather than a definitive conclusion. In the study by Gschwind et al [50], although the Step Mat Training (SMT) group demonstrated significant improvements in sit-to-stand times compared to the control group, no significant differences were found between the 2 intervention groups or between the intervention group and the control group. This limited improvement could suggest that Kinect training may not have been as effective in enhancing strength and balance. The intensity and specificity of exercises delivered through VR interventions are likely contributing factors. While some studies may have used tailored and progressively challenging tasks to enhance training efficacy, others may have implemented more

XSL•FO
RenderX

generic routines that did not adequately target the specific mobility deficits of participants.

In contrast to mobility outcomes, home-based VR training's effect on the TUG showed variation in the included studies. The lack of significant improvement could be attributed to the small sample size [49,50], which limited the ability to detect changes. This variation in effect on TUG could also be due to the specific design and characteristics of the VR interventions implemented in these studies. For instance, 1 study [49] used short-duration interventions of 10‐20 minutes per session, while another [51] used longer-duration interventions of 30 minutes per session. Similarly, the VR effect on knee extension strength did not show significant improvements in the 2 studies [49,50]. In the study by Schoene et al [49], the intervention did not specifically target knee extension strength or range of motion, focusing more on stepping performance and cognitive parameters related to fall risk. In contrast, Gschwind et al [50] found no significant differences between the 2 intervention groups or between the SMT group and the control group. This could be due to the SMT program focusing on proprioception, cognitive processing, and balance, which may have shifted the emphasis away from targeting strength. In contrast, the Kinect program may have placed more emphasis on strength-building exercises [50].

The current meta-analysis clearly highlights 2 aspects: the frail and prefrail status of participants and the home-based delivery of VR training. The importance of targeting older adults with frailty or prefrailty originates from the increased vulnerability among this demographic to functional decline and falls, emphasizing the need for tailored rehabilitation interventions [61,62]. Previous systematic reviews and trials examining VR interventions in rehabilitation centers among older populations, rather than specifically individuals with frailty or prefrailty, have reported varied findings. For instance, a recent systematic review [59] showed significant improvements in balance, strength, and mobility among older adults participating in center-based VR programs. Similarly, an RCT [60] delivered in nursing homes showed improvements in balance outcomes, specifically on the BBS, following structured VR interventions. However, such site-specific studies often comprise higher levels of supervision and formal instruction from a therapist, which are not necessarily available in home-based settings. This key distinction highlights the importance of evaluating home-based VR effectiveness independently. Notably, this review found challenges in standardizing intervention intensity and ensuring adherence and safety without direct professional oversight. This is in line with recent findings [63], which reported technology use challenges among older adults without supervision, potentially limiting the effectiveness of home-based VR interventions. Furthermore, studies conducted among older adults often report increased baseline functional ability compared to populations with frailty and prefrailty, possibly overestimating intervention effectiveness. For instance, the study by Donath et al. [64] reported significant improvements in functional performance following a supervised VR training program in healthy older adults. In contrast, our review's targeted populations with frailty and prefrailty showed smaller effect sizes in mobility outcomes (TUG), possibly due to lower initial functional status.

Overall, our study revealed that VR exercises significantly enhanced balance among older adults compared to those who engaged in regular exercises or remained inactive. However, there were no clear and significant differences in strength and functional mobility. This suggests that while VR training holds potential, its design and implementation require careful consideration to ensure progressive difficulty and proper evaluation through controlled studies.

Limitations

First, only studies published in English were included, raising the possibility of language bias. This might have caused the exclusion of related studies. While this decision was made to ensure precise analysis of findings, future reviews could compensate for this limitation by using translation resources or multilingual reviewers.

Second, there was considerable heterogeneity in intervention design across the included studies. Variations in VR training, session duration, frequency, and outcome measures made it difficult to directly compare results. The heterogeneity in study protocols could have been a factor in the inconsistent findings. Future studies are advised to design standardized intervention protocols to mitigate these differences.

Third, the small sample sizes in multiple studies limited the statistical power to identify significant outcomes. While we used rigorous quality assessment (eg, PEDro scale) to identify high-quality studies, the small sample size may have resulted in underpowered analyses. Future studies should include larger, adequately powered samples to strengthen the validity of the outcomes.

Fourth, this review excluded gray literature and unpublished studies, which may have introduced publication bias. Although we conducted a comprehensive search across multiple databases to mitigate this risk, future reviews should consider searching clinical trial registries and preprint servers to identify ongoing or unpublished work.

Future Directions

We should aim to standardize VR protocols. Future studies should develop structured VR interventions with gradual difficulty and standardized outcomes to enhance consistency and comparability across trials.

We should aim to assess long-term effects. Future studies should assess adherence, sustainability of outcomes, and cost-effectiveness over extended periods.

We should aim to enhance engagement and adherence. VR's interactive nature offers potential benefits beyond physical improvement, including motivation and adherence.

We should aim to explore clinical adoption. Future research should investigate health care providers' perceptions, barriers, and facilitators to implementing VR-based training in routine geriatric care.

XSL•FO
RenderX

Conclusion

This systematic review aimed to evaluate the effectiveness of home-based VR training in improving balance, strength, and mobility among older adults with frailty and prefrailty. The findings suggest that VR interventions are consistently effective in improving balance but show limited evidence for improving strength and mobility. These outcomes were influenced by variability in intervention design, duration, and intensity across studies.

Acknowledgments

We acknowledge the use of the generative artificial intelligence tool ChatGPT by OpenAI, which was used to assist in refining the language of early manuscript drafts. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Authors' Contributions

HA handled the conceptualization (lead), methodology (lead), formal analysis (lead), writing of the original draft (lead), and review and editing of the writing (equal). EA, LB, GK, and YA worked on the data curation (equal), investigation (equal), and review and editing of the writing (equal). AA and MAA assisted with the supervision (equal), validation (equal), and review and editing of the writing (equal).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Detailed overview of the search terms and strategies used.
[DOCX File, 20 KB - games_v13i1e67146_app1.docx]

Multimedia Appendix 2
Search terms used and description of VR. VR: virtual reality.
[DOCX File, 16 KB - games_v13i1e67146_app2.docx]

Checklist 1
PRISMA item checklist. PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.
[DOCX File, 22 KB - games_v13i1e67146_app3.docx]

References

1. Crocker TF, Brown L, Clegg A, et al. Quality of life is substantially worse for community-dwelling older people living with frailty: systematic review and meta-analysis. Qual Life Res 2019 Aug;28(8):2041-2056. [doi: 10.1007/s11136-019-02149-1] [Medline: 30875008]
2. Chang CH, Wei CC, Lien WC, et al. The usability and effect of a novel intelligent rehabilitation exergame system on quality of life in frail older adults: prospective cohort study. JMIR Serious Games 2025 Jan 21;13:e50669. [doi: 10.2196/50669] [Medline: 39841584]
3. Fried LP, Tangen CM, Walston J, et al. Frailty in older adults: evidence for a phenotype. J Gerontol A Biol Sci Med Sci 2001 Mar;56(3):M146-M156. [doi: 10.1093/gerona/56.3.m146] [Medline: 11253156]
4. Collard RM, Boter H, Schoevers RA, Oude Voshaar RC. Prevalence of frailty in community-dwelling older persons: a systematic review. J Am Geriatr Soc 2012 Aug;60(8):1487-1492. [doi: 10.1111/j.1532-5415.2012.04054.x] [Medline: 22881367]
5. Walston J, Hadley EC, Ferrucci L, et al. Research agenda for frailty in older adults: toward a better understanding of physiology and etiology: summary from the American Geriatrics Society/National Institute on Aging Research Conference on Frailty in Older Adults. J Am Geriatr Soc 2006 Jun;54(6):991-1001. [doi: 10.1111/j.1532-5415.2006.00745.x] [Medline: 16776798]
6. Davidson SL, Lee J, Emmence L, et al. Systematic review and meta-analysis of the prevalence of frailty and pre-frailty amongst older hospital inpatients in low- and middle-income countries. Age Ageing 2025 Jan 6;54(1):afae279. [doi: 10.1093/ageing/afae279] [Medline: 39757939]
7. Polidori MC, Ferrucci L. Frailty from conceptualization to action: the biopsychosocial model of frailty and resilience. Aging Clin Exp Res 2023 Apr;35(4):725-727. [doi: 10.1007/s40520-022-02337-z] [Medline: 36944846]
8. Davies K, Maharani A, Chandola T, O'Neill TW, Todd C, Pendleton N. A prospective analysis examining frailty remission and the association with future falls risk in older adults in England. Age Ageing 2023 Feb 1;52(2):afad003. [doi: 10.1093/ageing/afad003] [Medline: 36821643]

XSL•FO
RenderX

9. Vaishya R, Vaish A. Falls in older adults are serious. Indian J Orthop 2020 Feb;54(1):69-74. [doi: 10.1007/s43465-019-00037-x] [Medline: 32257019]
10. Montero-Odasso M, van der Velde N, Martin FC, et al. World guidelines for falls prevention and management for older adults: a global initiative. Age Ageing 2022 Sep 2;51(9):afac205. [doi: 10.1093/ageing/afac205] [Medline: 36178003]
11. Knight T, Kamwa V, Atkin C, et al. Acute care models for older people living with frailty: a systematic review and taxonomy. BMC Geriatr 2023 Dec 5;23(1):809. [doi: 10.1186/s12877-023-04373-4] [Medline: 38053044]
12. Zak M, Sikorski T, Wasik M, Courteix D, Dutheil F, Brola W. Frailty syndrome-fall risk and rehabilitation management aided by virtual reality (VR) technology solutions: a narrative review of the current literature. Int J Environ Res Public Health 2022 Mar 3;19(5):2985. [doi: 10.3390/ijerph19052985] [Medline: 35270677]
13. Nagaraju SP, Shenoy SV, Gupta A. Frailty in end stage renal disease: current perspectives. Nefrologia (Engl Ed) 2022;42(5):531-539. [doi: 10.1016/j.nefroe.2021.05.008] [Medline: 36792307]
14. Checa-López M, Oviedo-Briones M, Pardo-Gómez A, et al. FRAILTOOLS study protocol: a comprehensive validation of frailty assessment tools to screen and diagnose frailty in different clinical and social settings and to provide instruments for integrated care in older adults. BMC Geriatr 2019 Mar 18;19(1):86. [doi: 10.1186/s12877-019-1042-1] [Medline: 30885132]
15. Heng MWY, Chan AWD, Man REK, et al. Individual and combined associations of sarcopenia, osteoporosis and obesity with frailty in a multi-ethnic asian older adult population. BMC Geriatr 2023 Dec 5;23(1):802. [doi: 10.1186/s12877-023-04500-1] [Medline: 38053025]
16. O'Caoimh R, Sezgin D, O'Donovan MR, et al. Prevalence of frailty in 62 countries across the world: a systematic review and meta-analysis of population-level studies. Age Ageing 2021 Jan 8;50(1):96-104. [doi: 10.1093/ageing/afaa219] [Medline: 33068107]
17. Rohrmann S. Epidemiology of frailty in older people. In: Advances in Experimental Medicine and Biology: Springer, Cham; 2020:21-27. [doi: 10.1007/978-3-030-33330-0_3] [Medline: 31894543]
18. Mangin D, Lawson J, Risdon C, et al. Association between frailty, chronic conditions and socioeconomic status in community-dwelling older adults attending primary care: a cross-sectional study using practice-based research network data. BMJ Open 2023 Feb 21;13(2):e066269. [doi: 10.1136/bmjopen-2022-066269] [Medline: 36810183]
19. Cowley A, Goldberg SE, Gordon AL, Kerr M, Logan P. Exploring rehabilitation potential in older people living with frailty: a qualitative focus group study. BMC Geriatr 2021 Mar 6;21(1):165. [doi: 10.1186/s12877-021-02107-y] [Medline: 33676401]
20. Alves E, Gonçalves C, Oliveira H, Ribeiro R, Fonseca C. Health-related outcomes of structured home-based rehabilitation programs among older adults: a systematic literature review. Heliyon 2024 Aug 15;10(15):e35351. [doi: 10.1016/j.heliyon.2024.e35351] [Medline: 39170553]
21. Chambel G, Tinga AM, Huybens DS, Dias G, Pinto JF. Exploring the impact of a gamified exercise platform to support healthy ageing: home-based study with older adults. Presented at: 2023 IEEE 11th International Conference on Serious Games and Applications for Health (SeGAH); Aug 28-30, 2023; Athens, Greece p. 1-7. [doi: 10.1109/SeGAH57547.2023.10253760]
22. Hurst C, Dismore L, Granic A, et al. Attitudes and barriers to resistance exercise training for older adults living with multiple long-term conditions, frailty, and a recent deterioration in health: qualitative findings from the Lifestyle in Later Life - Older People's Medicine (LiLL-OPM) study. BMC Geriatr 2023 Nov 24;23(1):772. [doi: 10.1186/s12877-023-04461-5] [Medline: 38001414]
23. Rúa-Alonso M, Bovolini A, Costa-Brito AR, et al. Exploring perceived barriers to physical activity among older adults living in low-population density regions: gender differences and associations with activity dimensions. Healthcare (Basel) 2023 Nov 11;11(22):2948. [doi: 10.3390/healthcare11222948] [Medline: 37998440]
24. Kwan RYC, Liu J, Sin OSK, et al. Effects of virtual reality motor-cognitive training for older people with cognitive frailty: multicentered randomized controlled trial. J Med Internet Res 2024 Sep 11;26:e57809. [doi: 10.2196/57809] [Medline: 39259959]
25. Chen J, Or CK, Li Z, Yeung EHK, Chen T. Perceptions of patients with stroke regarding an immersive virtual reality-based exercise system for upper limb rehabilitation: questionnaire and interview study. JMIR Serious Games 2025 Jan 1;13:e49847. [doi: 10.2196/49847] [Medline: 39742513]
26. Béraud-Peigné N, Perrot A, Maillot P. Active video games training for older adults: comparative study of user experience, workload, pleasure, and intensity. JMIR Serious Games 2025 May 26;13:e67314. [doi: 10.2196/67314] [Medline: 40418855]
27. Kaiser W, de Bruin ED, Manser P. Domain-specific evaluation of exergame metrics among older adults with mild neurocognitive disorder: secondary analysis of 2 randomized controlled trials. JMIR Serious Games 2025 May 21;13:e65878. [doi: 10.2196/65878] [Medline: 40397948]
28. Mo N, Feng JY, Liu HX, Chen XY, Zhang H, Zeng H. Effects of exergaming on musculoskeletal pain in older adults: systematic review and meta-analysis. JMIR Serious Games 2023 Apr 25;11:e42944. [doi: 10.2196/42944] [Medline: 37097717]

XSL•FO
RenderX

29. Everard G, Declerck L, Lejeune T, et al. A self-adaptive serious game to improve motor learning among older adults in immersive virtual reality: short-term longitudinal pre-post study on retention and transfer. JMIR Aging 2025 Mar 3;8:e64004. [doi: 10.2196/64004] [Medline: 40053708]
30. Rico-Olarte C, Lopez DM, Eskofier BM, Becker L. Electrophysiological insights in exergaming-electroencephalography data recording and movement artifact detection: systematic review. JMIR Serious Games 2025 Apr 7;13:e50992. [doi: 10.2196/50992] [Medline: 40194274]
31. Peebles AT, van der Veen S, Stamenkovic A, France CR, Pidcoe PE, Thomas JS. A virtual reality game suite for graded rehabilitation in patients with low back pain and a high fear of movement: within-subject comparative study. JMIR Serious Games 2022 Mar 23;10(1):e32027. [doi: 10.2196/32027] [Medline: 35319471]
32. Maheta B, Kraft A, Interrante N, et al. Using virtual reality to improve outcomes related to quality of life among older adults with serious illnesses: systematic review of randomized controlled trials. J Med Internet Res 2025 Feb 26;27:e54452. [doi: 10.2196/54452] [Medline: 40009834]
33. Brazil CK, Rys MJ. The effect of VR on fine motor performance by older adults: a comparison between real and virtual tasks. Virtual Real 2024;28(2):113. [doi: 10.1007/s10055-024-01009-9]
34. Wang YH. Understanding senior adults' needs, preferences, and experiences of commercial exergames for health: usability study. JMIR Serious Games 2024 Apr 5;12:e36154. [doi: 10.2196/36154] [Medline: 38578674]
35. Du Q, Wei Y, Ma Y, et al. Efficacy of virtual reality-based interventions on cognitive function in patients with neuropsychiatric disorders: systematic review and meta-analysis of randomized controlled trials. JMIR Serious Games 2025 May 8;13:e67501. [doi: 10.2196/67501] [Medline: 40341171]
36. Yu D, Li X, Lai FHY. The effect of virtual reality on executive function in older adults with mild cognitive impairment: a systematic review and meta-analysis. Aging Ment Health 2023 Apr;27(4):663-673. [doi: 10.1080/13607863.2022.2076202] [Medline: 35635486]
37. Wong AKC, Zhang MQ, Bayuo J, et al. The effect of young people-assisted, individualized, motion-based video games on physical, cognitive, and social frailty among community-dwelling older adults with frailty: randomized controlled trial. JMIR Serious Games 2024 Nov 20;12:e57352. [doi: 10.2196/57352] [Medline: 39622701]
38. Zak M, Sikorski T, Krupnik S, et al. Physiotherapy programmes aided by VR solutions applied to the seniors affected by functional capacity impairment: randomised controlled trial. Int J Environ Res Public Health 2022 May 15;19(10):6018. [doi: 10.3390/ijerph19106018] [Medline: 35627554]
39. Zanotto T, Galperin I, Kumar DP, et al. Effects of a 6-week treadmill training with and without virtual reality on frailty in people with multiple sclerosis. Arch Phys Med Rehabil 2025 Feb;106(2):187-194. [doi: 10.1016/j.apmr.2024.09.010] [Medline: 39341443]
40. Alhasan H, Alshehri MA, Wheeler PC, Fong DTP. Effects of interactive videogames on postural control and risk of fall outcomes in frail and pre-frail older adults: a systematic review and meta-analysis. Games Health J 2021 Apr;10(2):83-94. [doi: 10.1089/g4h.2020.0009] [Medline: 33651955]
41. Lee YH, Lin CH, Wu WR, Chiu HY, Huang HC. Virtual reality exercise programs ameliorate frailty and fall risks in older adults: a meta-analysis. J Am Geriatr Soc 2023 Sep;71(9):2946-2955. [doi: 10.1111/jgs.18398] [Medline: 37165743]
42. Daly RM, Gianoudis J, Hall T, Mundell NL, Maddison R. Feasibility, usability, and enjoyment of a home-based exercise program delivered via an exercise app for musculoskeletal health in community-dwelling older adults: short-term prospective pilot study. JMIR mHealth uHealth 2021 Jan 13;9(1):e21094. [doi: 10.2196/21094] [Medline: 33439147]
43. Meredith SJ, Cox NJ, Ibrahim K, et al. Factors that influence older adults' participation in physical activity: a systematic review of qualitative studies. Age Ageing 2023 Aug 1;52(8):afad145. [doi: 10.1093/ageing/afad145] [Medline: 37595070]
44. Abdallah L, Stolee P, Lopez KJ, Whate A, Boger J, Tong C. The impact of COVID-19 on older adults' perceptions of virtual care: qualitative study. JMIR Aging 2022 Oct 20;5(4):e38546. [doi: 10.2196/38546] [Medline: 36054599]
45. Healy D, Flynn A, Conlan O, McSharry J, Walsh J. Older adults' experiences and perceptions of immersive virtual reality: systematic review and thematic synthesis. JMIR Serious Games 2022 Dec 6;10(4):e35802. [doi: 10.2196/35802] [Medline: 36472894]
46. Li Y, Wilke C, Shiyanov I, Muschalla B. Impact of virtual reality-based group activities on activity level and well-being among older adults in nursing homes: longitudinal exploratory study. JMIR Serious Games 2024 Mar 29;12:e50796. [doi: 10.2196/50796] [Medline: 38551635]
47. Virtual reality as a training for frail or pre-frail older adults to improve strength, postural control, and reduce the risk of falls. PROSPERO. URL: https://www.crd.york.ac.uk/PROSPERO/view/CRD42023478330 [accessed 2025-06-21]
48. Vestergaard S, Kronborg C, Puggaard L. Home-based video exercise intervention for community-dwelling frail older women: a randomized controlled trial. Aging Clin Exp Res 2008 Oct;20(5):479-486. [doi: 10.1007/BF03325155] [Medline: 19039291]
49. Schoene D, Lord SR, Delbaere K, Severino C, Davies TA, Smith ST. A randomized controlled pilot study of home-based step training in older people using videogame technology. PLoS ONE 2013;8(3):e57734. [doi: 10.1371/journal.pone.0057734] [Medline: 23472104]

XSL•FO
RenderX

50. Gschwind YJ, Schoene D, Lord SR, et al. The effect of sensor-based exercise at home on functional performance associated with fall risk in older people - a comparison of two exergame interventions. Eur Rev Aging Phys Act 2015;12:11. [doi: 10.1186/s11556-015-0156-5] [Medline: 26865875]
51. Karahan AY, Tok F, Taşkın H, Kuçuksaraç S, Başaran A, Yıldırım P. Effects of exergames on balance, functional mobility, and quality of life of geriatrics versus home exercise programme: randomized controlled study. Cent Eur J Public Health 2015 Nov;23 Suppl:S14-S18. [doi: 10.21101/cejph.a4081] [Medline: 26849537]
52. Yeşilyaprak SS, Yıldırım M, Tomruk M, Ertekin Ö, Algun ZC. Comparison of the effects of virtual reality-based balance exercises and conventional exercises on balance and fall risk in older adults living in nursing homes in Turkey. Physiother Theory Pract 2016;32(3):191-201. [doi: 10.3109/09593985.2015.1138009] [Medline: 27049879]
53. Geraedts HAE, Dijkstra H, Zhang W, et al. Effectiveness of an individually tailored home-based exercise rogramme for pre-frail older adults, driven by a tablet application and mobility monitoring: a pilot study. Eur Rev Aging Phys Act 2021 Jun 21;18(1):10. [doi: 10.1186/s11556-021-00264-y] [Medline: 34154524]
54. Moseley AM, Herbert RD, Sherrington C, Maher CG. Evidence for physiotherapy practice: a survey of the Physiotherapy Evidence Database (PEDro). Aust J Physiother 2002;48(1):43-49. [doi: 10.1016/s0004-9514(14)60281-6] [Medline: 11869164]
55. Herd CR, Meserve BB. A systematic review of the effectiveness of manipulative therapy in treating lateral epicondylalgia. J Man Manip Ther 2008;16(4):225-237. [doi: 10.1179/106698108790818288] [Medline: 19771195]
56. Thompson SG, Higgins JPT. How should meta-regression analyses be undertaken and interpreted? Stat Med 2002 Jun 15;21(11):1559-1573. [doi: 10.1002/sim.1187] [Medline: 12111920]
57. Fritz CO, Morris PE, Richler JJ. Effect size estimates: current use, calculations, and interpretation. J Exp Psychol Gen 2012 Feb;141(1):2-18. [doi: 10.1037/a0024338] [Medline: 21823805]
58. Kouvonen A, Kemppainen T, Taipale S, Olakivi A, Wrede S, Kemppainen L. Health and self-perceived barriers to internet use among older migrants: a population-based study. BMC Public Health 2022 Mar 23;22(1):574. [doi: 10.1186/s12889-022-12874-x] [Medline: 35321678]
59. Chen PJ, Hsu HF, Chen KM, Belcastro F. VR exergame interventions among older adults living in long-term care facilities: a systematic review with meta-analysis. Ann Phys Rehabil Med 2023 Apr;66(3):101702. [doi: 10.1016/j.rehab.2022.101702] [Medline: 36028201]
60. Ren Y, Lin C, Zhou Q, Yingyuan Z, Wang G, Lu A. Effectiveness of virtual reality games in improving physical function, balance and reducing falls in balance-impaired older adults: a systematic review and meta-analysis. Arch Gerontol Geriatr 2023 May;108:104924. [doi: 10.1016/j.archger.2023.104924] [Medline: 36680968]
61. Kannan L, Sahu U, Subramaniam S, et al. Gaming-based tele-exercise program to improve physical function in frail older adults: feasibility randomized controlled trial. J Med Internet Res 2024 Nov 27;26:e56810. [doi: 10.2196/56810] [Medline: 39602215]
62. Randriambelonoro M, Franck CP, Herrmann F, et al. Gamified physical rehabilitation for older adults with musculoskeletal issues: pilot noninferiority randomized clinical trial. JMIR Rehabil Assist Technol 2023 Mar 6;10:e39543. [doi: 10.2196/39543] [Medline: 36877563]
63. Lee DR, Lo JC, Ramalingam N, Gordon NP. Understanding the uptake of digital technologies for health-related purposes in frail older adults. J Am Geriatr Soc 2021 Jan;69(1):269-272. [doi: 10.1111/jgs.16841] [Medline: 33011975]
64. Donath L, Rössler R, Faude O. Effects of virtual reality training (exergaming) compared to alternative exercise training and passive control on standing balance and functional mobility in healthy community-dwelling seniors: a meta-analytical review. Sports Med 2016 Sep;46(9):1293-1309. [doi: 10.1007/s40279-016-0485-1] [Medline: 26886474]

Abbreviations

BBS: Berg Balance Scale
CS: Chair Stand
MD: mean difference
PEDro: Physiotherapy Evidence Database
PICOS: Population, Intervention, Comparison, Outcome, and Study Design
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analysis
RCT: randomized controlled trial
SMT: Step Mat Training
TUG: Timed Up and Go
VR: virtual reality

XSL•FO
RenderX

Edited by A Coristine; submitted 04.10.24; peer-reviewed by G Torkaman, PE Roos; revised version received 02.06.25; accepted 02.06.25; published 18.07.25.

Please cite as:
Alhasan H, Alandijani E, Bahamdan L, Khudary G, Aburaya Y, Awali A, Alshehri MA
Home-Based Virtual Reality Training for Enhanced Balance, Strength, and Mobility Among Older Adults With Frailty: Systematic Review and Meta-Analysis
JMIR Serious Games 2025;13:e67146
URL: https://games.jmir.org/2025/1/e67146
doi:10.2196/67146

© Hammad Alhasan, Elaf Alandijani, Lara Bahamdan, Ghofran Khudary, Yara Aburaya, Abdulaziz Awali, Mansour Abdullah Alshehri. Originally published in JMIR Serious Games (https://games.jmir.org), 18.7.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Effectiveness of Gamification Interventions to Improve Physical Activity and Sedentary Behavior in Children and Adolescents: Systematic Review and Meta-Analysis

Min Wang[1], MEd; Jisheng Xu[2], MEd; Xulin Zhou[2], BEd; Xingchen Li[3], MEd; Yu Zheng[1], PhD

[1]School of Economics and Management, Chengdu Sport University, 1942 Huanhu North Road, Eastern New Area, Chengdu, China

[2]School of Sports Medicine and Health, Chengdu Sport University, Chengdu, China

[3]School of Physical Education, Sichuan Normal University, Chengdu, China

Corresponding Author:
Yu Zheng, PhD
School of Economics and Management, Chengdu Sport University, 1942 Huanhu North Road, Eastern New Area, Chengdu, China

Abstract

Background: Physical activity (PA) is critically linked to the health outcomes of children and adolescents. Gamification interventions represent a promising approach to promote PA engagement. However, the effects of these interventions on both PA and sedentary behavior (SB) in this population remain controversial. This review seeks to clarify this controversy.

Objective: This systematic review aimed to evaluate the effectiveness of gamification interventions in enhancing PA and reducing SB in children and adolescents, while identifying potential moderators for PA promotion.

Methods: We systematically searched PubMed, Web of Science, Embase, EBSCO, and Cochrane Library databases for randomized controlled trials (RCTs) published between January 1, 2010, and August 1, 2024. Included RCTs examined gamification interventions targeting PA, SB, daily step counts, and BMI in children and adolescents. Random-effects meta-analyses were performed using RevMan 5.4 (Cochrane) and Stata 18.0 (StataCorp), with subgroup analyses assessing moderating effects of theoretical paradigms, game elements, and intervention duration. Methodological robustness was evaluated via the Egger regression test, sensitivity analyses (leave-one-out method), and funnel plot inspection for publication bias.

Results: A total of 16 RCTs involving 7472 children and adolescents (age range 6 - 18 y) were included. Our findings showed that the gamification interventions significantly increased moderate-to-vigorous physical activity (MVPA; standardized mean difference [SMD] 0.15, 95% CI 0.01 to 0.29; *P*=.04) and reduced BMI (SMD 0.11, 95% CI 0.05 to 0.18; *P*<.001). However, there was no significant improvement in SB (SMD 0.07, 95% CI −0.07 to 0.22; *P*=.33), vigorous physical activity (SMD 0.12, 95% CI -0.3 to 0.55; *P*=.56), moderate physical activity (SMD 0.16, 95% CI −0.2 to 0.53; *P*=.38), light physical activity (SMD −0.00, 95% CI −0.49 to 0.48; *P*>.99), and daily step count (SMD 0.22, 95% CI −0.51 to 0.94; *P*=.55). Subgroup analyses revealed significant moderation effects for MVPA improvement by theoretical paradigm, game elements, intervention duration, and study setting.

Conclusions: This meta-analysis confirms that gamification interventions effectively increased MVPA in children and adolescents, with sustained effects persisting beyond follow-up. The efficacy of these interventions is significantly moderated by theoretical paradigms, game elements, and intervention duration. However, blinding infeasibility contributed to prevalent performance bias, potentially introducing detection bias for subjective SB and PA metrics. Future research should strengthen blinding protocols for outcome assessors, enhance allocation concealment reporting, and validate conclusions through high-quality RCTs.

Trial Registration: PROSPERO CRD42024580995; https://www.crd.york.ac.uk/PROSPERO/view/CRD42024580995

(JMIR Serious Games 2025;13:e68151) doi:10.2196/68151

KEYWORDS

gamification; physical activity; behavior change; eHealth; health behavior; intervention; mobile phone; meta-analysis; systematic review; PRISMA

XSL•FO
RenderX

Introduction

Background

Regular physical activity (PA) is associated with numerous health benefits. However, most children and adolescents fail to meet global physical activity guidelines [1-3]. The World Health Organization (WHO) recommends that children and adolescents accumulate ≥60 minutes of daily moderate-to-vigorous physical activity (MVPA) [4]. Yet global surveillance indicates only 19.3% of 11‐17-year-olds meet this target, with average MVPA levels declining to 43.7 minutes/day in high-income countries [5]. Approximately 324 million adolescents (18% of the global adolescent population) live with overweight or obesity worldwide. Furthermore, insufficient PA contributes to rising rates of poor health outcomes and chronic diseases [6-9]. Conversely, regular PA can avert the risk of chronic diseases [10,11] and reduce the development of noncommunicable risk factors [12,13]. Epidemiological studies demonstrate a "dose-response" relationship between sedentary behavior (SB) and mortality [14]. High-level PA has been shown to attenuate or even eliminate the deleterious effects of SB on early adolescent mortality [14]. In this context, developing interventions to enhance PA and reduce SB represents a critical public health imperative. Given this pervasive deficit, gamification emerges as a targeted strategy to bridge the gap between adolescents' digital behaviors and health needs.

Gamification based on electronic devices and immersive technologies is considered a promising intervention for changing health behaviors [15,16]. Leveraging adolescents' fondness for digital products (with 95% of individuals in this age group using smartphones daily [17]), gamification strategies aim to transform PA into a goal-driven and rewarding experience. Defined as "the application of game-design elements in non-game contexts" [16], this approach incorporates gamification elements (eg, rewards, feedback, and social interactions) to create game-inspired experiences that enhance PA engagement [16].

However, current research predominantly focuses on adults and chronic conditions (eg, rheumatic diseases and cancer survivors) [18-22], where efficacy is established. In contrast, gamification's impact on children and adolescents remains understudied. While existing literature examines gamification's impact on PA or SB in healthy adolescents, intervention outcomes remain inconsistent [23,24]. Direito et al [25] suggest that underlying factors may influence gamification interventions for PA, including individual characteristics, environmental facilitators, and implementation fidelity. Although youth studies show limited total physical activity (TPA) or MVPA effects [26-28], these findings likely reflect methodological constraints rather than inherent limitations of PA interventions. Systematic reviews emphasize scarce high-quality studies and the need for rigorous randomized controlled trials (RCTs) to isolate gamification-specific effects [23,24,29]. Paradoxically, when using stringent experimental controls, gamification's impact may attenuate [24]. Consequently, evidence regarding gamification's effectiveness in promoting PA and reducing SB remains inconclusive, highlighting the need for meta-analyses of high-quality RCTs. Furthermore, gamification interventions based on theoretical paradigms may be more conducive to increasing PA in children and adolescents [30]. Several theoretical paradigms have been used for gamification interventions [31], including self-determination theory (SDT), transtheoretical models (TTMs), behavior change techniques (BCTs), and social cognitive theory (SCT). Different factors, such as game elements and theoretical paradigms, are crucial in determining the effectiveness of gamification interventions. Therefore, systematic evaluation is essential.

This Study

This systematic review aims to assess the effectiveness of gamification interventions in promoting PA and reducing SB in children and adolescents. Specifically, (1) the aim is to evaluate the efficacy of a gamification intervention in improving MVPA, vigorous physical activity (VPA), moderate physical activity (MPA), light physical activity (LPA), daily step count, BMI, and reducing SB in children and adolescents. (2) The aim is to examine the effectiveness of gamification interventions across diverse subgroups, considering the moderating effects of variables, such as theoretical framework, game elements, and intervention cycle. In addition, unlike existing reviews, this review will assess the effectiveness of gamification interventions in promoting PA and reducing SB in children and adolescents based on rigorous research components derived from RCTs. Subgroup analyses will examine the influence of crucial moderating variables, including theoretical framework, game elements, and intervention duration, on the effectiveness of gamification interventions. This approach will improve our understanding of the factors that influence the effectiveness of these interventions and facilitate the accurate quantification of intervention programs.

Methods

Registration

This study was conducted in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [32,33]. The detailed PRISMA checklist is provided in Checklist 1. The systematic review was registered on PROSPERO (International Prospective Register of Systematic Reviews; CRD42024580995).

Search Strategy and Information Sources

MW and XL conducted independent literature searches according to the PRISMA 2020 statement. A comprehensive search of multiple databases was performed, including PubMed, Web of Science, Ebsco, Embase, and the Cochrane Library. The search period was limited to January 1, 2010 (2010 being the date of the widespread adoption of the term "gamification" [34]), and August 1, 2024. A literature search strategy was developed by 2 investigators using Boolean operators ("AND" and "OR") to combine subject terms with open-access terms. The search terms used were "child" or "adolescent," "gamification," "exergaming," "exert-gaming," "gamified," "gameful," and "physical activity," "PA," "MVPA," and "sedentary behavior." The search terms "exergaming," "exergaming," "gamified," and "graceful" were combined with the search terms "physical activity," "PA," "MVPA," "sedentary

XSL•FO
RenderX

behavior," "SB," and "sedentary behavior." In addition, the search terms "sedentary lifestyle," "sedentary behavior," and "sedentary behavior" were included. A detailed description of the PubMed search strategy can be found in Multimedia Appendix 1. In addition, we supplemented our search results with references cited in previous reviews.

Inclusion and Exclusion Criteria

Inclusion Criteria

The PICOS (Population, Intervention, Comparison, Outcome, and Study Design) principles developed the screening criteria, and the inclusion criteria for this study were as follows: (1) The study comprised children and adolescents, aged 6 to 18 years, with no gender or health status restrictions. Participants with chronic conditions (eg, obesity, diabetes) were included if their condition permitted safe physical activity participation under medical guidance. Individuals were excluded only if they had physical impairments contraindicating PA (eg, uncontrolled cardiomyopathy), had intellectual or cognitive disabilities affecting intervention adherence, and if they had medical conditions precluding safe exercise (eg, acute fractures). (2) Gamification interventions systematically apply game design elements (eg, points, leaderboards, and badges) within nongame contexts to enhance PA participation or reduce SB in pediatric populations. This approach explicitly differs from serious games (eg, Pokémon Go)—defined as purpose-built digital experiences with nonentertainment objectives such as health education or skill training [35]. Whereas serious games create immersive online environments, gamification leverages motivational mechanics embedded in existing activities, exemplified by wearable-tracked PA metrics synced with mobile apps to trigger reward systems [36]. To resolve definitional ambiguities highlighted in the literature, 3 researchers (MW, JX, and XL) established consensus criteria through the Delphi methodology. (3) Including regular exercise methods, a control group, a no-intervention group, and gamified interventions. (4) Primary outcome measures included MVPA, VPA, MPA, LPA, BMI, SB, and daily steps. These outcomes were obtained through objective measurements using accelerometers, pedometers, or smartphone apps. Due to the increased risk of data bias associated with subjective measures obtained via self-report questionnaires, objective data are consistently preferred over self-report questionnaires in analyses [37]. Consequently, outcome measures derived from self-report questionnaires were excluded from this study. (5) The study design was an RCT.

Exclusion Criteria

The exclusion criteria were as follows: (1) unavailability of original articles, review studies, conference abstracts, repetitive publications, animal experiments, and nonrandomized controls; (2) incomplete data and fruitless attempts to contact the authors of the articles; (3) non–English-language literature; and (4) literature in which the outcome metrics were not related to PA and SB. Textbox 1 shows the summary of the inclusion and exclusion criteria.

Textbox 1. Summary of the inclusion and exclusion criteria.

Inclusion criteria:

- Participants: Children and adolescents (ages 1-18 y), regardless of gender or health status, including those with chronic conditions, overweight, obesity, or other health issues not preventing participation.
- Intervention: Gamification-based interventions incorporating at least one game design element (eg, points, leaderboards, and badges), delivered through wearable devices, mobile apps, or related platforms.
- Comparator: Usual care, no intervention, waitlist controls, or active nongamified interventions (eg, standard exercise programs).
- Outcomes: Objectively measured MVPA, VPA, MPA, LPA, SB, BMI, and daily steps (eg, measured with accelerometers such as ActiGraph GT3X+, pedometers, or smartphone apps).
- Study design: Full-text English RCTs or cluster RCTs (2010 - 2024).
- Data quality: Complete datasets.

Exclusion criteria:

- Participants: Adults, individuals with physical impairments or medical contraindications to physical activity (eg, uncontrolled cardiomyopathy or acute fractures), and children or adolescents with intellectual disabilities or cognitive impairments that hinder adherence to interventions.
- Intervention: Nongamified digital interventions or purely entertainment-focused games without systematic gamification elements.
- Comparator: Gamified control groups, noncomparative designs.
- Outcomes: Studies reporting only self-reported PA or SB outcomes, or those in which outcomes were unrelated to PA and SB.
- Study design: Nonrandomized trials, conference abstracts, review studies, repetitive publications, animal studies, and non-English literature.
- Data quality: Studies with incomplete datasets where attempts to contact authors were unsuccessful.

Study Selection

The article screening process uses the PRISMA 2020 statement flowchart paradigm to document the studies selected or excluded at each step [33]. The search results were summarized and imported into EndNote 21 literature management software (Clarivate Analytics, Inc), where MW performed the initial screening by sorting and removing duplicates. To ensure the inclusion of original studies that met the established criteria, authors XL and JX independently assessed the title and abstract

XSL•FO
RenderX

of the articles using EndNote 21 software. In addition, 2 researchers subjected all articles potentially meeting the criteria to a full-text review. Any discrepancies arising during the review process will be discussed and resolved with MW.

Data Extraction

Furthermore, 2 authors (XZ and XL) used a standardized data abstraction form to extract the data. Any discrepancies were discussed and resolved with the second author (MW and YZ). The extracted content included the following elements: (1) Basic information (authors, publication year, country, and study design); (2) Population characteristics (sample size and age distribution); (3) Interventions: Gamification elements (points, leaderboards, and badges); (4) Intervention duration (weeks); (5) Control group protocols; (6) Outcome metrics (mean [SD] for MVPA, VPA, MPA, LPA, SB, step counts, and BMI); (7) Theoretical frameworks and measurement instruments (eg, ActiGraph model specifications).

Risk of Bias Assessment

For each eligible trial, 2 reviewers (MW and XZ) used the specially constructed Cochrane Risk of Bias tool (outlined in the Cochrane Handbook for Evaluation of Intervention Systems [38]) to assess the risk of bias. The risk of bias was assessed in 7 domains: sequence generation, allocation concealment, blinding of subjects and personnel, blinding of outcome assessment, incomplete outcome data, selective outcome reporting, and other risks of bias. In this meta-analysis, we comprehensively assessed the potential risk of bias based on the content of the included literature. We categorized the studies as high risk, low risk, or unclear (if relevant information was unavailable). The results of these risk-of-bias assessments were documented in the review and included in the final analysis. To assess interrater agreement, we calculated the κ statistic. After a comprehensive assessment, we determined that the included studies were strikingly similar in design and that blinding was not feasible for this intervention. Consequently, we rated the risk of blinding bias for these studies as high. Given these considerations, the increased risk of bias associated with blinded programs is particularly evident in the risk of bias assessment plot.

Statistical Analysis

A meta-analysis was performed using RevMan 5.4 software (Cochrane) and Stata17 software (StataCorp). This included calculating effect size combinations, performing heterogeneity tests, generating forest plots, performing subgroup analyses, assessing publication bias (using funnel plots and the Egger test [39]), and performing sensitivity analyses. The original study data set included continuous variables. Therefore, the standardized mean difference (SMD) and its 95% CI were selected as the effect measures for the pooled analysis. In addition, PA and SB data, including means and SDs, were extracted independently from each study. For some of the studies that did not provide these data directly, we supplemented them using the estimation method of Hozo et al [39]. The heterogeneity test of this study was performed using I^2 and P values. When I^2 was less than 50%, and P was equal to or greater than .10, the data were analyzed using a fixed-effects model. Conversely, a meta-analysis used a random-effects model when I^2 exceeded 50%, and P was less than .10 [40]. Subgroup analyses will be performed to gain a deeper understanding of the impact of other potential factors on the outcome indicators. The potential for publication bias in the literature was assessed using funnel plots and the Egger test. Publication bias was inferred if the funnel plots showed asymmetry or if $P<.05$ in the Egger test. Otherwise, the possibility of publication bias was considered low. The effect size interpretation criteria were SMD<0.2 for negligible effect sizes, 0.2≤SMD<0.5 for small effect sizes, 0.5≤SMD<0.8 for moderate effect sizes, and SMD≥0.8 for large effect sizes. A Z statistic of $P<.05$ was used to assess the significance of the overall effect.

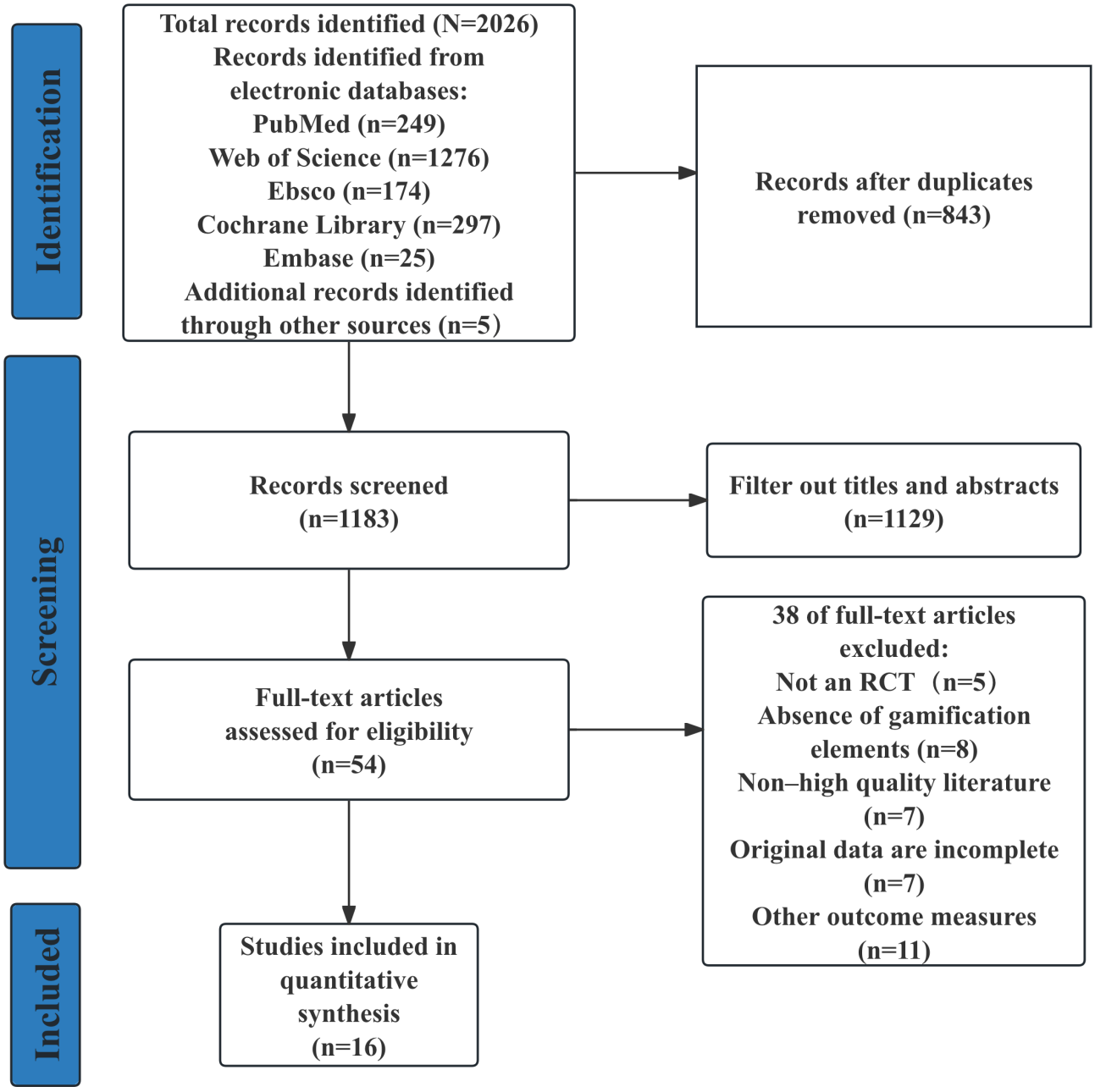
Results

Study Selection

A total of 2021 relevant articles were retrieved from PubMed (n=249), Web of Science (n=1276), EBSCO (n=174), Cochrane Library (n=297), and Embase (n=25), and 5 records were retrieved from the supplementary register. Figure 1 illustrates the study selection workflow. After removing 843 duplicate records using EndNote 21, 1183 unique records underwent title and abstract screening. This process excluded 1129 irrelevant publications, yielding 54 articles for full-text assessment. Following full-text review, 38 articles were excluded due to predefined criteria (eg, non-RCT design and irrelevant outcomes), resulting in 16 eligible studies included for final analysis [41-56].

XSL•FO
RenderX

Figure 1. Flowchart of the study selection. RCT: randomized controlled trial.



Characteristics of the Included Studies

The 16 English-language papers included in this study were published from 2012 to 2024 and contained 7472 subjects, ranging from 13 to 1544 subjects per study [41-56]. The included studies had the following essential characteristics. Participants ranged in age from 1 to 18 years. Study countries included the United States (5/16) [45,50,53-55], Canada (4/16) [46-48,56], New Zealand (2/16) [43,44], Finland (1/16) [49], Chile (1/16) [51], Northern Ireland (1/16) [42], Singapore (1/16) [52], and the United Kingdom (1/16) [41]. A total of 9 studies were tested in school settings [41,42,44-47,51,53,54], 5 in home settings [50,52,53,55,56], and 3 in settings where families and communities worked together [43,48,49]. Furthermore, 12 studies were RCTs [43,45-50,52-56] and 4 studies were cluster RCTs [41,42,44,51]. In addition, 7 studies used belt-worn accelerometers [43-45,50,53-55], 5 studies used wristband activity trackers [41,42,49,51,56], 3 studies used foot sensors as activity trackers [46-48], and 1 study used the MapMyFitness app (Under Armour) on a smartphone, eliminating the need for an additional wearable activity tracker [52]. The intervention period was 1 - 48 weeks, and the follow-up was 24 - 96 weeks. Outcome measures included MVPA, VPA, MPA, LPA, SB, and daily steps. A total of 14 studies used behavior change research theories, including SCT [53,55], SDT [41,42,45-48,52,54], behavior change wheel (BCW) [51], BCT [43], and TTM [49,56]. The trials have been registered in the Clinical Trials Registry [41-51,53-56], and 1 trial is not registered [52]. Furthermore, 14 studies have been funded [41-44,46-51,53-56] and 2 studies did not receive any funding [45,52]. In addition, 1 completed follow-up 48 to 96 weeks (mean 12 weeks) after completing the intervention [44], and 3 studies completed follow-up assessments 12 - 28 weeks after the intervention ended [42,51,55]. In terms of how gamification interventions were implemented, 14 studies used mobile apps

XSL•FO
RenderX

(eg, smartphone apps, iPod touch, Facebook groups, and Fitbit mobile apps) in addition to wearable PA trackers [41-43,45-50,52-56], and 2 studies focused on building school gamification environments [44,51]. The number of game elements used in gamification interventions ranged from 1 to 8, with most including 4 or more game elements. The most used game elements were socialization, interaction, rewards, points, feedback, challenges, and achievements. In addition, intervention durations ranged from 1 week to 96 weeks, with 5 (31.3%) studies having a follow-up period. This suggests that further evaluation of PA gamification is needed to determine the effectiveness and sustainability of game element interventions (Table 1).

XSL•FO
RenderX

Table . Summary of the intervention characteristics of the included studies.

Study	Sample size		Participants (age range in years)	Interventions	Comparator	Theoretical paradigm	Duration (weeks)	Game elements	Main outcomes	Country
	N	Male, n (%)								
Maloney et al [50], 2012	64	53 (82.8)	Adolescents with obesity or overweight (9-17)	Pedometer plus DDR ('Dance Revolution')	Pedometer only	No theory mentioned	12	Interaction	VPA[a], MPA[b], LPA[c], and daily steps	The United States
Garde et al [48], 2015	47	16 (34)	Adolescents (8-13)	"Mobile Kids Monster Manor "was a mobile exergame synchronized with an external activity monitor	Feedback through tractility	SDT[d]	2	Points, rewards, feedback, social interaction, challenges, and achievements	Daily steps	Canada
Garde et al [47], 2016	56	35 (62)	Adolescents (10-13)	"Mobile Kids Monster Manor (MKMM) combines a game with activity monitors."	No intervention	SDT	1	Points, rewards, feedback, social interaction, challenges, and achievements	Daily steps	Canada
Garde et al [46], 2018	37	16 (43)	Adolescents (9-11)	"MobileKids Monster Manor (MKMM) combines a game with activity monitors."	No intervention	SDT	1	Points, rewards, feedback, social interaction, challenges, and achievements	Daily steps	Canada
Fu et al [45], 2018	65	34 (52)	Preschool children (4-6)	"Go Noodle" is an online active video game program that provides children with a variety of PA and exercise choices	30 min active free-play sessions	SDT	12	Social interaction	Daily steps	The United States
Staiano [53], 2018	46	23 (50)	Children with obesity or overweight (10-12)	Exergaming group (Kinect and Xbox 360 gaming console, exergames, telehealth coaching)	No intervention	SCT[e]	24	Social interaction	zBMI[f]	The United States

XSL•FO
RenderX

Study	Sample size		Participants (age range in years)	Interventions	Comparator	Theoretical paradigm	Duration (weeks)	Game elements	Main outcomes	Country
	N	Male, n (%)								
Staiano et al [54], 2016	42	0 (0)	Adolescent girls (14-18)	Dance-based motion game intervention, featuring a structured gaming environment	No intervention	SDT	12	Social interaction	SB[g], LPA, MPA, and VPA	The United States
Pena et al [51], 2021	2320	1398 (60.2)	Students (10-12)	Nutrition and physical activity interventions using gamification strategies	Standard education curriculum	BCW[h]	16 - 28	Points, leaderboards, badges, and rewards	zBMI	Chile
Corepal et al [42], 2019	224	105 (46)	Adolescents (12-14)	Step Smart Challenge (Basic pedometer competition for schools, including gamified strategies)	No intervention	SDT	22 - 52	Feedback and rewards	MVPA[i], LPA, VPA, MPA, SB, and daily steps	Northern Ireland
Leinonen et al [49], 2017	496	496 (100)	Adolescents (17-18)	The intervention group used a gamified application, web-based mobile body and social activation service	No intervention	TTM[j]	24	Challenges, rewards, and competition	MVPA	Finland
Corder et al [41], 2020	2862	1491 (52)	Adolescents (13-14)	Go Active intervention and gained points and rewards	Regular school activities	SDT	40	Point	MVPA	The United Kingdom
Direito et al [43], 2015	51	22 (43)	Adolescents (14-17)	Use an immersive app (Zombies, Run) and a non-immersive app (Get Running).	Regular behavior group	BCTs	8	Points and challenges	MVPA, VPA, MPA, LPA, and SB	New Zealand
Tugault-Lafleure et al [56], 2023	214	104 (48.5)	Adolescents (10-17)	Use the Aim2Be app and get guidance from in-person coaches.	Only accept guidance from regular monitoring	TTM	12	Storytelling and narrative	TPA, zBMI, SB, and daily steps	Canada

XSL•FO
RenderX

Study	Sample size		Participants (age range in years)	Interventions	Comparator	Theoretical paradigm	Duration (weeks)	Game elements	Main outcomes	Country
	N	Male, n (%)								
Seah et al [52], 2020	36	0 (0)	High school girls (14-16)	Use the MapMyFitness (MMF) mobile app to track your workouts and get feedback.	Regular weekend activities	SDT	4	Tasks and challenges, rewards (currency earning and unlockable rewards), and social features	Daily steps	Singapore
Farmer et al [44], 2017	840	422 (50.2)	Children (1-8)	Redesigned play environments to encourage imaginative and independent free play	No intervention	1	48 - 96	Feedback, incentives, and social features	MVPA and zBMI	New Zealand
Staiano et al [55], 2022	72	31 (43)	Children (3-5)	Motor Skills app: instructional lessons, peer modeling videos, and behavioral scaffolding	Free Play app: lessons and videos promoting	SCT	12 - 24	Social interaction	MVPA, LPA, and SB	The United States

[a]VPA: vigorous physical activity.

[b]MPA: moderate physical activity.

[c]LPA: light physical activity.

[d]SDT: self-determination theory.

[e]SCT: social cognitive theory.

[f]zBMI: BMI *z*-score.

[g]SB: sedentary time or behavior.

[h]BCW: behavioral change wheel.

[i]MVPA: moderate-to-vigorous physical activity.

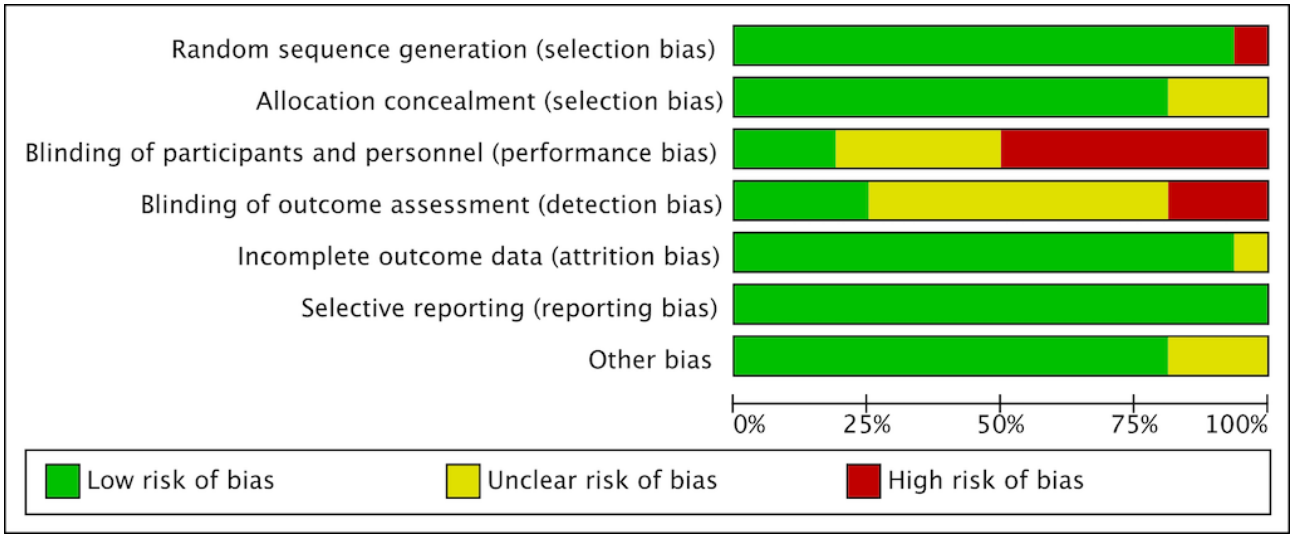
[j]TTM: the transtheoretical model.

Risk of Bias Assessment

The κ coefficient of the risk of bias data extracted by both authors was 0.73, which is in excellent agreement. The authors' ratings for each risk of bias item are presented as percentages in the meta-analysis, and the different biases for each study are summarized. Overall, 15 trials adequately reported randomized sequence generation, and 1 trial with randomized sequence generation was considered high risk because the trials were assigned sequentially in the order of recruitment [48], 13 trials (81%) described programs designed to conceal [41,43,45-53,55,56]. The blinding of the subjects and the personnel in a total of 5 trials is not known [41,46,49,52,53], while 8 trials had unclear blinding of the outcome assessments [43,44,46,47,49,50,52,54]. A total of 15trials provided data on subjects lost to follow-up. In contrast, 1 trial had unclear information [54], and 2 trials were at high risk of other bias [42,52]. In addition, 1 crossover trial was conducted without a blinding condition [47], and 1 trial has not yet been registered for publication, so there is uncertainty [56] (Figures 2 and 3).

XSL•FO
RenderX

Figure 2. Risk of bias graph.



XSL•FO
RenderX

Figure 3. Risk of bias summary [41-56].

XSL•FO
RenderX

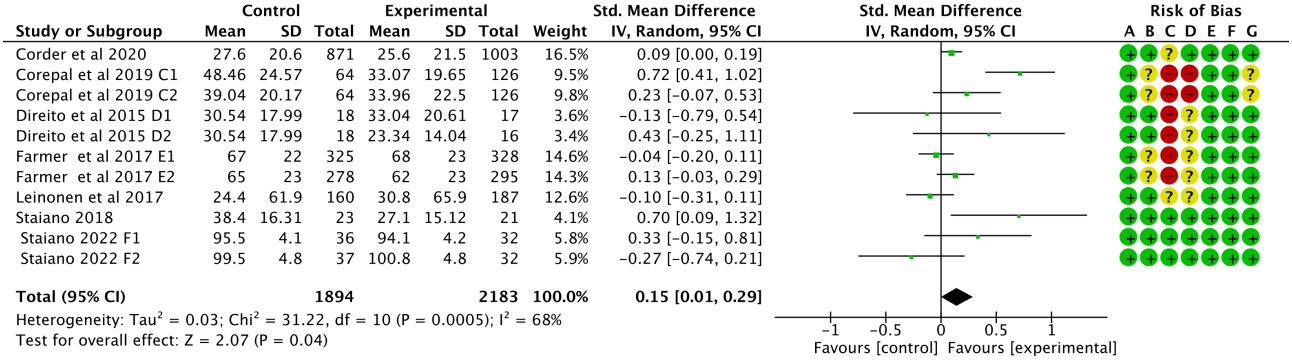
Results of the Meta-Analysis

Of the 16 studies, 11 (69%) were examined to assess the impact of gamification interventions on MVPA. The heterogeneity test showed a significant heterogeneity among the studies (I^2=68%; P<.1), so a random effects model was used for the analysis. The combined effect sizes showed that the gamification intervention significantly improved MVPA in children and adolescents compared to the control group (SMD 0.15, 95% CI 0.01 to 0.29; P=.04); see Figure 4 [41-44,49,53,55].

Of the 16 studies, 6 (38%) were examined to assess the impact of gamification interventions on VPA. The heterogeneity test revealed heterogeneity between studies (I^2=78%; P<.001), which required analysis using a random effects model. The combined effect size indicated that gamification interventions were ineffective in improving VPA in children and adolescents compared to controls (SMD 0.12, 95% CI −0.3 to 0.55; P=.56); see Figure 5 [42,43,50,54].

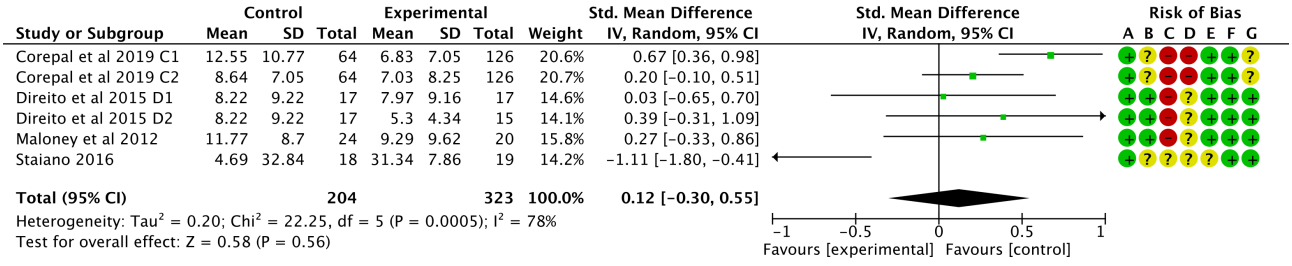
Of the 16 studies, 6 (38%) were examined to assess the impact of gamification interventions on MPA. Tests for heterogeneity showed more significant heterogeneity between studies (I^2=70%; P<.001), requiring analysis using random effects models. The meta-analysis found no significant differences in the effects of gamification interventions on MPA in children and adolescents compared with controls (SMD 0.16, 95% CI −0.2 to 0.53; P=.38); see Figure 6 [42,43,50,53].

Figure 4. Forest plot of the effect of gamification interventions on increasing moderate to vigorous physical activity [41-44,49,53,55].



Risk of bias legend
(**A**) Random sequence generation (selection bias)
(**B**) Allocation concealment (selection bias)
(**C**) Blinding of participants and personnel (performance bias)
(**D**) Blinding of outcome assessment (detection bias)
(**E**) Incomplete outcome data (attrition bias)
(**F**) Selective reporting (reporting bias)
(**G**) Other bias

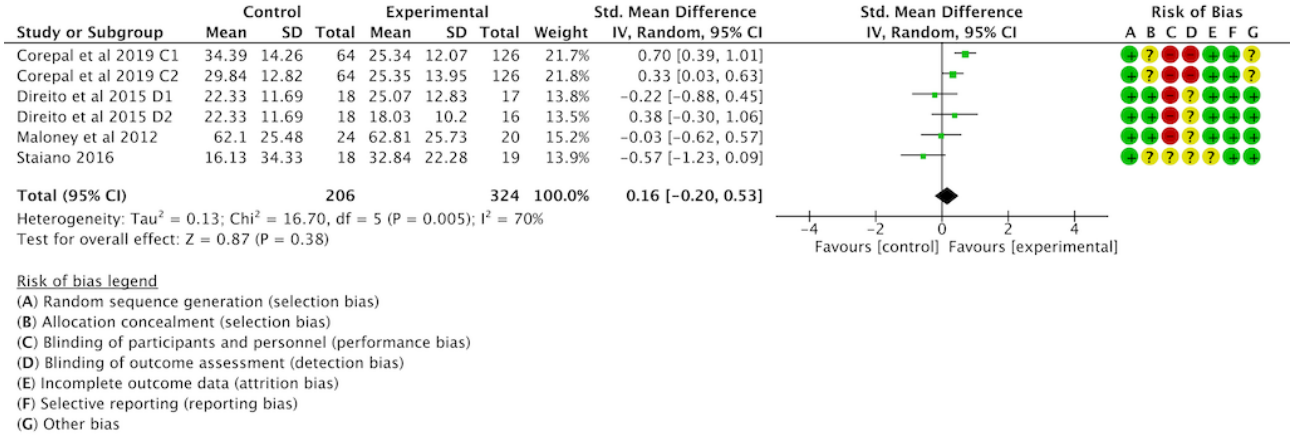
Figure 5. Forest plot of the effect of gamification interventions on increasing vigorous physical activity [42,43,50,54].



Risk of bias legend
(**A**) Random sequence generation (selection bias)
(**B**) Allocation concealment (selection bias)
(**C**) Blinding of participants and personnel (performance bias)
(**D**) Blinding of outcome assessment (detection bias)
(**E**) Incomplete outcome data (attrition bias)
(**F**) Selective reporting (reporting bias)
(**G**) Other bias

XSL•FO
RenderX

Figure 6. Forest plot of the effect of gamification interventions on increasing moderate physical activity [42,43,50,53].



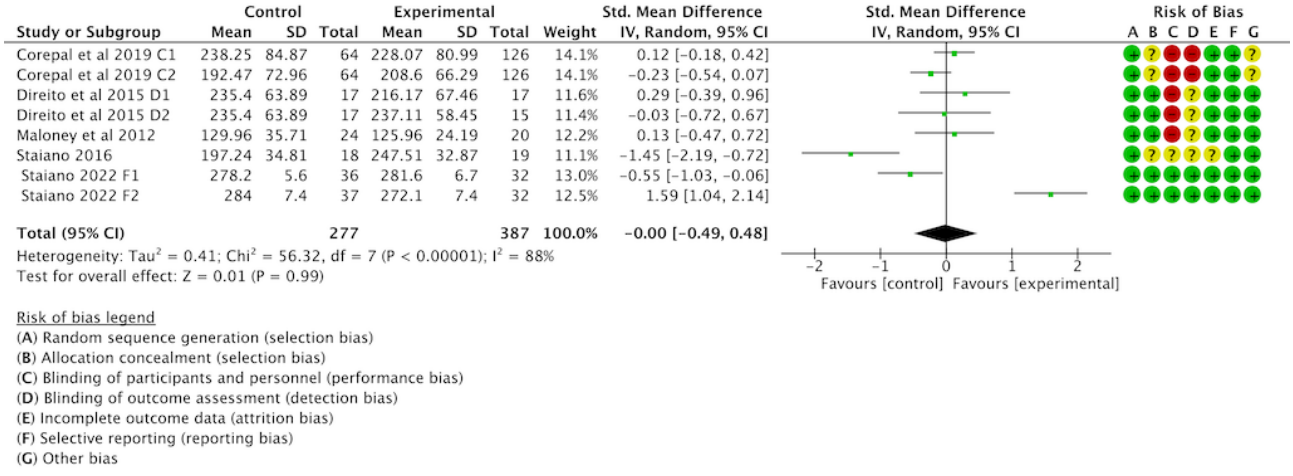
Of the 16 studies, 8 (50%) were examined to assess the impact of gamification interventions on LPA. The heterogeneity test revealed heterogeneity between studies (I^2=88%; P<.001), which required analysis using a random effects model. The combined effect sizes showed that the difference in LPA between the gamification intervention and the control group was insignificant (SMD −0.00, 95% CI −0.49 to 0.48; P=.99); see Figure 7 [42,43,50,54,55].

Of the 16 studies, 7 (44%) were examined to assess the impact of gamification interventions on SB. Heterogeneity tests showed heterogeneity between studies (I^2=19%; P>0.01), so fixed effects models were used for analyses. A meta-analysis found that gamification interventions did not significantly reduce SB in children and adolescents (SMD 0.07, 95% CI −0.07 to 0.22; P=.33); see Figure 8 [42,43,55,56].

Of the 16 studies, 12 (75%) were examined to assess the impact of gamification interventions on daily step counts. Tests for heterogeneity showed more significant heterogeneity between studies (I^2=95%; P<.001), requiring analysis using random effects models. A meta-analysis found that gamification interventions did not increase daily steps among children and adolescents (SMD 0.22,95% CI −0.51 to 0.94; P=.55); see Figure 9 [42,45-48,50,52,54,56].

Of the 16 studies, 4 (25%) were examined to assess the impact of gamification interventions on BMI. Heterogeneity tests showed little heterogeneity between studies (I^2=0%; P=.72), which required analysis using a fixed-effects model. A meta-analysis found that gamification interventions improved BMI in children and adolescents (SMD 0.11, 95% CI 0.05 to 0.18; P<.001); see Figure 10 [51,53,56].

Figure 7. Forest plot of the effect of gamification interventions on increasing light physical activity [42,43,50,54,55].



XSL•FO
RenderX

Figure 8. Forest plot of the effect of gamification interventions on decreasing sedentary behavior [42,43,55,56].

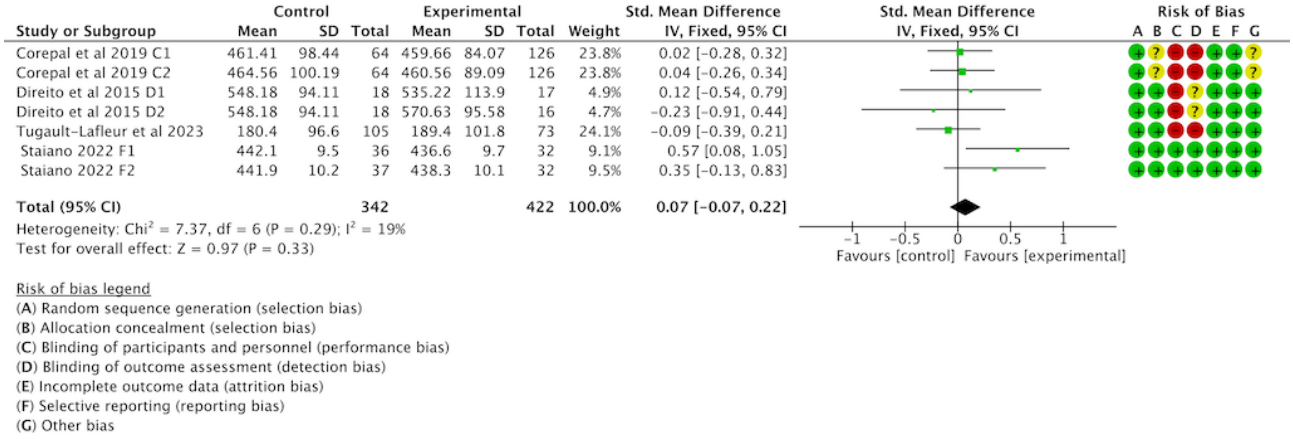


Figure 9. Forest plot of the effect of gamification interventions on daily steps [42,45-48,50,52,54,56].

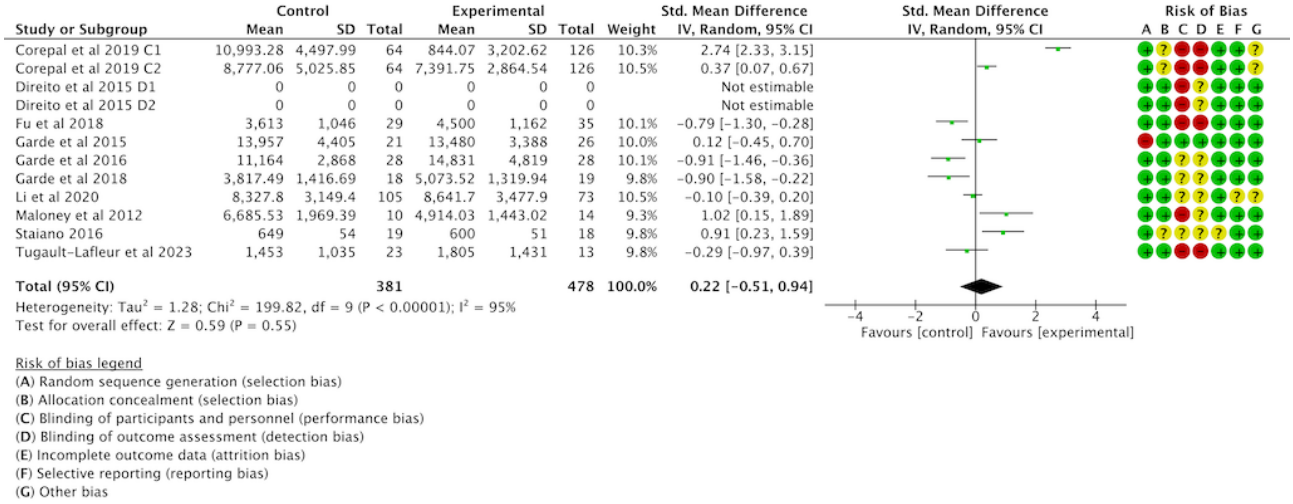
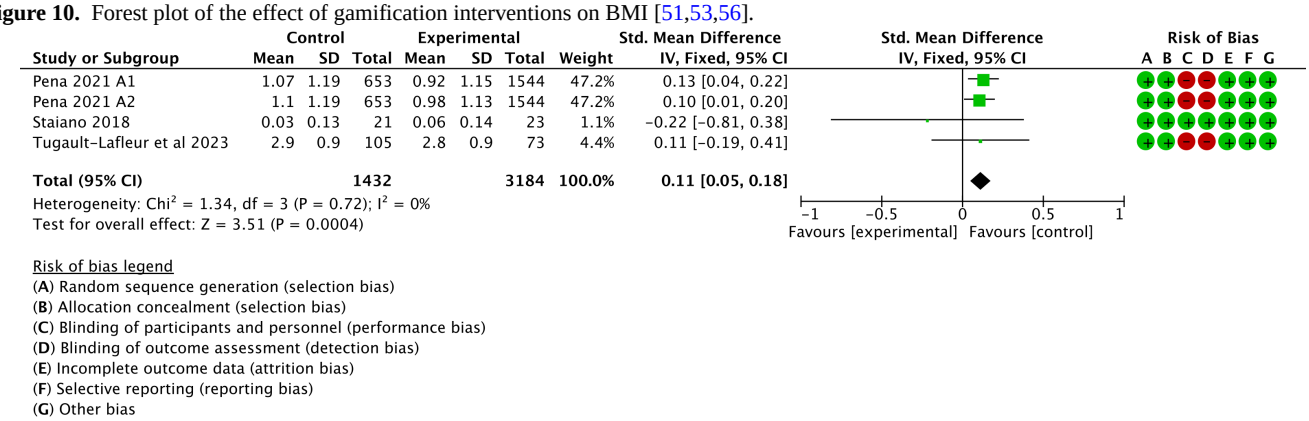


Figure 10. Forest plot of the effect of gamification interventions on BMI [51,53,56].



Sensitivity Analysis

In this study, sensitivity analyses of MVPA, LPA, and SB were conducted using Stata 18.0 to assess the robustness and reliability of the results. The sensitivity analyses showed that excluding any of the studies did not affect the magnitude of the effect of the gamification intervention on the outcomes of LPA (Multimedia Appendix 2) and SB (Multimedia Appendix 3), suggesting the robustness and reliability of the findings [12]. However, for MVPA (Multimedia Appendix 4), the sensitivity analysis identified 4 studies as outliers [41,42,53,55]. Excluding one of these studies changed the overall effect size, suggesting that the results were not robust enough and should be interpreted cautiously.

Subgroup Analyses

To further explore the effects of other components on outcome indicators, this study focused on subgroup analyses of MVPA, including age, theoretical paradigm, game elements, intervention duration, study setting, position of the PA monitor, and sample size. Only 4 articles assessed BMI as an outcome. Subgroup analyses were not performed for this outcome due to the small sample size. In addition, the results of all subgroup analyses for

XSL•FO
RenderX

SB and other PA indicators were not significant and were, therefore, not analyzed in detail.

A Subgroup Analysis of the Effects of a Gamification Intervention on MVPA

The results of the subgroup analyses of the gamification intervention on MVPA are presented in Multimedia Appendix 5, and there were no significant differences in age, wearer position, or sample size regarding improvements in MVPA in children and adolescents. The theoretical paradigm, intervention during, and game elements significantly moderate the effectiveness of gamification interventions in enhancing MVPA. Subgroup analyses based on theoretical paradigms showed a more pronounced effect of SDT-based interventions on MVPA promotion (SMD 0.39, 95% CI 0.01 to 0.77; I^2=91%). However, the effect of other theoretical paradigms based on SCT (SMD −0.19, 95% CI −0.76 to 0.39; I^2=72.6%) and BCTs (SMD 0.01, 95% CI −0.11 to 0.13; I^2=25%) was negligible. Subgroup analyses based on game elements showed significantly larger effect sizes for interventions based on elements such as rewards and feedback (SMD 0.19, 95% CI 0.01 to 0.37; I^2=86%) compared to interventions based primarily on social interaction (SMD −0.07, 95% CI −0.46 to 0.33; I^2=58%), with more significant heterogeneity between the two groups, but effect sizes 95% CI overlapped. Subgroup analyses based on intervention duration showed a more significant MVPA promotion effect for interventions >12 weeks compared to interventions ≤12 weeks (SMD 0.02, 95% CI −0.15 to 0.19; I^2=76.9% and SMD 0.14, 95% CI 0.02 to 0.26; I^2=1.1%), less heterogeneity. There was no overlap in the 95% CIs of effect sizes. Subgroup analyses based on trial setting showed that school-based interventions (SMD 0.18, 95% CI 0.01 to 0.36; I^2=79%) had larger effect sizes than home-based or combined home- and community-based interventions (SMD −0.08, 95% CI −0.35 to 0.20; I^2=48%). Heterogeneity was also more significant, and there was no overlap in the 95% CI for effect sizes.

Reporting Biases

Qualitative and quantitative methods were used to assess publication bias. Funnel plots were used to assess publication bias for the effects of gamification interventions on MVPA, LPA, and SB in children and adolescents. Funnel plots showed a mostly symmetrical pattern across the 3 studies (Multimedia Appendices 6-8). The Egger test was used to assess the size of publication bias of the included studies for all 3 studies (Multimedia Appendices 9-11). The Egger test was performed for MVPA (t_{11}=0.05; P=.97), LPA (t_8=0.01; P=.99), and SB (t_7=–0.83; P=.44). Therefore, all the literature included was judged to be free of publication bias in a comprehensive manner.

Discussion

Principal Findings

This paper systematically reviews the effects of gamification interventions on MVPA, VPA, MPA, LPA, SB, daily steps, and BMI in children and adolescents through a meta-analysis of 16 RCTs. It also responds to the call for future research needs by Mazeas et al [15]. The absence of significant publication bias for all outcomes in the included articles enhances the reliability of the evidence. The results showed that the gamification intervention group had a positive effect on improving MVPA but a more minor effect on reducing SB than the control group. Age, theoretical paradigm, game elements, and intervention duration were identified as essential moderators associated with the increased effectiveness of gamification interventions on MVPA. However, the effects on effect sizes were not entirely consistent.

Impact of Gamification Interventions on PA

The study found that the gamification intervention significantly increased MVPA in children and adolescents but did not show positive effects on VPA, MPA, LPA, daily steps, and reduction in SB, with an overall small, combined effect size. Although our findings are an extension of the most recently published systematic review, they are inconsistent with previous findings [15]. Previous research has demonstrated that gamification interventions can increase levels of PA and have shown efficacy in promoting PA in nonhealthy individuals as well as those with chronic disease, particularly in adults and older adults, as evidenced by significant increases in steps per day and low-intensity PA [57,58]. This increase was particularly evident in studies of adults with overweight or obesity, cancer survivors, and patients with chronic cardiometabolic disease [58-60]. In addition, people who are overweight or obese may have increased their PA levels because of weight loss goals or specific weight loss programs. However, compared with these studies, the results showed more minor and nonsignificant effects with lower quality of evidence. This may be because increasing PA in children and adolescents has proven challenging, whether through home-based gamification interventions or interventions targeting the school environment or curriculum [27,61,62]. This reduces individual adherence to some extent [60]. A total of 6 studies in the included literature showed poor adolescent adherence during the intervention, which affected the final study results [42,46-48,55,56]. Alternatively, the lack of benefit we observed for the intervention may be because part of the target population was already sufficiently active, which leaves little room for an increase in MVPA [44]. In addition, 5 of these studies had baseline activity levels greater than 150 min/d [42,43,45,48,55]. The paper also found daily step goal orientation essential for promoting increased daily step counts [63]. Research has shown that setting higher PA goals typically motivates individuals to achieve higher PA [64,65]. This effect is powerful when combined with goal setting, self-monitoring, and rewarding feedback [66]. Interventions targeting step goal setting were more limited in the included literature. This further explains the limited role of gamification interventions in promoting daily steps and low-intensity PA in healthy adolescents. At the same time, given the limited amount of relevant literature on both, the validity of this finding requires further validation.

MVPA subgroup analyses indicated that interventions using reward-feedback game mechanics (eg, points systems and performance badges) demonstrated superior efficacy. Compared to interventions emphasizing social interaction elements,

XSL•FO
RenderX

reward-based approaches significantly enhanced PA engagement (SMD 0.19, 95% CI 0.01 to 0.37; *P*=.04). Intervention duration and theoretical frameworks emerged as critical moderators: Long-term interventions (>12 weeks) yielded substantially greater MVPA (SMD 0.14, 95% CI 0.02 to 0.26; *P*=.02) improvement than short-term programs (≤12 weeks), confirming the sustainability of gamification effects and aligning with existing evidence on behavioral maintenance [15]. Theoretical paradigms guided 88% (14/16) of the included studies. SDT and SCT were the predominant frameworks, consistent with previous systematic reviews highlighting their use in PA interventions [67]. SDT is a well-established theory of motivation that has become a critical framework for health behavior interventions. An individual's motivation, considered the primary driver of behavior change, is often viewed as an extrinsic driver of PA activity. While elements such as rewards and feedback, primarily goal-directed, are considered intrinsic drivers, it is difficult to distinguish whether the effect is due to the sum of the internal and external drivers or an independent driver intervention. Subgroup analyses in this study indicated that game elements such as rewards and feedback can promote MVPA and that interventions from the SDT theoretical paradigm are effective in promoting MVPA. However, due to the limited amount of literature included, it was not possible to explore the importance of combining both on PA further. Therefore, further coordinated interventions are recommended for the future, such as increasing the weight of elements such as rewards and feedback in gamification while synergizing the SDT theoretical paradigm.

Impact of Gamification Interventions on SB

Meta-analysis revealed that gamification interventions failed to significantly reduce SB (SMD 0.07,95% CI −0.07 to 0.22; *P*=.33) in children and adolescents, with low-quality evidence aligning with previous studies [68,69]. Critically, this nonsignificant outcome more likely reflects methodological constraints rather than intervention inefficacy. Key evidence includes: (1) Pivotal moderating role of family environment: Family settings served as the primary context for gamification interventions [70,71], with 5 out of 7 included studies (71%) implemented in home environments. For instance, a Canadian eHealth intervention study demonstrated that family environment—particularly parenting styles, caregiving approaches, and household income—predicted a substantial proportion of variance in adolescents' intervention adherence [72]. Tugault-Lafleur and colleagues' [56] data also confirmed a strong relationship between parental adherence and youth participation. Only 1 trial [42] in the included literature assessed sedentary behavior in a school setting, and due to the small sample size, subgroup analyses of this outcome were not conducted. (2) Limitations in measurement tool adherence: Accelerometer-based PA monitoring demonstrated suboptimal participant adherence [73], while variations in wear time significantly biased SB estimates [74]. Notably, 2 included studies reported a decline in accelerometer adherence during interventions [43,44], with no intergroup differences (*P*>.05), potentially compromising outcome validity. (3) Efficacy gap in intervention design; Interventions targeting SB are more effective than interventions promoting PA [75]. Paradoxically, most included studies (4/7) lacked explicit SB-reduction objectives or personalized strategies [43,49,51,53]. Although 2 studies incorporated SB-focused components [42,52], these relied on passive advisory methods (eg, Facebook groups or VFB ecosystems). Such approaches likely failed to engage adolescents' active attention to device-generated feedback and elicit timely behavioral responses.

Moderating Variables on the Effects of Gamification Interventions for PA

Theoretical Paradigm

Several types of BCTs are currently commonly used in gamification interventions, and they have also been shown to be effective in influencing intervention outcomes [42]. Only 2 of 16 studies (12.5%) did not use a theoretical paradigm in their gamification interventions [44,50]. The majority was SDT (8/16). SDT is the most used theoretical paradigm in gamification intervention trials. SDT-based interventions are effective in improving MVPA (SMD 0.39, 95% CI 0.01 to 0.77; *P*=.04), in children and adolescents [23]. SDT is a well-established theory of motivation that has become a critical framework for health behavior interventions because of the belief that an individual's motivation is the primary driver of behavior change [76]. However, intrinsic or extrinsic motivation has different effects on behavior change. Existing research suggests that intrinsic motivation promotes more stable behavior change and psychological and social well-being [23]. Therefore, future research could use gamification to promote intrinsic motivation and increase PA engagement. Second, SCT is another theoretical framework widely adopted in PA promotion interventions. SCT proposes that a combination of personal factors, the social environment, and the physical environment influences an individual's level of PA. Among these factors, self-efficacy, self-regulation, and efficacy expectations are vital personal factors that promote PA and reduce SB [77]. In conclusion, SCTor SDT-based interventions may produce more effective outcomes than gamification interventions. However, due to the limited number of included studies, further validation through many high-quality studies is needed to support these conclusions.

Intervention Duration

Subgroup analysis of intervention duration showed an association between the effect of gamification on MVPA and intervention duration. Compared to intervention ≤12 weeks, intervention cycles of more than 12 weeks (including the follow-up period) were more likely to promote MVPA (SMD 0.14, 95% CI 0.02 to 0.26; *P*=.02), with a mean duration of 32.67 weeks (SD 25.8 weeks), suggesting that the effects of gamification persisted beyond the end of the intervention. Smartphone apps have been claimed to have a significant positive effect on PA only when used for a short period of <3 months [46], which is inconsistent with the current meta-analysis results. This may be due to the playful nature of gamification, where the effectiveness of the intervention is not only influenced by the intervention period but also by age, theoretical paradigm, game elements, and the type of measurement tool. It is worth noting that gamification interventions are not just a novelty effect and should have some long-term effects. However, given

XSL•FO
RenderX

the limited amount of literature included, it is still worth exploring whether this long-term effect diminishes after the intervention has ended.

Game Elements

Achievement-based and goal-oriented game elements, such as rewards and feedback, were the most frequently used game elements in the systematic evaluation, consistent with the results of the previous evaluation [36,78]. Analyses of the PA subgroup showed that a gamification intervention with rewards and feedback at its core significantly increased MVPA, effectively combining techniques to promote behavior change and motivate participants to be active. However, it has also been argued that extrinsic motivation (rewards, etc) may be detrimental to the long-term maintenance of intervention effects compared to intrinsic motivation (SDT or SCT) [79]. The effectiveness of socialization and interactivity, the second most used game element, is significantly influenced by the type of social incentive and the application method [80]. Gamification strategies significantly increased PA levels when used collaboratively between families. In contrast, they were less effective when used with individuals who did not know each other as intervention targets [80]. Therefore, future studies must explore the effectiveness of gamification interventions that integrate different social incentives to increase PA engagement and assess the applicability and impact of these interventions in different social relationship contexts. This will help to develop more precise and personalized gamified health promotion programs.

Limitations

This study has several limitations. First, the included literature was in English, which may have led to selection bias, and some studies had small sample sizes and poor compliance, which increased the risk of publication bias. Second, due to the nature of the intervention, the included literature did not implement blinding in data processing and analysis, as well as differences in recruitment settings, which may affect the results of the meta-analysis. Third, the outcome indicators used in the included literature studies had inconsistent units, and using SMD as an effect indicator requires caution in interpreting the results. Results suggest heterogeneity in some studies, and the source of this heterogeneity is unclear. The effectiveness of gamification interventions may vary among children and adolescents depending on demographics (eg, ethnic background, region, gender, and BMI), economic level, type of intervention (parent-centered vs child-centered), and degree of application personalization. These factors may have contributed to the observed heterogeneity. However, these factors were not analyzed in this study due to limitations imposed by the number and characteristics of the included studies. In conclusion, the results of this systematic review and meta-analysis should be interpreted with caution.

Conclusions

Gamification interventions effectively increase MVPA in children and adolescents, and this effect persists beyond the follow-up period. This suggests that the effects of gamification interventions are not solely due to their novelty but have long-term effects. However, the effects of gamification interventions on improving VPA, MPA, LPA, daily step count, and SB are not yet apparent. The number and quality of studies limit them and should be interpreted cautiously. In addition, theoretical paradigms, game elements, and intervention duration may be associated with the effectiveness of gamification interventions, and further research is needed on the optimal implementation of game elements and theoretical features to maximize PA engagement. In conclusion, due to the limited number and quality of included studies, the above findings need to be validated by additional high-quality studies.

Acknowledgments

The authors would like to thank the staff and faculty at the Institute of Sports Medicine and Health at Chengdu Sport University, which offered the funding support and the place to conduct the study. In addition, the authors acknowledge the use of the generative artificial intelligence tool DeepSeek by OpenAI, which was used to assist in refining the language of early manuscript drafts. This study was funded by the National Social Science Fund (grant 24XTY002), the Sichuan Provincial Philosophy and Social Science Foundation (grant SCJJ24ND206), and the Key Project of Sichuan Provincial Philosophy and Social Science Foundation (grant SCJJ23ND77).

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

MW designed the research protocol and drafted the report. XZ and XL searched the literature, analyzed the data, and interpreted the results. JX participated in the research protocol and reviewed the manuscript. YZ reviewed and revised the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Literature search strategy.

XSL•FO
RenderX

[DOC File, 35 KB - games_v13i1e68151_app1.doc]

Multimedia Appendix 2
Sensitivity analyses results on light physical activity.
[PDF File, 315 KB - games_v13i1e68151_app2.pdf]

Multimedia Appendix 3
Sensitivity analyses result on sedentary behavior.
[PDF File, 129 KB - games_v13i1e68151_app3.pdf]

Multimedia Appendix 4
Sensitivity analyses result on moderate-to-vigorous physical activity.
[PDF File, 178 KB - games_v13i1e68151_app4.pdf]

Multimedia Appendix 5
Summary of subgroup analysis results of gamification interventions for moderate-to-vigorous physical activity.
[DOC File, 53 KB - games_v13i1e68151_app5.doc]

Multimedia Appendix 6
Funnel plot of moderate-to-vigorous physical activity.
[PNG File, 146 KB - games_v13i1e68151_app6.png]

Multimedia Appendix 7
Funnel plot of sedentary behavior.
[PNG File, 52 KB - games_v13i1e68151_app7.png]

Multimedia Appendix 8
Funnel plot of light physical activity.
[PNG File, 47 KB - games_v13i1e68151_app8.png]

Multimedia Appendix 9
Egger test results for light physical activity.
[PNG File, 45 KB - games_v13i1e68151_app9.png]

Multimedia Appendix 10
Egger test results for sedentary behavior.
[PNG File, 48 KB - games_v13i1e68151_app10.png]

Multimedia Appendix 11
Egger test results for moderate-to-vigorous physical activity.
[PNG File, 145 KB - games_v13i1e68151_app11.png]

Checklist 1
PRISMA 2020 checklist.
[DOC File, 74 KB - games_v13i1e68151_app12.doc]

References

1. Centers for Disease Control and Prevention (CDC). Physical activity levels of high school students --- United States, 2010. MMWR Morb Mortal Wkly Rep 2011 Jun 17;60(23):773-777. [Medline: 21681173]
2. Lee IM, Shiroma EJ, Lobelo F, et al. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. Lancet 2012 Jul 21;380(9838):219-229. [doi: 10.1016/S0140-6736(12)61031-9] [Medline: 22818936]
3. Riddoch CJ, Mattocks C, Deere K, et al. Objective measurement of levels and patterns of physical activity. Arch Dis Child 2007 Nov;92(11):963-969. [doi: 10.1136/adc.2006.112136] [Medline: 17855437]
4. Bull FC, Al-Ansari SS, Biddle S, et al. World Health Organization 2020 guidelines on physical activity and sedentary behaviour. Br J Sports Med 2020;54:1451-1462. [doi: 10.1136/bjsports-2020-102955]

XSL•FO
RenderX

5. Guthold R, Stevens GA, Riley LM, Bull FC. Global trends in insufficient physical activity among adolescents: a pooled analysis of 298 population-based surveys with 1·6 million participants. The Lancet Child & Adolescent Health 2020 Jan;4(1):23-35. [doi: 10.1016/S2352-4642(19)30323-2]
6. Erskine HE, Moffitt TE, Copeland WE, et al. A heavy burden on young minds: the global burden of mental and substance use disorders in children and youth. Psychol Med 2015 May;45(7):1551-1563. [doi: 10.1017/S0033291714002888] [Medline: 25534496]
7. Kieling C, Baker-Henningham H, Belfer M, et al. Child and adolescent mental health worldwide: evidence for action. Lancet 2011 Oct 22;378(9801):1515-1525. [doi: 10.1016/S0140-6736(11)60827-1] [Medline: 22008427]
8. Lascar N, Brown J, Pattison H, Barnett AH, Bailey CJ, Bellary S. Type 2 diabetes in adolescents and young adults. Lancet Diabetes Endocrinol 2018 Jan;6(1):69-80. [doi: 10.1016/S2213-8587(17)30186-9] [Medline: 28847479]
9. Song P, Zhang Y, Yu J, et al. Global prevalence of hypertension in children: a systematic review and meta-analysis. JAMA Pediatr 2019 Dec 1;173(12):1154-1163. [doi: 10.1001/jamapediatrics.2019.3310] [Medline: 31589252]
10. Manson JE, Greenland P, LaCroix AZ, et al. Walking compared with vigorous exercise for the prevention of cardiovascular events in women. N Engl J Med 2002 Sep 5;347(10):716-725. [doi: 10.1056/NEJMoa021067] [Medline: 12213942]
11. Wolin KY, Yan Y, Colditz GA, Lee IM. Physical activity and colon cancer prevention: a meta-analysis. Br J Cancer 2009 Feb 24;100(4):611-616. [doi: 10.1038/sj.bjc.6604917] [Medline: 19209175]
12. Duclos M, Oppert JM, Verges B, et al. Physical activity and type 2 diabetes. Recommandations of the SFD (Francophone Diabetes Society) diabetes and physical activity working group. Diabetes Metab 2013 May;39(3):205-216. [doi: 10.1016/j.diabet.2013.03.005] [Medline: 23643351]
13. Pedersen BK, Saltin B. Exercise as medicine - evidence for prescribing exercise as therapy in 26 different chronic diseases. Scand J Med Sci Sports 2015 Dec;25 Suppl 3:1-72. [doi: 10.1111/sms.12581] [Medline: 26606383]
14. Ekelund U, Tarp J, Steene-Johannessen J, et al. Dose-response associations between accelerometry measured physical activity and sedentary time and all cause mortality: systematic review and harmonised meta-analysis. BMJ 2019 Aug 21;366:l4570. [doi: 10.1136/bmj.l4570] [Medline: 31434697]
15. Mazeas A, Duclos M, Pereira B, Chalabaev A. Evaluating the effectiveness of gamification on physical activity: systematic review and meta-analysis of randomized controlled trials. J Med Internet Res 2022 Jan 4;24(1):e26779. [doi: 10.2196/26779] [Medline: 34982715]
16. McConnell MV, Turakhia MP, Harrington RA, King AC, Ashley EA. Mobile health advances in physical activity, fitness, and atrial fibrillation: moving hearts. J Am Coll Cardiol 2018 Jun 12;71(23):2691-2701. [doi: 10.1016/j.jacc.2018.04.030] [Medline: 29880130]
17. Anderson M, Faverio M, Gottfried J. Teens, social media and technology 2023. : Pew Research Center; 2023 URL: https://abfe.issuelab.org/resources/43096/43096.pdf [accessed 2025-08-23]
18. Alam S, Zhang M, Harris K, Fletcher LM, Reneker JC. The impact of consumer wearable devices on physical activity and adherence to physical activity in patients with cardiovascular disease: a systematic review of systematic reviews and meta-analyses. Telemed J E Health 2023 Jul;29(7):986-1000. [doi: 10.1089/tmj.2022.0280] [Medline: 36445750]
19. Cai X, Qiu SH, Yin H, et al. Pedometer intervention and weight loss in overweight and obese adults with type 2 diabetes: a meta-analysis. Diabet Med 2016 Aug;33(8):1035-1044. [doi: 10.1111/dme.13104] [Medline: 26926674]
20. Ocagli H, Agarinis R, Azzolina D, et al. Physical activity assessment with wearable devices in rheumatic diseases: a systematic review and meta-analysis. Rheumatology (Oxford) 2023 Mar 1;62(3):1031-1046. [doi: 10.1093/rheumatology/keac476] [Medline: 36005834]
21. Singh B, Zopf EM, Howden EJ. Effect and feasibility of wearable physical activity trackers and pedometers for increasing physical activity and improving health outcomes in cancer survivors: a systematic review and meta-analysis. J Sport Health Sci 2022 Mar;11(2):184-193. [doi: 10.1016/j.jshs.2021.07.008]
22. Wong SH, Tan ZYA, Cheng LJ, Lau ST. Wearable technology-delivered lifestyle intervention amongst adults with overweight and obese: a systematic review and meta-regression. Int J Nurs Stud 2022 Mar;127:104163. [doi: 10.1016/j.ijnurstu.2021.104163]
23. Johnson D, Deterding S, Kuhn KA, Staneva A, Stoyanov S, Hides L. Gamification for health and wellbeing: a systematic review of the literature. Internet Interv 2016 Nov;6:89-106. [doi: 10.1016/j.invent.2016.10.002]
24. Koivisto J, Hamari J. Gamification of physical activity: a systematic literature review of comparison studies. Presented at: Proceedings of the 3rd International GamiFIN Conference; Apr 8-10, 2019; Levi, Finland URL: https://urn.fi/URN:NBN:fi:tuni-201910244075
25. Direito A, Carraça E, Rawstorn J, Whittaker R, Maddison R. mHealth technologies to influence physical activity and sedentary behaviors: behavior change techniques, systematic review and meta-analysis of randomized controlled trials. Ann Behav Med 2017 Apr;51(2):226-239. [doi: 10.1007/s12160-016-9846-0] [Medline: 27757789]
26. Love R, Adams J, van Sluijs EMF. Are school-based physical activity interventions effective and equitable? A meta-analysis of cluster randomized controlled trials with accelerometer-assessed activity. Obes Rev 2019 Jun;20(6):859-870. [doi: 10.1111/obr.12823] [Medline: 30628172]

XSL•FO
RenderX

27. Metcalf B, Henley W, Wilkin T. Effectiveness of intervention on physical activity of children: systematic review and meta-analysis of controlled trials with objectively measured outcomes (EarlyBird 54). BMJ 2012 Sep 27;345(sep27 1):e5888. [doi: 10.1136/bmj.e5888] [Medline: 23044984]
28. Sims J, Scarborough P, Foster C. The effectiveness of interventions on sustained childhood physical activity: a systematic review and meta-analysis of controlled studies. PLoS One 2015;10(7):e0132935. [doi: 10.1371/journal.pone.0132935] [Medline: 26193472]
29. Ning Y, Jia Z, Zhu R, Ding Y, Wang Q, Han S. Effect and feasibility of gamification interventions for improving physical activity and health-related outcomes in cancer survivors: an early systematic review and meta-analysis. Support Care Cancer 2022 Dec 30;31(1):92. [doi: 10.1007/s00520-022-07550-0] [Medline: 36585478]
30. Fiedler J, Eckert T, Wunsch K, Woll A. Key facets to build up eHealth and mHealth interventions to enhance physical activity, sedentary behavior and nutrition in healthy subjects - an umbrella review. BMC Public Health 2020 Oct 23;20(1):1605. [doi: 10.1186/s12889-020-09700-7] [Medline: 33097013]
31. Yang Q, Van Stee SK. The comparative effectiveness of mobile phone interventions in improving health outcomes: meta-analytic review. JMIR mHealth uHealth 2019 Apr 3;7(4):e11244. [doi: 10.2196/11244] [Medline: 30942695]
32. Quintana DS. From pre-registration to publication: a non-technical primer for conducting a meta-analysis to synthesize correlational data. Front Psychol 2015;6:1549. [doi: 10.3389/fpsyg.2015.01549] [Medline: 26500598]
33. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097. [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
34. Deterding S, Dixon D, Khaled R, Nacke L. From game design elements to gamefulness: defining “gamification”. Presented at: MindTrek ’11: Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments; Sep 28-30, 2011; Tampere Finland p. 9-15. [doi: 10.1145/2181037.2181040]
35. Johnson D, Horton E, Mulcahy R, Foth M. Gamification and serious games within the domain of domestic energy consumption: a systematic review. Renewable and Sustainable Energy Reviews 2017 Jun;73:249-264. [doi: 10.1016/j.rser.2017.01.134]
36. Cugelman B. Gamification: what it is and why it matters to digital health behavior change developers. JMIR Serious Games 2013 Dec 12;1(1):e3. [doi: 10.2196/games.3139] [Medline: 25658754]
37. Helmerhorst HJF, Brage S, Warren J, Besson H, Ekelund U. A systematic review of reliability and objective criterion-related validity of physical activity questionnaires. Int J Behav Nutr Phys Act 2012 Aug 31;9. [doi: 10.1186/1479-5868-9-103] [Medline: 22938557]
38. Stovold E, Beecher D, Foxlee R, Noel-Storr A. Study flow diagrams in Cochrane systematic review updates: an adapted PRISMA flow diagram. Syst Rev 2014 May 29;3:54. [doi: 10.1186/2046-4053-3-54] [Medline: 24886533]
39. Hozo SP, Djulbegovic B, Hozo I. Estimating the mean and variance from the median, range, and the size of a sample. BMC Med Res Methodol 2005 Apr 20;5:15840177. [doi: 10.1186/1471-2288-5-13] [Medline: 15840177]
40. Cumpston M, Li T, Page MJ, et al. Updated guidance for trusted systematic reviews: a new edition of the Cochrane Handbook for Systematic Reviews of Interventions. Cochrane Database Syst Rev 2019 Oct 3;10(10):ED000142. [doi: 10.1002/14651858.ED000142] [Medline: 31643080]
41. Corder K, Sharp SJ, Jong ST, et al. Effectiveness and cost-effectiveness of the GoActive intervention to increase physical activity among UK adolescents: a cluster randomised controlled trial. PLoS Med 2020 Jul;17(7):e1003210. [doi: 10.1371/journal.pmed.1003210] [Medline: 32701954]
42. Corepal R, Best P, O’Neill R, et al. A feasibility study of “The StepSmart Challenge” to promote physical activity in adolescents. Pilot Feasibility Stud 2019;5:132. [doi: 10.1186/s40814-019-0523-5] [Medline: 31832227]
43. Direito A, Jiang Y, Whittaker R, Maddison R. Apps for IMproving FITness and increasing physical activity among young people: the AIMFIT pragmatic randomized controlled trial. J Med Internet Res 2015 Aug 27;17(8):e210. [doi: 10.2196/jmir.4568] [Medline: 26316499]
44. Farmer VL, Williams SM, Mann JI, Schofield G, McPhee JC, Taylor RW. The effect of increasing risk and challenge in the school playground on physical activity and weight in children: a cluster randomised controlled trial (PLAY). Int J Obes (Lond) 2017 May;41(5):793-800. [doi: 10.1038/ijo.2017.41] [Medline: 28186099]
45. Fu Y, Burns RD, Constantino N, Zhang P. Differences in step counts, motor competence, and enjoyment between an exergaming group and a non-exergaming group. Games Health J 2018 Oct;7(5):335-340. [doi: 10.1089/g4h.2017.0188] [Medline: 29989430]
46. Garde A, Chowdhury M, Rollinson AU, et al. A multi-week assessment of a mobile exergame intervention in an elementary school. Games Health J 2018 Feb;7(1):1-8. [doi: 10.1089/g4h.2017.0023] [Medline: 29394109]
47. Garde A, Umedaly A, Abulnaga SM, et al. Evaluation of a novel mobile exergame in a school-based environment. Cyberpsychol Behav Soc Netw 2016 Mar;19(3):186-192. [doi: 10.1089/cyber.2015.0281] [Medline: 26882222]
48. Garde A, Umedaly A, Abulnaga SM, et al. Assessment of a mobile game (“MobileKids Monster Manor”) to promote physical activity among children. Games Health J 2015 Apr;4(2):149-158. [doi: 10.1089/g4h.2014.0095] [Medline: 26181809]

XSL•FO
RenderX

49. Leinonen AM, Pyky R, Ahola R, et al. Feasibility of gamified mobile service aimed at physical activation in young men. JMIR mHealth uHealth 2017;5(10). [doi: 10.2196/mhealth.6675] [Medline: 29017991]
50. Maloney AE, Threlkeld KA, Cook WL. Comparative effectiveness of a 12-week physical activity intervention for overweight and obese youth: exergaming with “Dance Dance Revolution”. Games Health J 2012 Apr;1(2):96-103. [doi: 10.1089/g4h.2011.0009] [Medline: 26193183]
51. Peña S, Carranza M, Cuadrado C, et al. Effectiveness of a gamification strategy to prevent childhood obesity in schools: a cluster controlled trial. Obesity (Silver Spring) 2021 Nov;29(11):1825-1834. [doi: 10.1002/oby.23165] [Medline: 34533295]
52. Seah MLC, Koh KT. The efficacy of using mobile applications in changing adolescent girls’ physical activity behaviour during weekends. Eur Phy Educ Rev 2021 Feb;27(1):113-131. [doi: 10.1177/1356336X20930741]
53. Staiano AE, Beyl RA, Guan W, Hendrick CA, Hsia DS, Newton RL Jr. Home-based exergaming among children with overweight and obesity: a randomized clinical trial. Pediatr Obes 2018 Nov;13(11):724-733. [doi: 10.1111/ijpo.12438] [Medline: 30027607]
54. Staiano AE, Beyl RA, Hsia DS, Katzmarzyk PT, Newton RL Jr. Twelve weeks of dance exergaming in overweight and obese adolescent girls: transfer effects on physical activity, screen time, and self-efficacy. J Sport Health Sci 2017 Mar;6(1):4-10. [doi: 10.1016/j.jshs.2016.11.005] [Medline: 28491483]
55. Staiano AE, Newton RL, Beyl RA, et al. mHealth intervention for motor skills: a randomized controlled trial. Pediatrics 2022 May 1;149(5):e2021053362. [doi: 10.1542/peds.2021-053362] [Medline: 35415743]
56. Tugault-Lafleur CN, De-Jongh González O, Macdonald J, et al. Efficacy of the Aim2Be intervention in changing lifestyle behaviors among adolescents with overweight and obesity: randomized controlled trial. J Med Internet Res 2023 Apr 25;25:e38545. [doi: 10.2196/38545] [Medline: 37097726]
57. Wang JW, Zhu Z, Shuling Z, et al. Effectiveness of mHealth app-based interventions for increasing physical activity and improving physical fitness in children and adolescents: systematic review and meta-analysis. JMIR mHealth uHealth 2024 Apr 30;12:e51478. [doi: 10.2196/51478] [Medline: 38687568]
58. Xu L, Li J, Zhang X, et al. Mobile health-based gamification intervention to increase physical activity participation among patients with coronary heart disease: study protocol of a randomised controlled trial. BMJ Open 2022 Jan 31;12(1):e054623. [doi: 10.1136/bmjopen-2021-054623] [Medline: 35105640]
59. Fanaroff AC, Patel MS, Chokshi N, et al. A randomized controlled trial of gamification, financial incentives, or both to increase physical activity among patients with elevated risk for cardiovascular disease: rationale and design of the be active study. Am Heart J 2023 Jun;260:82-89. [doi: 10.1016/j.ahj.2023.02.014] [Medline: 36870551]
60. Fang Y, Ma Y, Mo D, Zhang S, Xiang M, Zhang Z. Methodology of an exercise intervention program using social incentives and gamification for obese children. BMC Public Health 2019 Jun 3;19(1):686. [doi: 10.1186/s12889-019-6992-x] [Medline: 31159776]
61. Biddle SJH, Braithwaite R, Pearson N. The effectiveness of interventions to increase physical activity among young girls: a meta-analysis. Prev Med 2014 May;62:119-131. [doi: 10.1016/j.ypmed.2014.02.009] [Medline: 24530611]
62. Brown HE, Atkin AJ, Panter J, Wong G, Chinapaw MJM, van Sluijs EMF. Family-based interventions to increase physical activity in children: a systematic review, meta-analysis and realist synthesis. Obes Rev 2016 Apr;17(4):345-360. [doi: 10.1111/obr.12362] [Medline: 26756281]
63. Casado-Robles C, Viciana J, Guijarro-Romero S, Mayorga-Vega D. Effects of consumer-wearable activity tracker-based programs on objectively measured daily physical activity and sedentary behavior among school-aged children: a systematic review and meta-analysis. Sports Med Open 2022 Jan 31;8(1):18. [doi: 10.1186/s40798-021-00407-6] [Medline: 35099630]
64. Anson D, Madras D. Do low step count goals inhibit walking behavior: a randomized controlled study. Clin Rehabil 2016 Jul;30(7):676-685. [doi: 10.1177/0269215515593782] [Medline: 26150019]
65. Chevance G, Baretta D, Golaszewski N, et al. Goal setting and achievement for walking: a series of N-of-1 digital interventions. Health Psychol 2021 Jan;40(1):30-39. [doi: 10.1037/hea0001044] [Medline: 33252961]
66. Miller BG, Livingston CP, Zerger HM, Valbuena DA, Miltenberger RG. Evaluating public posting, goal setting, and rewards to increase physical activity in children. Behav Interv 2023 Feb;38(1):105-117. [doi: 10.1002/bin.1902]
67. Brickwood KJ, Watson G, O’Brien J, Williams AD. Consumer-based wearable activity trackers increase physical activity participation: systematic review and meta-analysis. JMIR Mhealth Uhealth 2019 Apr 12;7(4):e11819. [doi: 10.2196/11819] [Medline: 30977740]
68. Lee D, Frey GC, Min A, et al. Usability inquiry of a gamified behavior change app for increasing physical activity and reducing sedentary behavior in adults with and without autism spectrum disorder. Health Informatics J 2020 Dec;26(4):2992-3008. [doi: 10.1177/1460458220952909] [Medline: 32951500]
69. Selles WL, Santos EC, Romero BD, Lunardi AC. Effectiveness of gamified exercise programs on the level of physical activity in adults with chronic diseases: a systematic review. Disabil Rehabil 2024 Dec;46(26):6231-6239. [doi: 10.1080/09638288.2024.2323614] [Medline: 38444153]
70. Golan M. Parents as agents of change in childhood obesity--from research to practice. Int J Pediatr Obes 2006;1(2):66-76. [doi: 10.1080/17477160600644272] [Medline: 17907317]
71. Mâsse LC, Watts AW, Barr SI, et al. Individual and household predictors of adolescents’ adherence to a web-based intervention. Ann Behav Med 2015 Jun;49(3):371-383. [doi: 10.1007/s12160-014-9658-z] [Medline: 25270826]

XSL•FO
RenderX

72. Tu AW, Watts AW, Chanoine JP, et al. Does parental and adolescent participation in an e-health lifestyle modification intervention improves weight outcomes? BMC Public Health 2017 Apr 24;17(1):352. [doi: 10.1186/s12889-017-4220-0] [Medline: 28438202]
73. Böhm B, Karwiese SD, Böhm H, Oberhoffer R. Effects of mobile health including wearable activity trackers to increase physical activity outcomes among healthy children and adolescents: systematic review. JMIR Mhealth Uhealth 2019 Apr 30;7(4):e8298. [doi: 10.2196/mhealth.8298] [Medline: 31038460]
74. Coll F, Cavalheri V, Gucciardi DF, Wulff S, Hill K. Quantifying the effect of monitor wear time and monitor type on the estimate of sedentary time in people with COPD: systematic review and meta-analysis. J Clin Med 2022 Apr 1;11(7):1980. [doi: 10.3390/jcm11071980] [Medline: 35407588]
75. Commissaris D, Huysmans MA, Mathiassen SE, Srinivasan D, Koppes LLJ, Hendriksen IJM. Interventions to reduce sedentary behavior and increase physical activity during productive work: a systematic review. Scand J Work Environ Health 2016 May 1;42(3):181-191. [doi: 10.5271/sjweh.3544] [Medline: 26683116]
76. Deci ER, Ryan RM. Motivation, personality, and development within embedded social contexts: an overview of self-determination theory. In: The Oxford Handbook of Human Motivation: Oxford University Press; 2012. [doi: 10.1093/iwc/iwae012]
77. Elizabeth Morghen S, Robert W M. Pilot study of social cognitive theory variables as correlates of physical activity among adolescents with pediatric multiple sclerosis. Mult Scler Relat Disord 2020 Jul;42:102026. [doi: 10.1016/j.msard.2020.102026] [Medline: 32272443]
78. Vermeir JF, White MJ, Johnson D, Crombez G, Van Ryckeghem DML. The effects of gamification on computerized cognitive training: systematic review and meta-analysis. JMIR Serious Games 2020 Aug 10;8(3):e18644. [doi: 10.2196/18644] [Medline: 32773374]
79. Hicks K, Gerling K, Richardson G, Pike T, Burman O, Dickinson P. Understanding the effects of gamification and juiciness on players. Presented at: 2019 IEEE Conference on Games (CoG); Aug 20-23, 2019; London, United Kingdom p. 1-8. [doi: 10.1109/CIG.2019.8848105]
80. Patel MS, Small DS, Harrison JD, et al. Effectiveness of behaviorally designed gamification interventions with social incentives for increasing physical activity among overweight and obese adults across the United States: the STEP UP randomized clinical trial. JAMA Intern Med 2019 Dec 1;179(12):1624-1632. [doi: 10.1001/jamainternmed.2019.3505] [Medline: 31498375]

Abbreviations

BCT: behavior change technique
BCW: behavior change wheel
LPA: light physical activity
MPA: moderate physical activity
MVPA: moderate-to-vigorous physical activity
PA: physical activity
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROSPERO: International Prospective Register of Systematic Reviews
SB: sedentary behavior
SCT: social cognitive theory
SDT: self-determination theory
SMD: standardized mean difference
TTM: transtheoretical model
VPA: vigorous physical activity
zBMI: BMI z-score

Edited by A Coristine; submitted 29.10.24; peer-reviewed by K Carr; revised version received 28.07.25; accepted 30.07.25; published 18.09.25.

Please cite as:
Wang M, Xu J, Zhou X, Li X, Zheng Y
Effectiveness of Gamification Interventions to Improve Physical Activity and Sedentary Behavior in Children and Adolescents: Systematic Review and Meta-Analysis
JMIR Serious Games 2025;13:e68151
URL: https://games.jmir.org/2025/1/e68151
doi:10.2196/68151

XSL•FO
RenderX

© Min Wang, Jisheng Xu, Xulin Zhou, Xingchen Li, Yu Zheng. Originally published in JMIR Serious Games (https://games.jmir.org), 18.9.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Use of Gamified Digital Tools in Daily Tasks of Health Care Workers: Scoping Review

Binita Paudel[1,2], BPH, MA; Hussein Al-Shehabi[1], MSc, PhD; Rita Dörner[1], MSc, PhD; Charbel El Bcheraoui[1], MSc, PhD; Tessa Lennemann[3], MD, PhD; Andrea Bernasconi[1,4], MSc, MD, DrPH

[1]Evidence-Based Public Health Unit, Centre for International Health Protection, Robert Koch Institute, Gricht Strasse 27, Berlin, Germany

[2]Institute of International Health, Charité – Universitätsmedizin Berlin, Berlin, Germany

[3]Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ), Bonn, Germany

[4]Unité Epidémiologie et Recherche Clinique, Réseau de l'Arc, Saint-Imier, Switzerland

Corresponding Author:
Charbel El Bcheraoui, MSc, PhD
Evidence-Based Public Health Unit, Centre for International Health Protection, Robert Koch Institute, Gricht Strasse 27, Berlin, Germany

Abstract

Background: The quality and effectiveness of health care service delivery are significantly influenced by the engagement and motivation of health care workers. Integrating gamified digital tools (GDTs) into health care workers' workflows presents a promising approach to enhancing them. However, there is currently a lack of evidence supporting the implementation of such interventions.

Objective: This scoping review aims to summarize existing evidence on the influence of GDTs on the daily tasks of health care workers.

Methods: A scoping review was conducted following the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines for scoping reviews. We conducted a comprehensive search across different databases (PubMed, EMBASE, *International Journal of Serious Games*, Cochrane Library, and Google Scholar) for peer-reviewed studies and (OpenAlex, GreyNet, and IEEE Xplore) for gray literature published between January 2010 and January 2024. Eligibility criteria, developed using the SPIDER (Sample, Phenomenon of Interest, Design, Evaluation, and Research) framework, included qualitative, quantitative, and mixed methods studies involving health care workers using GDTs for daily tasks. Studies in English, French, Spanish, or Italian were eligible. Keywords and medical subject headings related to gamification and health care workers were used. The studies were screened, eligibility was assessed, and data were extracted. A narrative synthesis was used to summarize and interpret the findings.

Results: Of 5844 studies, 12 met the inclusion criteria and were included in the analysis. These studies exhibited considerable heterogeneity in the application of gamification. Feedback, competition, and dashboard features were the most common gamification elements identified. The implementation of these elements led to enhanced engagement, increased motivation, improved task completion, and promoted healthy competition among staff across various health care settings.

Conclusions: Integrating GDTs into health care workers' tasks holds significant potential to enhance engagement and motivation. However, empirical evidence is still lacking. Comparative studies are needed to gain comprehensive insights into the benefits and limitations of gamification in health care.

(JMIR Serious Games 2025;13:e70480) doi:10.2196/70480

KEYWORDS

game elements; gamification; gamified digital tools; health care workers

Introduction

The quality and effectiveness of delivering health care services greatly depend on the engagement and motivation of health care workers (HCWs), as their willingness to perform tasks in the workplace directly influences the quality of care offered [1,2]. Surveys conducted in several countries indicated a lack of motivation as the second most critical issue within the health care workforce, preceded only by the problem of staff shortage [2]. Many daily tasks performed by HCWs, such as medical data collection or routine physical examinations, are particularly prone to motivation lapses. Yet, these tasks are essential for generating reliable data that underpin public health decisions, improving evidence-based patient management, and, consequently, delivering effective and high-quality health care [3,4].

XSL•FO
RenderX

The growing integration of information and communication technologies (ICTs) into the health care sector has increasingly supported HCWs in clinical practice. According to the World Health Organization, by 2021, 80% (n=120) of member states had developed national digital health strategies [5]. Furthermore, global spending on digital health was estimated at approximately USD 288 billion in 2024, largely driven by investments in telemedicine, artificial intelligence, and mobile health (mHealth) solutions [6].

The advent of ICT presents an opportunity to address staff motivation through gamified digital tools (GDTs). GDTs are software applications or platforms that integrate game mechanics, such as points, badges, levels, leaderboards, or challenges, into nongame contexts to enhance user motivation, engagement, and learning outcomes [7].

The term "gamification" was first introduced in 2002 to enhance the learning process in the consumer goods industry by incorporating playful elements into software [8-10]. The concept of gamification has gained popularity in recent years [11]. Defined by Deterding et al [12] as "the incorporation of game design elements into nongame contexts," gamification entails integrating game-like elements, such as points and leaderboards, into environments that traditionally do not fit within the conventional scope of gaming [13]. For Zichermann and Cunningham [14], gamification is "the process of game-thinking and game mechanics to engage users and solve problems," while Bunchball [15] describes it as applying game elements to nongame activities, aiming to alter people's behaviors. Bunchball also recognizes gamification as a powerful strategy for influencing and motivating various groups of people, such as children, adults, or workers [15]. In fact, the primary purpose of gamification is to encourage positive behavior in individuals while also enhancing their motivation and participation in specific activities [16]. Thus, gamification has the potential to motivate employees while performing monotonous and repetitive tasks, boost their involvement, foster a positive approach to their jobs, and, as a result, elevate their overall productivity [17].

The application of gamification is today widespread across various sectors, including education, business, marketing, and health care. In education, gamification can make learning activities more engaging and motivate students to achieve their goals [18,19]. For instance, Duolingo, one of the most common mobile-assisted language learning platforms, incorporates numerous game-like elements to enhance students' engagement and motivation to learn new languages [20], and Kim et al [21] reported that college students in a gamified class outperformed their counterparts in traditional lecture-based classes by 40%. In business and marketing, gamification encourages employees to complete tasks or increase customer engagement with a product or service [22]. In health, gamification strategies are designed to motivate patients to follow treatment plans to maintain healthy lifestyles [23]. "Heart Game" [24] and "Heart Health" [25], which are gamified disease self-management apps for cardiovascular health, are examples of these health behaviors change support systems [11]. Along with the high penetration of smartphones and mobile devices into society [26], gamification has the potential for substantial growth. Its market, estimated at USD 15.4 billion for 2024, is expected to reach USD 48.7 billion by 2029 [27].

The basic structure of any gamification approach is the "gamification element." A gamification element refers to a specific feature or component derived from games that are used to enhance engagement, motivation, and participation in nongame contexts. These elements are incorporated into various activities or systems to make them more appealing and enjoyable. Examples of game elements are points, badges, leaderboards, and rewards [28]. Game elements like badges, feedback, and achievements foster a sense of competence, while the inclusion of teams and social networks enhances feelings of connectedness. Tailored features, such as avatars, empower individuals by providing control and choice over their actions, thereby reinforcing their sense of autonomy [29,30]. A comprehensive categorization of gamification elements was compiled by reviewing several authors' works [16,28,31-40], and it is provided in Table 1.

XSL•FO
RenderX

Table . A comprehensive list of gamification elements.

Elements of gamification	Description
Badges/trophies	A visual symbol representing a predetermined accomplishment and acknowledgement
Feedback	Immediate feedback on users' performance and achievements which induces users' sense of progress and achievement during tasks, increasing awareness of their own progress and enhancing their experiences
Competition	Team-based competition encourages users to develop the vested interest in the task, thereby sustaining their engagement and involvement throughout the activity. A set of tasks that incentivize who finishes early and quickly
Time pressure/countdown timer/ clock/deadlines	Time itself used to pressurize the users to finish a task, thereby enhancing engagement and challenging users within a defined timeframe
Leaderboards/scoreboard/ dashboard/rank	A leaderboard in gamification motivates competition and fosters engagement by showcasing and ranking participants based on their performance
Rewards	Serve as incentives, driving user motivation and fostering continued engagement by acknowledging and reinforcing positive actions
Avatar	Avatars in gamification allow individuals to represent themselves. Digital self-representation of users
Progress	Progress in gamification helps to track the progress made thus far, assess achievements and select desired badges or trophies, providing a clear overview of accomplished milestones and guiding future goals
Level	Levels represent the position of users within tasks, with each completed task potentially advancing them to the next level, introducing new challenges and difficulties
Points/scores	A quantification of achievements following the successful completion of tasks and responsibilities
Quest	A quest in gamification refers to the task or challenge that motivates users to engage with the application or system
Notification	Notifications prompt users to perform desired activities
Narrative story	A narrative story creates a scenario or theme inspired by a gamified system

Recent studies have increasingly employed structured taxonomies to categorize gamification mechanics in different domains, emphasizing the importance of aligning specific game elements with contextual objectives. For instance, Hervás et al [41] proposed a taxonomy comprising 6 fundamental categories tailored to the health sector, particularly aimed at fostering behavioral change. These categories are goals (direction and purpose), status (recognition via points or badges), randomness (surprise through chance), appointment (timed engagement), progression (tracking advancement), and social (interaction and competition). Schmidt-Kraepelin et al [42] also developed a taxonomy of gamification concepts for health apps. This taxonomy comprises 12 dimensions, each characterized by 2-3 mutually exclusive attributes. Toda et al [43] later presented a gamification taxonomy in an educational environment and identified 5 different categories: performance (acknowledgment, level, points, progression, and stats), ecological (chance, choice, rarity, and time pressure), social (competition, cooperation, reputation, and social pressure), personal (novelty, purpose, challenges, renovation, and sensation), and fictional (narrative and storytelling).

The mechanism behind gamification could be simplified into extrinsic and intrinsic motivations. Extrinsic motivation refers to the use of external rewards or incentives to encourage people to engage in certain behaviors or actions. In contrast, intrinsic motivation is driven by internal rewards such as personal satisfaction or enjoyment [28].

Despite the widespread success of gamification in promoting learning behaviors, advancing health education [44], and promoting healthy habits [45], the extent of its effectiveness in enhancing the activities of HCWs is still underexplored. This gap persists even amid the rapid and ongoing digital transformation of health systems worldwide. While ICT enhances efficiency by reducing administrative burden and improving clinical decision-making, its adoption could present challenges as frontline HCWs must navigate complex digital interfaces, manage large volumes of patient data, and adapt to evolving systems. Poorly designed tools can lead to cognitive overload, workflow disruptions, and increased stress, impacting job satisfaction, burnout rates, and, finally, patient care quality [46-48].

This study aims to bridge this knowledge gap by examining how GDTs could shape the daily work experience of HCWs, ensuring that technological advancements enhance rather than hinder their ability to provide high-quality care.

Through this unique focus, this study can provide evidence of gamification in HCW's environment and practical insights for

XSL•FO
RenderX

policymakers, health care institutions, and digital tool developers.

This review is the first to explore the integration of GDTs into the daily workflows of HCWs and to present evidence of their influence on their daily tasks.

Methods

For this review, we adopted the PRISMA ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) [49] (Multimedia Appendix 1).

Search Strategy and Selection Criteria

We conducted a comprehensive literature search, utilizing various databases and search terms. PubMed, EMBASE, *International Journal of Serious Games*, Cochrane Library, and Google Scholar were searched for peer-reviewed research, while OpenAlex, GreyNet, and IEEE Xplore were searched for gray literature known to index technical reports, conference proceedings, and non-peer-reviewed research relevant to gamification. Our search strategy included Medical Subject Headings and keywords such as “gamification,” “videogame,” “healthcare worker,” and “healthcare professional” to ensure the inclusion of all relevant studies from both peer-reviewed and gray literature. We employed Boolean operators like “OR” and “AND,” along with truncation (*), to account for variation in word endings, such as gamif* for gamification or gamified. An expert librarian supported this rigorous search to ensure the inclusion of all relevant studies. The detailed search strategy is available in Multimedia Appendix 2. After an initial screening based on titles and abstracts, we conducted a second screening on the full text retrieved by 2 authors (BP and HAS) through Rayyan [50], with a third author (AB) consulted in case of disagreement. The weighted kappa coefficient of Cohen was calculated to assess the degree of agreement between the 2 reviewers, and it was applied as suggested by Landis and Koch [51]. Additionally, we conducted a supplementary manual search through the reference lists of the included studies.

Eligibility Criteria

The SPIDER (Sample, Phenomenon of Interest, Design, Evaluation, and Research) framework [52] determined the inclusion and exclusion criteria as outlined in Table 2.

Table . Eligibility criteria according to the SPIDER (Sample, Phenomenon of Interest, Design, Evaluation, and Research) framework.

Criterion	Inclusion	Exclusion
Sample (S)	Health care workers including community health workers, community health volunteers, and data clerks	Patients, students, non-health care workers
Phenomenon of interest (PI)	Gamified digital tools considered on the basis of the elements of gamification embedded (points, badge, feedback, etc) used in the daily activities	Gamified tools not involving health care delivery or lacking gamification features
Design (D)	Peer-reviewed sources and gray literature	Informal sources without clear institutional affiliation or academic relevance
Evaluation (E)	Studies reporting on tool features, usability, engagement, health care worker experience, or outcomes related to work performance	
Research type (R)	Qualitative, quantitative, mixed methodology	Editorial, commentaries, editorials, opinion pieces, or protocols
Others		
Publication year	Published between January 2010 and January 2024	
Language	English, French, Italian, Spanish	Publications in other languages
Technology format	Tools applying gamification through ICT[a] (eg, mobile apps, electronic health records, telemedicine platforms)	Tools not using ICT (eg, physical board games)
Publication type	Peer-reviewed journal articles, conference papers, and abstracts	Unpublished material, blogs
Subject focus	Gamified digital tools used for health care service delivery or operational/clinical support tasks	Tools used solely for training, disease self-management, or in non-health care contexts
Gamified tool user group	Health care workers as direct users of the gamified digital tool	Tools used exclusively by patients, students, or non-health care professionals

[a]ICT: information and communication technology.

Only interventions involving GDTs for daily health care tasks were selected for inclusion, provided they featured at least 1 gamification element. For this review, HCWs were defined broadly to include not only clinical staff but also volunteers, data clerks, and health administrators who used GDTs in their daily tasks, for example, by entering clinical data into databases

XSL•FO
RenderX

within a health care setting. Studies in which gamification elements were not implemented within ICT (eg, board games) were omitted, as well as studies in which GDTs were primarily aimed at training, education, and self-care disease management, or not related to health care.

In our study, we considered studies published between January 2010 and January 2024, as the term gamification gained widespread recognition after 2010 [10-12], and to capture the progression of digital health since the release of the "Principles for Digital Development" [53]. Additionally, studies of all designs (eg, qualitative, quantitative, and mixed methods studies) published in English, French, Spanish, or Italian were included.

Data Extraction and Quality Appraisal

One author (BP) conducted independent data extraction using a spreadsheet prepared by consensus by the reviewers using Microsoft Excel [54], subsequently revised by 2 authors (AB and HAS). For each included study, we extracted information on the year of publication, country of implementation, study objective, design, data collection method, target population, sample size, type of digital intervention, application context, gamification elements used, and main findings.

Two authors (BP and AB) conducted the quality assessment. A third author (HAS) was consulted in case of disagreement. The quality assessment was based on the method outlined by Sardi et al [11], which involved utilizing a set of five closed-ended questions to assess the relevance and quality of the publication: (1) whether the paper provides a detailed description of the game elements employed, (2) if the study presented empirical results, (3) if the benefits and (4) limitations of gamification were explicitly addressed, and (5) the relevance of the source where the study was published investigated through Scimago Journal & Country Rank [55]. The score assigned to each study could range from 0 (lowest) to 6 (highest). Based on the average score, the studies were classified into 3 categories: high (5 or above), average (below 5 and above or equal to 3), and poor (below 3).

The quality assessment checklist and results are given in Multimedia Appendix 3. To assess the degree of agreement between the raters, a weighted kappa coefficient of Cohen was calculated.

Data Synthesis

Given the heterogeneity of the studies, we employed a narrative approach to summarize the findings and address our research question. As GDTs have been integrated into various digital health interventions, a classification based on the different working contexts and the professional background of health staff was deemed most suitable for conceptualizing and describing their application. This format also aligns with the notion that digital tools should be designed according to the profiles of their users, as stated by the "Principles for Digital Development" [53]. We, therefore, classified users into three main categories: (1) community health care workers (CHWs), frontline workers engaging with communities outside of health facilities; (2) HCWs working in primary health care (PHC); and (3) HCWs working in secondary health facilities, such as hospitals. Finally, as suggested by several authors [11,16,40,56-59], we summarized the benefits of the GDTs on the HCWs as follows: increased engagement, enhanced motivation, task completion and behavior change, and heightened healthy competition.

Increased engagement refers to HCWs' higher involvement and commitment to their tasks, leading to improved performance and better patient care [11,56-58]. Enhanced motivation means a stronger drive to perform duties effectively, resulting in higher job satisfaction and productivity [11,16,56-58]. Task completion involves the consistent and timely fulfillment of assigned duties based on a behavior change [56,59]. Healthy competition is a positive rivalry that encourages staff to excel, share best practices, and support each other's growth, enhancing overall care quality [40].

Results

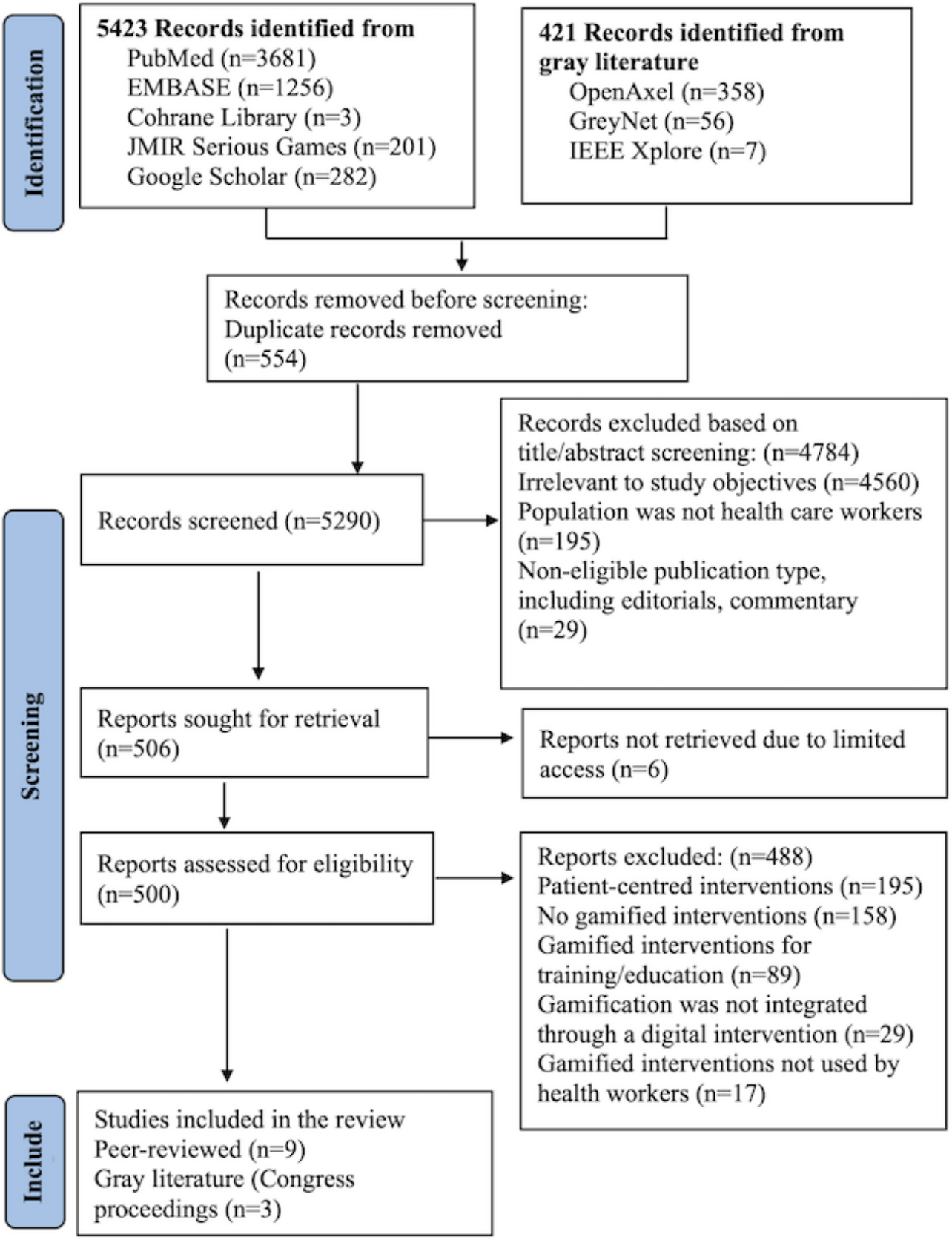
Overview

A total of 5844 records were identified through a comprehensive search of multiple databases. After removing 554 duplicates, 5290 studies were screened by title and abstract. Of these, 4784 were excluded for not meeting the study objectives, primarily because they did not address gamification. An additional 195 studies were excluded for not focusing on health care, and 29 were removed based on publication type (eg, editorial or commentary). Subsequently, 506 studies were sought for retrieval. Of these, 6 could not be accessed as full text and were therefore excluded. The final 500 studies underwent a second screening. At this stage, 488 were excluded: 195 focused on patient-centered intervention rather than HCWs, 158 did not incorporate gamified components, 89 described gamified tools solely for training and educational purposes, 29 included gamification elements not integrated into a digital platform, and 17 involved GDTs not intended for use by HCWs. Finally, based on the eligibility criteria, 12 studies were selected for the final review, comprising 9 studies that had been peer-reviewed and 3 Congress proceedings.

The interrater reliability for selecting the included studies, measured using Cohen's weighted kappa coefficient [51], was 0.72, suggesting a substantial agreement between raters. The PRISMA flow diagram of the study selection process is presented in Figure 1.

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram illustrating the study selection process.



Quality of the Included Studies

The average quality score of the selected studies, based on the assessment method proposed by Sardi et al [11], was 3.4, with scores ranging from 1 (minimum) to 5 (maximum). A total of 5 studies were classified as high quality, 4 as average quality, and 3 as poor quality. Most of the studies (n=10, 83.3%) lacked a discussion concerning the limitations of the GDTs, and many (n=5, 41.7%) did not discuss any benefit. A total of 8 studies (66.7%) were published or presented in recognized journals or

XSL•FO
RenderX

Congresses. The Cohen weighted kappa coefficient among the raters was very good: 0.83 [51].

Basic Information of Included Studies

Of the 12 included studies, 9 (75%) were peer-reviewed in scientific journals, while the remaining 3 (25%) were sourced from gray literature (Congress proceedings). A total of 6 (50%) studies originated from high-income countries (HICs): Canada [60], Australia [61], the United States [62-64], and Portugal [65]; 6 (50%) from low- and middle-income countries (LMICs): Nigeria [66], India [67], Pakistan [68], Kenya [69], and Ethiopia [70]. One study was a multi-country intervention carried out in Bangladesh, Burkina Faso, and Ecuador. A total of 3 interventions focus on CHWs' use of GDTs, 3 are implemented in PHC settings, and 6 in hospitals. The majority of the studies was published before 2020 (n=10, 83.3%). The most common study design was quasi-experimental (n=5, 41.7%), followed by qualitative studies (n=2, 16.7%). Of the remaining 5 studies, 1 was a cross-sectional study, 1 a case study, 1 an interventional pilot study, 1 a diagnostic accuracy study, and 1 an observational study. The characteristics of each study are presented in Table 3.

XSL•FO
RenderX

Table . Overview of the characteristics of the included studies.

Author Year Country/s	Objectives	Study type Sample size	Digital intervention	Health domain or application context	Results	Context
Luedtke et al [66] 2023 Nigeria	to improve SAP[a] compliance to guidelines	Interventional Surgeons (n=20)	Surgeons received real-time feedback from a virtual mentor concerning the most appropriate SAP	SAP	321 SAP prescriptions recorded. 12% adjusted to guidelines after app feedback (P<.01). 10% changed decision on SAP necessity (P<.01)	Hospital
Finette et al [71] 2019 Bangladesh Burkina Faso, Ecuador	To improve care quality by enhancing iCCM[b] protocol compliance	Validation study FLWs[c] (n=49), LHPs[d] (n=22), Children aged 2-60 mo (n=861)	The platform guides FLWs through clinical assessments with embedded demos and animated GIFs[e] to improve data quality	iCCM	FLW assessments with the tool showed 84% - 99% specificity correlation with experts; triage recommendations also aligned. The tool was appreciated for usability, learning, and support	FLWs
Shah et al [67] 2019 India	To assess the uptake, feasibility, effectiveness, and perceptions of a mobile phone app. To increase the coverage of maternal, newborn, and child health services	Mixed methods study Mothers of 1 - to 4-month-old infants (n=99), Mothers with 6- to 9-month-old infants (n=187), FLWs (n=15)	FLWs (ASHAs[f]) register pregnant women and children via the app, which schedules home visits with reminders and urgency indicators. Checklists and algorithms support diagnosis and treatment. A web interface enables real-time supervision and tracks high-risk cases, inventory, and incentives	Mother and child health care	Coverage improved in the intervention area versus control for newborn care (56% vs 10%), exclusive breastfeeding (44% vs 23%), and care-seeking for maternal (77% vs 57%), neonatal (78% vs 27%), and pneumonia screening (41% vs 24%). FLWs showed sufficient app competency with an 88% login rate during the study	FLWs
Zaidi et al [68] 2020 Pakistan	To improve data quality in EPI[g]	Qualitative study Vaccinators (n=26), government district managers (n=7)	Vaccinators use an app to record data, track stocks, and send schedule reminders. Real-time data informs monthly meetings, supports performance recognition, and promotes best practices	EPI	Digital immunization tracking was well-received; data were more valid than manual entry. Key motivators: ease of use, data quality, recognition, field support, and empowerment	PHC[h]

XSL•FO
RenderX

Author Year Country/s	Objectives	Study type Sample size	Digital intervention	Health domain or application context	Results	Context
Mckeown et al [60] 2016 Canada	To improve sepsis identification and management by increasing guidelines compliance	Quasi-experimental study Clinicians in emergency departments (n=not stated)	The "150 Lives in 150 Days" campaign engaged clinicians in adopting sepsis protocol. Patient data were collected via apps, websites, or papers. Game-like elements (countdown, lives saved) created urgency, while feedback and learning modules boosted engagement	Sepsis identification and management	Severe sepsis mortality dropped to 6.4% (Oct 2013-Mar 2014) from 21.1% in earlier periods	Hospital
Orchard et al [61] 2019 Australia	AF[i] screening in general practice	Qualitative study GPs[j] (n=21), nurses (n=13), practice managers (n=11)	The app used EMRs[k] to guide evidence-based AF treatment per guidelines. Feedback to practices covered screening totals, results, AF diagnoses, and treatment compliance	AF screening	Users appreciated the tools, though technical issues occasionally disrupted screening. Time constraints were the main barrier	PHC
Orwoll et al [62] 2018 USA	To prevent CLABSI[l]	Quasi-experimental Nurses (n=105)	During shifts, nurses use the app for line care and to track protocol adherence and performance. A gamification engine enables team competitions	CLABSI prevention	Compared to the previous year, CLABSI rates in the intervention group (9886 line-days) dropped by 48% (P=.03), with no significant change in controls (7879 line-days)	Hospital
Wenner et al [69] 2014 Kenya	To design an intuitive, fun, and efficient interactive tool for data collection	Case study FLWs (n=not stated)	Intuitive inputs (sliders, dials, buttons) support valid data entry, while visual cues (eg, height slider with tall/short images) enhance clarity. Sensory feedback (lights, sounds) creates an engaging, accessible user experience	Data collection	Leveraging tangible computing reduces the technical literacy required to use a device, crucial in low-resource settings	FLWs
Little et al [70] 2013 Ethiopia	To support HCWs[m] and midwives in mother and child health care service	Mixed method: HCWs (n=20), midwives (n=12)	An analytics dashboard and mobile scorecard enable health workers and supervisors to monitor their progress and key performance indicators	Mother and child health care	HCWs quickly adapted to touchscreens with minimal support. Gamification fostered ownership and empowerment	PHC

XSL•FO
RenderX

Author Year Country/s	Objectives	Study type Sample size	Digital intervention	Health domain or application context	Results	Context
Owens et al [64] 2018 USA	To reduce delay in data collection	Quasi-experimental study HCWs (n=not stated)	A computer app calculates delays from daily EMR extracts, displaying a real-time color-coded dashboard to alert staff about patient delays	Data collection	Patient delays were reduced by 27% in the >60-min category	Hospital
Collins et al [63] 2013 USA	To measure the impact of an automated hand hygiene monitoring system	Quasi-experimental study HCWs (n=60)	Hand hygiene programs focusing on solution dispensing, compliance, and infection rates, featuring team competition with weekly compliance graphs by unit, and individual feedback through point-of-care units and posted compliance reports	Hand hygiene compliance	75.4% increase in total hand hygiene solution dispensing, 96.4% improvement in monthly hand hygiene compliance, and health care–associated infections reduction by 11.1%	Hospital
Marques et al [65] 2017 Portugal	To improve hand hygiene compliance	Cross-sectional study Nurses (n=6)	Real-time data collected and feedback from an indoor location system regarding each nurses' hand hygiene compliance	Hand hygiene compliance	The impact of this intervention is still under evaluation	Hospital

[a]SAP: surgical antibiotic prophylaxis.

[b]iCCM: Integrated Community Case Management.

[c]FLW: front line worker.

[d]LHP: local health care professional.

[e]GIF: graphics interchange format.

[f]ASHA: Accredited Social Health Activist.

[g]EPI: Expanded Programme on Immunization.

[h]PHC: primary health care.

[i]AF: atrial fibrillation.

[j]GP: general practitioner.

[k]EMR: electronic medical record.

[l]CLABSI: central line-associated bloodstream infections.

These 12 studies address a variety of health topics, including compliance with hand hygiene [63,65], mother and childcare [67,70], health-related data collection [64,69], surgical antibiotic prophylaxis (SAP) [66], Integrated Community Case Management (iCCM) [71], and Expanded Programme on Immunization (EPI) implementation [68], sepsis diagnosis and therapy [60], atrial fibrillation (AF) screening [61], and central line-associated bloodstream infections prevention [62]. Table 3 presents a comprehensive overview of the included studies.

The most common gamification elements integrated into the GDTs were the use of feedback (n=9, 75%), use of competition (n=6, 50%), leaderboards/dashboards/scoreboards (n=6, 50%), badges (n=3, 25%), and time pressure elements (n=3, 25%). A total of 3 (25%) digital interventions [67,69,71] incorporated a single gamification element, whereas most (n=9, 75%) integrated at least 2. Table 4 presents a summary of the elements of gamification adopted in these studies.

XSL•FO
RenderX

Table . Elements of gamification presented in the reviewed studies.

Elements of gamification/studies	Luedtke et al [66]	Finette et al [71]	Shah et al [67]	Zaidi et al [68]	Mckeown et al [60]	Orchard et al [61]	Orwoll et al [62]	Wenner et al [69]	Little et al [70]	Owens et al [64]	Collins et al [63]	Marques et al [65]
Feedback	✓[a]	✓		✓	✓	✓		✓	✓		✓	✓
Badges	✓				✓							✓
Time pressure/countdown clock			✓		✓					✓		
Competition				✓	✓	✓	✓				✓	✓
Narrative or story					✓							
Challenges	✓				✓							
Progress					✓							
Leaderboard/scoreboard/dashboard					✓		✓		✓	✓	✓	✓
Award					✓		✓					
Social networking					✓		✓					
Points							✓					✓
Level												✓
Virtual goods/rewards/prize					✓							✓
Cooperation												✓
Emotional relationship												✓

[a]✓: elements of gamification presented.

Use of Gamified Digital Tools by Community Health Workers

We identified 3 studies [67,69,71] that involved CHWs using GDTs, all conducted in LMICs.

One intervention in Kenya focuses on improving data collection by introducing a custom-made GDT device with intuitive input mechanisms like sliders and dials, complemented by visual aids and sensory feedback (eg, when collecting height data, the device featured images of a tall individual at the top and a shorter individual at the bottom of a height slider). This approach was reported to enhance CHWs' engagement, but it does not present any evidence of increased effectiveness in data collection [69].

In India, a quasi-experimental study evaluated an intervention with a smartphone app for health activists and a web interface for PHC staff to improve maternal and childcare. The app used time pressure as a gamification element, turning text red for overdue tasks to prompt faster completion. The intervention increased home-based newborn care from 10% to 56%, exclusive breastfeeding from 23% to 44%, care-seeking for maternal and neonatal complications from 57% to 77% and 27% to 78%, respectively, and pneumonia care from 24% to 41%. The time pressure likely improved task completion, though the study did not quantify its specific impact [67].

MEDSINC, a mobile health platform, integrates Integrated Community Case Management guidelines to support CHWs in the clinical assessment, triage, and treatment of children aged between 2 months and 6 years. This tool includes demonstrations

XSL•FO
RenderX

and animated graphics interchange formats to enhance data collection quality and accuracy. Children's clinical assessments by CHWs using MEDSINC showed a specificity correlation of 84%-99% with local health care experts, and their triage recommendations were highly correlated [71]. The tool received positive feedback for its user-friendliness, educational value, and enhancement of job performance, showing that gamified data visualization can motivate data collection.

Use of Gamified Digital Tools in Primary Health Care

A total of 3 interventions evaluated GDTs in PHC: two in LMICs and one in an HIC [61,68,70].

An observational study in Ethiopia [70] evaluated an Open Data Kit platform for maternal and childcare case management. The tool included an analytic dashboard and a scorecard to monitor key performance indicators. The intervention was well-received by HCWs, who quickly adapted to smartphones and required minimal technical support. However, while visualizing progress appeared to motivate HCWs' performance, the study did not establish a causal link between the heightened motivation and the overall impact of the intervention.

In Pakistan, an app was developed to monitor routine immunization by recording and uploading data, synchronizing profiles with central databases, and managing vaccine inventories [68]. The tool used real-time data to monitor vaccinations, track incomplete cases, plan outreach, adjust micro plans, and allocate fuel allowances. It demonstrated greater accuracy than manual data entry and used gamification to highlight top performers (leaderboard), reward high achievers, and foster healthy competition. Also, individual vaccine performance feedback increased vaccinators' engagement.

A qualitative study investigated the use of an eHealth tool within an atrial fibrillation screening program in PHC in Australia [61]. The app extracted EMR data to provide real-time screening prompts and treatment recommendations. Additionally, a smartphone electrocardiograph tool was used for diagnosis. The app generated stroke risk scores and treatment advice, with monthly individual de-identified data reports for quality improvement. These reports included screening rates, new diagnoses, and treatment adherence. HCWs appreciated the eHealth tools, and regular feedback was beneficial for quality improvement, increased the proportion of eligible people to be screened for atrial fibrillation, and maintained motivation. Moreover, the feedback encouraged app usage by fostering healthy competition.

Use of Gamified Digital Tools in Hospital Settings

A total of 6 interventions were carried out in hospital care, 5 [60,62,63,65,66] focusing on reducing health care–associated infections by increasing adherence to guidelines and procedures, and 1 study [64] was carried out on patient flow management.

Two interventions [63,65] incorporated gamification elements to improve hand hygiene in intensive care units. In the United States, a multifaceted strategy included hand washing automated monitoring machines, communication, and leadership initiatives [63]. Key features included hospital-wide competitions, public compliance postings, and electronic point-of-care performance reporting. This approach led to a 75% increase in hand hygiene solution dispensing, a 96% improvement in hand hygiene compliance, and an 11% reduction in health care–associated infections, demonstrating that individual feedback and team competition can significantly impact health outcomes.

A similar intervention in Portugal used an automated monitoring system based on indoor location detection to assess nurses' proximity to patients and handwashing habits. Initial trials faced accuracy issues, preventing a full evaluation of the gamification elements. However, the intervention aimed to provide real-time feedback through dashboards, email notifications, and competitions with avatars and leaderboards, with its impact on hand hygiene compliance still under evaluation [65].

To prevent nosocomial infections, another study evaluated the use of the CLABSI (Central Line-Associated Bloodstream Infections) app to engage nurses in preventing infections caused by central catheters in intensive care units through a quasi-experimental design [62]. This app allowed nurses to monitor protocol compliance and compare their performance against benchmarks and their units. Over 12 months, the intervention group showed a 48% decline in CLABSI rates compared to the previous year, while the control group showed no significant change. The app included gamification elements such as leaderboards, social networking, points, team competitions, awards, and prizes. A total of 3 competitions were held during the study period. Self-assessment completions were significantly higher, with a median of 43 self-assessments per unit per month versus two in non-contest months. Nurse engagement with the app was also 10 times higher during contest months.

Prevention was at the core of another study conducted in Nigeria, aiming to change surgeons' behavior regarding SAP [66]. A virtual mentor provided real-time feedback on surgical risk, antibiotic choice, and therapy duration. A WhatsApp group supported app usage, and 321 surgeries were recorded. The study found that 12% (n=37) of surgeons adjusted their decisions to align with guidelines, and 10% (n=33) changed their antibiotic prophylaxis decisions. Gamification elements such as challenges and digital badges increased engagement, though the app's lack of EMR integration limited its effectiveness.

In 2015, Canada launched a 150-day campaign to enhance sepsis treatment in emergency departments using a gamification approach to increase the adoption of sepsis protocol [60]. Clinicians organized themselves as teams, and clinical data from sepsis patients were collected online, via the app, or on paper and reviewed for protocol adherence. The campaign featured a countdown clock, a daily tally of saved lives, awards, and a feedback module. Positive extrinsic motivators, such as feedback and rewards, along with intrinsic motivators, including mastery, autonomy, relatedness, and purpose, were incorporated into this life-saving campaign, which showed a reduction in severe sepsis mortality from 21% to 6%. The healthy competition among clinicians was crucial to the campaign's success.

Lastly, 1 US-based intervention was not strictly connected to clinical care but focused on patient flow management. A radiology department implemented a tailored dashboard to address patient data collection delays. This application

XSL•FO
RenderX

calculated delays using daily EMR data and displayed real-time, color-coded dashboards to alert staff. A quasi-experimental study showed that the intervention reduced patient delays over 60 minutes by 27%, with time pressure as the gamification element [64].

Discussion

Principal Findings

The primary purpose of this review was to collate existing evidence on how incorporating GDTs into the daily workflow of HCWs affects their behavior and performance in health-related tasks. Of the 12 reviewed studies, 5 [60,62,66,68,69] reported that integrating different game elements such as challenges, prizes, badges, feedback, leaderboards, social networks, competition, and time pressure helps to enhance HCWs' engagement; 3 [60,61,71] emphasized the role of gamification in increasing motivation; 3 [62,66,67] promoted task completion and behavior change; and 4 [60,61,63,68] boosted healthy competition among users.

The reviewed studies suggested that GDTs can improve hand hygiene compliance, sick child assessment, maternal and child health care, sepsis management and control, and CLABSI rates. However, we observed significant heterogeneity across the interventions, due to differences in health care settings, the purposes of the digital tools, and user characteristics. These differences make it challenging to identify a consistent impact of GDTs. Furthermore, gamification was rarely the primary focus of the research, as most studies prioritized a comprehensive evaluation of the broader ICT strategy. As a result, isolating the specific effects of gamification from the overall digital intervention remains difficult, limiting the ability to establish clear causal links.

In other contexts, literature is rich with contributions from authors who highlighted the positive effect of gamification on education and learning [44,72,73], as well as health and wellness [45]. However, to the best of our knowledge, this review is the first to explore the integration of GDTs into the daily workflows of HCWs.

The reviewed studies offer several important insights. Evidence from HICs suggests that GDTs are most effective when integrated into a broader organizational framework that includes workflow reorganization and improved intra-team communication. Additionally, the integration of EMRs plays a key role by enabling real-time access to patient data, thereby facilitating more effective performance feedback and the delivery of timely clinical prompts. However, these structural elements require continuous attention to ensure seamless integration and to maximize the impact of GDTs, particularly when they are strategically incorporated into comprehensive public health initiatives.

On the other hand, for HCWs with limited digital literacy, particularly in LMICs, the reviewed studies suggest that a simple and engaging user interface plays a particularly important role in facilitating adoption and sustained use. Many CHWs may face challenges navigating complex systems, making it essential for the interface to be intuitive, visually appealing, and engaging. Poorly designed or overly complicated interfaces can act as significant barriers to adoption and may ultimately undermine the effectiveness of the intervention.

Among the 12 studies reviewed, only 2 [65,69] explicitly addressed the challenges associated with integrating gamification elements into broader ICT interventions. One study noted that incorporating gamification elements, which are visual and acoustic, into a digital data collection tool can drain the device's power rapidly, potentially reducing its appeal [69]. Another study in Portugal found that requiring HCWs to engage with gamification elements like badges, virtual goods, and content unlocking outside working hours decreased their enthusiasm [65]. The same study suffered significant setbacks due to technological issues, which hampered an evaluation of the benefits of gamification.

The difficulties encountered in these studies exemplify the importance of careful design in GDTs: a technologically flawed product can compromise all the benefits of gamification, and a solution not designed with users in mind can lead to resistance from them. Researchers and developers must ensure that the gamification elements are relevant and engaging for diverse user groups, as many factors can influence attitudes toward GDTs, including age, education level, work experience, culture, and computer skills [74]. Although gamification design is fundamentally user-centered [33], none of the interventions reviewed mentioned involving users in the design process.

Comparison to Prior Work

In this review, among the various gamification elements utilized and embedded in digital health interventions, feedback emerged as the most commonly used component. This is consistent with previous reviews on the use of gamification in other domains [11,75]. Sardi et al [11] and Edwards et al [75] reported that feedback was used in over 90% (n=60) of the studies they reviewed on eHealth and health promotion, respectively. The present review also highlights that performance feedback enhances HCWs' experience and engagement in their tasks, leading to behavioral change [40,75].

Competition emerged as the second most utilized element of gamification in 50% (n=6) of the included interventions, corroborating findings from a previous systematic review on gamification for health profession education [18]. Dashboards, leaderboards, and scoreboards were the most commonly used elements to convey feedback and foster competition. However, the use of external rewards, such as feedback, competitions, and leaderboards, to motivate HCWs can present several important limitations, as suggested by several studies. Although such incentives may be effective in generating short-term engagement, they risk becoming repetitive or losing their appeal over time. More critically, they may fail to foster sustained engagement or promote intrinsic motivation, which are essential for long-term behavioral change and integration into routine practice [11,12,28,76]. These extrinsic motivators are highly effective at attracting initial participation, but they might not support enduring involvement as well as intrinsic motivators. Therefore, in the design of GDTs, striking a balance between extrinsic and intrinsic motivations is pivotal to maintaining engagement and fostering meaningful participation [28].

XSL•FO
RenderX

Regrettably, none of the studies analyzed in this review investigate the long-term impact of gamification or identify which specific gamification elements are most effective compared to others, limiting our ability to draw conclusions about the sustained benefits and potential limitations of GDTs.

Furthermore, although gamification is no longer a new concept in ICT, none of the studies reviewed addressed potential ethical issues. Although gamification is generally well received by HCWs, gamification may lead to unintended consequences, including addiction, cheating, or compromised data quality, especially when points, badges, or leaderboards overshadow the intended purpose [77,78]. For example, rewarding speed over accuracy may result in errors. Competition, one of the most common gamification elements adopted, fosters a spirit of healthy competition among HCWs [79]. However, it also carries the risk of demoralizing workers due to feelings of irrelevance and oppression by other users and of moving the focus excessively on measurable outcomes rather than quality [77]. To mitigate this, competition should promote teamwork through group-based formats rather than individual performance. Additionally, gamification raises ethical concerns related to user autonomy and consent [74].

Despite ethical considerations being increasingly recognized in ICTs [80], the ethical discourse within the gamification field remains underdeveloped [80-82]. Authors like Kim and Werbach [80] urge caution to avoid manipulative or harmful design practices.

Strengths and Limitations

This study has 2 main strengths. First, by examining the current literature and available tools, we provided an initial overview of the potential effects of integrating gamified digital tools into the daily tasks of HCWs. Second, by conducting a scoping review to map the literature broadly, we have paved the way for future work to identify and select key outcomes for systematic reviews and meta-analyses in this domain more easily.

Our review also has certain limitations. First, it was challenging to identify studies that applied gamification to the daily tasks of HCWs, as many studies do not mention gamification explicitly. To locate potential studies, we thoroughly examined the methodologies and ICT interventions to pinpoint elements of gamification accurately. Second, most interventions researched and included in our review were pilots or case studies, characterized by small sample sizes and brief follow-ups. As a result, these studies have several limitations in evaluating long-term benefits that can affect their reliability and generalizability. Third, many studies do not rigorously describe or assess changes in motivation, behavior, or health outcomes. Fourth, given our goal to capture all relevant evidence and examples, we included a broad range of study designs, as is typical in a scoping review. Due to the rapid growth in digital health, prioritizing evidence based on quality criteria to capture scientifically based findings is challenging. Therefore, we applied the criteria developed by Sardi et al [11] to assess publications focusing on GDTs, ensuring the inclusion of evidence obtained through the most scientifically sound methodologies. Fifth, although our search was not limited to English, we were unable to find evidence in Spanish, French, or Italian. Lastly, only 1 author performed the data extraction, which may have introduced errors or biases. However, to ensure the accuracy of data extraction, all authors discussed which information should be extracted before the extraction process began. Subsequently, 2 authors carefully reviewed the extracted data to verify their accuracy.

Future Directions

Our study highlighted that no study to date has directly compared gamified digital tools with their nongamified counterparts to assess their impact on HCWs' daily tasks. Future research would benefit from adopting a 3-arm study design to better isolate the effects of gamification: (1) a control group without any digital tool, (2) a group using a nongamified digital tool, and (3) a group using a gamified digital tool. Furthermore, studies should incorporate longer follow-up periods and larger sample sizes to overcome the evident limitations identified in this scoping review, namely small sample sizes, lack of comparative groups, and short study durations. Extended follow-ups and broader participant inclusion would not only strengthen the robustness of findings but also allow the detection of potential tool misuse and ethical breaches that may emerge over time. Such an approach would enable a more comprehensive evaluation of whether gamification truly adds value beyond the use of digital tools alone and whether its effects are sustainable and ethically sound.

Conclusion

The integration of GDTs into the daily workflows of HCWs holds significant potential to enhance engagement, motivation, and overall performance across a range of health-related tasks. However, this area remains largely underexplored, with limited empirical evidence to support its systematic implementation. This gap needs more robust comparative studies to provide precise insights into the benefits and limitations of GDTs.

Therefore, due to the limited evidence base, gamification remains less developed than other components of digital health programming, such as scalability, interoperability, or digital health landscape assessment, which are often supported by established frameworks and tools. Currently, there are no formal guidelines or standardized frameworks to assist developers and policymakers in effectively integrating GDTs into broader digital strategies.

Additionally, the lack of such guidance raises the risk of ethical misuse, such as manipulation, coercion, or unintended negative effects. Therefore, clear design principles and ethical safeguards are necessary to ensure that gamification boosts engagement without harming professional integrity or user autonomy.

XSL•FO
RenderX

Acknowledgments

The authors would like to acknowledge Katharina Heldt for her guidance in supporting rigorous data searches to ensure the inclusion of all relevant studies. Funding for the development of this manuscript was received from the Deutsche Gesellschaft für Internationale Zusammenarbeit (German Corporation for International Cooperation, GIZ) as part of the Digital Innovation in Pandemic Control (DIPC) Initiative, financed by the Bundesministerium für Wirtschaftliche Zusammenarbeit (Federal Ministry for Economic Cooperation and Development, BMZ).

Data Availability

All data generated or analyzed during this study are included in this paper and its multimedia appendix files.

Authors' Contributions

Conceptualization: AB (Lead), BP (Lead), HAS (supporting)
Data curation: BP
Formal analysis: AB (supporting), BP (lead), HAS (supporting)
Funding acquisition: TL, CEB
Investigation: AB (supporting), BP (lead), HAS (supporting)
Methodology: AB (supporting), BP (lead), HAS (supporting)
Project administration: AB, CEB
Resources: AB
Supervision: AB
Validation: AB, HAS
Visualization: BP (lead)
Writing – original draft: AB (supporting), BP (lead)
Writing – review & editing: AB (supporting), BP (lead), CEB (supporting), HAS (supporting), RD (supporting), TL (supporting)

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA-ScR checklist.
[DOCX File, 20 KB - games_v13i1e70480_app1.docx]

Multimedia Appendix 2
Search strategy.
[DOCX File, 17 KB - games_v13i1e70480_app2.docx]

Multimedia Appendix 3
The quality assessment checklist and results.
[DOCX File, 20 KB - games_v13i1e70480_app3.docx]

References

1. Alhassan RK, Spieker N, van Ostenberg P, Ogink A, Nketiah-Amponsah E, de Wit TFR. Association between health worker motivation and healthcare quality efforts in Ghana. Hum Resour Health 2013 Aug 14;11(1-11):37. [doi: 10.1186/1478-4491-11-37] [Medline: 23945073]
2. Karaferis D, Aletras V, Raikou M, Niakas D. Factors influencing motivation and work engagement of healthcare professionals. Mater Sociomed 2022 Sep;34(3):216-224. [doi: 10.5455/msm.2022.34.216-224] [Medline: 36310751]
3. Abdul-Rahman T, Ghosh S, Lukman L, et al. Inaccessibility and low maintenance of medical data archive in low-middle income countries: mystery behind public health statistics and measures. J Infect Public Health 2023 Oct;16(10):1556-1561. [doi: 10.1016/j.jiph.2023.07.001] [Medline: 37566992]
4. Ibrahim H, Liu X, Zariffa N, Morris AD, Denniston AK. Health data poverty: an assailable barrier to equitable digital health care. Lancet Digit Health 2021 Apr;3(4):e260-e265. [doi: 10.1016/S2589-7500(20)30317-4] [Medline: 33678589]
5. World health organization. Global strategy on digital health 2020-2025. 2021. URL: https://iris.who.int/handle/10665/344249 [accessed 2025-04-18]
6. Digital health market size, share & trends analysis report by technology (healthcare analytics, mhealth), by component (hardware, software, services), by application, by end use, by region, and segment forecasts, 2025 - 2030. : Grand View Research URL: https://www.grandviewresearch.com/industry-analysis/digital-health-market [accessed 2025-04-18]

XSL•FO
RenderX

7. Koivisto J, Hamari J. The rise of motivational information systems: a review of gamification research. Int J Inf Manage 2019 Apr;45(45):191-210. [doi: 10.1016/j.ijinfomgt.2018.10.013]
8. Dreimane S. Gamification before its definition—an overview of its historical development. Presented at: 15th International Technology, Education and Development Conference; Mar 8-9, 2021; Online Conference p. 7187-7193. [doi: 10.21125/inted.2021.1434]
9. Khaitova NF. History of gamification and its role in the educational process. IJMMU 2021;8(5):212. [doi: 10.18415/ijmmu.v8i5.2640]
10. Kim B. The Popularity of Gamification in the Mobile and Social Era: Library Technology Reports; 2015, Vol. 51:5-9 URL: https://journals.ala.org/index.php/ltr/article/view/5628 [accessed 2024-02-10]
11. Sardi L, Idri A, Fernández-Alemán JL. A systematic review of gamification in e-health. J Biomed Inform 2017 Jul;71(31-48):31-48. [doi: 10.1016/j.jbi.2017.05.011] [Medline: 28536062]
12. Deterding S, Khaled R, Nacke L. Gamification: toward a definition. In: Dixon D, editor. Presented at: CHI 2011 Gamification Workshop Proceedings 2011 Annual Conference on Human Factors in Computing Systems (CHI'11); May 7-12, 2011; Vancouver, Canada URL: http://gamification-research.org/wp-content/uploads/2011/04/02-Deterding-Khaled-Nacke-Dixon.pdf [accessed 2024-01-12]
13. Brühlmann F, Mekler E, Opwis K. Gamification from the Perspective of Self-Determination Theory and Flow: University of Basel; 2013. [doi: 10.13140/RG.2.1.1181.8080]
14. Zichermann G, Cunningham C. Gamification by Design: Implementing Game Mechanics in Web and Mobile Apps, 1st edition: O'Reilly Media, Inc; 2011.
15. Bunchball I. Gamification 101: an introduction to the use of game dynamics to influence behavior. White paper 2010:1-18 [FREE Full text]
16. Darejeh A, Salim SS. Gamification solutions to enhance software user engagement—a systematic review. Int J Hum Comput Interact 2016 Aug 2;32(8):613-642. [doi: 10.1080/10447318.2016.1183330]
17. Swacha J, editor. Gamification in enterprise information systems: what, why and how. Presented at: 2016 Federated Conference on Computer Science and Information Systems; Sep 11-14, 2016; Gdansk, Poland. [doi: 10.15439/2016F460]
18. van Gaalen AEJ, Brouwer J, Schönrock-Adema J, Bouwkamp-Timmer T, Jaarsma ADC, Georgiadis JR. Gamification of health professions education: a systematic review. Adv in Health Sci Educ 2021 May;26(2):683-711. [doi: 10.1007/s10459-020-10000-3]
19. Xu M, Luo Y, Zhang Y, Xia R, Qian H, Zou X. Game-based learning in medical education. Front Public Health 2023;11(1113682):1113682. [doi: 10.3389/fpubh.2023.1113682] [Medline: 36935696]
20. Shortt M, Tilak S, Kuznetcova I, Martens B, Akinkuolie B. Gamification in mobile-assisted language learning: a systematic review of Duolingo literature from public release of 2012 to early 2020. Comput Assist Lang Learn 2023 Mar 4;36(3):517-554. [doi: 10.1080/09588221.2021.1933540]
21. Kim S, Song K, Lockee B, Burton J. What Is Gamification in Learning and Education? Gamification in Learning and Education: Enjoy Learning like Gaming: Springer; 2017:25-38. [doi: 10.1007/978-3-319-47283-6_4]
22. Conaway R, Garay MC. Gamification and service marketing. Springerplus 2014;3(1):653. [doi: 10.1186/2193-1801-3-653] [Medline: 25392812]
23. Berglund A, Jaarsma T, Berglund E, Strömberg A, Klompstra L. Understanding and assessing gamification in digital healthcare interventions for patients with cardiovascular disease. Eur J Cardiovasc Nurs 2022 Aug 29;21(6):630-638. [doi: 10.1093/eurjcn/zvac048] [Medline: 35709297]
24. Dithmer M, Rasmussen JO, Grönvall E, et al. "The Heart Game": using gamification as part of a telerehabilitation program for heart patients. Games Health J 2016 Feb;5(1):27-33. [doi: 10.1089/g4h.2015.0001] [Medline: 26579590]
25. Radhakrishnan K, Toprac P, O'Hair M, et al. Interactive digital e-health game for heart failure self-management: a feasibility study. Games Health J 2016 Dec;5(6):366-374. [doi: 10.1089/g4h.2016.0038] [Medline: 27976955]
26. Chen Q, Ji H, Shang Q. The association between the diversity of online activities on smartphones and cognitive function among middle-aged and elderly Chinese adults. BMC Public Health 2024;24(1):38383336. [doi: 10.1186/s12889-024-17932-0]
27. Mordor Intelligence. Gamification market—industry trends, size & report 2024. 2024 Feb 2. URL: https://www.mordorintelligence.com/industry-reports/gamification-market [accessed 2024-02-08]
28. Seaborn K, Fels DI. Gamification in theory and action: a survey. Int J Hum Comput Stud 2015 Feb;74:14-31. [doi: 10.1016/j.ijhcs.2014.09.006]
29. Jahn K, Kordyaka B, Machulska A, et al. Individualized gamification elements: the impact of avatar and feedback design on reuse intention. Comput Human Behav 2021 Jun;119:106702. [doi: 10.1016/j.chb.2021.106702]
30. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000 Jan;55(1):68-78. [doi: 10.1037//0003-066x.55.1.68] [Medline: 11392867]
31. Anunpattana P, Khalid MNA, Iida H, Inchamnan W. Capturing potential impact of challenge-based gamification on gamified quizzing in the classroom. Heliyon 2021 Dec;7(12):e08637. [doi: 10.1016/j.heliyon.2021.e08637] [Medline: 35005284]

XSL•FO
RenderX

32. Arya V, Sambyal R, Sharma A, Dwivedi YK. Brands are calling your AVATAR in Metaverse—a study to explore XR ‐ based gamification marketing activities & consumer ‐ based brand equity in virtual world. J of Consumer Behaviour 2024 Mar;23(2):556-585. [doi: 10.1002/cb.2214]
33. Binsar F, Riantono IE, Kartono R, Bandur A. Gamification to grow motivation for interactive engagement of health nurses in using health information systems: a conceptual framework. Presented at: The Asia Pacific Computer Systems Conference; Feb 18, 2023; North Carolina, USA. [doi: 10.1007/978-981-19-7904-0_10]
34. Dale S. Gamification: making work fun, or making fun of work? Bus Inf Rev 2014;31(2):82-90. [doi: 10.1177/0266382114538350]
35. Mendoza K, Villalobos-Daniel VE, Jáuregui A, et al. Development of a crowdsourcing- and gamification-based mobile application to collect epidemiological information and promote healthy lifestyles in Mexico. Sci Rep 2024 Mar 14;14(1):6174. [doi: 10.1038/s41598-024-56761-4] [Medline: 38486091]
36. Miller AS, Cafazzo JA, Seto E. A game plan: gamification design principles in mHealth applications for chronic disease management. Health Informatics J 2016 Jun;22(2):184-193. [doi: 10.1177/1460458214537511] [Medline: 24986104]
37. Pesare E, Roselli T, Corriero N, Rossano V. Game-based learning and gamification to promote engagement and motivation in medical learning contexts. Smart Learn Environ 2016 Dec;3(1):1-21. [doi: 10.1186/s40561-016-0028-0]
38. Jun M, Lucas T. Gamification elements and their impacts on education: a review. Multidiscip Rev 2025;8(5):2025155. [doi: 10.31893/multirev.2025155]
39. Tuah NM, Ahmedy F, Gani A, Yong LN. A survey on gamification for health rehabilitation care: applications, opportunities, and open challenges. Information 2021;12(2):91. [doi: 10.3390/info12020091]
40. Wang YF, Hsu YF, Fang K. The key elements of gamification in corporate training—the Delphi method. Entertain Comput 2022 Jan;40:100463. [doi: 10.1016/j.entcom.2021.100463] [Medline: 35991001]
41. Hervás R, Ruiz-Carrasco D, Mondéjar T, Bravo J. Gamification mechanics for behavioral change: a systematic review and proposed taxonomy. Presented at: Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare; May 23, 2017; Barcelona, Spain p. 395-404. [doi: 10.1145/3154862.3154939]
42. Schmidt-Kraepelin M, Thiebes S, Tran MC, Sunyaev A. What's in the game? Developing a taxonomy of gamification concepts for health apps. Presented at: Hawaii International Conference on System Sciences; Jan 3-6, 2018; Waikoloa Village, HI p. 1217-1226. [doi: 10.24251/HICSS.2018.150]
43. Toda AM, Klock ACT, Oliveira W, et al. Analysing gamification elements in educational environments using an existing gamification taxonomy. Smart Learn Environ 2019 Dec;6(1):1-4. [doi: 10.1186/s40561-019-0106-1]
44. Subhash S, Cudney EA. Gamified learning in higher education: a systematic review of the literature. Comput Human Behav 2018 Oct;87:192-206. [doi: 10.1016/j.chb.2018.05.028]
45. Pereira P, Duarte E, Rebelo F, Noriega P, editors. A review of gamification for health-related contexts. Presented at: Design, User Experience, and Usability User Experience Design for Diverse Interaction Platforms and Environments: Third International Conference, DUXU 2014, Held as Part of HCI International 2014; Jun 22-27, 2014; Heraklion, Crete, Greece. [doi: 10.1007/978-3-319-07626-3_70]
46. Wirkkala M, Wijk K, Larsson AC, Engström M. Technology frustration in healthcare—does it matter in staff ratings of stress, emotional exhaustion, and satisfaction with care? A cross-sectional correlational study using the job demands-resources theory. BMC Health Serv Res 2024 Dec 6;24(1):1557. [doi: 10.1186/s12913-024-11906-z] [Medline: 39643868]
47. Pavuluri S, Sangal R, Sather J, Taylor RA. Balancing act: the complex role of artificial intelligence in addressing burnout and healthcare workforce dynamics. BMJ Health Care Inform 2024 Aug 24;31(1):e101120. [doi: 10.1136/bmjhci-2024-101120] [Medline: 39181545]
48. Shan Y, Shang J, Yan Y, Ye X. Workflow interruption and nurses' mental workload in electronic health record tasks: an observational study. BMC Nurs 2023 Mar 9;22(1):63. [doi: 10.1186/s12912-023-01209-9] [Medline: 36890555]
49. Tricco AC, Lillie E, Zarin W, et al. PRISMA extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 2;169(7):467-473. [doi: 10.7326/M18-0850] [Medline: 30178033]
50. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan—a web and mobile app for systematic reviews. Syst Rev 2016 Dec 5;5(1):210. [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
51. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159-174. [doi: 10.2307/2529310] [Medline: 843571]
52. Cooke A, Smith D, Booth A. Beyond PICO: the SPIDER tool for qualitative evidence synthesis. Qual Health Res 2012 Oct;22(10):1435-1443. [doi: 10.1177/1049732312452938] [Medline: 22829486]
53. Principles for digital development. The principles for digital development: a compass for those working to promote sustainable and inclusive development in today's complex digital landscape. 2024 Mar 25. URL: https://digitalprinciples.org [accessed 2024-02-27]
54. Held B, Moriarty B, Richardson T. Microsoft Excel functions and formulas with Excel 2019/Office 365: mercury learning and information. 2019. URL: https://terrorgum.com/tfox/books/microsoftexcelfunctionsandformulas_5e.pdf [accessed 2024-05-27]
55. SJR. Scimago Journal & Country Rank 2022. 2022 May 20. URL: https://www.scimagojr.com [accessed 2024-05-27]

XSL•FO
RenderX

56. Raikhel AV, Blau K, Alberty K, et al. Vax the max, a gamification intervention for COVID-19 vaccination task engagement in the inpatient setting. Am J Med Qual 2023;38(1):47-56. [doi: 10.1097/JMQ.0000000000000094] [Medline: 36472420]
57. Abdul Rahman MH, Ismail Yusuf Panessai I, Mohd Noor NAZ, Mat Salleh NS. Gamification elements and their impacts on teaching and learning—a review. IJMA 2018 Dec 31;10(6):37-46 [FREE Full text] [doi: 10.5121/ijma.2018.10604]
58. Krishnamurthy K, Selvaraj N, Gupta P, et al. Benefits of gamification in medical education. Clin Anat 2022 Sep;35(6):795-807. [doi: 10.1002/ca.23916] [Medline: 35637557]
59. King CE, Kells A, Trout L, Yirinec A, Zhou S, Zurca AD. Gamification educational intervention improves pediatric nurses' comfort and speed drawing up code-dose epinephrine. J Pediatr Nurs 2023;71(55-9):55-59. [doi: 10.1016/j.pedn.2023.03.013] [Medline: 37004310]
60. McKeown S, Krause C, Shergill M, Siu A, Sweet D. Gamification as a strategy to engage and motivate clinicians to improve care. Healthc Manage Forum 2016 Mar;29(2):67-73. [doi: 10.1177/0840470415626528] [Medline: 26872801]
61. Orchard J, Li J, Gallagher R, Freedman B, Lowres N, Neubeck L. Uptake of a primary care atrial fibrillation screening program (AF-SMART): a realist evaluation of implementation in metropolitan and rural general practice. BMC Fam Pract 2019 Dec 6;20(1):170. [doi: 10.1186/s12875-019-1058-9] [Medline: 31810441]
62. Orwoll B, Diane S, Henry D, et al. Gamification and microlearning for engagement with quality improvement (GAMEQI): a bundled digital intervention for the prevention of central line-associated bloodstream infection. Am J Med Qual 2018;33(1):21-29. [doi: 10.1177/1062860617706542] [Medline: 28482161]
63. Collins A, Moore T, Davenport P. Targeting the competitive spirit to diminish nosocomial infection markers. Am J Infect Control 2013 Jun;41(6):S51-S52. [doi: 10.1016/j.ajic.2013.03.109] [Medline: 22939417]
64. Owens A, Demmert A, Abdollahian D, et al. 3:00 PM Abstract No. 141 Interventional radiology patient throughput: application of lean sigma methodology and real-time delay dashboard to improve operations and patient experience. J Vasc Interv Radiol 2018 Apr;29(4):S63-S64. [doi: 10.1016/j.jvir.2018.01.160]
65. Marques R, Gregório J, Pinheiro F, Póvoa P, da Silva MM, Lapão LV. How can information systems provide support to nurses' hand hygiene performance? Using gamification and indoor location to improve hand hygiene awareness and reduce hospital infections. BMC Med Inform Decis Mak 2017 Jan 31;17(1):15. [doi: 10.1186/s12911-017-0410-z] [Medline: 28143613]
66. Luedtke S, Wood C, Olufemi O, et al. Gamified antimicrobial decision support app (GADSA) changes antibiotics prescription behaviour in surgeons in Nigeria: a hospital-based pilot study. Antimicrob Resist Infect Control 2023 Dec 6;12(1):141. [doi: 10.1186/s13756-023-01342-9] [Medline: 38053212]
67. Shah P, Madhiwala N, Shah S, et al. High uptake of an innovative mobile phone application among community health workers in rural India: an implementation study. Natl Med J India 2019;32(5):262-269. [doi: 10.4103/0970-258X.295956] [Medline: 32985439]
68. Zaidi S, Shaikh SA, Sayani S, et al. Operability, acceptability, and usefulness of a mobile app to track routine immunization performance in rural Pakistan: interview study among vaccinators and key informants. JMIR Mhealth Uhealth 2020 Feb 13;8(2):e16081. [doi: 10.2196/16081] [Medline: 32053115]
69. Wenner G, Marino M, Obeysekare E, Mehta K. Making data collection in low-resource contexts intuitive, fun, and interactive. In: Mehta K, editor. Presented at: 2014 IEEE Global Humanitarian Technology Conference (GHTC); Oct 10-13, 2014; San Jose, CA. [doi: 10.1109/GHTC.2014.6970260]
70. Little A, Medhanyie A, Yebyo H, Spigt M, Dinant GJ, Blanco R. Meeting community health worker needs for maternal health care service delivery using appropriate mobile technologies in Ethiopia. PLoS One 2013;8(10):e77563. [doi: 10.1371/journal.pone.0077563] [Medline: 24204872]
71. Finette BA, McLaughlin M, Scarpino SV, et al. Development and initial validation of a frontline health worker mhealth assessment platform (MEDSINC®) for children 2-60 months of age. Am J Trop Med Hyg 2019 Jun;100(6):1556-1565. [doi: 10.4269/ajtmh.18-0869] [Medline: 30994099]
72. Bai S, Hew KF, Huang B. Does gamification improve student learning outcome? Evidence from a meta-analysis and synthesis of qualitative data in educational contexts. Educational Research Review 2020 Jun;30:100322. [doi: 10.1016/j.edurev.2020.100322]
73. Zeybek N, Saygı E. Gamification in education: why, where, when, and how?—A systematic review. Games and Culture 2024 Mar;19(2):237-264. [doi: 10.1177/15554120231158625]
74. Kipturgo MK, Kivuti-Bitok LW, Karani AK, Muiva MM. Attitudes of nursing staff towards computerisation: a case of two hospitals in Nairobi, Kenya. BMC Med Inform Decis Mak 2014 Apr 29;14(1-8):35. [doi: 10.1186/1472-6947-14-35] [Medline: 24774008]
75. Edwards EA, Lumsden J, Rivas C, et al. Gamification for health promotion: systematic review of behaviour change techniques in smartphone apps. BMJ Open 2016 Oct 4;6(10):e012447. [doi: 10.1136/bmjopen-2016-012447] [Medline: 27707829]
76. Hamari J, Koivisto J, Sarsa H. Does gamification work?—A literature review of empirical studies on gamification. In: Hamari J, Koivisto J, Sarsa H, editors. Presented at: 2014 47th Hawaii International Conference on System Sciences (HICSS); Jan 6-9, 2014; Waikoloa, HI. [doi: 10.1109/HICSS.2014.377]

XSL•FO
RenderX

77. Andrade FR, Mizoguchi R, Isotani S. The bright and dark sides of gamification. Presented at: Intelligent Tutoring Systems: 13th International Conference, ITS 2016; Jun 7-10, 2016; Zagreb, Croatia. [doi: 10.1007/978-3-319-39583-8_17]
78. Hadi Mogavi R, Guo B, Zhang Y, Haq EU, Hui P, Ma X. When gamification spoils your learning: a qualitative case study of gamification misuse in a language-learning app. Presented at: Proceedings of the Ninth ACM Conference on Learning@ Scale; Jun 1, 2022; New York City NY USA. [doi: 10.1145/3491140.3528274]
79. Rigby S, Ryan RM. Glued to Games: How Video Games Draw Us in and Hold Us Spellbound: Bloomsbury Publishing USA; 2011:79.
80. Kim TW, Werbach K. More than just a game: ethical issues in gamification. Ethics Inf Technol 2016 Jun;18(2):157-173. [doi: 10.1007/s10676-016-9401-5]
81. Shahri A, Hosseini M, Phalp K, Taylor J, Ali R. Towards a code of ethics for gamification at enterprise. Presented at: The Practice of Enterprise Modeling: 7th IFIP WG 81 Working Conference, PoEM 2014; Nov 12-13, 2014; Manchester, UK. [doi: 10.1007/978-3-662-45501-2_17]
82. Mora A, Riera D, González C, Arnedo-Moreno J. Gamification: a systematic review of design frameworks. J Comput High Educ 2017 Dec;29(3):516-548. [doi: 10.1007/s12528-017-9150-4]

Abbreviations

AF: atrial fibrillation
CHW: community health care worker
CLABSI: Central Line-Associated Bloodstream Infections
EMR: electronic medical record
EPI: Expanded Programme on Immunization
GDT: gamified digital tools
HCW: health care worker
HIC: high-income country
ICT: information and communication technology
LMIC: low- and middle-income countries
PHC: primary health care
SAP: surgical antibiotic prophylaxis

Edited by A Coristine; submitted 23.12.24; peer-reviewed by JH Song, R Hervás; revised version received 29.08.25; accepted 29.08.25; published 20.10.25.

Please cite as:
Paudel B, Al-Shehabi H, Dörner R, El Bcheraoui C, Lennemann T, Bernasconi A
Use of Gamified Digital Tools in Daily Tasks of Health Care Workers: Scoping Review
JMIR Serious Games 2025;13:e70480
URL: https://games.jmir.org/2025/1/e70480
doi:10.2196/70480

© Binita Paudel, Hussein Al-Shehabi, Rita Dörner, Charbel El Bcheraoui, Tessa Lennemann, Andrea Bernasconi. Originally published in JMIR Serious Games (https://games.jmir.org), 20.10.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Efficacy of Virtual Reality–Based Interventions on Cognitive Function in Patients With Neuropsychiatric Disorders: Systematic Review and Meta-Analysis of Randomized Controlled Trials

Qiujing Du[1*]; Yuhan Wei[1*]; Yuexuan Ma[2]; Changqing Liu[1]; Shanshan Du[1]; Qi Zhang[2]; Xiaotong Gong[2]; Jiaju Yang[2]; Qijie Li[1*]; Ka Li[1*]

[1]Medicine and Engineering Interdisciplinary Research Laboratory of Nursing & Materials, West China Hospital/West China School of Nursing, Sichuan University, 37 Guoxue Lane, Wuhou District, Chengdu, China

[2]West China Hospital/West China School of Nursing, Sichuan University, Chengdu, China

[*]these authors contributed equally

Corresponding Author:
Ka Li
Medicine and Engineering Interdisciplinary Research Laboratory of Nursing & Materials, West China Hospital/West China School of Nursing, Sichuan University, 37 Guoxue Lane, Wuhou District, Chengdu, China

Abstract

Background: Virtual reality (VR) technology has emerged as a promising tool for cognitive rehabilitation in patients with neuropsychiatric disorders. These patients often endure significant cognitive impairments, which are associated with decreased quality of life and increased disease burden. Traditional treatments are limited in improving cognitive functions, making VR-based interventions an area of growing interest.

Objective: This meta-analysis aims to evaluate the efficacy of VR-based interventions on cognitive function in patients with neuropsychiatric disorders by synthesizing data from randomized controlled trials (RCTs).

Methods: Following PRISMA guidelines, we conducted a comprehensive search across PubMed, Web of Science, MEDLINE, EMBASE, and Cochrane Library for RCTs from January 2010 to December 2024. Studies were included if they evaluated the impact of VR-based interventions on cognitive outcomes in patients with neuropsychiatric disorders. Data extraction and risk of bias assessment were performed independently by 2 researchers. Meta-analyses were conducted using random-effects models, and standardized mean differences (SMDs) as effect size.

Results: A total of 21 RCTs involving 1051 participants were included. Overall, VR-based interventions significantly improved cognitive functions of patients with neuropsychiatric disorders (SMD 0.67, 95% CI 0.33-1.01, z=3.85; P<.001). Subgroup analyses revealed significant benefits for cognitive rehabilitation training (SMD 0.75, 95% CI 0.33-1.17, z=3.53; P<.001), exergame-based training (SMD 1.09, 95% CI 0.26-1.91, z=2.57; P=.01), and telerehabilitation and social functioning training (SMD 2.21, 95% CI 1.11-3.32, z=3.92; P<.001). Conversely, immersive cognitive training, music attention training, and vocational and problem-solving skills training did not yield significant improvements (z=1.86, P=.06; z=0.35, P=.72; z=0.88, P=.38; respectively). Disease-type subgroup analyses indicated significant improvements in schizophrenia (SMD 0.92, 95% CI 0.22-1.62, z=2.58; P=.01), and mild cognitive impairment (SMD 0.75, 95% CI 0.16-1.35, z=2.47; P=.01), but not in brain injuries, Parkinson disease, or stroke (z=0.34, P=.73; z=1.26, P=.21; z=1.16, P=.24; respectively).

Conclusions: This meta-analysis revealed that VR-based interventions can improve cognitive functions among individuals with neuropsychiatric disorders, with notable improvements observed in cognitive rehabilitation training, exergame-based training, and tele-rehabilitation and social functioning training. These results offer valuable evidence supporting the use of VR technology in rehabilitation for neuropsychiatric conditions and inform the optimization of future intervention approaches.

Trial Registration: PROSPERO CRD42023445000; https://www.crd.york.ac.uk/PROSPERO/view/CRD42023445000

(JMIR Serious Games 2025;13:e67501) doi:10.2196/67501

KEYWORDS

virtual reality; neuropsychiatric disorders; cognitive function; systematic review; meta-analysis; mobile phone

XSL•FO
RenderX

Introduction

"Patients with neuropsychiatric disorders" refer to individuals enduring neurological or psychiatric conditions, or those exhibiting symptoms pertinent to both neurology and psychiatry [1]. The illnesses of these patients typically involve dysfunction within the nervous system (including the brain), accompanied by psychological or behavioral abnormalities [2]. Cognitive impairment, a prevalent symptom in these patients with an incidence exceeding 70%, is characterized by a marked and measurable decline in cognitive abilities compared to previous levels of function. This can manifest as deficits in higher-order brain functions such as learning, memory, problem-solving, and decision-making [3,4]. Research has confirmed that the occurrence of cognitive impairments in patients with neuropsychiatric disorders is significantly associated with decreased quality of life and increased disease burden [5,6], which may lead to more complex health challenges and heightened care requirements, imposing substantial pressure on the patients themselves, their families, and society at large. Therefore, there is an urgent need to develop relevant strategies to address these significant challenges, aiming to improve cognitive function and enhance the quality of life for these patients.

Traditional intervention strategies for cognitive function in patients with neuropsychiatric disorders primarily include pharmacological treatment and cognitive rehabilitation. Pharmacological treatments are commonly used to alleviate psychiatric symptoms but have limited efficacy in improving cognitive functions [7]. Cognitive rehabilitation is considered a promising nonpharmacological intervention, aiming to enhance patients' cognitive functions through systematic training and practice. Studies have shown that cognitive rehabilitation can effectively improve cognitive domains such as attention, executive function, and working memory in these patients [8]. However, the implementation of traditional cognitive rehabilitation may face challenges in resource-limited settings, and the transfer effect of cognitive enhancements to functional results in individuals' everyday lives is generally insufficient. Moreover, a lack of motivation among participants represents one of the key barriers to achieving effective therapeutic goals in traditional cognitive rehabilitation [9,10].

Virtual reality (VR) and augmented reality (AR) are emerging technologies with significant potential to address the challenges associated with cognitive rehabilitation. VR technology is an advanced system that integrates information and computer technologies to create lifelike web-based environments using computer-generated imagery and digital interactive devices. It is designed to provide users with an immersive experience, enabling them to not only observe passively but to also interact actively with the digital environment [11]. VR technology could provide a fully controllable and secure setting, supporting remote access and offering a more ecologically valid and effective environment for cognitive rehabilitation [12]. Second, it can simulate multimodal scenarios highly similar to those that patients might encounter in their daily lives, making it easier to integrate cognitive training with everyday functioning [13]. Third, its high level of interactivity and game-inspired design can help increase patient participation [14]. Compared to VR, fewer studies have explored the use of AR in cognitive rehabilitation. AR, by integrating virtual elements with the physical world, carries a lower risk of adverse events, such as falls or dizziness, making it a safer alternative in some contexts [15]. However, immersive VR provides a higher level of engagement and cognitive stimulation, which is particularly beneficial in exergame-based training. To address the safety concerns associated with immersive VR, specific measures are implemented, such as having medical personnel present during sessions and limiting session duration to mitigate dizziness [16-18]. These precautions ensure the safe and effective use of immersive VR in cognitive rehabilitation. A recent meta-analysis has shown that VR-based interventions have a beneficial effect on overall cognition and executive function in older adults [19]. Additionally, a previous narrative review indicated promising results of VR-based cognitive rehabilitation in executive function and visuospatial abilities [20], nevertheless, this review focused on patients with neurological disorders and did not conduct a quality assessment or quantitative synthesis of the studies reviewed. Moreover, different types of cognitive rehabilitation interventions have varying impacts on cognitive functions, and their interaction effects with VR technology differ [21].

Based on this, our study consolidates data from randomized controlled trials (RCTs) of VR-based interventions aimed at improving cognitive function in patients with neuropsychiatric disorders, to evaluate the efficacy of interventions and quality of these RCTs. The objectives of our study are (1) to quantitatively investigate the efficacy of VR-based interventions on the cognitive function of patients with neuropsychiatric disorders and (2) to assess the effects of different types of interventions and analyze possible reasons for any differences observed.

Methods

Study Design

The design, execution, and reporting of this meta-analysis adhered to the PRISMA guidelines of the 2020 statement [22]. This study has been registered at PROSPERO, and the registration number is CRD42023445000.

Ethical Considerations

No humans or animals were involved in this meta-analysis, so ethical approval was waived.

Search Strategy

An exhaustive search was systematically performed across multiple databases, including PubMed, Web of Science, MEDLINE, EMBASE, and Cochrane Library. In light of the progressive maturation of VR technology subsequent to 2010 and the increased scholarly attention toward exploring VR as a modality for supporting psychological and cognitive interventions during this period [23], the timeframe for our literature search shall be delimited from January 2010 to December 2024. The used search methodology was grounded in the PICOS framework, encompassing the specific elements of Patient or Population, Intervention, Comparison, Outcome,

XSL•FO
RenderX

and Study type. No specific restrictions were applied regarding demographic variables such as age, gender, education, etc. However, for the language, the search was conducted exclusively in the English language. The search methodology used multiple strategies, including examination of titles and abstracts, use of relevant keywords, and incorporation of free and medical subject headings terms, etc. The specific retrieval strategies for each database are outlined in Multimedia Appendix 1.

Criteria for Study Eligibility

Inclusion criteria for our review encompassed the following criteria: (1) participants were diagnosed with neuropsychiatric diseases; (2) interventions used the VR technology; (3) the control groups were assigned to either receive conventional treatment or were placed on a waitlist for comparison purposes; (4) the study quantitatively assessed the cognitive function. No restrictions were imposed on the specific measurement tools or instruments used to evaluate the indices; (5) the literature was published in English; and (6) the included studies were designed per the principles of RCTs. Exclusion criteria for this study encompassed the following criteria: (1) qualitative study, conference paper, letter, review, commentary, or protocol; or (2) studies where access to full-text articles was not available or where relevant full data were not available.

Study Selection and Data Extraction

All identified study titles and abstracts were retrieved and imported into Zotero 7.0 (Corporation for Digital Scholarship). Duplicate papers were subsequently removed from the dataset. Two researchers independently assessed the titles and abstracts of the studies in parallel to ensure compliance with the inclusion criteria. Subsequently, the full texts of each study were reviewed per the predetermined eligibility criteria. Discrepancies were resolved through a consensus-seeking process involving discussion and consultation with a third investigator.

Two researchers independently performed data extraction from the selected studies and documented the relevant information within a pre-established data extraction form, encompassing the following essential components: the first author, the published year, the country or region where the trial was conducted, the disease type, the ages (mean and SD), the years of education (mean and SD); the sample size (experimental or control group), and the gender; the type, content, hardware, software, immersive or nonimmersive approach and detailed session information of the intervention; and the assessment time and the instruments of outcome measures.

Quality Assessment

Using the Cochrane RoB2 (Risk of Bias 2) tool, 2 investigators carried out a separate evaluation of the methodological rigor and potential bias of the selected studies. RoB2 is a systematic review tool developed by the Cochrane Risk of Bias Working Group aimed at assessing the methodological quality of RCTs [24]. This tool identifies potential sources of bias that may affect the reliability of study outcomes through a series of guiding questions that cover key areas such as the randomization process, deviations from intended interventions, missing outcome data, outcome measurement, and selective reporting. It provides a final judgment on the risk of bias. Third-party consultation was used to resolve disagreements.

Data Synthesis and Analysis

We used Review Manager 5.3 (Cochrane) to conduct our statistical analyses. For continuous outcome indexes, the mean difference along with a 95% CI was used when all selected studies reported outcomes using the same unit and magnitude; otherwise, the standard mean difference was used to express the results.

Initially, the chi-square test was used to assess the statistical heterogeneity among the outcomes reported in each individual study, accompanied by the calculation of the corresponding *P* value; to quantitatively evaluate the magnitude of heterogeneity, the I^2 test was used [25]. Given the likelihood of functional disparities among studies conducted independently by different researchers, a random-effects meta-analysis was conducted for the outcome to account for potential heterogeneity across the included studies. Additionally, we conducted a subgroup analysis based on different types of interventions. To explore the potential reasons for the differences in the integrated outcomes of various intervention types based on VR technology, we conducted another subgroup analysis for different types of neuropsychiatric disorders and carried out sensitivity analyses by excluding certain studies. We constructed a funnel plot to evaluate potential publication bias systematically across the studies. In the final stage, forest plots were generated to visually present the effect estimates of outcome variables across all included studies. A level of statistical significance was established at $P<.05$.

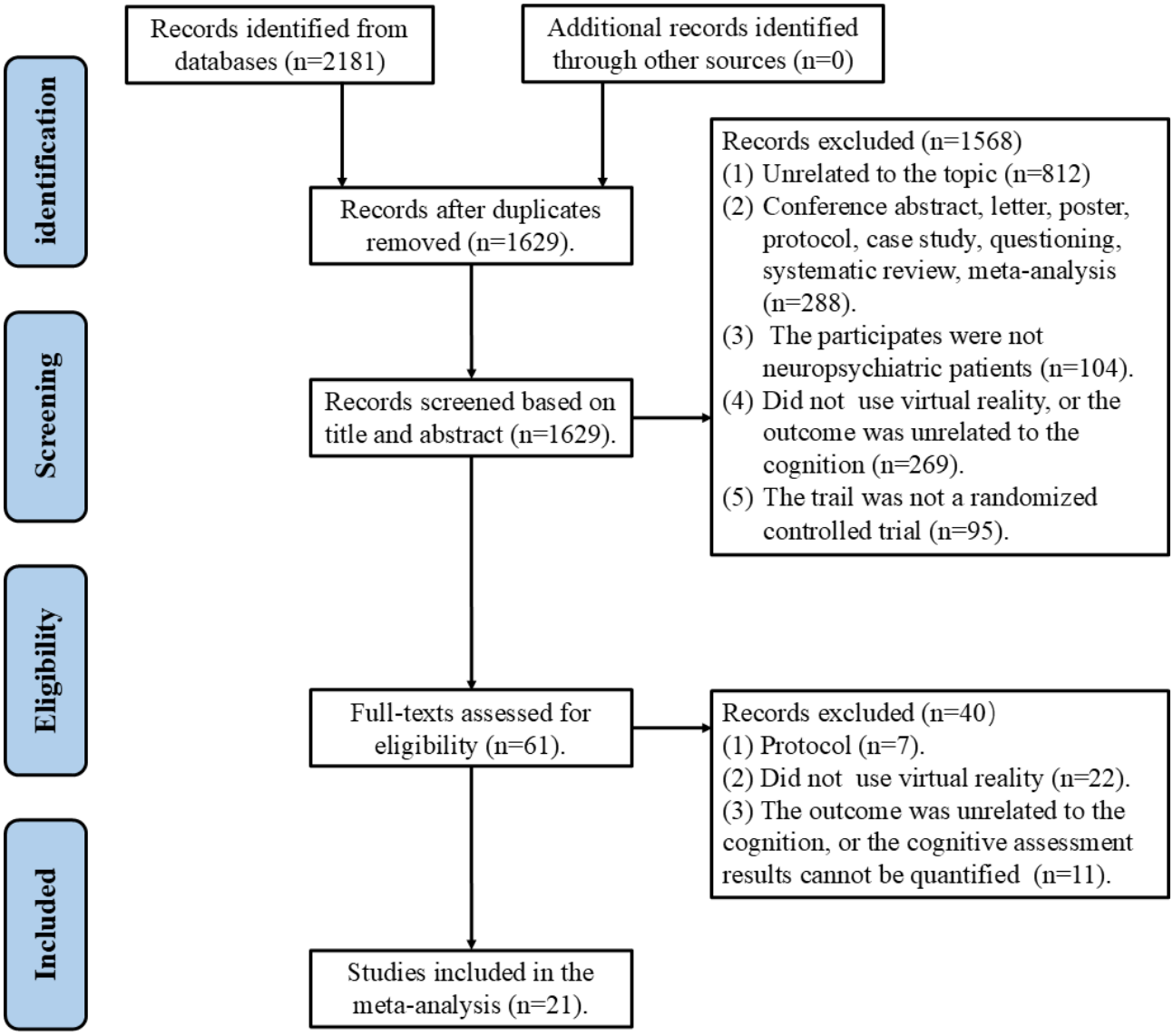
Results

Search Results and Selection

According to the 5 electronic databases, the initial search yielded a total of 2181 documents. Following a deduplication process, 552 redundant documents were eliminated. The exclusion of 1568 articles was based on a thorough evaluation of their titles and abstracts. Subsequently, the full texts of the remaining 61 articles underwent rigorous examination. Forty studies were excluded after full-text assessment for the following reasons: 7 were protocol studies, 22 did not employ VR technology, and 11 had outcomes unrelated to cognition or produced cognitive assessment results that could not be quantified. Consequently, 21 studies were included in the meta-analysis. A visual representation of the search and selection processes is provided in Figure 1. The unweighted Cohen κ values for study selection (0.86), data extraction (0.84), and risk of bias assessment (0.81) indicate an excellent level of interrater agreement, as per Cochrane Handbook criteria.

XSL•FO
RenderX

Figure 1. PRISMA flowchart. PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.



Description of Studies in This Review

Study Characteristics

The salient attributes of the 21 studies that were incorporated into the analysis are documented in Multimedia Appendix 2. Each of these studies adhered to the rigorous design of RCTs and was published within the time frame spanning from 2010 to 2024. The review encompassed a total of 12 nations or regions, among which Korea, Hong Kong, China, and Taiwan, China, were notably included.

Characteristics of Patients

A total of 1051 patients were part of our research, with sample sizes varying from 17 [26] to 293 [17] individuals in each respective study. The mean age of the patients exhibited considerable variation, ranging from 14.81 (SD 1.74) years [27] to 77.35 (SD 6.75) years [18]. The ratio of males to females among the participants was nearly 1:2, with 356 males and 695 females; and the participants' mean years of education spanned from 4.76 (SD 3.25) years [18] to 15.09 (SD 4.74) years [28] across the included studies. The participants' disease types included schizophrenia (n=2, 9.5%), brain injury (n=2, 9.5%), Parkinson disease (n=3, 14.3%), mild cognitive impairment (n=8, 38.1%), stroke (n=3, 14.3%), and other neuropsychiatric disorders (n=3, 14.3%).

Characteristics of VR-Based Interventions

According to the goals and content of training or treatment, the intervention types in this review can be divided into 6 categories, involving cognitive rehabilitation training (n=8, 38.1%), immersive cognitive training (n=6, 28.6%), exergame-based training (n=4, 19.0%), music attention training (n=1, 4.8%), telerehabilitation and social functioning training (1, 4.8%), and vocational and problem-solving skills training (n=1, 4.8%).

For cognitive rehabilitation training, all exercises adopt a task-oriented learning model, meaning that specific goals and tasks are designed that patients need to complete to achieve rehabilitation objectives. Moreover, one of the characteristics of this type of intervention is the multimodal stimulation empowered by VR technology, implying that training is not limited to a single sensory input (such as visual) but integrates various sensory stimuli (such as visual, auditory, and tactile inputs) [26-33]. For instance, in the "ball and bird" activity of the IREX system (GESTURETEK HEALTH), users not only

XSL•FO
RenderX

see the color of the ball but also experience different responses based on varying contact pressures [31]. Some training programs particularly emphasize combining physical movement with cognitive tasks, such as reciting poetry or doing mathematics problems while walking, promoting overall cognitive rehabilitation [32].

For immersive cognitive training, all exercises use highly immersive VR equipment [18,34-38]. The equipment primarily consists of head-mounted displays and wireless controllers or joysticks. Devices such as the Oculus Quest 2 (Meta Platforms, Inc) and HTC Vive (HTC Corporation) provide high-resolution visual output, capable of creating realistic virtual worlds. Wireless controllers or joysticks are used for interaction within the virtual environment, such as grabbing objects, operating machines, or navigating. Through haptic feedback (vibration or other physical signals), these devices simulate a genuine sense of touch, offering users a more direct operational experience within the virtual environment.

For exergame-based training, all exercises provide a variety of physical activities, including yoga and balance exercises, cycling control, and rowing training [17,39-41]. A key feature is the use of VR technology to achieve precise motion tracking. For instance, in rowing training, on-screen arrow indicators can help users understand how to correctly adjust their direction [41]; in cycling control games, wirelessly connected motion sensors can monitor the user's riding posture, ensuring safety and effectiveness [17].

For music attention training, the study by Jeong et al [42] introduces an application based on a VR environment and the Unity platform (Unity Technologies Inc) that enhances attention and cognitive health through interactive music experiences. Participants drum according to the indication of falling colored balls on the screen, thereby exercising their focus, hand-eye coordination, and sense of rhythm while enjoying music. Regarding telerehabilitation and social functioning training, the research by Maggio et al [43] offers participants a platform enabled by VR technology that allows them to complete social challenges remotely via a smartphone. The core content of this training lies in simulating real-life social scenarios, such as initiating conversations and managing personal finances, with the aim of promoting cognitive rehabilitation and improving social skills. For vocational and problem-solving skills training, the study by Man et al [44] provides an artificial intelligence VR-based vocational training system training, which is based on VR technology and simulates various vocational skill exercises in an office environment setting, aiming to enhance cognitive functions. The aforementioned 6 types of interventions primarily simulate real-world scenarios, which can assist patients in transferring the skills they have learned to real-life environments, enhancing the transfer effect of cognitive improvements on activities of daily living. Meanwhile, all intervention types emphasize the importance of immediate interaction and feedback.

The technological implementations across 21 studies encompassed diverse hardware configurations. These included multiple VR headset models (HTC Vive, Oculus Rift CV1 [Meta Platforms, Inc], Oculus Quest 2, and Samsung Gear VR), video wall systems, portable computing devices, and supplementary rehabilitation equipment such as physical therapy instruments, Nintendo Wii balance boards (Nintendo Co, Ltd), and robotic exoskeletons. Complementing these hardware configurations, software solutions consisted of 3 main types: purpose-built applications developed for specific training protocols, modified commercial software packages, and unaltered off-the-shelf programs. Regarding methodological approaches, 6 studies adopted immersive methods and 15 studies used nonimmersive methods or unclear and ambiguous classifications. Notably, intervention sessions exhibited significant variation in frequency (2 - 5 times weekly), session duration (15 - 60 min), total session count (6 - 40 sessions), and overall intervention duration (2 wk to 8 mo).

Characteristics of Controls

The control groups were subjected to random assignments, wherein they were allocated to either the conventional intervention or waitlist conditions for the entire study duration.

Outcome Measures

Multiple outcome measures were administered, and participants were assessed at various time points, with evaluations conducted after the intervention in all cases. Nine studies used the Montreal Cognitive Assessment scale to evaluate participants' cognitive function, while 6 studies used the Mini-Mental State Examination scale for cognitive assessment. The remaining studies used various tools including the Loewenstein Occupational Therapy Cognitive Assessment-Geriatric, Repeatable Battery for the Assessment of Neuropsychological Status, Cognistat, the Wechsler Adult Intelligence Scale-Revised Block Design Test, the Das-Naglieri Cognitive Assessment System, and the Wechsler Memory Scale-Fourth Edition.

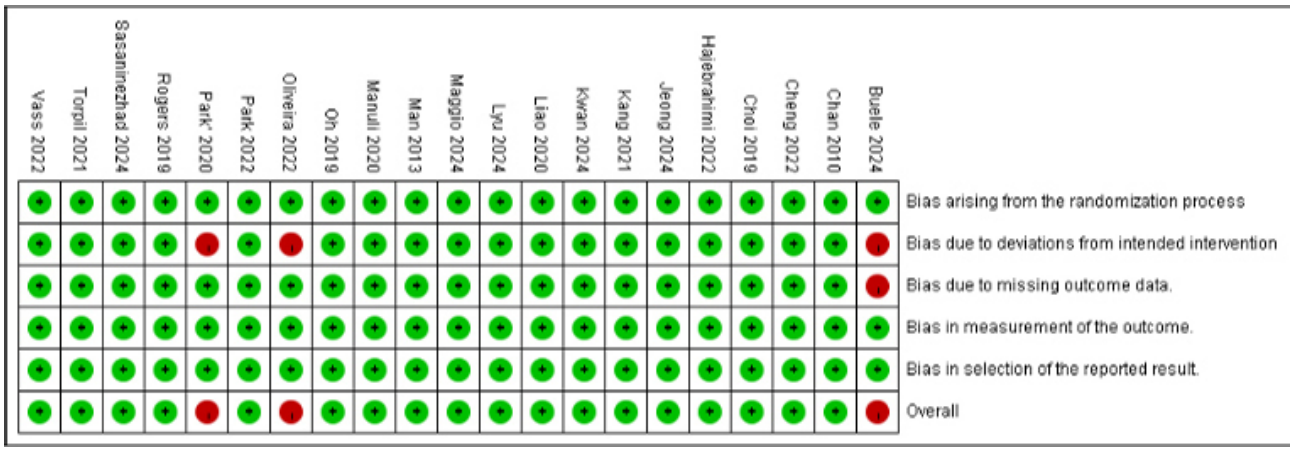
Both the Montreal Cognitive Assessment and Mini-Mental State Examination are widely used cognitive screening instruments that have demonstrated good reliability and validity across multiple studies [17,18,26,29,32-34,36,38-44]. The Loewenstein Occupational Therapy Cognitive Assessment-Geriatric is a cognitive assessment tool specifically designed for elderly individuals, focusing on cognitive functions relevant to daily activities [30]. Repeatable Battery for the Assessment of Neuropsychological Status is a standardized test battery intended for rapid screening and assessment of neuropsychological functioning in adults, encompassing 5 domains of cognition [28]. Cognistat is a clinically used tool for the quick screening and evaluation of an individual's cognitive function, with a distinctive feature being its inclusion of an affective state screen and a preliminary assessment of the individual's level of consciousness [31]. The Wechsler Adult Intelligence Scale-Revised Block Design Test aims to assess spatial visual ability, visual constructional skills, and problem-solving abilities [35]. Das-Naglieri Cognitive Assessment System is grounded in the PASS (Planning, Attention, Simultaneous, and Successive Processing) theory and is suitable for assessing cognitive functions in children and adolescents [27]. The Wechsler Memory Scale-Fourth Edition primarily evaluates memory functions within cognitive aspects [37].

XSL•FO
RenderX

Risk of Bias

The methodological quality of the 21 selected studies is depicted in Figure 2. Most studies (n=18, 85.7%) were overall rated as low risk. However, the study by Buele et al [18] was assessed as high risk overall due to a lack of specific information on whether all participants adhered strictly to the intended intervention. Additionally, during the study, 5 participants withdrew. The text does not mention how these missing data were handled, leading to a high-risk rating due to deviations from intended intervention and bias arising from missing outcome data. Furthermore, the studies by Oliveira et al [26] and Park et al [38] were also rated as high risk overall because they did not specify if blinding was used and did not clearly state whether all participants adhered strictly to the intended intervention, resulting in a high risk of bias due to deviations from the intended intervention.

Figure 2. The risk of bias assessment from Cochrane RoB2 [17,18,26-44]. RoB2: Risk of Bias 2.

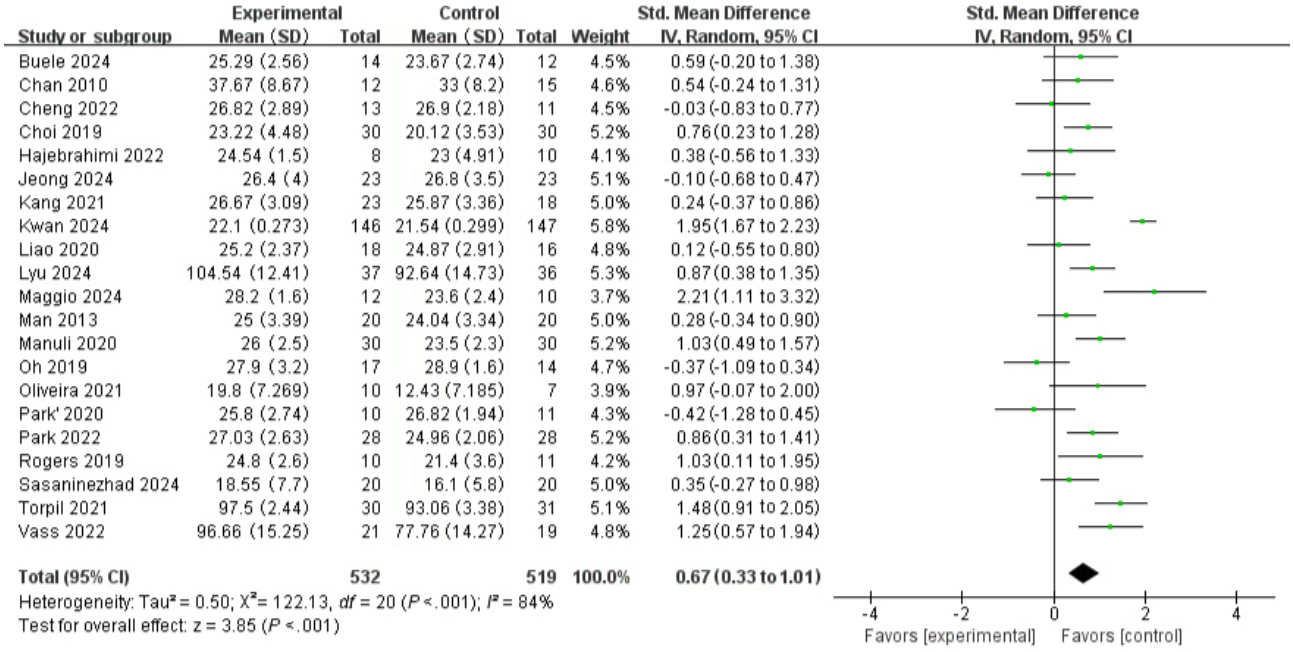


Meta-Analysis Results

Overall Effect

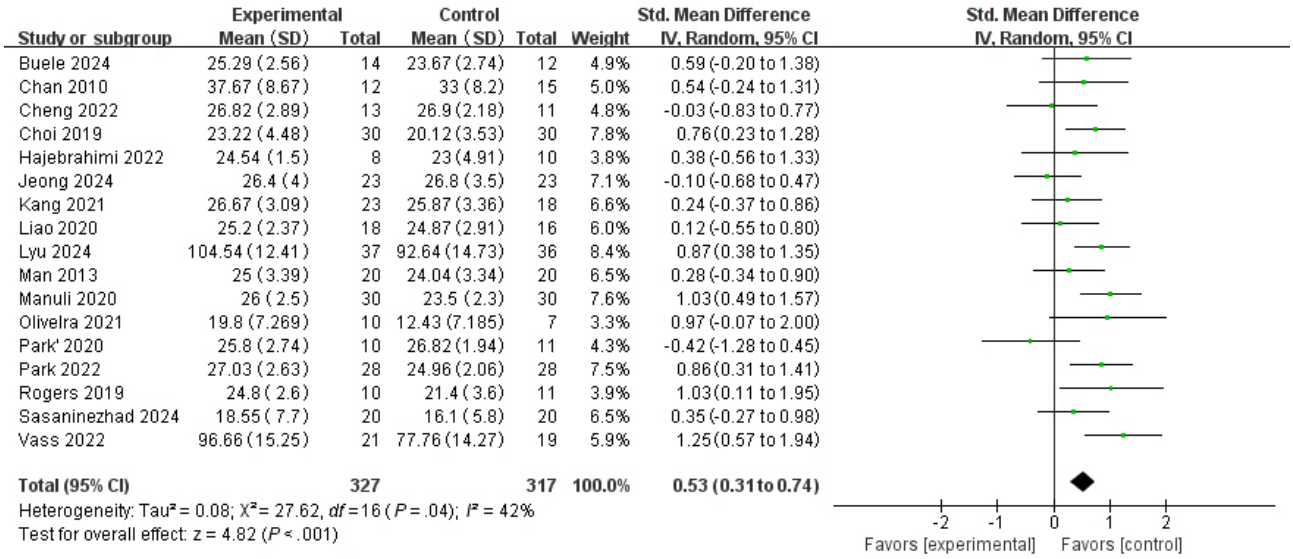
Regarding cognitive function, a random-effects model was used, revealing a notable level of heterogeneity among the 21 included studies (I^2=84%, P<.001). The meta-analysis' overall effect indicated that interventions based on VR technology could improve cognitive functions in patients with neuropsychiatric disorders (standardized mean difference [SMD] 0.67, 95% CI 0.33-1.01, z=3.85, P<.001; Figure 3). For problems with heterogeneity >50%, we conducted a sensitivity analysis by removing the 4 studies [17,30,33,43] and recalculated the combined estimate on remaining studies, and a significant decrease in heterogeneity was found (I^2 decreased from 84% to 42%). The overall effect of the meta-analysis demonstrated a similar result as above (SMD 0.53, 95% CI 0.31-0.74, z=4.82; P<.001), confirming the robustness of the conclusion. (Figure 4). After re-examining these 4 studies, we identified that the potential sources of heterogeneity may include disparities in baseline participant characteristics, divergent sample sizes, variability in assessment approaches, differences in intervention durations, etc.

Figure 3. The overall effect of VR-based interventions on cognitive function in patients with neuropsychiatric disorders [17,18,26-44]. Std: standard; VR: virtual reality.



XSL•FO
RenderX

Figure 4. After removing 4 studies that may be sources of heterogeneity, the overall effect of VR-based interventions on cognitive function in patients with neuropsychiatric disorders [18,26-29,31,32,34-42,44] . Std: standard; VR: virtual reality.

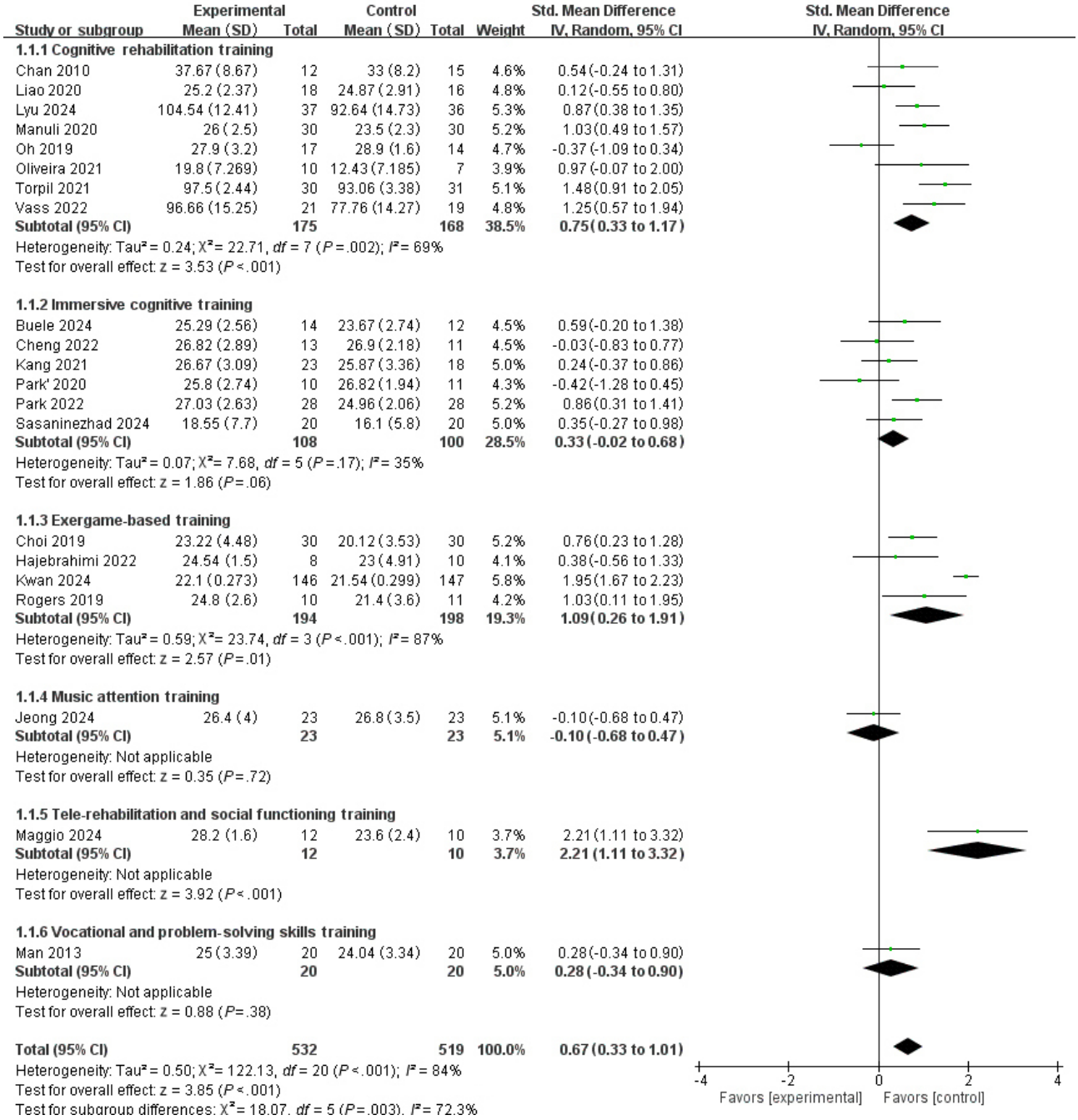


Subgroup Analysis Based on Intervention Types

According to the types of interventions, the data were stratified into 6 distinct groups for subgroup analysis, including cognitive rehabilitation training, immersive cognitive training, exergame-based training, music attention training, telerehabilitation and social functioning training, and vocational and problem-solving skills training. The results showed that based on VR, cognitive rehabilitation training, exergame-based training, and telerehabilitation and social functioning training could significantly improve cognitive functions in patients with neuropsychiatric disorders (SMD 0.75, 95% CI 0.33-1.17, z=3.53, P<.001; SMD 1.09, 95% CI 0.26-1.91, z=2.57, P=.01; SMD 2.21, 95% CI 1.11-3.32, z=3.92, P<.001; respectively), whereas immersive cognitive training, music attention training, and vocational and problem-solving skills training do not yield significant improvements in cognitive functions for these patients (SMD 0.33, 95% CI –0.02 to 0.68, z=1.86, P=.06; SMD −0.1, 95% CI −0.68 to 0.47, z=0.35, P=.72; SMD 0.28, 95% CI −0.34 to 0.90, z=0.88, P=.38; respectively). Additionally, the groups of cognitive rehabilitation training and exergame-based training exhibited high heterogeneity, with I^2 values of 69% and 87%, respectively (Figure 5).

XSL•FO
RenderX

Figure 5. The effect of subgroup analysis based on intervention types [17,18,26-44]. Std: standard.



Subgroup Analysis Based on Disease Types

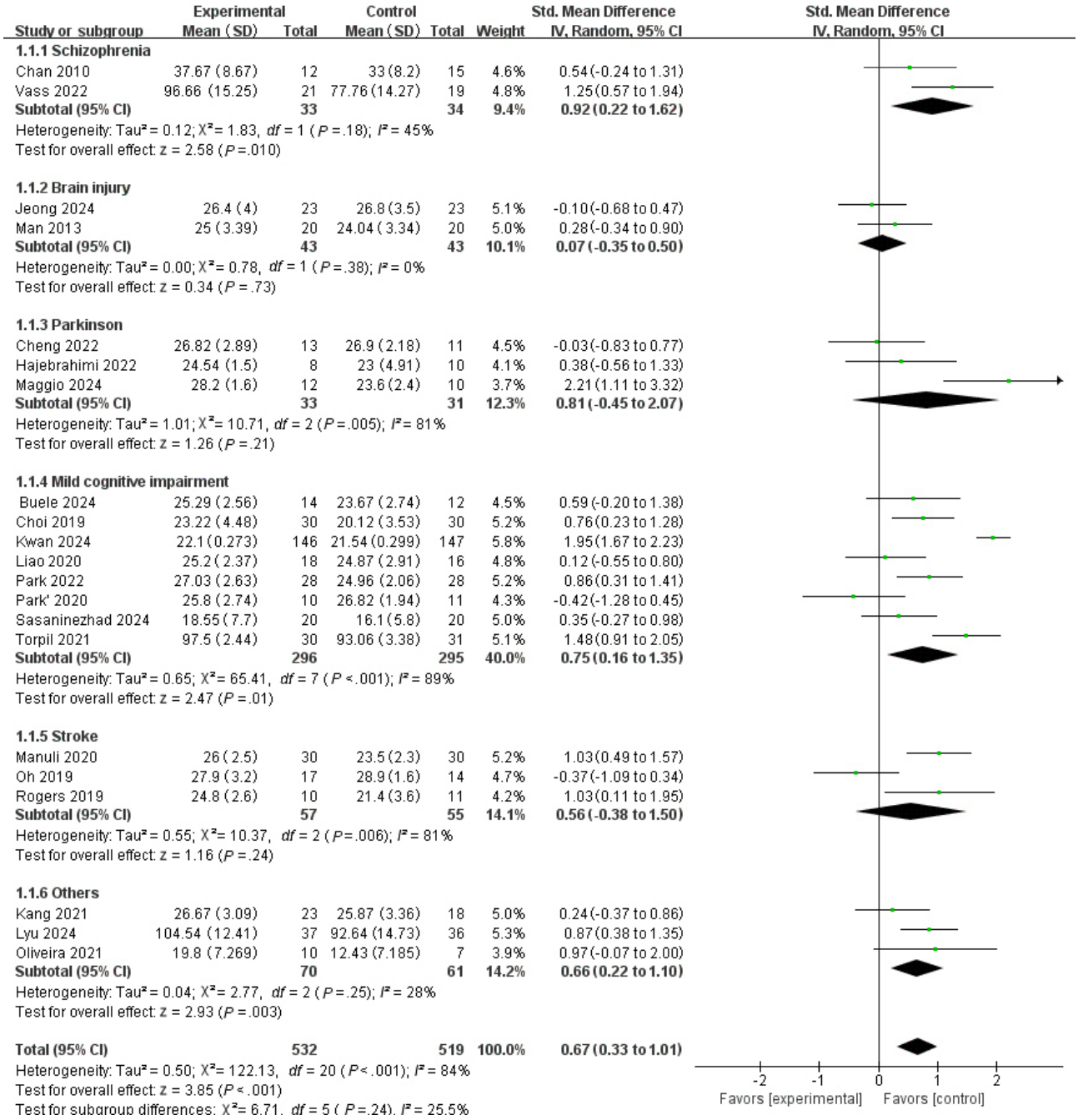
To investigate the potential reasons for differences in the impact of various intervention types based on VR technology on the cognitive functions of patients with neuropsychiatric disorders, besides the reasons related to the intervention types, we also proposed to analyze potential causes from the perspective of the types of neuropsychiatric disorders. The data was divided into 6 subgroups for subgroup analyses: schizophrenia, brain injury, Parkinson disease, mild cognitive impairment, stroke, and others. We found that VR-based interventions significantly improved cognitive functions in individuals with schizophrenia, mild cognitive impairment, and other neuropsychiatric disorders (SMD 0.92, 95% CI 0.22-1.62, z=2.58, P=.01; SMD 0.75, 95% CI 0.16-1.35, z=2.47, P=.01; SMD 0.66, 95% CI 0.22-1.10, z=2.93, P=.003; respectively), but did not yield significant improvements for those with brain injuries, Parkinson disease, or stroke (SMD 0.07, 95% CI −0.35 to 0.50, z=0.34, P=.73; SMD 0.81, 95% CI −0.45 to 2.07, z=1.26, P=.21; SMD 0.56, 95% CI −0.38 to 1.50, z=1.16, P=.24; respectively). High heterogeneity was observed in the Parkinson disease, mild cognitive impairment, and stroke groups, with I^2 values of 81%, 89%, and 81%, respectively (Figure 6).

We found that among the 6 studies classified as immersive cognitive training, 5 focused on patients with mild cognitive impairment or predementia. Subgroup analyses consistently indicated that interventions based on VR technology could effectively improve cognitive function in these patients. Notably, the study by Cheng et al [34] targeted patients with Parkinson disease, and subgroup analyses showed that VR-based interventions had no effect on improving cognitive function in

XSL•FO
RenderX

these patients; therefore, we speculated that this study may be a potential reason for the lack of statistically significant differences in the outcomes of immersive cognitive training. Subsequently, we conducted a sensitivity analysis by excluding the study by Cheng et al [34], which resulted in immersive cognitive training, showing statistically significant improvements in cognitive function. This further confirmed our hypothesis. Music attention training and vocational and problem-solving skills training both focused on patients with brain injuries, and subgroup analyses showed that VR-based interventions did not improve cognitive function in patients with brain injuries.

Figure 6. The effect of subgroup analysis based on disease types [17,18,26-44]. Std: standard.

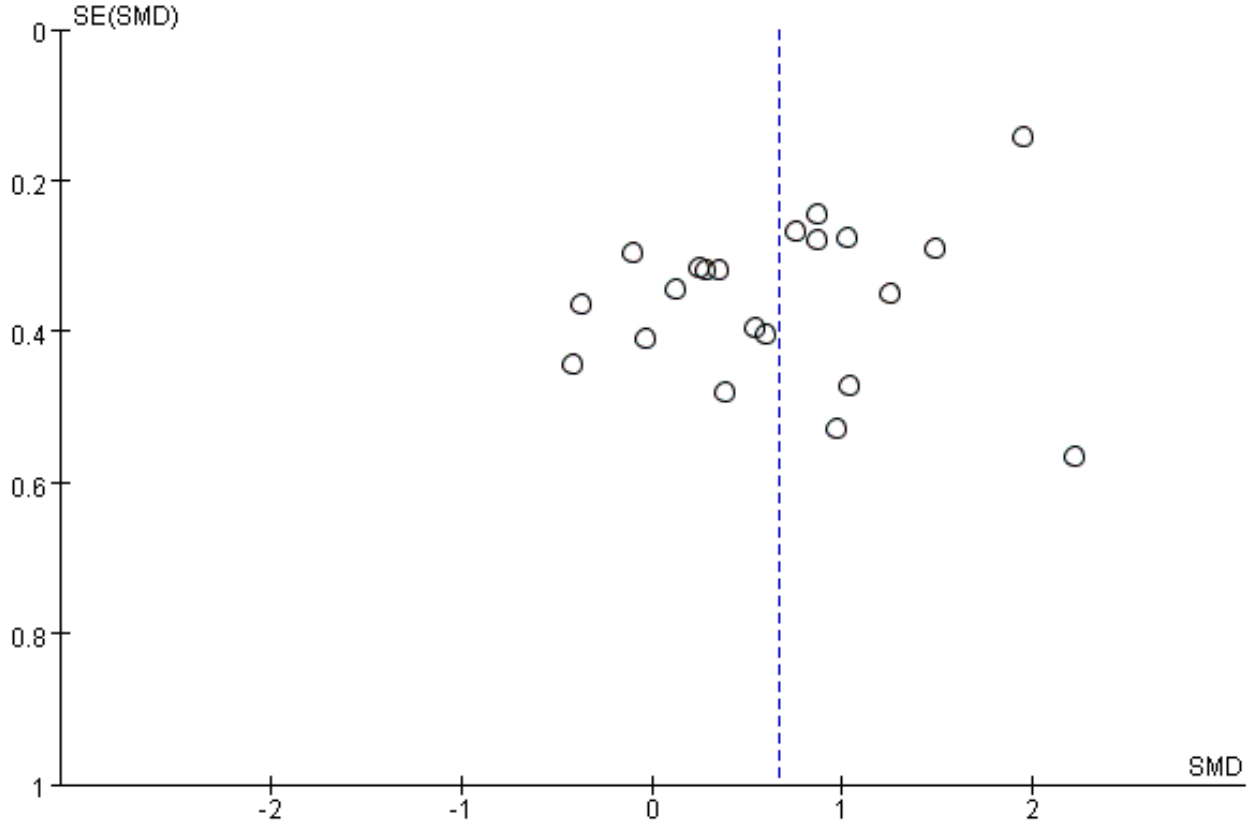


Publication Bias

To assess publication bias among the studies, a funnel plot is drawn. In this plot, the distribution of points is not entirely symmetrical, particularly on the left side, which could be an indication of publication bias. However, since most of the studies included in the meta-analysis report were smaller sample sizes, they may not accurately reflect publication bias (Figure 7).

XSL•FO
RenderX

Figure 7. The funnel plot of publication bias. SMD: standardized mean difference.



Discussion

Principal Findings

The principal aim of this meta-analysis was to quantitatively assess the efficacy of VR-based interventions on cognitive function among individuals with neuropsychiatric disorders. Through subgroup analyses, we sought to elucidate the efficacy of distinct intervention types and to investigate potential reasons for differences in results. Our findings demonstrated that VR-based interventions yield a significant positive effect on cognitive function in patients with neuropsychiatric disorders. Specifically, subgroup analyses revealed that cognitive rehabilitation training, exergame-based training, and telerehabilitation and social functioning training were associated with improvements in cognitive performance. Conversely, immersive cognitive training, music attention training, and vocational and problem-solving skills training did not significantly enhance cognitive functions within this patient population. Further analysis from the standpoint of disease categorization suggested that the unique characteristics of conditions for Parkinson disease and brain injury may underlie the lack of statistical significance observed in the meta-analytic synthesis for immersive cognitive training, music attention training, and vocational and problem-solving skills training, which underscored the importance of considering the specific etiology and pathology of neuropsychiatric disorders when designing VR-based therapeutic interventions.

Interpretation of Results

This study is the first to use quantitative analysis methods to evaluate the impact of VR-based interventions on cognitive functions in patients with neuropsychiatric disorders, providing preliminary evidence of potential efficacy in this therapeutic area. It represents an advancement in personalized medicine and the development of digital health tools. We included data from several relatively high-quality RCTs, revealing the efficacy and potential of VR-based interventions in improving cognitive functions, which are critical for the rehabilitation of neuropsychiatric disorders.

In our study, VR-based interventions featured immersive and highly interactive characteristics that can stimulate neuroplasticity in the brain. This property allows the brain to adapt to new information and skills by forming new synaptic connections or strengthening existing ones [45,46]. In VR environments, the simultaneous provision of visual, auditory, and tactile sensory inputs promotes synchronous activity among different sensory processing areas of the brain. This multisensory integration process is essential for constructing a coherent cognitive representation and can, to some extent, compensate for deficits caused by damage to certain sensory channels [47,48]. Different types of VR experiences can specifically activate particular regions of the brain. For instance, visual-spatial navigation tasks may increase activity in the hippocampus, whereas complex social interaction scenarios could impact the prefrontal cortex and other brain areas associated with higher cognitive functions [49,50]. This targeted stimulation may aid in repairing or optimizing the functions of these brain regions. Furthermore, VR-based interventions can

XSL•FO
RenderX

also influence the brain's reward system. When users complete challenges or achieve goals within a VR environment, they typically receive immediate feedback and reward signals, such as increased scores or virtual prizes. This type of positive reinforcement mechanism can activate the reward centers in the brain, such as the nucleus accumbens, triggering the release of dopamine, which in turn can increase motivation levels and enhance learning outcomes [51,52]. Another crucial point is that the VR-based interventions in this review primarily simulate real-world life scenarios, which is beneficial for improving the transfer effect of cognitive enhancements to activities of daily living. This approach aids in social reintegration and boosts the ability to live independently [53].

Through subgroup analysis of intervention types, we found that immersive cognitive training, music attention training, and vocational and problem-solving skills training based on VR technology did not improve cognitive functions in patients with neuropsychiatric disorders. This is inconsistent with previous research findings [54]. In previous studies, the aforementioned interventions primarily targeted patients with mild cognitive impairment or dementia; in contrast, our meta-analysis also included patients with Parkinson disease and brain injury. Second, subgroup analyses in this study revealed that exergame-based training was associated with cognitive improvement. Emerging evidence suggests that body position moderates the effectiveness of exergame-based training on global cognition, with stepping movement training in a predominantly standing position showing greater benefits compared to weight shifting or exergaming in a sitting position [55]. Consequently, the positive effects of exergame-based training on cognition may have been underestimated in our study. Notably, among the 4 exergame-based training studies included in this analysis, 3 used a seated position, while only 1 involved a standing position [17,39-41]. The limited sample size prevented further detailed and effective analysis. Therefore, future research should investigate the influence of body posture on the efficacy of different types of VR-based training to develop more comprehensive and effective cognitive intervention programs. Subgroup analysis by disease types revealed that patients with Parkinson disease and brain injury did not show significant cognitive improvements from VR-based interventions. After excluding the study on Parkinson disease [34] through sensitivity analysis, immersive cognitive training demonstrated statistically significant improvement effects. Parkinson disease, characterized by resting tremors, muscle rigidity, bradykinesia, and postural instability, can limit a patient's ability to interact effectively within a VR environment, thereby reducing the efficacy of training [54]. For example, decreased fine motor control may impair the ability to complete tasks requiring hand-eye coordination [56]. Furthermore, abnormalities in cortico-basal ganglia circuits in Parkinson disease may exacerbate executive dysfunction, further hindering cognitive training outcomes [57]. Similarly, different types of brain injuries can lead to either localized or diffuse brain tissue damage. Localized injuries typically affect specific functional areas, whereas diffuse injuries disrupt broader neural networks. When critical neural pathways are severed, simple repetitive training may be insufficient to rebuild these connections, thereby limiting the effectiveness of VR-based cognitive training [58,59]. However, emerging evidence suggests that while VR-based training often presumes baseline executive functioning, such as planning and organization [57], recent trials demonstrate its adaptability to populations with neurological impairments. For example, Dubbeldam et al [60] reported that technology use did not lead to higher dropout rates, and adherence was significantly higher in tailored interventions. This finding aligns with Swinnen et al [61], who conducted a systematic review and found high attendance adherence rates even in institutionalized individuals with major neurocognitive disorder. Thus, the limited efficacy observed in our study may reflect limitations in the intervention design rather than being solely attributable to inherent neuroanatomical constraints. These findings underscore the importance of tailoring VR-based interventions to the specific needs and capabilities of patient populations, as well as optimizing the design to maximize therapeutic outcomes.

Advantages and Prospects

The strength of this review lies in its novel use of meta-analysis to quantify the effects of VR-based interventions on cognitive functions in patients with neuropsychiatric disorders, significantly enhancing statistical power and providing a more comprehensive and robust estimate of intervention efficacy [62]. Simultaneously, subgroup analyses were conducted to evaluate the effects of different VR-based intervention types and to examine potential reasons for efficacy variations from the perspective of disease types. These findings help identify potentially effective VR interventions and technical components for improving cognitive functions in patients with neuropsychiatric disorders and suggest differences in responses to various interventions across patient groups. Consequently, this provides a scientific basis for the development of personalized treatment plans and promotes the precise application and advancement of VR technology in the rehabilitation of neuropsychiatric disorders.

Limitations

This study acknowledges several limitations that merit consideration. First, the studies included in our review exhibited certain heterogeneity due to differences among participants' baseline characteristics, sample sizes, assessment methods, intervention durations, etc. Second, only original research articles published in English were included, which may lead to language-related selection bias. Additionally, there may be some publication bias in this study. Last, despite using a broad search strategy across 5 major electronic databases, some studies have incomplete or entirely missing information sources. Therefore, our conclusions should be interpreted with caution.

Implications for Research

As VR-based interventions advance within personalized medicine paradigms, future research must adopt structured, user-centered frameworks to align technological development with the specific needs, preferences, and capabilities of populations with neuropsychiatric disorders. Frameworks such as the Generative Co-Design Framework [63] and the Multidisciplinary Iterative Design of Exergames [64] emphasize participatory, iterative design involving end users, clinicians,

XSL•FO
RenderX

and industry partners. These methodologies, exemplified by projects such as Brain-IT (information technology) help ensure that interventions are grounded in real-world usability and adherence while integrating established guidelines such as the Medical Research Council Framework for Complex Interventions [65,66]. This approach prioritizes multidisciplinary collaboration and continuous feedback to enhance translational impact. Beyond efficacy validation, research must investigate the mechanisms through which VR enhances cognitive function, such as neuroplasticity and attentional engagement, to identify potential commonalities or differential pathways across disorders. Mechanistic clarity will inform the optimization of VR components and protocols to better target therapeutic effects. Thus, by integrating user-centered design, mechanistic investigation, and technological innovation, VR has the potential to evolve into a robust, evidence-based tool for neuropsychiatric rehabilitation, optimizing recovery trajectories and long-term patient outcomes.

Conclusions

This meta-analysis showed that interventions based on VR technology have the potential to improve cognitive functions in patients with neuropsychiatric disorders, particularly in areas such as cognitive rehabilitation training, exergame-based training, and telerehabilitation and social functioning training. The findings provide valuable evidence supporting the application of VR technology in the rehabilitation of neuropsychiatric disorders and highlight possible directions for future intervention strategies.

Acknowledgments

We extend our sincere gratitude to all researchers whose studies contributed to this meta-analysis, as well as to our collaborators, advisors, and funding bodies. This research was funded by the Key Projects of the National Natural Science Foundation of China (grant U22A20334).

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

QD and YW contributed equally to this paper. QL and KL contributed equally to this work as cocorresponding authors. KL, QD, YW, and QL were responsible for the conceptualization and methodology. QD, YW, YM, CL, SD, QZ, XG, and JY contributed to the data curation and original draft preparation. KL supervised the project. QD, YW, YM, and QL were responsible for software development and validation. KL, QL, QD, and YW reviewed and edited this paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Sources searched and search strategies.
[DOCX File, 22 KB - games_v13i1e67501_app1.docx]

Multimedia Appendix 2
The characteristics of included studies.
[DOCX File, 37 KB - games_v13i1e67501_app2.docx]

Checklist 1
PRISMA 2020 checklist. PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.
[PDF File, 90 KB - games_v13i1e67501_app3.pdf]

References

1. Estes ML, McAllister AK. Maternal immune activation: implications for neuropsychiatric disorders. Science 2016 Aug 19;353(6301):772-777. [doi: 10.1126/science.aag3194] [Medline: 27540164]
2. Fabrazzo M, Russo A, Camerlengo A, et al. Delirium and cognitive impairment as predisposing factors of COVID-19 infection in neuropsychiatric patients: a narrative review. Medicina (Kaunas) 2021 Nov 14;57(11):1244. [doi: 10.3390/medicina57111244] [Medline: 34833462]
3. Huang YC, Lee Y, Lee CY, et al. Defining cognitive and functional profiles in schizophrenia and affective disorders. BMC Psychiatry 2020 Jan 31;20(1):39. [doi: 10.1186/s12888-020-2459-y] [Medline: 32005199]
4. Liu G, Zhang X, Huo X, Li W. Prevalence, influencing factors, and clinical characteristics of cognitive impairment in elderly patients with schizophrenia. Front Psychiatry 2022;13:910814. [doi: 10.3389/fpsyt.2022.910814] [Medline: 35733805]

XSL•FO
RenderX

5. de Raykeer RP, Hoertel N, Blanco C, et al. Effects of depression and cognitive impairment on quality of life in older adults with schizophrenia spectrum disorder: results from a multicenter study. J Affect Disord 2019 Sep 1;256:164-175. [doi: 10.1016/j.jad.2019.05.063] [Medline: 31176189]
6. Foguet-Boreu Q, Sancho AG, Lopez JMS, et al. Association between cognitive impairment and cardiovascular burden in patients with severe mental disorder. Cogn Neuropsychiatry 2020 Jan;25(1):1-13. [doi: 10.1080/13546805.2019.1676709] [Medline: 31599190]
7. Lepage M, Guimond S, Raedler T, et al. Strategies for achieving better cognitive health in individuals with schizophrenia spectrum: a focus on the canadian landscape: stratégies pour atteindre une meilleure santé cognitive chez les personnes souffrant du spectre de la schizophrénie: un regard sur le paysage canadien. Can J Psychiatry 2025 Feb;70(2):85-97. [doi: 10.1177/07067437241261928] [Medline: 39051555]
8. Rodríguez-Blanco L, Lubrini G, Vidal-Mariño C, Ríos-Lago M. Efficacy of cognitive rehabilitation of attention, executive functions, and working memory in psychotic disorders: a systematic review. Actas Esp Psiquiatr 2017 Jul;45(4):167-178. [Medline: 28745389]
9. Jensen JH, Knorr U, Vinberg M, Kessing LV, Miskowiak KW. Discrete neurocognitive subgroups in fully or partially remitted bipolar disorder: associations with functional abilities. J Affect Disord 2016 Nov 15;205:378-386. [doi: 10.1016/j.jad.2016.08.018] [Medline: 27573491]
10. Lewandowski KE, Sperry SH, Cohen BM, et al. Treatment to Enhance Cognition in Bipolar Disorder (TREC-BD): efficacy of a randomized controlled trial of cognitive remediation versus active control. J Clin Psychiatry 2017;78(9):e1242-e1249. [doi: 10.4088/JCP.17m11476] [Medline: 29045770]
11. Hu Y, Yuan X, Ye P, et al. Virtual reality in clinical nursing practice over the past 10 years: umbrella review of meta-analyses. JMIR Serious Games 2023 Nov 23;11:e52022. [doi: 10.2196/52022] [Medline: 37997773]
12. Tieri G, Morone G, Paolucci S, Iosa M. Virtual reality in cognitive and motor rehabilitation: facts, fiction and fallacies. Expert Rev Med Devices 2018 Feb;15(2):107-117. [doi: 10.1080/17434440.2018.1425613] [Medline: 29313388]
13. Bohil CJ, Alicea B, Biocca FA. Virtual reality in neuroscience research and therapy. Nat Rev Neurosci 2011 Nov 3;12(12):752-762. [doi: 10.1038/nrn3122] [Medline: 22048061]
14. Makransky G, Andreasen NK, Baceviciute S, Mayer RE. Immersive virtual reality increases liking but not learning with a science simulation and generative learning strategies promote learning in immersive virtual reality. J Educ Psychol 2021;113(4):719-735. [doi: 10.1037/edu0000473]
15. Muurling M, de Boer C, Vairavan S, et al. Augmented reality versus standard tests to assess cognition and function in early Alzheimer's disease. NPJ Digit Med 2023 Dec 18;6(1):234. [doi: 10.1038/s41746-023-00978-6] [Medline: 38110486]
16. Kim SH, Cho SH. Benefits of virtual reality program and motor imagery training on balance and fall efficacy in isolated older adults: a randomized controlled trial. Medicina (Kaunas) 2022 Oct 28;58(11):1545. [doi: 10.3390/medicina58111545] [Medline: 36363502]
17. Kwan RYC, Liu J, Sin OSK, et al. Effects of virtual reality motor-cognitive training for older people with cognitive frailty: multicentered randomized controlled trial. J Med Internet Res 2024 Sep 11;26:e57809. [doi: 10.2196/57809] [Medline: 39259959]
18. Buele J, Avilés-Castillo F, Del-Valle-Soto C, Varela-Aldás J, Palacios-Navarro G. Effects of a dual intervention (motor and virtual reality-based cognitive) on cognition in patients with mild cognitive impairment: a single-blind, randomized controlled trial. J Neuroeng Rehabil 2024 Aug 1;21(1):130. [doi: 10.1186/s12984-024-01422-w] [Medline: 39090664]
19. Kantola M, Ilves O, Honkanen S, et al. The effects of virtual reality training on cognition in older adults: a systematic review, meta-analysis, and meta-regression of randomized controlled trials. J Aging Phys Act 2024 Jun 1;32(3):321-349. [doi: 10.1123/japa.2023-0217] [Medline: 38242114]
20. Riva G, Mancuso V, Cavedoni S, Stramba-Badiale C. Virtual reality in neurorehabilitation: a review of its effects on multiple cognitive domains. Expert Rev Med Devices 2020 Oct;17(10):1035-1061. [doi: 10.1080/17434440.2020.1825939] [Medline: 32962433]
21. Chen JW, Du WQ, Zhu K. Network meta-analysis of the effects of different cognitive trainings on the cognitive function of patients with mild cognitive impairment. J Psychiatr Res 2024 Jun;174:26-45. [doi: 10.1016/j.jpsychires.2024.03.051] [Medline: 38608550]
22. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71. [doi: 10.1136/bmj.n71] [Medline: 33782057]
23. Moon HJ, Han S. Perspective: present and future of virtual reality for neurological disorders. Brain Sci 2022 Dec 9;12(12):1692. [doi: 10.3390/brainsci12121692] [Medline: 36552152]
24. Sterne JAC, Savović J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019 Aug 28;366:l4898. [doi: 10.1136/bmj.l4898] [Medline: 31462531]
25. Veroniki AA, Jackson D, Viechtbauer W, et al. Methods to estimate the between-study variance and its uncertainty in meta-analysis. Res Synth Methods 2016 Mar;7(1):55-79. [doi: 10.1002/jrsm.1164] [Medline: 26332144]
26. Oliveira J, Gamito P, Souto T, et al. Virtual reality-based cognitive stimulation on people with mild to moderate dementia due to Alzheimer's disease: a pilot randomized controlled trial. Int J Environ Res Public Health 2021 May 16;18(10):5290. [doi: 10.3390/ijerph18105290] [Medline: 34065698]

XSL•FO
RenderX

27. Lyu S, Zhong S, Luo Y, et al. Effects of virtual reality-based cognitive training for adolescents with depressive episodes: a pilot randomized controlled study. Psychiatry Res 2024 Oct;340:116144. [doi: 10.1016/j.psychres.2024.116144] [Medline: 39167866]
28. Vass E, Simon V, Csukly G, Fekete Z, Kis B, Simon L. Virtual reality-based theory of mind intervention in schizophrenia: preliminary efficacy results. Compr Psychiatry 2022 Nov;119:152350. [doi: 10.1016/j.comppsych.2022.152350] [Medline: 36272240]
29. Manuli A, Maggio MG, Latella D, et al. Can robotic gait rehabilitation plus vrtual reality affect cognitive and behavioural outcomes in patients with chronic stroke? A randomized controlled trial involving three different protocols. J Stroke Cerebrovasc Dis 2020 Aug;29(8):104994. [doi: 10.1016/j.jstrokecerebrovasdis.2020.104994] [Medline: 32689601]
30. Torpil B, Şahin S, Pekçetin S, Uyanık M. The effectiveness of a virtual reality-based intervention on cognitive functions in older adults with mild cognitive impairment: a single-blind, randomized controlled trial. Games Health J 2021 Apr;10(2):109-114. [doi: 10.1089/g4h.2020.0086] [Medline: 33058735]
31. Chan CLF, Ngai EKY, Leung PKH, Wong S. Effect of the adapted virtual reality cognitive training program among Chinese older adults with chronic schizophrenia: a pilot study. Int J Geriatr Psychiatry 2010 Jun;25(6):643-649. [doi: 10.1002/gps.2403] [Medline: 19806599]
32. Liao YY, Tseng HY, Lin YJ, Wang CJ, Hsu WC. Using virtual reality-based training to improve cognitive function, instrumental activities of daily living and neural efficiency in older adults with mild cognitive impairment. Eur J Phys Rehabil Med 2020 Feb;56(1):47-57. [doi: 10.23736/S1973-9087.19.05899-4] [Medline: 31615196]
33. Oh YB, Kim GW, Han KS, et al. Efficacy of virtual reality combined with real instrument training for patients with stroke: a randomized controlled trial. Arch Phys Med Rehabil 2019 Aug;100(8):1400-1408. [doi: 10.1016/j.apmr.2019.03.013] [Medline: 31002812]
34. Cheng TC, Huang SF, Wu SY, Lin FG, Lin WS, Tsai PY. Integration of virtual reality into transcranial magnetic stimulation improves cognitive function in patients with Parkinson's disease with cognitive impairment: a proof-of-concept study. J Parkinsons Dis 2022;12(2):723-736. [doi: 10.3233/JPD-212978] [Medline: 34897103]
35. Park JH. Effects of virtual reality-based spatial cognitive training on hippocampal function of older adults with mild cognitive impairment. Int Psychogeriatr 2022 Feb;34(2):157-163. [doi: 10.1017/S1041610220001131] [Medline: 32616109]
36. Kang JM, Kim N, Lee SY, et al. Effect of cognitive training in fully immersive virtual reality on visuospatial function and frontal-occipital functional connectivity in predementia: randomized controlled trial. J Med Internet Res 2021 May 6;23(5):e24526. [doi: 10.2196/24526] [Medline: 33955835]
37. Sasaninezhad M, Moradi A, Farahimanesh S, Choobin MH, Almasi-Dooghaee M. Enhancing cognitive flexibility and working memory in individuals with mild cognitive impairment: exploring the impact of virtual reality on daily life activities. Geriatr Nurs 2024;56:32-39. [doi: 10.1016/j.gerinurse.2023.12.008] [Medline: 38211369]
38. Park JH, Liao Y, Kim DR, et al. Feasibility and tolerability of a culture-based virtual reality (VR) training program in patients with mild cognitive impairment: a randomized controlled pilot study. Int J Environ Res Public Health 2020 Apr 27;17(9):3030. [doi: 10.3390/ijerph17093030] [Medline: 32349413]
39. Hajebrahimi F, Velioglu HA, Bayraktaroglu Z, Helvaci Yilmaz N, Hanoglu L. Clinical evaluation and resting state fMRI analysis of virtual reality based training in Parkinson's disease through a randomized controlled trial. Sci Rep 2022 May 16;12(1):8024. [doi: 10.1038/s41598-022-12061-3] [Medline: 35577874]
40. Rogers JM, Duckworth J, Middleton S, Steenbergen B, Wilson PH. Elements virtual rehabilitation improves motor, cognitive, and functional outcomes in adult stroke: evidence from a randomized controlled pilot study. J Neuroeng Rehabil 2019 May 15;16(1):56. [doi: 10.1186/s12984-019-0531-y] [Medline: 31092252]
41. Choi W, Lee S. The effects of virtual kayak paddling exercise on postural balance, muscle performance, and cognitive function in older adults with mild cognitive impairment: a randomized controlled trial. J Aging Phys Act 2019 Apr 29;27(6):861-870. [doi: 10.1123/japa.2018-0020] [Medline: 31034303]
42. Jeong E, Ham Y, Lee SJ, Shin JH. Virtual reality-based music attention training for acquired brain injury: a randomized crossover study. Ann N Y Acad Sci 2024 Nov;1541(1):151-162. [doi: 10.1111/nyas.15249] [Medline: 39476208]
43. Maggio MG, Luca A, Cicero CE, et al. Effectiveness of telerehabilitation plus virtual reality (Tele-RV) in cognitive e social functioning: a randomized clinical study on Parkinson's disease. Parkinsonism Relat Disord 2024 Feb;119:105970. [doi: 10.1016/j.parkreldis.2023.105970] [Medline: 38142630]
44. Man DWK, Poon WS, Lam C. The effectiveness of artificial intelligent 3-D virtual reality vocational problem-solving training in enhancing employment opportunities for people with traumatic brain injury. Brain Inj 2013;27(9):1016-1025. [doi: 10.3109/02699052.2013.794969] [Medline: 23662639]
45. Gangemi A, De Luca R, Fabio RA, et al. Effects of virtual reality cognitive training on neuroplasticity: a quasi-randomized clinical trial in patients with stroke. Biomedicines 2023 Dec 6;11(12):3225. [doi: 10.3390/biomedicines11123225] [Medline: 38137446]
46. Ismail FY, Fatemi A, Johnston MV. Cerebral plasticity: windows of opportunity in the developing brain. Eur J Paediatr Neurol 2017 Jan;21(1):23-48. [doi: 10.1016/j.ejpn.2016.07.007] [Medline: 27567276]
47. Friese U, Daume J, Göschl F, König P, Wang P, Engel AK. Oscillatory brain activity during multisensory attention reflects activation, disinhibition, and cognitive control. Sci Rep 2016 Sep 8;6:32775. [doi: 10.1038/srep32775] [Medline: 27604647]

XSL•FO
RenderX

48. Benetti S, Zonca J, Ferrari A, Rezk M, Rabini G, Collignon O. Visual motion processing recruits regions selective for auditory motion in early deaf individuals. Neuroimage 2021 Apr 15;230:117816. [doi: 10.1016/j.neuroimage.2021.117816] [Medline: 33524580]
49. Salgado-Pineda P, Landin-Romero R, Portillo F, et al. Examining hippocampal function in schizophrenia using a virtual reality spatial navigation task. Schizophr Res 2016 Apr;172(1-3):86-93. [doi: 10.1016/j.schres.2016.02.033] [Medline: 26924620]
50. Ainsworth M, Sallet J, Joly O, et al. Viewing ambiguous social interactions increases functional connectivity between frontal and temporal nodes of the social brain. J Neurosci 2021 Jul 14;41(28):6070-6086. [doi: 10.1523/JNEUROSCI.0870-20.2021] [Medline: 34099508]
51. Murty VP, Adcock RA. Enriched encoding: reward motivation organizes cortical networks for hippocampal detection of unexpected events. Cereb Cortex 2014 Aug;24(8):2160-2168. [doi: 10.1093/cercor/bht063] [Medline: 23529005]
52. Mohebi A, Pettibone JR, Hamid AA, et al. Dissociable dopamine dynamics for learning and motivation. Nature New Biol 2019 Jun 6;570(7759):65-70. [doi: 10.1038/s41586-019-1235-y]
53. Buele J, Varela-Aldás JL, Palacios-Navarro G. Virtual reality applications based on instrumental activities of daily living (iADLs) for cognitive intervention in older adults: a systematic review. J Neuroeng Rehabil 2023 Dec 19;20(1):168. [doi: 10.1186/s12984-023-01292-8] [Medline: 38110970]
54. Kalia LV, Lang AE. Parkinson's disease. Lancet 2015 Aug 29;386(9996):896-912. [doi: 10.1016/S0140-6736(14)61393-3] [Medline: 25904081]
55. Manser P, Herold F, de Bruin ED. Components of effective exergame-based training to improve cognitive functioning in middle-aged to older adults - a systematic review and meta-analysis. Ageing Res Rev 2024 Aug;99:102385. [doi: 10.1016/j.arr.2024.102385] [Medline: 38914262]
56. Dahdal P, Meyer A, Chaturvedi M, et al. Fine motor function skills in patients with Parkinson disease with and without mild cognitive impairment. Dement Geriatr Cogn Disord 2016;42(3-4):127-134. [doi: 10.1159/000448751] [Medline: 27643700]
57. Benazzouz A, Mamad O, Abedi P, Bouali-Benazzouz R, Chetrit J. Involvement of dopamine loss in extrastriatal basal ganglia nuclei in the pathophysiology of Parkinson's disease. Front Aging Neurosci 2014;6:87. [doi: 10.3389/fnagi.2014.00087] [Medline: 24860498]
58. Wolf JA, Koch PF. Disruption of network synchrony and cognitive dysfunction after traumatic brain injury. Front Syst Neurosci 2016;10:43. [doi: 10.3389/fnsys.2016.00043] [Medline: 27242454]
59. Bai L, Yin B, Lei S, et al. Reorganized hubs of brain functional networks after acute mild traumatic brain injury. J Neurotrauma 2023 Jan;40(1-2):63-73. [doi: 10.1089/neu.2021.0450] [Medline: 35747994]
60. Dubbeldam R, Stemplewski R, Pavlova J. Technology-assisted physical activity interventions for older people in their home-based environment: a scoping review. JMIR Preprints. Preprint posted online on Aug 25, 2024. [doi: 10.2196/preprints.65746]
61. Swinnen N, Vandenbulcke M, Vancampfort D. Exergames in people with major neurocognitive disorder: a systematic review. Disabil Rehabil Assist Technol 2022 May;17(4):376-389. [doi: 10.1080/17483107.2020.1785566] [Medline: 32697614]
62. Noma H, Sugasawa S, Furukawa TA. Robust inference methods for meta-analysis involving influential outlying studies. Stat Med 2024 Sep 10;43(20):3778-3791. [doi: 10.1002/sim.10157] [Medline: 38899515]
63. Bird M, McGillion M, Chambers EM, et al. A generative co-design framework for healthcare innovation: development and application of an end-user engagement framework. Res Involv Engagem 2021 Mar 1;7(1):12. [doi: 10.1186/s40900-021-00252-7] [Medline: 33648588]
64. Li Y, Munoz J, Mehrabi S, Middleton L, Cao S. Multidisciplinary iterative design of exergames (MIDE): a framework for supporting the design, development, and evaluation of exergames for health. 2020 Presented at: 2nd International Conference on HCI in Games (HCI-Games) Held as Part of the 22nd International Conference on Human-Computer Interaction (HCII); Jun 29 to Jul 4, 2024; Washington, DC, United States. [doi: 10.1007/978-3-030-50164-8_9]
65. Manser P, de Bruin ED. Making the best out of IT: design and development of exergames for older adults with mild neurocognitive disorder - a methodological paper. Front Aging Neurosci 2021;13:734012. [doi: 10.3389/fnagi.2021.734012] [Medline: 34955806]
66. Skivington K, Matthews L, Simpson SA, et al. A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. BMJ 2021 Sep 30;374:n2061. [doi: 10.1136/bmj.n2061] [Medline: 34593508]

Abbreviations

AR: augmented reality
PASS: Planning, Attention, Simultaneous, and Successive Processing
PICOS: Patient or Population, Intervention, Comparison, Outcome, and Study
RCT: randomized controlled trial
RoB2: Risk of Bias 2

XSL•FO
RenderX

SMD: standardized mean difference
VR: virtual reality

Edited by A Coristine; submitted 13.10.24; peer-reviewed by A Ramzi, P Manser; revised version received 24.02.25; accepted 25.02.25; published 08.05.25.

Please cite as:
Du Q, Wei Y, Ma Y, Liu C, Du S, Zhang Q, Gong X, Yang J, Li Q, Li K
Efficacy of Virtual Reality–Based Interventions on Cognitive Function in Patients With Neuropsychiatric Disorders: Systematic Review and Meta-Analysis of Randomized Controlled Trials
JMIR Serious Games 2025;13:e67501
URL: https://games.jmir.org/2025/1/e67501
doi:10.2196/67501

© Qiujing Du, Yuhan Wei, Yuexuan Ma, Changqing Liu, Shanshan Du, Qi Zhang, Xiaotong Gong, Jiaju Yang, Qijie Li, Ka Li. Originally published in JMIR Serious Games (https://games.jmir.org), 8.5.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Efficacy of Motion-Sensing Game–Assisted Pulmonary Rehabilitation in Patients With Chronic Obstructive Pulmonary Disease: Systematic Review and Meta-Analysis of Randomized Controlled Trials

Cuirong Hu*, BS; Xia Lan*, MS; Yanwen Luo, MBA; Shu Zhu, BS; Guilan Cheng, BS

Division of Internal Medicine, Institute of Integrated Traditional Chinese and Western Medicine, West China Hospital, Sichuan University/West China School of Nursing, Sichuan University, No.37 Guoxue Street, Wuhou District, Chengdu, Sichuan Province, China

*these authors contributed equally

Corresponding Author:
Guilan Cheng, BS
Division of Internal Medicine, Institute of Integrated Traditional Chinese and Western Medicine, West China Hospital, Sichuan University/West China School of Nursing, Sichuan University, No.37 Guoxue Street, Wuhou District, Chengdu, Sichuan Province, China

Abstract

Background: While motion-sensing game (MSG) platforms provide immersive, real-time feedback environments for rehabilitation, research findings on their effectiveness in chronic obstructive pulmonary disease (COPD) pulmonary rehabilitation remain heterogeneous.

Objective: This study aims to systematically evaluate the efficacy of MSG-assisted pulmonary rehabilitation for patients with COPD.

Methods: This meta-analysis was conducted in accordance with the *Cochrane Handbook for Systematic Reviews of Interventions* and the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 statement. Eight electronic databases (PubMed, Cochrane Library, Embase, CINAHL, Web of Science, Scopus, China National Knowledge Infrastructure, and Wanfang) were systematically searched from inception to March 2025. Randomized controlled trials (RCTs) comparing MSG-assisted versus conventional pulmonary rehabilitation in patients with COPD were included. Data analysis was performed using Review Manager 5.3 (Cochrane Collaboration) and Stata 17 (StataCorp LLC). Mean differences (MDs) and odds ratios (ORs) with 95% CIs were calculated. Risk of bias was assessed using the revised Cochrane Risk of Bias tool (Cochrane Collaboration), and evidence quality was evaluated using the Grading of Recommendations Assessment, Development, and Evaluations (GRADE) approach.

Results: Analysis of 12 RCTs (n=776) demonstrated that MSG-assisted pulmonary rehabilitation, compared with conventional pulmonary rehabilitation, significantly improved 6-Minute Walk Distance (MD 23.23, 95% CI 14.47 - 31.99; *P*<.001), upper limb strength via 30-Second Arm Curl Test (MD 1.83, 95% CI 0.63 - 3.03; *P*=.003), balance performance (Balance Evaluation Systems Test; MD 2.34, 95% CI 1.52 - 3.17; *P*<.001), and exercise adherence (OR 3.00, 95% CI 1.38 - 6.52; *P*=.005). Additionally, significant improvements were observed in dyspnea severity (MD −0.25, 95% CI −0.48 to −0.02; *P*=.03), health-related quality of life (MD −6.00, 95% CI −10.96 to −1.04; *P*=.02), and psychological outcomes including anxiety (MD −2.41, 95% CI −3.42 to −1.39; *P*<.001) and depression (MD −1.40, 95% CI −2.69 to −0.42; *P*=.03). The overall methodological quality of the included studies was suboptimal with most evidence rated as “low” or “very low” quality.

Conclusions: MSG-assisted pulmonary rehabilitation demonstrates significant improvements in exercise capacity, respiratory symptoms, quality of life, and psychological well-being among patients with COPD. Despite potential benefits, the predominance of low-quality evidence highlights the necessity for risk-benefit assessment before clinical implementation. Future research priorities should include larger, methodologically rigorous RCTs, standardized intervention protocols, investigation of sustained therapeutic effects, and cost-effectiveness analyses to establish definitive evidence for optimal implementation of gaming technology in pulmonary rehabilitation programs.

Trial Registration: PROSPERO CRD42024603044; https://www.crd.york.ac.uk/PROSPERO/view/CRD42024603044

(JMIR Serious Games 2025;13:e69562) doi:10.2196/69562

KEYWORDS

motion-sensing game; chronic obstructive pulmonary disease; pulmonary rehabilitation; exercise capacity; meta-analysis

XSL•FO
RenderX

Introduction

Chronic obstructive pulmonary disease (COPD) is characterized by persistent airflow limitation resulting from a complex interplay of airway inflammation, parenchymal destruction, and systemic manifestations [1]. This pathophysiological process extends beyond the lungs to induce skeletal muscle dysfunction, which is characterized by reduced muscle mass, strength, and endurance, and significantly contributes to exercise intolerance and impaired quality of life [1]. COPD represents a significant global health challenge, affecting over 400 million individuals worldwide and is projected to become the third leading cause of death by 2030 [2]. The economic burden of COPD continues to rise, with annual direct health care costs exceeding US $50 billion in high-income countries [3]. While pulmonary rehabilitation (PR), particularly exercise training, has been established as a cornerstone intervention that significantly enhances exercise capacity, alleviates symptoms, and improves quality of life [4,5], its implementation faces considerable challenges. Traditional programs struggle with issues of accessibility and adherence, reaching only 10% - 20% of eligible patients, primarily due to geographical barriers, limited health care resources, and lack of motivation [6,7].

The recent digital health revolution, accelerated by the COVID-19 pandemic, has catalyzed innovative approaches to health care delivery [7]. Motion-sensing game–assisted pulmonary rehabilitation (PR+MSG) has emerged as a promising solution, leveraging advanced technologies such as Nintendo Wii, Xbox Kinect, and Subor systems [8]. These platforms offer unique advantages, including real-time movement tracking and immediate feedback capabilities, providing superior accessibility and user-friendliness compared with conventional mobile apps, telehealth systems, and complex virtual reality setups [9,10]. The immersive and gamified nature of motion-sensing game (MSG)–assisted rehabilitation creates an engaging environment that effectively addresses both the physical and psychological aspects of rehabilitation, potentially overcoming the longstanding challenges of traditional rehabilitation programs by enhancing motivation and adherence [11].

While several randomized controlled trials (RCTs) have demonstrated the potential benefits of PR+MSG in the management of COPD [12], research results remain inconsistent [13]. Furthermore, previous systematic reviews and meta-analyses have examined technology-enhanced exercise interventions—including mobile apps, web-based platforms, telehealth systems, and virtual reality interfaces—in chronic respiratory conditions. However, they either encompassed a broader range of technologies or focused on specific outcomes [14,15]. As of the study period, no comprehensive meta-analysis has specifically evaluated the efficacy of PR+MSG across multiple outcomes in patients with COPD. Therefore, this meta-analysis aims to assess the effects of PR+MSG on exercise capacity, symptoms, quality of life, and psychological outcomes in patients with COPD. The findings will provide crucial evidence to inform the implementation of MSG in PR programs.

Methods

Search Strategy

We conducted a comprehensive search of PubMed, Cochrane Library, Embase, CINAHL, Web of Science, Scopus, China National Knowledge Infrastructure, and Wanfang databases from inception to March 2025, without language restrictions. Additionally, we manually searched the references of included studies and relevant published reports to identify eligible studies. A search strategy was developed for PubMed and the other databases (Table S1 in Multimedia Appendix 1).

Inclusion and Exclusion Criteria

We selected the studies based on the following inclusion and exclusion criteria (Textbox 1).

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria

- Randomized controlled trials (RCTs) evaluating motion-sensing game (MSG)–assisted exercise training.
- Participants with chronic obstructive pulmonary disease (COPD) diagnosed according to the Global Initiative for Chronic Obstructive Lung Disease guidelines.
- Intervention group using MSGs from any MSG system (including Nintendo Wii, Subor, EyeToy, Kinect, etc) during pulmonary rehabilitation (PR) and control group receiving routine PR without game assistance.
- Measured outcomes including: (1) exercise capacity (6-Minute Walk Distance, 30-Second Arm Curl Test, Brief Balance Evaluation Systems Test, and completion rate of exercise training), (2) health-related quality of life (HRQL) measured using disease-specific or generic HRQL instruments, for example, COPD Assessment Test score, Chronic Respiratory Disease Questionnaire, and St. George's Respiratory Questionnaire, (3) dyspnea severity (eg, the Medical Research Council scale and the modified Medical Research Council scale), and (4) psychological state (eg, Hospital Anxiety and Depression Scale, Self-Rating Depression Scale, and Self-Rating Anxiety Scale).

Exclusion criteria

- Participants with non-COPD respiratory diseases or other significant conditions.
- Abstract-only publications without full text or efficacy reporting.
- Studies with incomplete data or protocols only.

XSL•FO
RenderX

Study Selection

The search was conducted by the first author and recorded in a Microsoft Excel file, from which duplicate entries were subsequently removed. Two independent reviewers (CH and XL) manually screened each title and abstract, and if necessary, a third reviewer (SZ) was consulted to resolve any disagreements. All full texts that met the inclusion criteria were included in the analysis.

Data Extraction

Two researchers (CH and XL) independently extracted and recorded the following information from the enrolled studies: (1) study characteristics (author, publication year, country, study design, sample size, age, sex, and PR exercise training program), (2) methodological characteristics, and (3) outcome indicators (exercise capacity, health-related quality of life (HRQL), severity of dyspnea, and psychological state). Any disagreements were resolved by a third researcher (SZ).

Quality Assessment

The risk of bias in individual studies was evaluated by two independent researchers (CH and XL) using the revised Cochrane Risk of Bias tool [16], which assesses the randomization process, adherence to interventions, completeness of outcome data, outcome measurement, and selection of reported results. The quality was rated as "low risk," "some concerns," or "high risk." The Grading of Recommendations Assessment, Development, and Evaluations (GRADE) approach was employed to assess the quality of evidence for individual outcomes [17]. The quality of evidence was classified as high, moderate, low, or very low by evaluating five domains: risk of bias, inconsistency, indirectness, imprecision, and publication bias. Any disagreements were resolved by a third researcher (SZ).

Statistical Analysis

Data analysis was conducted using Review Manager 5.3 (The Cochrane Collaboration) and Stata 17 (StataCorp LLC). To address potential baseline imbalances, we performed a pooled analysis using change scores, defined as the difference between baseline and postintervention values [18]. A random-effects model was used to estimate the aggregated outcomes. For dichotomous variables, we calculated odds ratios (ORs) with 95% CIs, while for continuous variables, mean differences (MDs) with 95% CIs were used. The overall results of the meta-analysis were visualized with forest plots. Forest plots visually summarize meta-analysis results by displaying individual study effect sizes (eg, risk ratios) with CIs, weighted summary effects, and heterogeneity metrics such as I^2. Heterogeneity among studies was assessed using Cochran Q test and quantified with the I^2 statistic. We interpreted I^2 values of 25%, 50%, and 75% as indicative of low, moderate, and high heterogeneity, respectively. Statistical significance was established at a 2-tailed P value of less than .05. The credibility of the evidence was evaluated using the GRADE approach, implemented through GRADEpro GDT software (Cochrane Collaboration). To assess potential publication bias, we generated funnel plots for outcomes with 10 or more studies and conducted Egger and Begg tests when appropriate. Subgroup analyses were performed based on the MSG system used (Xbox Kinect, Nintendo Wii, or Subor) to explore potential sources of heterogeneity and evaluate the consistency of treatment effects across different platforms.

Ethical Considerations

This systematic review and meta-analysis was conducted in accordance with the *Cochrane Handbook for Systematic Reviews of Interventions* and the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 statement (Checklist 1). The protocol was registered with the International Prospective Register of Systematic Reviews (PROSPERO; CRD42024603044). Since the data for this study were derived from previously published research, ethical approval and informed consent were not required.

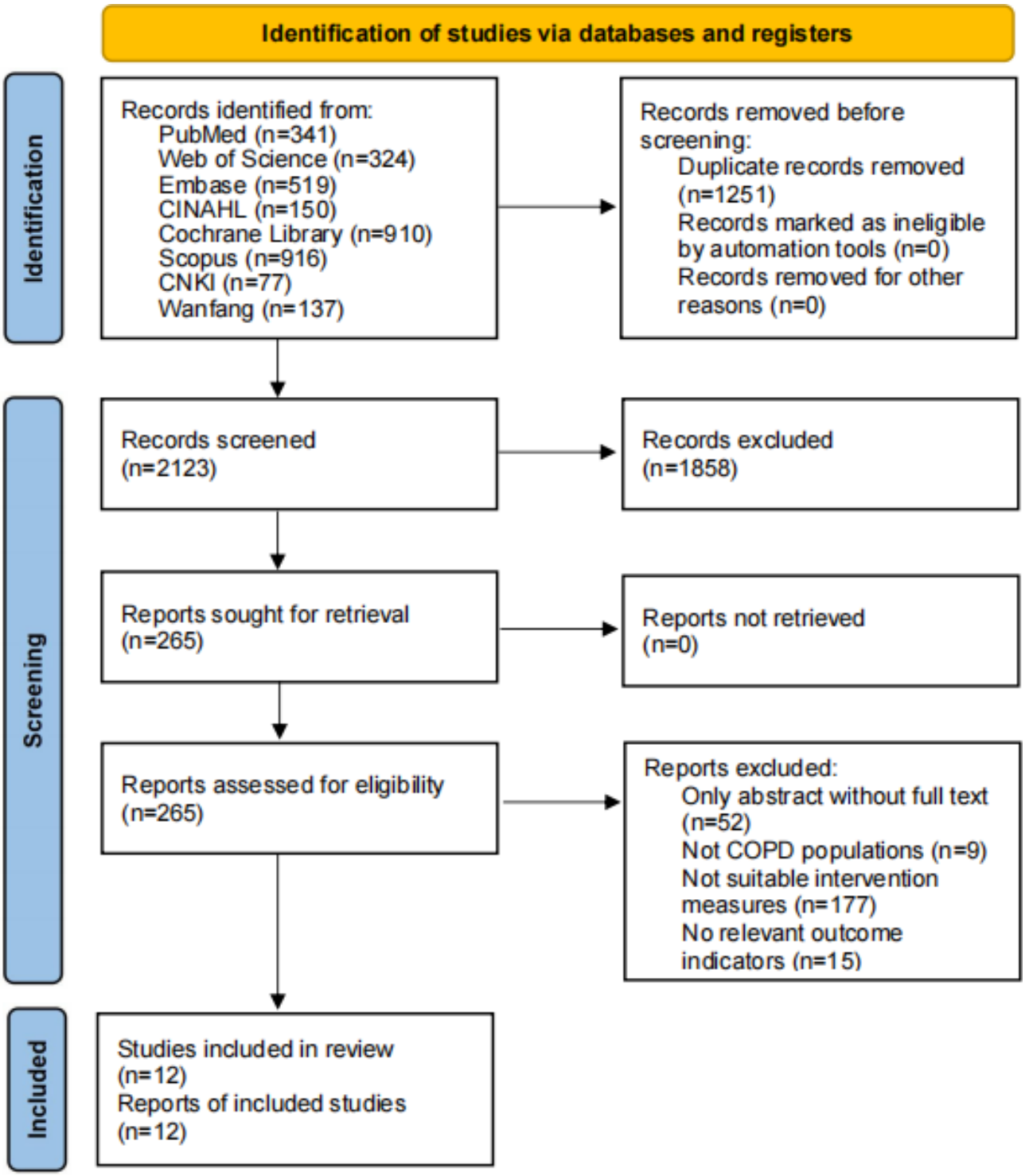
Results

Flow and Characteristics of Included Studies

The flow diagram of the selection procedure is presented in Figure 1. Initially, 3374 papers were identified through database searches. After the removal of duplicates, 2123 papers remained for the first screening. Following a review of titles and abstracts, 265 papers were selected for full-text evaluation, resulting in the exclusion of 1858 papers. Ultimately, 12 trials involving 776 participants met the inclusion criteria and were incorporated into the quantitative synthesis (Figure 1). Tables 1 and 2 summarize the characteristics of the included studies, which were published between 2014 and 2025. The majority of these studies were conducted in China (n=8) [12,19-25], while the remainder were distributed across Poland (n=2) [13,26], Indonesia (n=1) [27], and Italy (n=1) [28]. Intervention durations ranging from 2 to 18 weeks, most commonly between 6 and 8 weeks. A total of 3 main gaming systems were used: the Subor systems (6 studies) [12,19-23], Kinect-based systems (4 studies) [13,24-26], and Nintendo Wii systems (2 studies) [27,28]. Sessions typically lasted 20 to 30 minutes and were conducted 5 times weekly under professional supervision with established safety monitoring protocols.

XSL•FO
RenderX

Figure 1. Flow diagram of the study selection procedure for the meta-analysis of randomized controlled trials investigating the efficacy of motion-sensing game–assisted pulmonary rehabilitation compared with routine pulmonary rehabilitation in patients with chronic obstructive pulmonary disease, following Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) guidelines. CNKI: China National Knowledge Infrastructure; COPD: chronic obstructive pulmonary disease.



XSL•FO
RenderX

Table . Characteristics of randomized controlled trials evaluating motion-sensing game–assisted pulmonary rehabilitation (PR+MSG) versus routine pulmonary rehabilitation in patients with chronic obstructive pulmonary disease. Data are presented as absolute values and means (SD).[a]

First author (publication year)	Country	Sample size, n	Experimental group: PR+MSG		Control group: routine pulmonary rehabilitation		Duration of intervention	Reported outcomes
			Sample size, n	Age (years), mean (SD)	Sample size, n	Age (years), mean (SD)		
Jin et al (2025) [12]	China	74	36	73.97 (8.69)	38	74.00 (10.01)	6 wks	①③
Rutkowski et al (2020) [13]	Poland	72	38	60.6 (4.3)	34	62.1 (2.9)	2 wks	①②
Rutkowski et al (2019) [26]	Poland	68	34	60.5 (4.3)	34	62.1 (2.9)	2 wks	①②
Sutanto et al 2019 [27]	Indonesia	20	10	65.1 (7.5)	10	65.6 (4.7)	6 wks	①⑤
Mazzoleni et al (2014) [28]	Italy	39	19	68.9 (11)	20	73.5 (9.2)	2-3 wks	①②⑤
Wei et al (2024) [23]	China	122	61	72.32 (6.38)	61	72.32 (6.38)	8 wks	①③⑤
Zhu et al (2021) [25]	China	43	22	65.16 (12.78)	21	64.54 (12.89)	16 wks	①④⑤⑥
Zhou et al (2021) [24]	China	119	61	70.54 (6.89)	58	71.38 (5.78)	18 wks	①④⑤
Jin (2021) [19]	China	69	35	73.97 (8.69)	34	74.00 (10.01)	6 wks	①②⑦
Geng (2020) [21]	China	40	21	68.66 (6.41)	19	68.52 (5.76)	8 wks	①⑤⑥
Geng et al (2020) [22]	China	44	22	68.66 (6.41)	22	68.52 (5.76)	8 wks	⑦
Jin et al (2020) [20]	China	66	33	74.15 (8.75)	33	73.97 (10.14)	Begin within 48 h of admission and continue until the day of discharge	③

[a]① 6-Minute Walk Distance; ② 30-Second Arm Curl Test; ③ Brief Balance Evaluation Systems Test; ④ completion rate of exercise training; ⑤ severity of dyspnea; ⑥ quality of life; and ⑦ psychological state.

XSL•FO
RenderX

Table . Detailed characteristics of the experimental group evaluating PR+MSG[a] in patients with chronic obstructive pulmonary disease.

First author (publication year)	PR+MSG	MSG[b] system	Game	Implementation
Jin et al (2025) [12]	Patients received a once-daily, 20-min MSG session based on the control group.	The Little King somatosensory gaming system (Subor) consisted of a host console, motion sensors, and a display screen. The system features dual sensor technology for enhanced motion capture accuracy and real-time feedback.	Three games ("Kitchen Sharp Knife," "Swimming Master," and "Table Tennis Master") were selected, and one-to-one instruction was implemented to ensure that the patients mastered the exercises correctly.	The exercise begins with a 3-min "Kitchen Sharp Knife" warm-up, continues for 5 min with the left hand only, 5 min with the right hand only, takes a 3-mins break, and concludes with a 4-min "Swimming Master" exercise. Patients can decide to sit or stand during exercise based on their comfort and overall health. Oxygen was administered if the patient's oxygen saturation fell below 88%, and exercise resumed after stabilizing oxygen saturation. The exercise was discontinued if the Borg score exceeded 6 or adverse events occurred.
Rutkowski et al (2020) [13]	Patients participated in traditional pulmonary rehabilitation, complemented with endurance exercise training and Kinect system–based virtual reality sessions.	The system utilized an Xbox 360 console, a Kinect motion sensor, and a projector with speakers. Session intensity was controlled through Kinect Adventures games (Microsoft Game Studios, Washington, US).	Four Kinect games were selected and implemented: "20,000 Leaks," "Curvy Creek," "Rally Ball," and "Reflex Ridge."	Each patient participated in the 4 games in a fixed sequence during a 20-min session, conducted 5 d weekly with consistent workloads. Heart rate monitoring ensured patients remained below their age-predicted maximal heart rate (208 - 0.7×age). A physiotherapist supervised all virtual reality training sessions. Exercise was modified or discontinued if patients showed signs of excessive exertion or if adverse events occurred.
Rutkowski et al (2019) [26]	Patients received standard pulmonary rehabilitation plus once-daily Kinect system–based exercise sessions.	The system consisted of an Xbox 360 console, Kinect motion sensor, and a projector with speakers. The console was positioned one meter high on a table, with the Kinect sensor mounted on the projector, projecting images onto a wall 2.5 meters away. The designated play area measured 1.8×1.8 meters and was situated 1.2 meters from the Kinect sensor.	Four Kinect Adventures games were implemented: rafting, cross-country running, ball hitting, and roller-coaster ride.	Patients performed specific movements in front of the motion sensor following Kinect Adventures minigame protocols. Each game session began with the manufacturer's instructions displaying game objectives and avatar control guidance.

XSL•FO
RenderX

First author (publication year)	PR+MSG	MSG[b] system	Game	Implementation
Sutanto et al (2019) [27]	Patients received standard hospital–based outpatient exercise training plus individualized 30-min Wii Fit gaming sessions in a dedicated room equipped with Nintendo Wii, balance board, and flat-screen TV.	The Nintendo Wii Fit gaming system included a Wii console, Wii balance board, and display screen.	Three specific Wii Fit exercises were implemented: yoga exercises (“Deep Breathing” and “Half Moon”), “Torso Twist,” and aerobic exercise (“Free Run”).	Each patient received initial one-to-one instruction for proper game execution. Sessions were performed standing on the Wii balance board (except during “Free Run”). Game duration, difficulty levels, and scores were documented. Exercise was discontinued if the pulse rate exceeded the Karnoven formula maximum, the respiratory rate exceeded 30/min, or SpO2 fell below 90%. All adverse events were recorded.
Mazzoleni et al (2014) [28]	Patients participated in traditional pulmonary rehabilitation complemented by seven daily 1-h Wii Fit Plus exercise sessions during the final wk of a 3-wk rehabilitation program. Sessions were conducted on an individual basis in a dedicated room equipped with a TV screen.	The Nintendo Wii Fit Plus system included a Wii console, Wii Balance Board as a haptic sensor–based interface, and display. The patient’s movements were translated to an “avatar” in the games, providing visual and auditory feedback.	Three Wii Fit Plus activities were selected and implemented in sequence: (1) “Yoga” activity involving deep breathing in standing position on the balance board (5 min at start and end), (2) “Jogging Plus” with 10-min feedback-aided running on spot, (3) “Twisting and Squat” comprising 10-min feedback-aided trunk twisting and arm/leg squatting exercises.	Each session followed a standardized protocol with continuous monitoring of ECG, pulse oximetry, and respiratory rate. Exercise intensity was maintained to elicit dyspnea levels of 4 - 6 on the modified Borg scale. Heart rate was kept below 80% of predicted maximum and oxygen saturation above 85%. Sessions were individually supervised by a physiotherapist. Exercise was modified or discontinued if heart rate exceeded limits, oxygen saturation fell below threshold, or if adverse events occurred.
Wei et al (2024) [23]	Patients received routine pulmonary rehabilitation combined with MSG sessions during hospitalization. Training was conducted individually in a dedicated room based on the patient’s condition in a semirecumbent, sitting, or standing position.	The Little King somatosensory gaming system (Subor) consisted of a host console, motion sensors, and a display screen. The system features dual sensor technology for enhanced motion capture accuracy and real-time feedback.	Around 3 specific games were selected: (1) “Kitchen Sharp Knife” for a 3-min warm-up, (2) “Table Tennis Master” for a 10-min exercise (5 min each for left and right hands), and (3) “Swimming Master” for a 4-min exercise.	Each session lasted approximately 20 min with continuous monitoring of oxygen saturation, breathing, blood pressure, and heart rate. Patients could rest for 1 min if experiencing fatigue or dyspnea and resume after symptoms resolved. Oxygen was administered if saturation was low; exercise resumed after stabilization. The physiatrist provided initial one-on-one instruction to ensure proper game execution. For home-based training after discharge, patients maintained exercise logs with twice-weekly telephone follow-up for supervision and guidance during the 8-wk intervention period.

XSL•FO
RenderX

First author (publication year)	PR+MSG	MSG[b] system	Game	Implementation
Zhu et al (2021) [25]	Patients received standard rehabilitation combined with Kinect 2.0–based somatosensory game training. Training sessions were conducted under the supervision of nurses and rehabilitation physicians with initial demonstration and guidance.	The system utilized an XBOX360 console with a Kinect 2.0 motion sensor connected to an E9i projector for display. The setup provided real-time motion tracking and visual feedback.	Around 3 specific Kinect games were implemented in sequence: (1) “Cross-country Running” engaging in cardio exercise, (2) “Fruit Ninja” involving upper limb coordination, and (3) “Obstacle Skiing” combining balance and full-body movements. Each game was demonstrated with a detailed movement breakdown before patient participation.	Exercise sessions were conducted once daily, 5 times per wk. Session duration was progressively increased from an initial 15 min to 35 min per d (adding 10 min weekly). Health care professionals provided real-time correction of improper movements during training. The three games were performed in a fixed sequence under professional supervision to ensure proper form and safety.
Zhou et al (2021) [24]	Patients received standard rehabilitation plus Kinect-based somatosensory game training. Training was conducted under the close supervision of rehabilitation physicians with vital signs monitoring through wearable pulse oximeters.	The system consisted of a control console, Kinect motion sensor, and a projector with speakers. The setup required specific positioning: console height at 1 meter, designated play area of 1.8×1.8 meters, positioned 1.2 meters from the Kinect sensor. Equipment calibration was performed before each session to ensure accurate motion tracking.	Around 4 Kinect games were implemented in sequence: (1) “Cross-country Running” for cardio exercise, (2) “Rafting” for balance training, (3) “Ball Hitting” for upper limb coordination, and (4) “Mountain Biking” for full-body movements. Each game sequence was preceded by professional demonstration and instruction.	Sessions lasted approximately 30 min, conducted once daily, 5 times per wk for 24 wks. A 1-min rest interval was provided between games. Training was performed under the arp Knifsupervision of health care professionals or family members with continuous monitoring through pulse oximetry connected to a home health management system. Patient vital signs were monitored in real-time, allowing physicians to review health status and adjust rehabilitation plans accordingly. Regular follow-up was maintained throughout the intervention period.
Jin (2021) [19]	Patients received standard care plus early rehabilitation with somatosensory game training based on individualized exercise prescriptions. Training sessions were conducted under one-on-one professional guidance starting within 48 h of admission.	The Little King somatosensory gaming system (Subor) consisted of a host console, motion sensors, and a display screen. The system features dual sensor technology for enhanced motion capture accuracy and real-time feedback.	About 3 specific games were implemented in sequence: (1) “Kitchen Sharp Knife” for 3-min warm-up, (2) “Table Tennis Master” for bilateral upper limb training (5 min each hand), and (3) “Swimming Master” for 4-min full-body exercise. Games were selected to target different aspects of rehabilitation.	Sessions lasted 20 min total, conducted once daily, 5 d per wk for 6 wks. Exercise position could be adjusted between standing, sitting, or semirecumbent based on patient condition. Brief rest periods (15 - 30 s) were allowed during games if patients experienced dyspnea or fatigue. Exercise intensity was maintained at Borg scale 3 - 6. Vital signs (heart rate, blood pressure, respiratory rate, and oxygen saturation) and Borg scores were monitored before and after exercise. Oxygen was administered if saturation fell below 88%, with exercise resuming after stabilization. Exercise was discontinued if the Borg score exceeded 6 or adverse events occurred.

XSL•FO
RenderX

First author (publication year)	PR+MSG	MSG[b] system	Game	Implementation
Geng (2020) [21]	Patients received standard guidance plus home-based upper limb training using somatosensory games. Initial instruction included game knowledge and operation methods with hands-on demonstration.	The Little King somatosensory gaming system (Subor) consisted of a host console, motion sensors, and a display screen. The system features dual sensor technology for enhanced motion capture accuracy and real-time feedback.	Three specific games were selected: (1) "Kitchen Sharp Knife," (2) "Table Tennis Master," and (3) "Boxing." Each game targeted different aspects of upper limb rehabilitation, with detailed operation guides provided to patients.	After baseline assessment, researchers installed gaming equipment in patients' homes. One-on-one instruction was provided until patients successfully demonstrated game mastery. Written operation manuals were distributed for reference. The 8-wk program consisted of 30-min sessions, 5 d per wk. The initial session was supervised for at least 30 min with SpO_2 monitoring and assessment of fatigue and dyspnea postexercise to ensure safety.
Geng et al (2020) [22]	Patients received standard health guidance plus home-based somatosensory game training. Training methods were individually instructed with regular follow-up through WeChat, phone calls, and home visits.	The Little King somatosensory gaming system (Subor) consisted of a host console, motion sensors, and a display screen. The system features dual sensor technology for enhanced motion capture accuracy and real-time feedback.	Three specific games were implemented: "Kitchen Sharp Knife" for warm-up and cool-down (3 min each, divided into 2×90 s sessions), followed by "Table Tennis Master" (15 min in 3×5 min sessions) and "Boxing" (9 min in 6×90 s sessions) as the main exercise components.	The 8-wk program consisted of daily 30-min sessions, conducted 5 d per wk. Exercise could be performed standing or seated, with brief rest periods (15 - 30 s) allowed during games if experiencing dyspnea or fatigue. Exercise intensity was self-regulated using the modified Borg scale (maintaining 3 - 5), with instructions to stop if experiencing shortness of breath, chest tightness, palpitations, or fatigue. Patients maintained exercise diaries documenting training date, duration, game types, intensity (Borg score), and subjective symptoms. Health care professionals regularly reviewed these diaries and provided individualized guidance through multiple communication channels.

XSL•FO
RenderX

First author (publication year)	PR+MSG	MSG[b] system	Game	Implementation
Jin et al (2020) [20]	Patients received standard care plus somatosensory game training guided by the early rehabilitation team. Training sessions were conducted under one-on-one professional instruction to ensure proper exercise technique.	The Little King somatosensory gaming system (Subor) consisted of a host console, motion sensors, and a display screen. The system features dual sensor technology for enhanced motion capture accuracy and real-time feedback.	Around 3 games were implemented in sequence: "Kitchen Sharp Knife" for warm-up (3 min), "Table Tennis Master" for bilateral upper limb training (5 min each hand), and "Swimming Master" for full-body exercise (4 min), with a 3-min rest period between activities.	Daily 20-min sessions could be performed in standing, sitting, or semirecumbent positions. Brief rest periods (15 - 30 s) were allowed during games if experiencing dyspnea or fatigue. Exercise intensity was maintained at modified Borg scale 3 - 6. Vital signs (heart rate, blood pressure, respiratory rate, and oxygen saturation) and the Borg scores were monitored before and after exercise. Oxygen was administered per physician orders if saturation fell below 88%, with exercise resuming after stabilization. Exercise was discontinued if Borg score exceeded 6 or adverse events occurred. The rehabilitation team performed daily assessments during hospitalization, documenting pre-exercise evaluation, exercise components, completion status, and postexercise assessment.

[a]PR+MSG: motion-sensing game–assisted pulmonary rehabilitation.

[b]MSG: motion-sensing game.

Risk of Bias in the 12 Included Studies

Based on the revised Cochrane Risk of Bias tool assessment of these 12 RCTs, all studies demonstrated "Some concerns" in the overall risk of bias evaluation, with none rated as high risk. In specific domains, the randomization process exhibited mixed quality, with 33.3% (4/12) of studies rated as low risk and 66.7% (8/12) as having some concerns. All studies (12/12, 100%) were rated as "Some concerns" for deviations from intended interventions, indicating potential issues with protocol adherence. However, all studies demonstrated a low risk of bias regarding missing outcome data and the selection of reported results, reflecting good follow-up rates and appropriate outcome reporting. The measurement of outcomes was generally appropriate, with 66.7% (8/12) of studies showing low risk and 33.3% (4/12) showing some concerns. Further details are provided in Figure S1A and S1B in Multimedia Appendix 2.

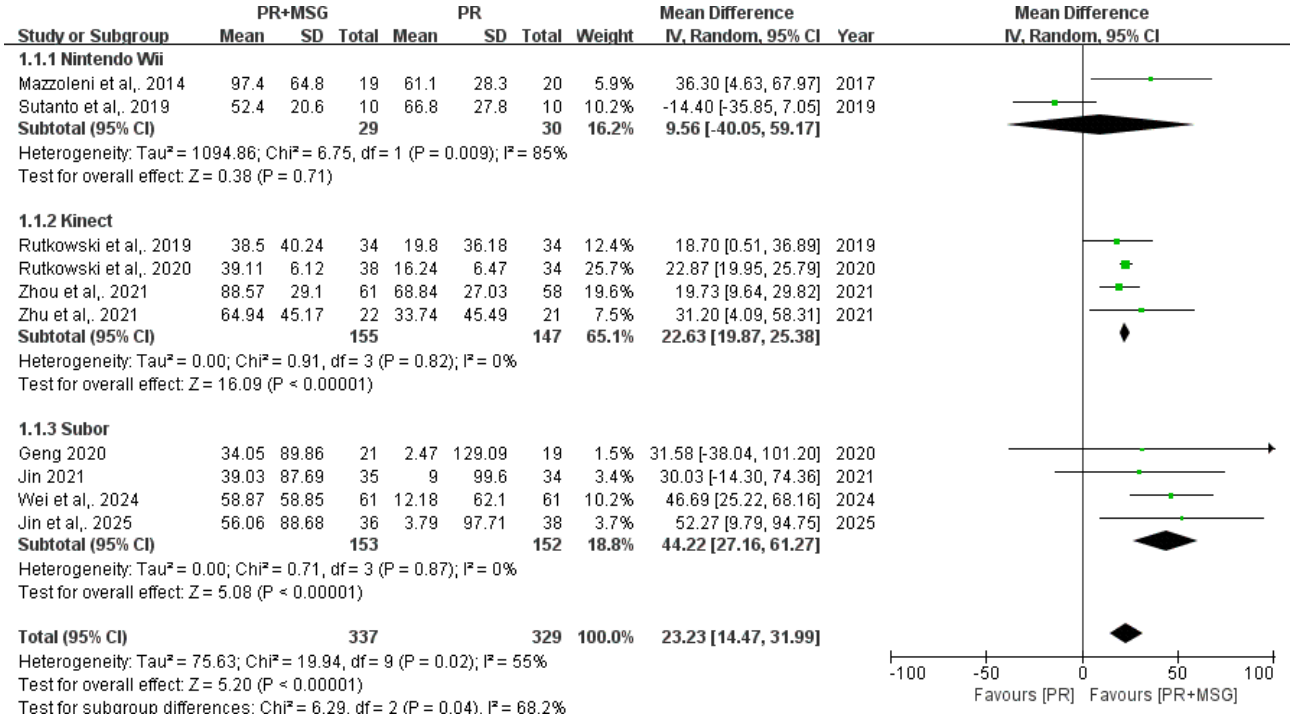
Outcomes

Exercise Capacity and Program Adherence

For the 6-Minute Walk Distance (6MWD), analysis of 10 RCTs (n=666) demonstrated that PR+MSG significantly improved walking distance compared with PR alone (MD 23.23, 95% CI 14.47-31.99 m; $P<.001$), with moderate heterogeneity ($I^2=55\%$). Subgroup analyses revealed that while Nintendo Wii–based interventions showed no significant advantage (MD 9.56, 95% CI −14.05 to 59.17 m; $P=.71$), both Kinect-based (MD 22.63, 95% CI 19.87-25.38 m; $P<.001$) and Subor-based systems (MD 44.22, 95% CI 27.16 to 61.27 m; $P<.001$) demonstrated significant improvements. Both Kinect and Subor subgroups display no heterogeneity ($I^2=0\%$) (Figure 2).

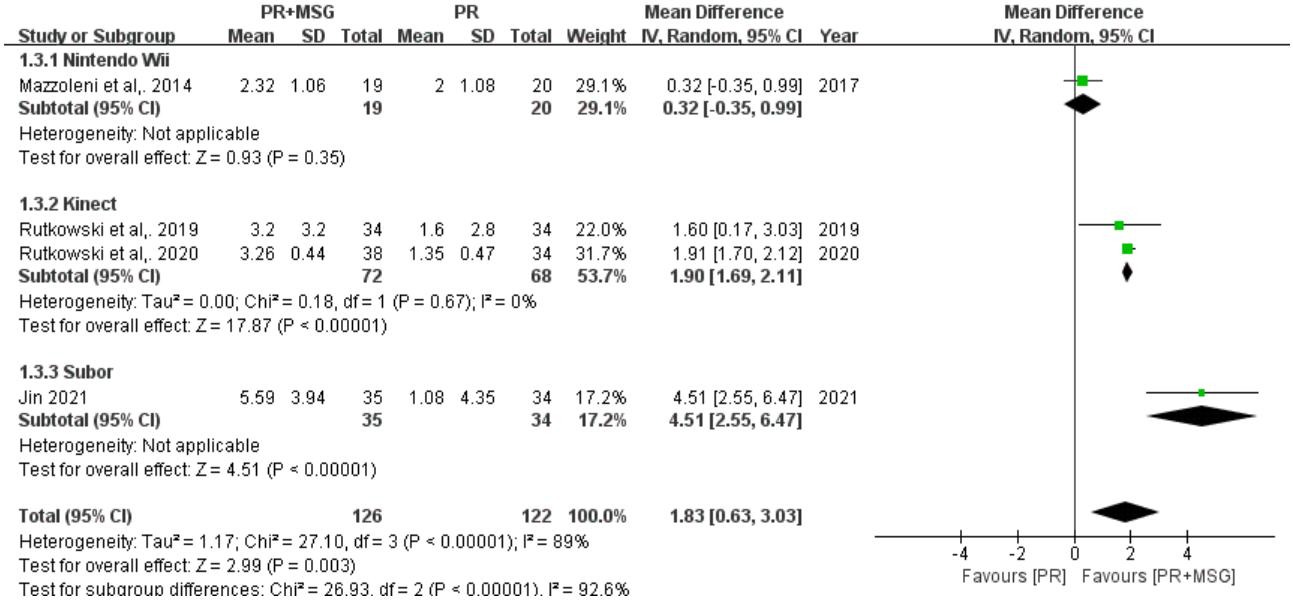
XSL•FO
RenderX

Figure 2. Forest plots depicting the effect of motion-sensing game–assisted pulmonary rehabilitation (PR+MSG) versus routine pulmonary rehabilitation (PR) on 6-Minute Walk Distance in patients with chronic obstructive pulmonary disease [12,13,19,21,23-28].



Analysis of the 30-Second Arm Curl Test from 4 RCTs (n=248) indicated that PR+MSG significantly enhanced upper limb muscle strength (MD 1.83, 95% CI 0.63-3.03 repetitions; *P*=.003), despite considerable heterogeneity (I^2=89%). In subgroup analyses, Nintendo Wii–based training showed no significant improvement (MD 0.32, 95% CI −0.35 to 0.99; *P*=.35), while both Kinect-based (MD 1.90, 95% CI 1.69 to 2.11; *P*<.001) and Subor-based systems (MD 4.51, 95% CI 2.55-6.47; *P*<.001) demonstrated significant benefits (Figure 3).

Figure 3. Forest plots depicting the effect of motion-sensing game–assisted pulmonary rehabilitation (PR+MSG) versus routine pulmonary rehabilitation (PR) on upper limb strength as measured by the 30-Second Arm Curl Test in patients with chronic obstructive pulmonary disease [13,19,26,28].



For balance function, assessed by the Brief Balance Evaluation Systems Test, analysis of 3 studies (n=262) revealed significant improvements with PR+MSG (MD 2.34, 95% CI 1.52-3.17 points; *P*<.001), with remarkable consistency across studies (I^2=0%; *P*=.84) (Figure 4). Additionally, examination of exercise training completion rates from 2 studies (n=162) showed that PR+MSG significantly increased adherence (OR 3.00, 95% CI 1.38-6.52; *P*=.005), with no heterogeneity (I^2=0%; *P*=.35). (Figure 5).

XSL•FO
RenderX

Figure 4. Forest plots depicting the effect of motion-sensing game–assisted pulmonary rehabilitation (PR+MSG) versus routine pulmonary rehabilitation (PR) on balance function assessed by the Brief Balance Evaluation Systems Test in patients with chronic obstructive pulmonary disease [12,20,23].

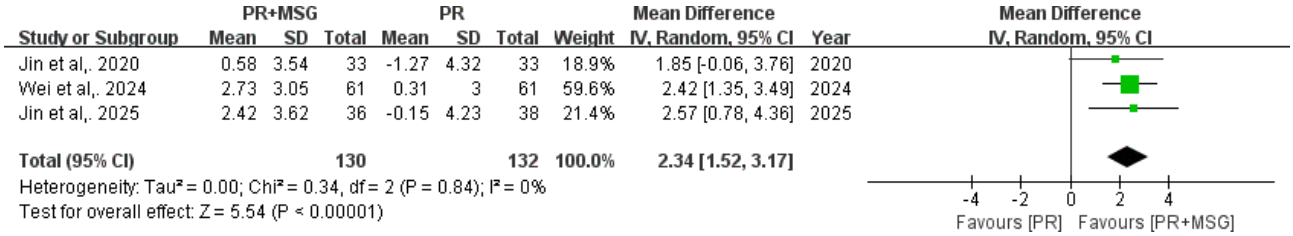
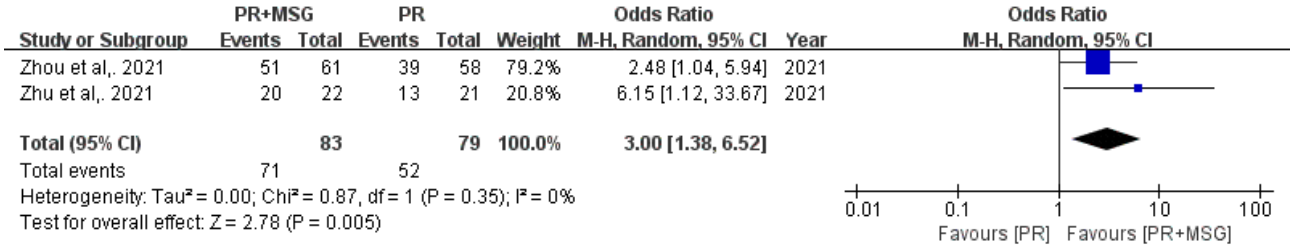


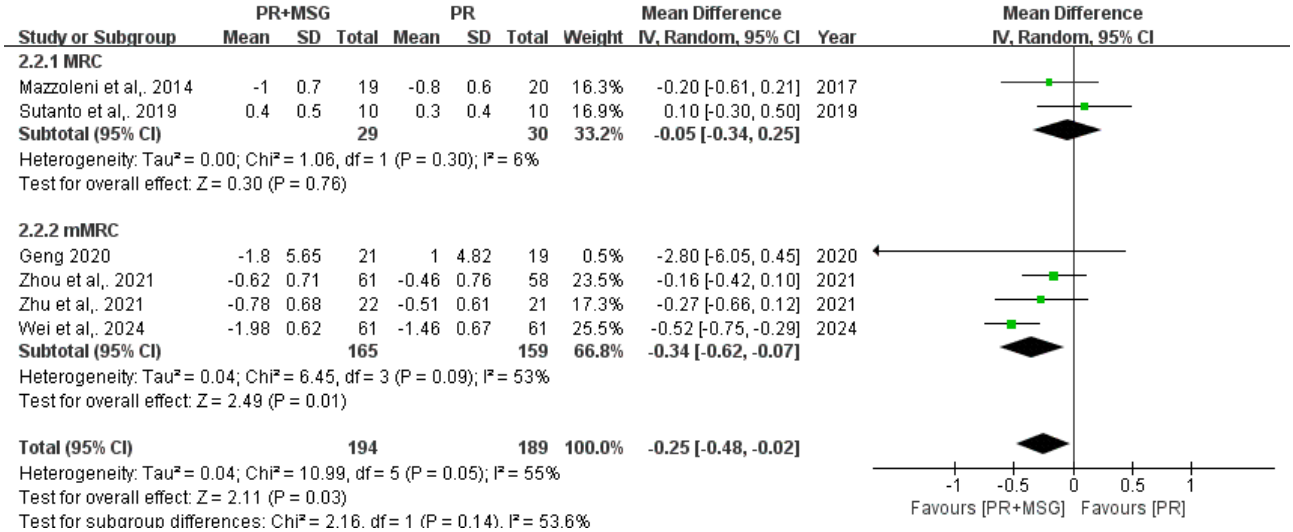
Figure 5. Forest plots illustrating the comparison of completion rates of exercise training between motion-sensing game–assisted pulmonary rehabilitation and routine pulmonary rehabilitation in patients with chronic obstructive pulmonary disease [24,25].



Severity of Dyspnea and Quality of Life

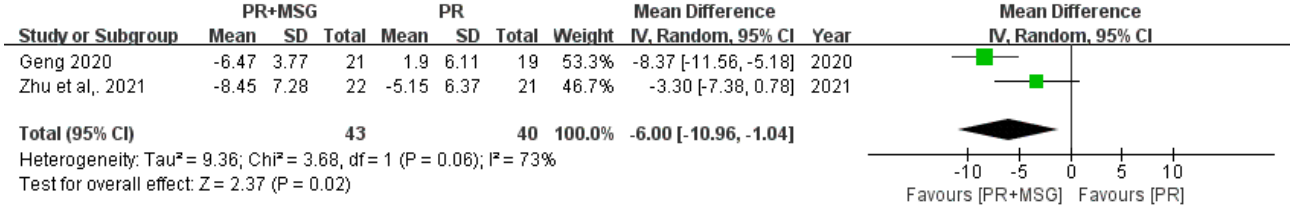
Analysis of dyspnea severity encompassing 6 studies (n=383) demonstrated that PR+MSG significantly alleviated dyspnea compared with routine PR alone (MD −0.25, 95% CI −0.48 to −0.02; *P*=.03), with moderate heterogeneity (I^2=55%). Subgroup analyses revealed differential effects based on assessment tools, while studies using the Medical Research Council (MRC) scale showed no significant improvement (MD −0.05, 95% CI −0.34 to 0.25; *P*=.76; I^2=6%, 2 studies), those using the modified MRC scale demonstrated significant benefits (MD −0.34, 95% CI −0.62 to −0.07; *P*=.01; I^2=53%, 4 studies) (Figure 6).

Figure 6. Forest plots depicting the effect of motion-sensing game–assisted pulmonary rehabilitation (PR+MSG) versus routine pulmonary rehabilitation (PR) on the severity of dyspnea in patients with chronic obstructive pulmonary disease [21,23-25,27,28].



For HRQL, analysis of 2 studies (n=83) using the COPD Assessment Test (CAT) showed significant improvements with PR+MSG (MD −6.00, 95% CI −10.96 to −1.04; *P*=.02). Despite moderate heterogeneity (I^2=73%), both studies consistently indicated benefits favoring PR+MSG intervention (Figure 7).

Figure 7. Forest plots depicting the effect of motion-sensing game–assisted pulmonary rehabilitation (PR+MSG) versus routine pulmonary rehabilitation (PR) on health-related quality of life in patients with chronic obstructive pulmonary disease [21,25].

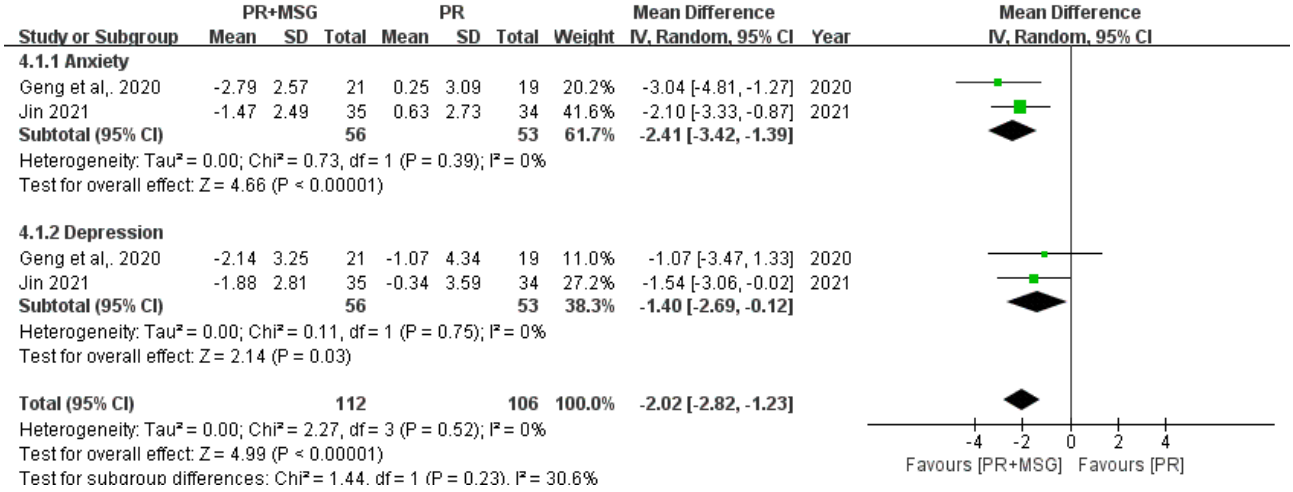


XSL•FO
RenderX

Psychological State

Assessment of psychological outcomes using the Hospital Anxiety and Depression Scale from 2 studies (n=218) revealed significant improvements with PR+MSG for both anxiety (MD −2.41, 95% CI −3.42 to −1.39; *P*<.001) and depression (MD −1.40, 95% CI −2.69 to −0.42; *P*=.03). Notably, both psychological domains showed consistent benefits across studies with no heterogeneity (I^2=0%), with anxiety demonstrating a more pronounced effect size compared with depression (Figure 8).

Figure 8. Forest plots depicting the effect of motion-sensing game–assisted pulmonary rehabilitation (PR+MSG) versus routine pulmonary rehabilitation (PR) on psychological outcomes in patients with chronic obstructive pulmonary disease [19,22].



Risk of Bias Within Outcomes

The quality of all synthesized evidence was assessed as “low” or “very low” using the GRADE approach (Table S2 in Multimedia Appendix 3). Sensitivity analysis used the selective model analysis method, wherein both fixed-effect and random-effect models were applied to the aforementioned experimental data, yielding consistent results. A leave-one-out analysis was conducted by sequentially excluding each study and re-estimating the analysis. This process revealed minimal changes in all indicators compared with the original analysis, suggesting low sensitivity and enhancing the robustness and reliability of the results. Additionally, the funnel plot for the 6MWD (10 studies, P_{Egger}=0.69) showed no apparent publication bias.

Discussion

Principal Findings

This meta-analysis demonstrates that PR+MSG shows advantages over routine PR in improving exercise capacity, balance function, symptom control, quality of life, and psychological outcomes across multiple health domains. These findings provide scientific evidence supporting the integration of gaming technology into rehabilitation programs.

Exercise Capacity and Program Adherence

Our analysis demonstrates that PR+MSG significantly enhances exercise capacity across multiple domains, consistently outperforming conventional rehabilitation programs. Of particular note, newer systems such as Kinect and Subor show superior outcomes, likely attributable to their advanced motion capture capabilities and comprehensive whole-body engagement features. These systems effectively address multiple aspects of physical function simultaneously, including cardiorespiratory endurance, upper limb strength, and balance function [29].

The 6MWD, a critical functional parameter in patients with COPD, reflects overall physical endurance and directly correlates with disease severity, mortality risk, and performance in activities of daily living [30]. Similarly, improvements in upper limb strength assessed by the 30-Second Arm Curl Test address a frequently overlooked aspect of COPD-related dysfunction. Patients with COPD commonly experience upper limb muscle weakness resulting from respiratory muscle overload, deconditioning, and systemic inflammation [31], which significantly impairs performance in essential daily activities requiring arm movements, including self-care and household tasks. Balance performance, evaluated through Balance Evaluation Systems Test, holds particular relevance as patients with COPD exhibit increased fall risk due to muscle weakness, altered body mechanics, oxygen desaturation, and medication side effects [32]. These balance impairments represent significant contributors to disability and reduced quality of life. The documented improvements in both upper limb strength and balance function therefore have substantial clinical implications, as they directly enhance patients’ functional independence, daily activity performance, and reduce fall risk [31]. These enhanced outcomes align with and extend existing evidence supporting both the effectiveness of motion-sensing game systems in improving postural control [33] and the superior benefits of comprehensive, whole-body engagement during PR [34]. The multidimensional improvements observed in our analysis suggest that MSG-based interventions provide a more holistic approach to addressing the complex functional limitations experienced by patients with COPD.

The significantly higher program completion rates represent a crucial advancement in addressing poor adherence, a persistent

XSL•FO
RenderX

challenge in PR [35]. This is particularly significant as patients with COPD typically demonstrate notoriously low adherence rates due to symptom burden, comorbidities, and motivational challenges [36]. The direct relationship between improved adherence and enhanced rehabilitation outcomes highlights the clinical importance of these findings. Interventions improving adherence directly enhance rehabilitation outcomes. This enhanced engagement can be attributed to the gaming elements' unique features: immediate feedback, achievement motivation, and progressive challenge levels [11]. The motion-controlled interaction promotes motor learning and functional movement patterns that directly translate to daily activities, addressing key limitations of traditional rehabilitation exercises [37]. These findings suggest that MSGs offer a promising approach to optimize exercise training outcomes while maintaining patient engagement.

Symptoms and Quality of Life

The significant reduction in dyspnea severity, particularly as measured by the modified MRC scale, indicates that PR+MSG offers enhanced symptom management benefits compared with conventional rehabilitation. The superior outcomes in modified MRC subgroups likely reflect this scale's greater sensitivity in detecting functional dyspnea changes [38]. The moderate effect size and low heterogeneity across studies, despite varying gaming platforms and protocols, demonstrate consistent improvement in breathlessness management [39,40]. Dyspnea severity, while primarily a respiratory symptom, profoundly affects physical activity participation, psychological well-being, and social engagement, making its improvement clinically meaningful given its substantial impact on patients' quality of life, daily activities, and health care utilization [3]. The enhanced symptom management can be attributed to the gaming environment's distinctive features: progressive exercise intensity with engagement through distraction, real-time feedback for breathing control and pacing, and interactive elements that boost self-efficacy in managing breathlessness [41,42]. This aligns with previous research confirming that interactive gaming interventions effectively reduce perceived dyspnea through improved breathing control and exercise tolerance [3,43], suggesting a mechanistic pathway by which MSG intervention translates to functional improvement beyond conventional rehabilitation approaches.

Regarding HRQL, our findings demonstrate that PR+MSG yields significant improvements in patients' perceived health status. This enhancement is particularly meaningful as HRQL represents a crucial patient-centered outcome that reflects the intervention's comprehensive impact on daily functioning and overall well-being [44]. These results align with existing evidence suggesting that engaging, interactive exercise modalities enhance both physical outcomes and quality of life measures [45]. While moderate heterogeneity in the analysis reflects variations in implementation approaches and patient characteristics, the improved CAT scores suggest that game-based interventions effectively address multiple health status domains, including symptom burden, activity limitation, and psychosocial impact [46].

Psychological Outcomes

Our meta-analysis demonstrated significant improvements in psychological well-being with PR+MSG intervention, with notably greater effects on anxiety compared with depression scores. These findings align with previous evidence suggesting that interactive gaming environments can effectively mitigate psychological distress in chronic disease management [47,48]. The gaming elements' distinctive features—achievement-oriented tasks, immediate feedback, and enhanced social interaction—appear particularly effective in reducing anxiety-related symptoms, which commonly impede exercise participation in patients with COPD [49-51]. This psychological benefit is especially relevant considering patients with COPD experience substantially higher rates of psychological distress compared with the general population, which directly contributes to disease progression, exacerbation risk, and mortality [52]. Additionally, reduced anxiety and depression levels contribute to better exercise adherence [53], while gamification elements help maintain long-term engagement through enhanced enjoyment and reduced perceived exertion [54]. These results support implementing comprehensive rehabilitation approaches that address both physical and psychological aspects of chronic respiratory diseases [3].

Strengths and Limitations

This meta-analysis represents the first comprehensive review specifically focusing on MSG-assisted exercise training in patients with COPD, distinguishing itself from previous reviews that broadly examined various technological interventions. By systematically evaluating multiple outcomes (exercise capacity, symptoms, quality of life, and psychological parameters) through RCTs, this study provides the first synthesis of evidence on the comprehensive benefits of MSG-based rehabilitation. These findings offer practical guidance for enhancing patient care in contemporary clinical settings while suggesting promising directions for future research in technology-enhanced rehabilitation strategies.

Several key limitations warrant consideration in this study. First, methodologically, effective blinding was challenging due to the distinctive nature of game-assisted rehabilitation compared with conventional approaches, potentially introducing implementation bias. The heterogeneity among included studies manifested primarily in the varying gaming platforms and rehabilitation protocols, as well as inconsistent intervention specifications (intensity, frequency, and duration), which may have introduced clinical heterogeneity and affected outcome comparability. While we identified these possible contributing factors, the relatively small number of eligible studies precluded us from conducting meaningful subgroup analyses to quantify their specific contributions to the observed heterogeneity. Additionally, most included studies were constrained by small sample sizes and relatively short follow-up periods, limiting both the assessment of long-term efficacy and internal validity. The overall methodological quality of the included studies was suboptimal, with most evidence rated as "low" or "very low" quality. Furthermore, this study did not systematically evaluate the cost-effectiveness of implementing game-assisted

XSL•FO
RenderX

rehabilitation systems in clinical settings. These limitations underscore the need for future research to conduct larger, well-designed RCTs with standardized intervention protocols, extended follow-up periods, and comprehensive cost-benefit analyses to provide more reliable theoretical guidance for clinical practice.

Implications for Future Research and Practices

The findings demonstrate that incorporating MSGs into PR programs offers a promising strategy to enhance both clinical outcomes and patient engagement. This approach may be particularly beneficial for three key patient populations: those with poor adherence to conventional rehabilitation protocols, patients reporting exercise-related anxiety or low motivation, and individuals requiring concurrent physical and psychological support. Health care providers can leverage this technology-enhanced rehabilitation method to create more personalized and engaging treatment plans, potentially improving long-term adherence and rehabilitation outcomes. Additionally, the game-based format may help clinicians better monitor patient progress and adjust intervention intensity in real-time, though implementation should consider individual patient preferences, physical capabilities, and facility resources.

Conclusion

This systematic review and meta-analysis suggests that PR+MSG may improve multiple clinical outcomes in patients with COPD, including exercise capacity, respiratory symptoms, HRQL, and psychological well-being. The observed heterogeneity in effectiveness across gaming platforms highlights the importance of personalized intervention selection based on individual patient characteristics and health care resource settings. While our results indicate potential benefits, the predominance of low-quality evidence underscores the need for risk-benefit assessment in clinical implementation of PR+MSG approaches. Future research priorities should include larger, methodologically rigorous RCTs, developing standardized intervention protocols, investigating sustained therapeutic effects, and conducting cost-effectiveness analyses to establish more definitive evidence for the optimal implementation of gaming technology in PR programs.

Data Availability

All data relevant to the study are included in this paper.

Authors' Contributions

CH, XL, YL, SZ, and GC conceptualized and designed the study. CH, XL, YL, SZ, and GC acquired, analyzed, and interpreted the data. CH and XL drafted the manuscript. CH, XL, and YL performed statistical analysis. SZ and GC supervised the study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Comprehensive search strategy used to identify randomized controlled trials evaluating motion-sensing game–assisted pulmonary rehabilitation in patients with chronic obstructive pulmonary disease.
[DOCX File, 19 KB - games_v13i1e69562_app1.docx]

Multimedia Appendix 2
Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included studies.
[DOCX File, 99 KB - games_v13i1e69562_app2.docx]

Multimedia Appendix 3
Grading of Recommendations Assessment, Development, and Evaluations evidence profile and summary of findings for primary outcomes comparing motion-sensing game–assisted pulmonary rehabilitation versus routine pulmonary rehabilitation in patients with chronic obstructive pulmonary disease.
[DOCX File, 21 KB - games_v13i1e69562_app3.docx]

Checklist 1
Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 checklist.
[PDF File, 138 KB - games_v13i1e69562_app4.pdf]

References

1. Agustí A, Celli BR, Criner GJ, et al. Global initiative for chronic obstructive lung disease 2023 report: GOLD executive summary. Eur Respir J 2023 Apr;61(4):2300239. [doi: 10.1183/13993003.00239-2023] [Medline: 36858443]
2. Global health estimates: life expectancy and leading causes of death and disability. World Health Organization. URL: https://www.who.int/data/gho/data/themes/mortality-and-global-health-estimates [accessed 2025-05-05]

XSL•FO
RenderX

3. Spruit MA, Singh SJ, Garvey C, et al. An official American Thoracic Society/European Respiratory Society statement: key concepts and advances in pulmonary rehabilitation. Am J Respir Crit Care Med 2013 Oct 15;188(8):e13-e64. [doi: 10.1164/rccm.201309-1634ST] [Medline: 24127811]
4. McCarthy B, Casey D, Devane D, et al. Pulmonary rehabilitation for chronic obstructive pulmonary disease. Cochrane Database Syst Rev 2015 Feb 23;2015(2):CD003793. [doi: 10.1002/14651858.CD003793.pub3] [Medline: 25705944]
5. Rochester CL, Vogiatzis I, Holland AE, et al. An official American Thoracic Society/European Respiratory Society Policy statement: enhancing implementation, use, and delivery of pulmonary rehabilitation. Am J Respir Crit Care Med 2015 Dec 1;192(11):1373-1386. [doi: 10.1164/rccm.201510-1966ST] [Medline: 26623686]
6. Cox NS, Oliveira CC, Lahham A, et al. Pulmonary rehabilitation referral and participation are commonly influenced by environment, knowledge, and beliefs about consequences: a systematic review using the Theoretical Domains Framework. J Physiother 2017 Apr;63(2):84-93. [doi: 10.1016/j.jphys.2017.02.002] [Medline: 28433238]
7. Dalko K, Elsuson HA, Kalter I, et al. Virtual reality applications for the implementation of domestic respiratory rehabilitation programs for patients with long COVID and post-COVID condition: scoping review. JMIR Serious Games 2024 May 31;12:e52309. [doi: 10.2196/52309] [Medline: 38819890]
8. Skrzatek A, Nuic D, Cherif S, et al. Brain modulation after exergaming training in advanced forms of Parkinson's disease: a randomized controlled study. J Neuroeng Rehabil 2024 Aug 5;21(1):133. [doi: 10.1186/s12984-024-01430-w] [Medline: 39103924]
9. Wüest S, Borghese NA, Pirovano M, et al. Usability and effects of an Exergame-Based Balance Training Program. Games Health J 2014 Apr 1;3(2):106-114. [doi: 10.1089/g4h.2013.0093] [Medline: 24804155]
10. Laver K, George S, Thomas S, et al. Virtual reality for stroke rehabilitation: an abridged version of a Cochrane review. Eur J Phys Rehabil Med 2015 Aug;51(4):497-506. [Medline: 26158918]
11. Meekes W, Stanmore EK. Motivational determinants of exergame participation for older people in assisted living facilities: mixed-methods study. J Med Internet Res 2017 Jul 6;19(7):e238. [doi: 10.2196/jmir.6841] [Medline: 28684385]
12. Jin X, Jin M, Zhang B, et al. The association of conventional therapy associated with somatosensory interactive game enhances the effects of early pulmonary rehabilitation for patients with acute exacerbation of chronic obstructive pulmonary disease: a randomized controlled trial. Games Health J 2025 Apr;14(2):127-135. [doi: 10.1089/g4h.2023.0095] [Medline: 39207252]
13. Rutkowski S, Rutkowska A, Kiper P, et al. Virtual reality rehabilitation in patients with chronic obstructive pulmonary disease: a randomized controlled trial. Int J Chron Obstruct Pulmon Dis 2020;15:117-124. [doi: 10.2147/COPD.S223592] [Medline: 32021150]
14. Bonnevie T, Smondack P, Elkins M, et al. Advanced telehealth technology improves home-based exercise therapy for people with stable chronic obstructive pulmonary disease: a systematic review. J Physiother 2021 Jan;67(1):27-40. [doi: 10.1016/j.jphys.2020.12.006] [Medline: 33358547]
15. Cox NS, Dal Corso S, Hansen H, et al. Telerehabilitation for chronic respiratory disease. Cochrane Database Syst Rev 2021 Jan 29;1(1):CD013040. [doi: 10.1002/14651858.CD013040.pub2] [Medline: 33511633]
16. Sterne JAC, Savović J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019 Aug 28;366:l4898. [doi: 10.1136/bmj.l4898] [Medline: 31462531]
17. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 2008 Apr 26;336(7650):924-926. [doi: 10.1136/bmj.39489.470347.AD] [Medline: 18436948]
18. Yan Q, Zhu C, Li L, et al. The effect of targeted palliative care interventions on depression, quality of life and caregiver burden in informal caregivers of advanced cancer patients: a systematic review and meta-analysis of randomized controlled trials. Int J Nurs Stud 2024 Dec;160:104895. [doi: 10.1016/j.ijnurstu.2024.104895] [Medline: 39305684]
19. Jin X. Application and effect evaluation of somatosensory interactive games in exercise intervention for inpatients with chronic obstructive pulmonary disease at the acute exacerbation stage [Master's thesis in Chinese]. : Soochow University; 2021 URL: https://d.wanfangdata.com.cn/thesis/ChhUaGVzaXNOZXdTMjAyNDA5MjAxNTE3MjUSCFkzODc2Nzg1GghkOGZrY3NweQ%3D%3D [accessed 2025-05-09]
20. Jin X, Zhang B, Niu M, et al. Somatosensory interactive games on the balance function of patients in the acute exacerbation stage of chronic obstructive pulmonary disease. Nurs J Chin PLA 2020;37(11):5-8. [doi: 10.3969/j.issn.1008-9993.2020.11.002]
21. Geng M. Application study of home-based upper limb exercise program based on somatosensory interactive games in patients with stable chronic obstructive pulmonary disease [master's thesis]. : Soochow University; 2020 URL: https://link.cnki.net/doi/10.27351/d.cnki.gszhu.2020.003909 [accessed 2025-05-27]
22. Geng M, Niu M, Han Y, et al. Effects of applying home-based somatosensory interactive game training for fatigue and negative emotions of patients with stable COPD. Nurs Pract Res 2020;17(13):68-71. [doi: 10.3969/j.issn.1672-9676.2020.13.026]
23. Wei H, Wang J, Wang X. Effects of somatosensory interactive games on dyspnea, balance function and exercise capacity in patients with acute exacerbation of chronic obstructive pulmonary disease: a randomized controlled trial. Maternal Child Care 2024;4(7):1724-1727 [FREE Full text]

XSL•FO
RenderX

24. Zhou T, Zhang Y, Zhu Y. Effects of exercise program based on somatosensory interactive games in elderly patients with stable COPD. Chin J Modern Nurs 2021;27(3):303-308. [doi: 10.3760/cma.j.cn115682-20200914-05351]
25. Zhu Y, Xie H, Zhang L. Clinical effect of applying somatosensory games based on Kinect 2.0 system to elderly patients with chronic obstructive pulmonary disease. Acta Academiae Medicinae Wannan 2021;40(3):291-294. [doi: 10.3969/j.issn.1002-0217.2021.03.024]
26. Rutkowski S, Rutkowska A, Jastrzębski D, et al. Effect of virtual reality-based rehabilitation on physical fitness in patients with chronic obstructive pulmonary disease. J Hum Kinet 2019 Oct;69:149-157. [doi: 10.2478/hukin-2019-0022] [Medline: 31666897]
27. Sutanto YS, Makhabah DN, Aphridasari J, et al. Videogame assisted exercise training in patients with chronic obstructive pulmonary disease: a preliminary study. Pulmonology 2019 Sep 30;25(5):275-282. [doi: 10.1016/j.pulmoe.2019.03.007]
28. Mazzoleni S, Montagnani G, Vagheggini G, et al. Interactive videogame as rehabilitation tool of patients with chronic respiratory diseases: preliminary results of a feasibility study. Respir Med 2014 Oct;108(10):1516-1524. [doi: 10.1016/j.rmed.2014.07.004] [Medline: 25087837]
29. Hernandez-Martinez J, Cid-Calfucura I, Herrera-Valenzuela T, et al. Xbox Kinect Sports vs. Nintendo Switch Sports and their effects on body composition and physical performance in older females: a randomized controlled trial. J Clin Med 2024 Aug 23;13(17):17. [doi: 10.3390/jcm13174987] [Medline: 39274198]
30. Janjua S, Banchoff E, Threapleton CJ, et al. Digital interventions for the management of chronic obstructive pulmonary disease. Cochrane Database Syst Rev 2021 Apr 19;4(4):CD013246. [doi: 10.1002/14651858.CD013246.pub2] [Medline: 33871065]
31. Karagiannis C, Savva C, Korakakis V, et al. The effects of upper limb exercise training on upper limb muscle strength in people with chronic obstructive pulmonary disease: a systematic review and meta-analysis of randomized controlled trials. Ther Adv Respir Dis 2023;17:17534666231170813. [doi: 10.1177/17534666231170813] [Medline: 37165688]
32. Paixão C, Rebelo P, Oliveira A, et al. Responsiveness and minimal clinically important difference of the Brief-BESTest in people with COPD after pulmonary rehabilitation. Phys Ther 2021 Nov 1;101(11):11. [doi: 10.1093/ptj/pzab209] [Medline: 34499161]
33. Delbressine JM, Vaes AW, Goërtz YM, et al. Effects of exercise-based interventions on fall risk and balance in patients with chronic obstructive pulmonary disease: a systematic review. J Cardiopulm Rehabil Prev 2020 May;40(3):152-163. [doi: 10.1097/HCR.0000000000000513] [Medline: 32355076]
34. Loughran KJ, Emerson J, Avery L, et al. Exercise-based interventions targeting balance and falls in people with COPD: a systematic review and meta-analysis. Eur Respir Rev 2024 Apr;33:172. [doi: 10.1183/16000617.0003-2024] [Medline: 38925795]
35. Schrijver J, Lenferink A, Brusse-Keizer M, et al. Self-management interventions for people with chronic obstructive pulmonary disease. Cochrane Database Syst Rev 2022 Jan 10;1(1):CD002990. [doi: 10.1002/14651858.CD002990.pub4] [Medline: 35001366]
36. Skibdal KM, Emme C, Hansen H. A mixed-methods study of thoughts and attitudes towards participation in pulmonary telerehabilitation among people with severe and very severe COPD who declined participation in pulmonary rehabilitation. Patient Prefer Adherence 2022;16:2781-2798. [doi: 10.2147/PPA.S380832]
37. Barry G, Galna B, Rochester L. The role of exergaming in Parkinson's disease rehabilitation: a systematic review of the evidence. J Neuroeng Rehabil 2014 Mar 7;11:33. [doi: 10.1186/1743-0003-11-33] [Medline: 24602325]
38. Crisafulli E, Clini EM. Measures of dyspnea in pulmonary rehabilitation. Multidiscip Respir Med 2010;5(3):202-210. [doi: 10.1186/2049-6958-5-3-202]
39. Pitta F, Probst VS, Kovelis D, et al. Validation of the Portuguese version of the London Chest Activity of Daily Living Scale (LCADL) in chronic obstructive pulmonary disease patients. Rev Port Pneumol 2008;14(1):27-47. [Medline: 18265916]
40. O'Shea SD, Taylor NF, Paratz JD. Progressive resistance exercise improves muscle strength and may improve elements of performance of daily activities for people with COPD: a systematic review. Chest 2009 Nov;136(5):1269-1283. [doi: 10.1378/chest.09-0029] [Medline: 19734323]
41. Jones S, Man WDC, Gao W, et al. Neuromuscular electrical stimulation for muscle weakness in adults with advanced disease. Cochrane Database Syst Rev 2016 Oct 17;10(10):CD009419. [doi: 10.1002/14651858.CD009419.pub3] [Medline: 27748503]
42. Tabak M, Vollenbroek-Hutten MM, van der Valk PD, et al. A telerehabilitation intervention for patients with chronic obstructive pulmonary disease: A randomized controlled pilot trial. Clin Rehabil 2014 Jun;28(6):582. [doi: 10.1177/0269215513512495] [Medline: 24293120]
43. Garrod R, Bestall JC, Paul EA, et al. Development and validation of a standardized measure of activity of daily living in patients with severe COPD: the London Chest Activity of Daily Living scale(LCADL). Respir Med 2000 Jun;94(6):589-596. [doi: 10.1053/rmed.2000.0786] [Medline: 10921765]
44. Bishop JA, Spencer LM, Dwyer TJ, et al. Effect of pulmonary rehabilitation duration on exercise capacity and health-related quality of life in people with chronic obstructive pulmonary disease (PuRe Duration Trial): a randomized controlled equivalence trial. Respirology 2025 Jan;30(1):41-50. [doi: 10.1111/resp.14820] [Medline: 39228164]

XSL•FO
RenderX

45. Butler SJ, Lee AL, Goldstein RS, et al. Active video games as a training tool for individuals with chronic respiratory diseases: a systematic review. J Cardiopulm Rehabil Prev 2019 Mar;39(2):85-90. [doi: 10.1097/HCR.0000000000000320] [Medline: 29485524]
46. Ammous O, Feki W, Lotfi T, et al. Inspiratory muscle training, with or without concomitant pulmonary rehabilitation, for chronic obstructive pulmonary disease (COPD). Cochrane Database Syst Rev 2023 Jan 6;1(1):CD013778. [doi: 10.1002/14651858.CD013778.pub2] [Medline: 36606682]
47. Li J, Theng YL, Foo S. Game-based digital interventions for depression therapy: a systematic review and meta-analysis. Cyberpsychol Behav Soc Netw 2014 Aug;17(8):519-527. [doi: 10.1089/cyber.2013.0481] [Medline: 24810933]
48. Baranowski T, Buday R, Thompson DI, et al. Playing for real: video games and stories for health-related behavior change. Am J Prev Med 2008 Jan;34(1):74-82. [doi: 10.1016/j.amepre.2007.09.027] [Medline: 18083454]
49. Yohannes AM, Alexopoulos GS. Depression and anxiety in patients with COPD. Eur Respir Rev 2014 Sep;23(133):345-349. [doi: 10.1183/09059180.00007813] [Medline: 25176970]
50. Fleming TM, Bavin L, Stasiak K, et al. Serious games and gamification for mental health: current status and promising directions. Front Psychiatry 2016;7:215. [doi: 10.3389/fpsyt.2016.00215] [Medline: 28119636]
51. Raz G, Davidovitch S, Halevi M, et al. Impact of movie and video game elements on tic manifestation in children. Eur J Neurol 2024 Feb;31(2):e16120. [doi: 10.1111/ene.16120] [Medline: 37946628]
52. Gordon CS, Waller JW, Cook RM, et al. Effect of pulmonary rehabilitation on symptoms of anxiety and depression in COPD: a systematic review and meta-analysis. Chest 2019 Jul;156(1):80-91. [doi: 10.1016/j.chest.2019.04.009] [Medline: 31034818]
53. Mikkelsen K, Stojanovska L, Polenakovic M, et al. Exercise and mental health. Maturitas 2017 Dec;106:48-56. [doi: 10.1016/j.maturitas.2017.09.003] [Medline: 29150166]
54. Lyons EJ, Hatkevich C. Prevalence of behavior changing strategies in fitness video games: theory-based content analysis. J Med Internet Res 2013 May 7;15(5):e81. [doi: 10.2196/jmir.2403] [Medline: 23651701]

Abbreviations

6MWD: 6-Minute Walk Distance
CAT: COPD Assessment Test
COPD: chronic obstructive pulmonary disease
GRADE: Grading of Recommendations Assessment, Development, and Evaluations
HRQL: health-related quality of life
MD: mean difference
MRC: Medical Research Council
MSG: motion-sensing game
OR: odds ratio
PR: pulmonary rehabilitation
PR+MSG: motion-sensing game–assisted pulmonary rehabilitation
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROSPERO: International Prospective Register of Systematic Reviews
RCT: randomized controlled trial

Edited by A Coristine; submitted 03.12.24; peer-reviewed by F Wu, QQ Shi; revised version received 31.03.25; accepted 31.03.25; published 29.05.25.

Please cite as:
Hu C, Lan X, Luo Y, Zhu S, Cheng G
Efficacy of Motion-Sensing Game–Assisted Pulmonary Rehabilitation in Patients With Chronic Obstructive Pulmonary Disease: Systematic Review and Meta-Analysis of Randomized Controlled Trials
JMIR Serious Games 2025;13:e69562
URL: https://games.jmir.org/2025/1/e69562
doi:10.2196/69562

© Cuirong Hu, Xia Lan, Yanwen Luo, Shu Zhu, Guilan Cheng. Originally published in JMIR Serious Games (https://games.jmir.org), 29.5.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic

XSL•FO
RenderX

information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Scrutinizing the Gateway Relationship Between Gaming and Gambling Disorder: Scoping Review With a Focus on the Southeast Asian Region

Kristiana Siste[1], MD, PhD; Daniel L King[2,3], PhD; Enjeline Hanafi[1], MD; Lee Thung Sen[1], MRES, MD; Adrian Adrian[1], MD; Belinda Julivia Murtani[1], MRES, MD

[1]Department of Psychiatry, Faculty of Medicine, Universitas Indonesia – Dr. Cipto Mangunkusumo National Referral Hospital, Jakarta, Indonesia

[2]College of Education, Psychology and Social Work, Flinders University, Adelaide, Australia

[3]School of Psychology, The University of Adelaide, Adelaide, Australia

Corresponding Author:
Kristiana Siste, MD, PhD
Department of Psychiatry, Faculty of Medicine, Universitas Indonesia – Dr. Cipto Mangunkusumo National Referral Hospital, Jakarta, Indonesia

Abstract

Background: The gaming and gambling overlap has intensified with new evidence emerging. However, the relationship between gaming and gambling in the digital space is still inconclusive, especially in resource-limited Asian countries.

Objective: This study aims to review available evidence on the possible interaction and focuses specifically on the gateway interaction between gambling and gaming. Additionally, this review delves into the state of evidence from the Southeast Asian region, providing an in-depth analysis of this underexplored area.

Methods: We performed a scoping review by sifting through the publications in five databases. We focused on the gateway interaction and provided a possible pathway model, while two other convergence relationships were provided for comparison.

Results: The scoping review identified a total of 289 publications, with the majority being empirical (n=181), although only 12 studies used longitudinal designs. A significant proportion of the publications (n=152) concentrated on the correlation or comorbidity between gaming and gambling. Most of the evidence has originated from Global North countries, with very limited research emerging from Southeast Asia (n=8). The most commonly studied gambling-like element in video games was loot boxes (n=105). Other elements investigated included esports betting, skin betting, token wagering, gambling advertisements, and gambling-like features. Several longitudinal studies have highlighted the risk of the gateway effect associated with gamblification involvement. However, emerging evidence suggests more nuanced underlying mechanisms that drive the transition from gaming to gambling.

Conclusions: Overall, there is early evidence of linkage between gambling and gaming, through shared structural and biopsychosocial characteristics. This association possibly extends beyond disparate comorbidity, as such engagement in one activity might influence the risk of partaking in the other behavior. The field requires further longitudinal data to determine the directionality and significant precipitating factors of the gateway effect, particularly evidence from Asia.

(JMIR Serious Games 2025;13:e59740) doi:10.2196/59740

KEYWORDS

behavioral addiction; convergence; gateway effect; gambling advertisement; gamblification; monetized gaming

Introduction

Video games have undergone substantial evolution since their integration with internet-based platforms, with a notable shift being the monetization of virtual items through chance-based activities [1]. Among the most widely debated mechanics is the “loot box” system [2]. This system permits players to purchase boxes containing randomized items of varying rarity, functionality, and monetary value. Highly coveted items, which enhance gameplay or provide unique aesthetic modifications, are often accessible exclusively through loot boxes [2,3]. Although the concept of probabilistic rewards for in-game items has been a feature of video games since the early days of role-playing games, it is evident that such mechanics have become increasingly monetized by capitalizing on the psychological loop between probability and instant gratification [4].

In addition to loot boxes, numerous games have implemented other forms of microtransactions, including downloadable content, purchasable upgrades, and play passes, which can be

XSL•FO
RenderX

acquired with real-world money or via intermediary in-game currencies [2,5]. The gambling-like features (GLFs) of video games have expanded to include both players and spectators, notably through "skin betting" and esports wagering. Esports, much as traditional sports, now allow spectators to place bets on teams and players [6], with websites emerging that facilitate match tracking and performance analysis (eg, Oddspedia). A survey conducted in the United Kingdom revealed that 8.5% of adults had engaged in betting on esports [7], while an Australian study found that 41.5% of respondents participated in esports cash betting, skin betting, or skin gambling [8]. Other platforms have broadened the scope of betting to include not only professional esports leagues but also everyday matches involving skilled players, with transactions occurring in both real-world currency and virtual items or skins, the latter referred to as "skin betting."

The *ICD-11* (*International Classification of Diseases, 11th Revision*) recently classified gaming and gambling disorders as conditions related to addictive behaviors, underscoring shared diagnostic criteria and symptomatology [9]. Global data indicate a past-year prevalence of gambling disorder ranging from 0.12% to 5.8% [10], while the prevalence of gaming disorder stands at approximately 3.3% [11]. Prior research has estimated the comorbidity between gaming and gambling disorders at approximately 15%. Moreover, emerging evidence has suggested a potential "gateway" effect, where individuals transition from gaming to gambling disorder [12,13]. However, contemporary discourse in this nascent field has shifted toward examining the underlying motivations that drive the initiation and persistence of the transition between the two disorders. It is also plausible that certain gamer populations are more frequently sensitized to gambling stimuli, partly due to the proliferation of targeted gambling advertisements—similar to those promoting other addictive substances [14] and particularly within Southeast Asia [15]. This sensitization places vulnerable adolescents and young adults at elevated risk for problematic gambling behaviors [16]. The increasingly aggressive nature of gambling advertisements, coupled with inadequate regulatory oversight, exacerbates this issue. The paucity of clinical evidence, particularly concerning the gateway effect, presents therapeutic challenges and underscores the urgent need for further research [17].

This study seeks to explore the forms of intersection between gaming and gambling disorders. To achieve this, we conducted a scoping review to systematically evaluate the existing evidence on the overlap between gaming and gambling disorders. The scoping review method was selected, as it offers an exploratory overview of the available literature [18], allowing for the identification of gaps in research. Our scoping review attempts to focus on the gateway effect, while also examining other associations, that is, embedded gambling mechanics within digital games and the clinical comorbidity between the two disorders.

Methods

Scoping Review Objectives and Outline

This scoping review was conceptualized as an exploratory approach to research synthesis, aimed at identifying key evidence and theoretical concepts within the field. While several significant areas were included in the search, the examination of gambling mechanics and clinical comorbidity served primarily as background, providing context for the current state of research on the gateway effect. The primary focus of the discussion is the gateway effect, specifically the transition from gaming to gambling disorder. In contrast to a systematic review, this study was not preregistered with a review registry [19,20]. Nevertheless, the review was conducted following the guidelines outlined in the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) checklist [18,21].

Search Strategy and Reference Screening

The search strategy focused on systematically examining existing literature on the shared mechanics between gaming and gambling, the concept of gaming as a "gateway" to gambling disorder, and the co-occurrence or comorbidity of gaming and gambling disorders. Detailed information regarding the search methodology is provided in Multimedia Appendix 1. To ensure thoroughness, the search incorporated terms from relevant medical subject headings. Databases including PubMed, ScienceDirect, ProQuest, Google Scholar, and Garuda (the Indonesian national scientific database) were searched using this strategy from their inception until July 31, 2024, with the search results summarized in the *Results* section. Following prior recommendations, the first 300 results from Google Scholar were included [22]. Additionally, a backward citation search was conducted to ensure the comprehensiveness of the studies identified. The results from the database search and the backward citation search were exported into a reference management software (EndNote, version 20, Clarivate).

Inclusion and Exclusion Criteria

This review encompassed all study types, including commentaries, editorials, letters, and reports [23], to capture the breadth of available evidence and highlight areas requiring further research. The inclusion criteria for nonempirical studies required a discussion of gaming and gambling characteristics. Empirical studies must include statistical data, such as overlapping prevalences, correlations, associations, regression analyses, or other methods that establish a relationship between gaming and gambling. Textbox 1 provides a comprehensive outline of the inclusion and exclusion criteria. The terms "video game(-ing)" and "game(-ing)" were defined based on previous research as "playable interactive digital entertainment that requires audiovisual apparatus, typically demands strategic play, and may involve narratives" [24,25] to avoid confusion stemming from the interchangeable use of "gaming" and "gambling" in some legal contexts [26]. Furthermore, this scoping defines gateway effect as the transitory relationship from video gaming to gambling and reverse gateway effect being the vice versa (gambling to video gaming) [27]. Other definitions of gateway effect are beyond the scope of this

XSL•FO
RenderX

scoping review. This paper explored multiple facets of video gaming that might propagate such effects including, but not limited to, microtransactions (ie, in-game payments made available from indirect virtual money or directly through real-life currencies [17]), loot boxes (ie, purchasable content that will give chance-based items [28]), and gambling advertising in video games. Titles and abstracts were screened for eligibility based on inclusion and exclusion criteria by two authors (LTS and AA), with disagreements resolved by a third author (EH). Full-text analysis and data extraction were conducted by two authors (LTS and AA) and subsequently verified by a third author (KS) to ensure methodological rigor. The overall extracted data and interpretation were reviewed by KS and DLK.

Textbox 1. Criteria for inclusion and exclusion of publications for review.

Inclusion criteria

- This review includes all types of published studies, such as commentaries, editorials, viewpoints, proceedings, reports, empirical studies, and theses.
- Nonempirical studies should discuss the relationship between gaming and gambling characteristics, including converging mechanics (structural features), combined epidemiological or natural histories, risk correlations, or predictive risks.
- Empirical studies should provide statistical data on the overlapping prevalence, clinical profiles, correlations, associations, regression analyses, or other analytical methods that measure the relationship between gaming and gambling.
- Accessible full-text records.

Exclusion criteria

- Studies that focus solely on gaming, without any direct reference to gambling elements or gambling disorder.
- Studies that focus solely on gambling, such as electronic gambling machines, video lottery terminals, cyber gambling, online gambling, and other forms of simulated gambling, without any direct reference to video gaming or gaming disorder.
- Studies that measure gaming disorder using nonspecific screening tools, such as those designed to assess general internet addiction, rather than tools specifically developed for gaming disorder.
- Studies published in languages other than English or Bahasa Indonesia.
- Studies that focus exclusively on comparing the adoption of diagnostic criteria, either *ICD-11* (*International Classification of Diseases, Eleventh Revision*) or *DSM-5* (*Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition*).

Data Extraction

The data extraction method was designed to synthesize evidence in alignment with this study's key research questions. Following the updated guidelines from JBI (Joanna Briggs Institute) for scoping reviews [18], as well as other relevant recommendations [19,21], and drawing from a prior scoping study [29], a comprehensive set of extraction variables was developed and organized into an Excel (Microsoft Corp) workbook. These variables aim to capture the types of studies, year of publication (prioritizing the date of first digital publication), study location (based on the sample rather than author affiliation), study design and temporality, objectives, study description, the relationship between gaming and gambling, and key findings.

Studies were classified into three categories: reviews (including systematic reviews, scoping reviews, literature reviews, narrative reviews, book chapters, and review reports), empirical studies (encompassing any primary qualitative or quantitative designs, as well as meta-analyses), and viewpoint or commentary pieces (including opinion papers, editorials, comments, and viewpoints). The overlap between gaming and gambling was further categorized into three distinct groups based on this study's focus: (1) converging mechanics (where the papers examined the shared mechanics between video games and gambling), (2) correlation or comorbidity (where the papers analyzed the overlap or association between problematic gaming and gambling behaviors), and (3) gateway effect (where the papers assessed the predictive relationship between gaming disorder, its gambling-like mechanics, and the development or escalation of gambling behavior). Findings were reported as descriptive data, in line with the recommended guideline [18].

Ethical Considerations

This scoping review does not require ethical clearance as it reviews published papers.

Results

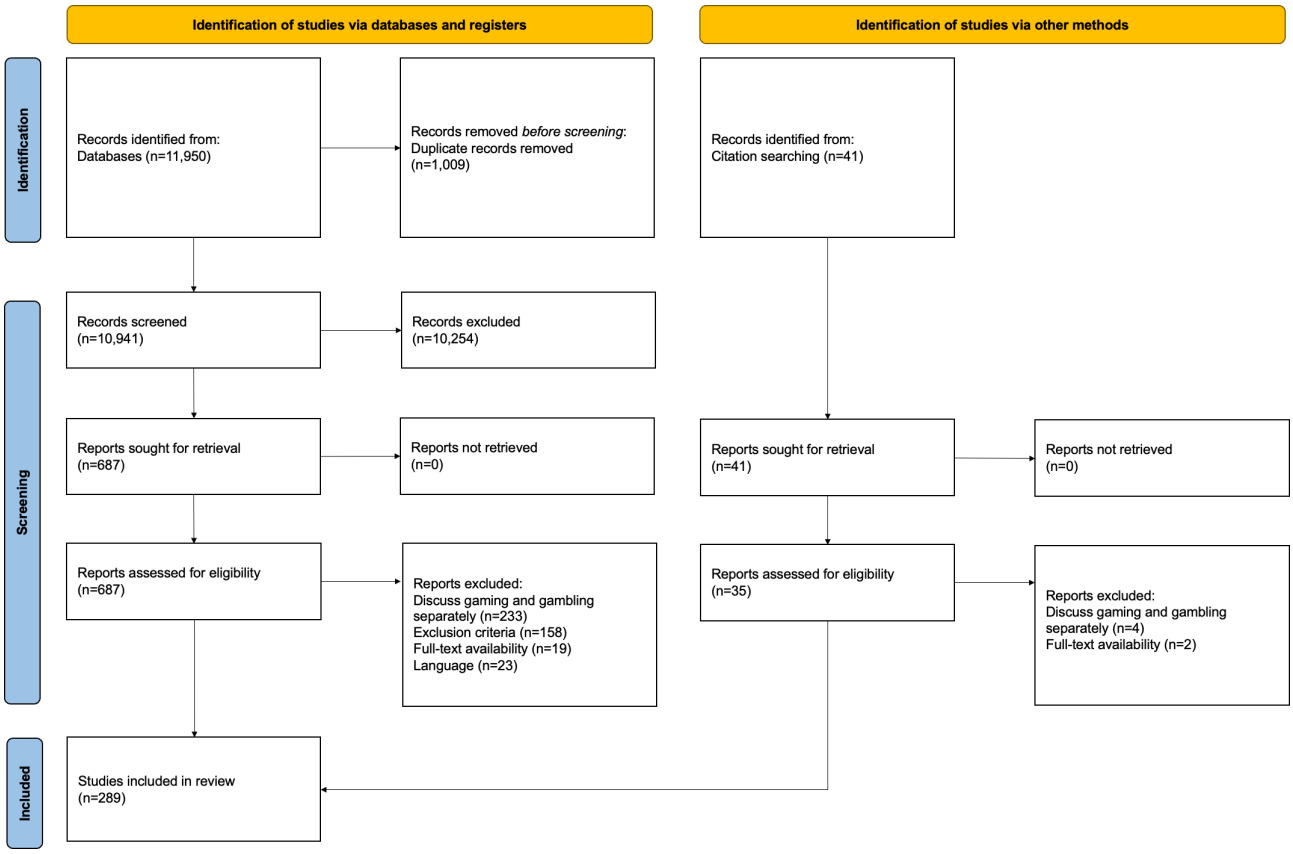
Publication Characteristics

The search strategy and inclusion or exclusion criteria yielded 289 relevant publications (see Multimedia Appendix 2). Figure 1 presents a detailed PRISMA-ScR flow diagram of the database search process. The earliest publication identified dates back to 1989, with an upward trend in research addressing gaming and gambling convergence observed since then (Figure 2), culminating in a peak of 49 publications in 2022. Publications spanned at least 36 countries (excluding those with multinational affiliations), with the highest numbers originating from the United Kingdom (n=53), Australia (n=52), and Canada (n=30). In Southeast Asia, this scoping review identified 5 studies from Indonesia, 2 from Singapore, and 1 from Malaysia. Of the 5 studies from Indonesia, only 2 are empirical [30,31], while the remaining studies consist of commentaries and reviews that examine the relationship between gaming and gambling in the context of Indonesian or Islamic law [32-34].

XSL•FO
RenderX

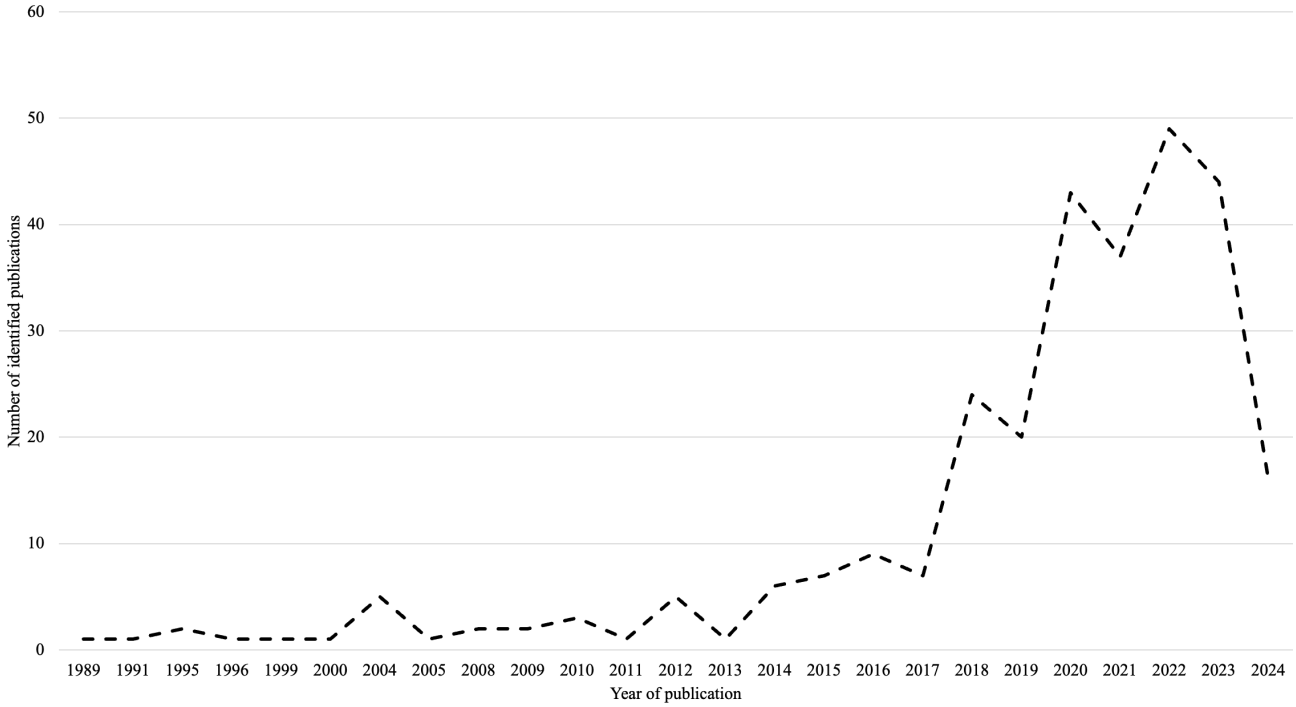
Among the studies analyzed, 181 are empirical, 69 are review publications, and 39 are viewpoint or commentary pieces. Sample sizes in the reviewed studies varied significantly, ranging from a single-patient case study to a large-scale study involving 16,196 participants. The majority of empirical studies (n=101) used adult samples (defined as aged older than 18 years), while 19.3% (n=35) focused on children and adolescents, and 14.4% (n=26) included both adults and adolescents. Approximately 10.5% (n=19) of the empirical studies either did not involve human participants or did not report age-specific data.

Figure 1. PRISMA-ScR flow diagram. The search across five databases initially yielded 11,950 publications. After removing 1009 duplicates, 10,254 records were excluded for not meeting the inclusion or exclusion criteria during the screening process. Subsequently, 687 publications were retrieved for full-text review. Additionally, 41 records were identified through reference back-searching. Finally, a total of 289 publications were included in this review. PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews.



XSL•FO
RenderX

Figure 2. Number of identified publications across the years. This scoping review identified a total of 289 published records spanning from 1989 to 2024. An increasing trend in the number of publications has been observed since 2013, with the peak occurring in 2022 when 49 publications were recorded.



Gaming-Gambling Interactions Overview

Looking at the area of convergence, 152 (55.5%) of publications examined the correlation or comorbidity between gaming and gambling, 35% (n=96) assessed the converging mechanics of video game gamblification, and 15% (n=41) explored the gateway effect of gaming to gambling. Gaming and gambling comorbidity ranged from 1.2% to 26.3% [35-38] and varying correlations, from weak [39,40], moderate [41,42] to strong [43-45]. Out of all identified publications, the form of microtransactions most studied was loot box (n=105), followed by esports betting (n=49), and then skins betting (n=29). Twenty-three studies focused on GLFs, 11 studies specifically examined the effects of gambling advertisements in and around video games, and 3 studies described token wagering. Contrastingly, several new research indicated nonsignificant association between loot box involvement and problem gambling [46-48]. Cryptogames have only recently emerged [49,50] representing a new frontier in the overlap between gaming and gambling. Some view cryptogames to be gambling-like activities, as they blend elements of skill and chance and are inherently linked to cryptocurrency and monetary liquefaction [49].

Longitudinal Studies

Out of all the empirical studies, 12 publications used a longitudinal design. Of the 12 longitudinal publications, five publications [51-55] focused on adolescent samples and scrutinized the gateway effect. These longitudinal studies on adolescents examined baseline involvement with microtransactions (eg, loot boxes and virtual currencies) and exposure to gambling advertisements. All studies found that these baseline factors significantly predicted the development of problematic gambling over time. Alternatively, 7 publications concentrated on the adult population [27,48,56-60]. However, only 4 investigated the gateway effect [27,48,59,60] focusing on baseline involvement to loot boxes, gambling advertisements, skin betting, and problematic gaming. Out of those 4 studies, 2 reported significant gaming to gambling transition [27,60], while the other 2 studies did not find significant gateway effect [48,59]. Furthermore, 2 longitudinal studies investigated both gateway and reverse gateway effect. Both studies reported that only the gateway effect demonstrated statistical significance, while the reverse gateway effect did not [27,53].

Risk Factors to Gateway Effect

Biological

Demographically, the younger populations are at greater risk for both gaming and gambling disorders. Some studies have observed the gateway effect from gaming to gambling disorder to be more frequent among youth [61-63]. In contrast, another group has demonstrated that esports bettors had a later onset age than traditional sports bettors and carry higher odds for problematic gambling [64]. Multiple studies indicate that male sex may be a risk factor for developing the gateway effect [28,62,65], though one study exhibited its potential as a protective factor for the gateway effect [53]. Paradoxically, emerging research suggest that female sex may be associated with an increased risk, particularly in the post-COVID-19 period [55,66,67].

Psychosocial Factors

Psychologically, novelty-seeking temperament, compulsivity, and impulsivity [68-70] are correlated to interacting with gambling features in video games, although findings have been inconsistent. The gateway effect was also observed to be predisposed by cognitive errors [28,47,54], for example, predictive bias, sunk-cost bias, or expectancy. A study by Spicer

XSL•FO
RenderX

et al [47] found that the transition from gaming to gambling is more accurately explained by the connection between gaming behaviors and gambling-related cognitions, rather than solely through involvement with loot boxes. Several studies proposed simulated gambling in social networking applications and having peers making similar microtransactions [54,71]. Additionally, a greater risk of developing pathological gambling was seen for gamers with average income but higher spending [72]. In spite of that, other studies were unable to replicate these findings [73,74].

Drivers of Gateway Effect

Structural Features

Cited intermediary pathways for the transition from gaming to gambling include priming of gambling stimuli that leads to normalization and also desensitization toward monetary losses [28,61]. Several studies [27,52,61] have described affinity, exposure to gambling features in video games, and gambling advertisement as factors increasing the risk, particularly among youth. Frequency of in-game transactions (eg, purchased or sold loot boxes), amount spent monthly, number of opening loot boxes or other microtransactions, and history of participation in traditional gambling were predictive of pathological gambling [52,62,75,76]. An overarching phenomenon, termed GLFs in modern video games, has been described [77-81]. GLFs are manifested as virtual currencies, chance-based purchases, and event features [78,79,81]. Other forms include meta-game or real-life rewards, daily log-in payouts, and near-miss features [79]. Some recent studies also noted the separate impact of gambling advertisements [52,54,82-85], which enhanced gambling interest, engagement in chance-based activities in-game, associated with earlier gambling onset, and higher gambling participation using real-world currency. These gambling advertisements appear on live streaming platforms [86] and social media apps, oftentimes trying to appeal to younger populations [85].

Motives for Shifting

Multiple studies have looked into motivations for either using skin or monetary instruments for esports betting [87], involvement in both gaming and gambling [76], using GLFs [31,88,89], and specifically to transition from gaming to gambling [28,90]. Motives for shifting from gaming to gambling include perceived similarity of structure between GLFs and gambling games, thus sense of normalization to shift or even a routine [28]. Players also seek excitement and emotions that are thought to be similar or more intense through gambling activities. Furthermore, they are aware of the addiction potential of GLFs and gambling games. They see gambling as another outlet to channel the compulsion. Subsequently, the positive experience from GLFs form cognitive biases such as illusion of control. Lastly, the video game players view virtual funds or currencies and subsequent real-life rewards as carrying social and financial benefits to help fund their gameplay [90]. Adolescents may view accessibility to underage gambling as a form of social prestige [91].

Harm and Preventive Measures for Gaming-Gambling Convergence

Harms and Consequences

Involvement in microtransactions and gamers with gambling engagement was shown with higher gambling harm categories [92-95]. Gambling with skins rather than traditional monetary form demonstrated a higher degree of experienced harm on the Short Gambling Harm Screen [96]. Experienced harms include financial loss, physical symptoms, psychological harm, social deficits, work or academic distress, or legacy harm [30,96]. For example, Drummond et al [97] reported a high degree of psychological distress among video gamers who purchased loot boxes. Intriguingly, a study in the United States found that students with esports as a career choice experienced stigmatization from school counselors. This might stem from the societal biases on the nontraditional career pathway and the association between esports to addiction, gambling, and overall negative judgment [98].

Preventive Measures

Some research has advocated for GLFs in video games to be described and incorporated as criteria for licensure or censorship [46,99]. In line, there are suggestions that this information is disclosed to parents and youth, allowing the making of informed decisions [46]. Some hold the view that parents should be equipped with knowledge about the risks associated with microtransactions, chance-based activities, and gambling imagery in video games [100]. Others have reviewed the potential for implementing several layers of protection, including age checking, payment security measures, and setting a maximum spending limit [101,102]. Several viewpoint papers have also expanded to call for accountability from the gaming and gambling industries [100,103], urging them to engage with regulators on consumption risks. Several jurisdictions have responded by banning specific microtransactions in video games, for instance, loot boxes [104]. A commentary paper had voiced concerns regarding the creation of black markets due to the outright banning of microtransactions [105]. Technically, major video game corporations could also implement access blocks for skins or items on unregulated third-party websites [106].

Discussion

This scoping review aimed to examine the existing body of literature, map the landscape of available publications, and analyze selected evidence. A specific relationship, known as the “gateway effect,” describes the transition from gaming to gambling disorders [107]. Our findings indicate an increase in studies seeking to clarify this relationship, highlighting the clinical challenges posed by the convergence of gaming and gambling platforms. This overlap potentially heightens awareness of the gateway relationship, leading to increased reports and complex analyses. Overall, the pooled data suggest a small to moderate correlation between gaming and gambling behaviors [108]. However, much of the evidence supporting the gateway effect is derived from cross-sectional studies, limiting causal inference. Additionally, studies varied widely in defining initial pathological gaming behaviors and subsequent

XSL•FO
RenderX

problematic gambling activity [8,27,28,53,63,68,69,75,76,92,94,109-111]. For instance, Vadlin et al [53] assessed predictive gaming behaviors using the Gaming Addiction Identification Test, whereas Molde et al [27] used the Gaming Addiction Scale. Exposure to GLFs in nongambling environments may also acclimate individuals to gambling [61,112]. Some recent studies have failed to replicate significant associations between in-game gambling activities (ie, loot boxes) and problematic gambling [46-48]. More nuanced analyses suggest that the association between loot box engagement and problematic gambling may be indirect, driven by gambling-related cognitive biases [47]. In addition, Xiao et al [46] propose that cultural factors, such as limited gambling availability and enforced transparency within the gambling industry, may contribute to the nonsignificant association observed in their Chinese-speaking sample.

Although estimates exist for gaming-gambling comorbidity, the rates of transition via the gateway effect remain unclear. Moreover, despite global reports of higher rates of gaming disorder [113], there is limited data from Asian countries on gaming-gambling convergence. In Indonesia, all forms of gambling, including loot boxes, are prohibited [32], which may contribute to a societal stigma surrounding such behaviors and thus related research. Additionally, Southeast Asian countries tend to prioritize enforcement measures, such as crackdowns on gambling [114], rather than investing in public health research on gambling [115]. In Indonesia, gambling and other behavioral addictions are not covered by the national health insurance system, further contributing to the scarcity of data and limited research in this area. Nonetheless, there has been a recent rise in empirical studies from Indonesia examining gambling elements within online video games [30,31]. The Lancet Public Health Commission's [116] initiative on global cooperation to address boundary-crossing gambling products should encourage research in Global South countries, such as Indonesia. Enhanced empirical evidence is essential for informing policies in these regions, as unique national gambling cultures [46] and differing jurisdictions [117] have been shown to influence gaming-gambling convergence in various ways.

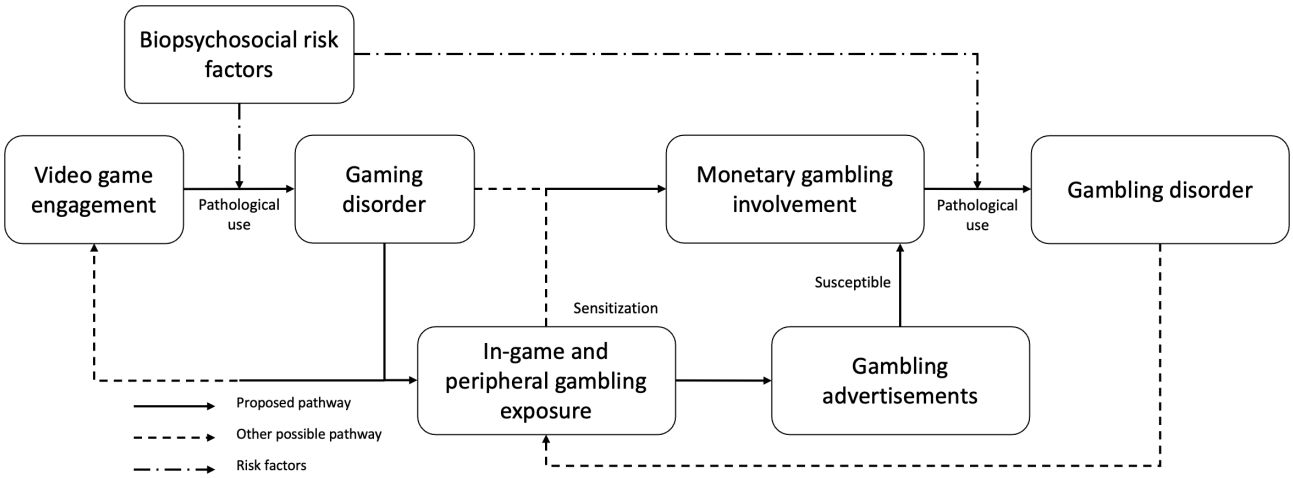
Some experts suggest that legal age attainment may amplify gambling participation [118], though this factor is inapplicable in countries where gambling is prohibited, such as Indonesia. Additionally, the gateway transition may be encouraged by increased economic access (eg, loans or buy-now-pay-later options [119]) upon reaching adulthood. Furthermore, the transitional mechanism may be reinforced by exposure to gambling advertisements [54]. However, such exposure and initial GLF involvement do not necessarily lead to problematic gambling behaviors. Multiple studies have indicated that additional biopsychosocial vulnerabilities (eg, gambling-related cognitive fallacies) are required for these behaviors to develop [47,48]. Preliminary findings indicate that certain biopsychosocial factors in the gateway effect mirror elements of the Interaction of Person-Affect-Cognition-Execution model [120], which describes risk factors for gaming and gambling disorders independently. These factors include younger age, specific temperamental traits, and impaired cognitive processes. Notably, the evidence linking sex to the gateway effect has become increasingly ambiguous. Female participation in video games and GLFs has been steadily rising [67], which corresponds with a heightened risk of problematic gambling among this demographic. Some researchers propose that this trend may resemble the telescoping phenomenon observed in gambling disorders, where the onset and progression of gambling-related issues occur more rapidly [55]. Concurrently, online gambling advertisements have increasingly targeted female audiences [121]. There is evidence of shared neurobiological disruptions to the prefrontal cortices and striatum, particularly within the default mode and reward circuits, in both gaming and gambling disorders [122]. However, definitive data linking these neural alterations to the gateway effect are still lacking.

Several identified studies also observed that the reverse effect (gambling to gaming) was not statistically significant [27,109,123]. This phenomenon raised the question of whether gambling has higher addictive potency than gaming. Some gamblers engage in video game chance-based activities as they are perceived to be "safer" [28]. Prior gambling participation did correlate to higher microtransaction involvement [124-126]. Moreover, several studies [52,54,82] have discovered the associative link between gambling marketing around video games to gambling behavior, though modest, and gambling industries have used these targeted advertisements to expand microtransactions (eg, loot box) usage [127].

These patterns might be sensitized with prior engagement to chance-based activities in-game and exposure to gambling advertisements. The two elements might represent a dual-hit pathway for the gateway transition mechanism, similar to the two-hit hypothesis of schizophrenia [128,129] that proposes genetic factors (first hit) sensitize an individual to the environmental insults (second hit) leading to schizophrenia. Adopting the concept, the in-game or peripheral participation in gambling-like activities primes the individual as the first hit, and the gambling advertisements act as the second insult leading to real-world gambling participation (see Figure 3; note that the figure is not meant to be exhaustive). An alternative pathway is presented that might circumvent the dual hit pathway, meaning only one of the two hits is experienced. For instance, the individual progresses from microtransactions to monetary gambling without exposure to gambling advertisements. The proposed dual hit pathway remains a hypothetical and theoretical model requiring empirical evidence to demonstrate if it poses a greater risk than the other pathway.

XSL•FO
RenderX

Figure 3. This is a simplified illustration of the dual-hit model proposed for the transition from gaming to gambling. An alternative pathway bypasses the dual-hit approach, where only one of the two hits is experienced. Peripheral exposure includes gambling-related features found on platforms linked to video games, such as, but not limited to, loot box involvement, watching betting activities on video game streaming services, or spectating esports betting. Gambling advertisements may be encountered through pop-up advertisements, video game chat applications, online forums, competitive events, or other evolving forms of interaction.



Consequently, the response to mitigate the migration to gambling behaviors and the associated harms has been insufficient. The rapid evolution of gaming and gambling mechanisms on digital platforms has outpaced regulatory efforts. Historical policy experiences indicate that outright bans may not be the most effective solution [105,127]. Emerging blended forms, such as cryptogames, may also introduce new legal challenges. Instead, industry and structural strategies could encompass implementing more stringent age restrictions [130], restricting API access by third-party platforms [131], instituting additional licensing requirements [104], and promoting self-regulation and ethical conduct within the industry [132]. In 2016, Indonesia introduced its game rating system, which permits gambling-like elements in games rated for players aged 13 years and older, provided they do not include cashing-out features. However, the age threshold should be raised to 18 years, as gambling-like elements present risks and are illegal in the physical realm. Despite the recent increase in research on the gateway effect from gaming to gambling, there remains a significant lack of empirical, clinical, and psychological preventive measures. Early detection should be implemented for all patients with gaming disorder by screening for involvement in microtransactions and other gambling-like activities. This approach will facilitate the application of harm reduction strategies, such as limiting access to real-world money and restricting transactions to virtual currencies that cannot be cashed out. Additionally, engaging families or significant others in the use of supervision applications can help restrict exposure to gambling advertisements. More targeted interventions should focus on clinical management to address cognitive biases and gambling fallacies, fostering awareness that chance-based activities constitute gambling and may have negative consequences.

Overall, the existing studies illustrate that children, adolescents, and emerging adults are particularly vulnerable to the convergence of gaming and gambling. Additional predisposing factors contributing to the transition toward risky behaviors in these groups, such as high impulsivity and sensation-seeking tendencies [76,133], warrant further research. Moreover, additional research is needed to substantiate the gateway effect, aligning with the perspectives presented in previous reviews and commentaries [17,25,107]. Most studies on the gateway effect have focused on traditional Western gambling activities, leaving region-specific gambling activities, such as pachinko in Japan or karambol in Indonesia, relatively underexplored. Further research is essential to clarify the epidemiology, specific mechanisms of transition, directionality of the transition, definitive risk factors, and effective preventive measures related to the gaming-gambling gateway effect.

This study has several limitations. First, the scoping review approach lacks the comprehensiveness of a systematic review, and the identified publications are not exhaustive (for instance, preprint databases were not included, and the search strategy was not preregistered). Consequently, the descriptive counts and proportions presented here reflect only the exploratory sample of identified publications, not the entire body of literature [18]. Second, this study did not address gamified gambling or the growing trend of skill-based gambling forms. Publications focusing solely on social casino gambling were also excluded, as these are traditional gambling activities delivered via the internet rather than video games [107]. Third, this scoping review incorporates some viewpoints and commentary pieces to map the publication landscape. Where feasible, references to these sources have been minimized or clearly identified. Readers are encouraged to be mindful of the varying levels of evidence these pieces provide and to refer to the supplementary materials or reference list for verification if needed. Fourth, the majority of empirical studies rely on self-reported data, which may impact accuracy and necessitates caution in interpretation. Finally, studies that compare gaming and gambling through distinct or combined samples, or with differing diagnostic criteria, were excluded, as such heterogeneity does not support direct examination of overlap or convergence. Nonetheless, these studies still offer valuable insights into the broader relationship between gaming and gambling disorders as behavioral addictions.

XSL•FO
RenderX

Acknowledgments

The publication of this paper is funded by the science promotion cooperation with the Corporate Social Responsibility Division of PT ("*Perseroan Terbatas*") Mineral Industri Indonesia (Persero; MIND ID). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of this paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOCX File, 15 KB - games_v13i1e59740_app1.docx]

Multimedia Appendix 2
Results of data extraction.
[DOCX File, 201 KB - games_v13i1e59740_app2.docx]

Checklist 1
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) checklist.
[PDF File, 62 KB - games_v13i1e59740_app3.pdf]

References

1. King DL, Delfabbro P, Griffiths MD. The convergence of gambling and digital media: implications for gambling in young people. J Gambl Stud 2010 Jun;26(2):175-187. [doi: 10.1007/s10899-009-9153-9]
2. Drummond A, Sauer JD. Video game loot boxes are psychologically akin to gambling. Nat Hum Behav 2018 Aug;2(8):530-532. [doi: 10.1038/s41562-018-0360-1] [Medline: 31209310]
3. Drummond A, Sauer JD, Hall LC, Zendle D, Loudon MR. Why loot boxes could be regulated as gambling. Nat Hum Behav 2020 Oct;4(10):986-988. [doi: 10.1038/s41562-020-0900-3] [Medline: 32601458]
4. King DL, Gainsbury SM, Delfabbro PH, Hing N, Abarbanel B. Distinguishing between gaming and gambling activities in addiction research. J Behav Addict 2015 Dec;4(4):215-220. [doi: 10.1556/2006.4.2015.045] [Medline: 26690615]
5. Davidovici-Nora M. Innovation in business models in the video game industry: free-to-play or the gaming experience as a service. Comput Game J 2013 Dec;2(3):22-51. [doi: 10.1007/BF03392349]
6. Bányai F, Griffiths MD, Király O, Demetrovics Z. The psychology of esports: a systematic literature review. J Gambl Stud 2019 Jun;35(2):351-365. [doi: 10.1007/s10899-018-9763-1] [Medline: 29508260]
7. Wardle H, Petrovskaya E, Zendle D. Defining the esports bettor: evidence from an online panel survey of emerging adults. Int Gambl Stud 2020 Oct 6;20(3):487-499. [doi: 10.1080/14459795.2020.1826559] [Medline: 37431491]
8. Greer N, Rockloff M, Russell AMT. Gambling and video games: are esports betting and skin gambling associated with greater gambling involvement and harm? : Victorian Responsible Gambling Foundation; 2020 URL: https://responsiblegambling.vic.gov.au/resources/publications/gambling-and-video-games-are-esports-betting-and-skin-gambling-associated-with-greater-gambling-involvement-and-harm-967/ [accessed 2025-01-08]
9. Stein DJ, Szatmari P, Gaebel W, et al. Mental, behavioral and neurodevelopmental disorders in the ICD-11: an international perspective on key changes and controversies. BMC Med 2020 Jan 27;18(1):21. [doi: 10.1186/s12916-020-1495-2] [Medline: 31983345]
10. Calado F, Griffiths MD. Problem gambling worldwide: an update and systematic review of empirical research (2000-2015). J Behav Addict 2016 Dec;5(4):592-613. [doi: 10.1556/2006.5.2016.073] [Medline: 27784180]
11. Kim HS, Son G, Roh EB, et al. Prevalence of gaming disorder: a meta-analysis. Addict Behav 2022 Mar;126:107183. [doi: 10.1016/j.addbeh.2021.107183] [Medline: 34864436]
12. Li W, Mills D, Nower L. The relationship of loot box purchases to problem video gaming and problem gambling. Addict Behav 2019 Oct;97:27-34. [doi: 10.1016/j.addbeh.2019.05.016] [Medline: 31129456]
13. Macey J, Hamari J. Investigating relationships between video gaming, spectating esports, and gambling. Comput Human Behav 2018 Mar;80:344-353. [doi: 10.1016/j.chb.2017.11.027]
14. Siste K, Sen LT. Ascending pattern of alcohol use and underage drinking in Asia: a commentary on the article by Assanangkornchai and Vichitkunakorn (2020). Alcohol Clin Exp Res 2021 Mar;45(3):496-499. [doi: 10.1111/acer.14556] [Medline: 33486766]

XSL•FO
RenderX

15. Lala G, Chandra S, Ogun N, Moody L, Third A. Online Safety Perceptions, Needs, and Expectations of Young People in Southeast Asia: Consultations with Young People in Indonesia: Western Sydney University; 2022. [doi: 10.26183/TZ74-EV38]
16. Lussier ID, Derevensky J, Gupta R, Vitaro F. Risk, compensatory, protective, and vulnerability factors related to youth gambling problems. Psychol Addict Behav 2014 Jun;28(2):404-413. [doi: 10.1037/a0034259] [Medline: 24274433]
17. Gibson E, Griffiths MD, Calado F, Harris A. The relationship between videogame micro-transactions and problem gaming and gambling: a systematic review. Comput Hum Behav 2022 Jun;131:107219. [doi: 10.1016/j.chb.2022.107219]
18. Peters MDJ, Marnie C, Tricco AC, et al. Updated methodological guidance for the conduct of scoping reviews. JBI Evd Synth 2020;18(10):2119-2126. [doi: 10.11124/JBIES-20-00167]
19. Colquhoun HL, Levac D, O'Brien KK, et al. Scoping reviews: time for clarity in definition, methods, and reporting. J Clin Epidemiol 2014 Dec;67(12):1291-1294. [doi: 10.1016/j.jclinepi.2014.03.013] [Medline: 25034198]
20. Winters N, Langer L, Geniets A. Scoping review assessing the evidence used to support the adoption of mobile health (mHealth) technologies for the education and training of community health workers (CHWs) in low-income and middle-income countries. BMJ Open 2018 Jul;8(7):e019827. [doi: 10.1136/bmjopen-2017-019827]
21. Tricco AC, Lillie E, Zarin W, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 2;169(7):467-473. [doi: 10.7326/M18-0850] [Medline: 30178033]
22. Haddaway NR, Collins AM, Coughlin D, Kirk S. The role of Google Scholar in evidence reviews and its applicability to grey literature searching. PLoS One 2015;10(9):e0138237. [doi: 10.1371/journal.pone.0138237] [Medline: 26379270]
23. Blanco D, Altman D, Moher D, Boutron I, Kirkham JJ, Cobo E. Scoping review on interventions to improve adherence to reporting guidelines in health research. BMJ Open 2019 May 9;9(5):e026589. [doi: 10.1136/bmjopen-2018-026589] [Medline: 31076472]
24. Esposito N. A short and simple definition of what a videogame is. Presented at: Digital Games Research Conference 2005, Changing Views: Worlds in Play; Jun 16-20, 2005; Vancouver, British Columbia, Canada.
25. King DL, Delfabbro PH. The convergence of gambling and monetised gaming activities. Curr Opin Behav Sci 2020 Feb;31:32-36. [doi: 10.1016/j.cobeha.2019.10.001]
26. Humphreys A, Latour KA. Framing the game: assessing the impact of cultural representations on consumer perceptions of legitimacy. J Consum Res 2013 Dec 1;40(4):773-795. [doi: 10.1086/672358]
27. Molde H, Holmøy B, Merkesdal AG, et al. Are video games a gateway to gambling? A longitudinal study based on a representative Norwegian sample. J Gambl Stud 2019 Jun;35(2):545-557. [doi: 10.1007/s10899-018-9781-z] [Medline: 29869768]
28. Spicer SG, Fullwood C, Close J, Nicklin LL, Lloyd J, Lloyd H. Loot boxes and problem gambling: investigating the "gateway hypothesis". Addict Behav 2022 Aug;131:107327. [doi: 10.1016/j.addbeh.2022.107327] [Medline: 35397261]
29. Montiel I, Basterra-González A, Machimbarrena JM, Ortega-Barón J, González-Cabrera J. Loot box engagement: a scoping review of primary studies on prevalence and association with problematic gaming and gambling. PLoS One 2022;17(1):e0263177. [doi: 10.1371/journal.pone.0263177] [Medline: 35085370]
30. Lubis AA, Saleh S, Marsa YJ. The phenomenon of online gambling under the guise of online games among college student. J Hum Soc Stud 2022 Oct 29;6(3):363-367. [doi: 10.33751/jhss.v6i3.6794]
31. Kesuma AE, Princes E. Antecedents of gacha gaming intention: extending UTAUT2 with structural video game characteristics. Comput Hum Behav Rep 2024 May;14:100405. [doi: 10.1016/j.chbr.2024.100405]
32. Putra MTR, Gunadi A. Legalitas sistem monetisasi lootbox dalam transaksi game online berdasarkan undang-undang nomor 11 tahun 2008 jo undang-undang nomor 19 tahun 2016. J Huk Adigama 2020;3(1):1480-1503. [doi: 10.24912/adigama.v3i1.9404]
33. Budiman R, Romadini NA, Herwandi Aziz MA, Pratama AG. The impact of online gambling among Indonesian teens and technology. ITSDI 2022;3(2):162-167 [FREE Full text] [doi: 10.34306/itsdi.v3i2.559]
34. Malik BW, Harisah H. An islamic perspective on e-sport competition. Islamuna J Stud Islam 2024;11(1):165-183. [doi: 10.19105/islamuna.v10i2.13620]
35. Sanders J, Williams R. Factors distinguishing problem gamblers, problem video-gamers, and dual problem gamblers/video-gamers [Abstract]. J Behav Addict 2017;6(Supplement-1):1-74. [doi: 10.1556/jba.6.2017.suppl.1]
36. Peter SC, Li Q, Pfund RA, Whelan JP, Meyers AW. Public stigma across addictive behaviors: casino gambling, esports gambling, and internet gaming. J Gambl Stud 2019 Mar;35(1):247-259. [doi: 10.1007/s10899-018-9775-x]
37. Szerman N, Basurte-Villamor I, Vega P, et al. Is there such a thing as gambling dual disorder? Preliminary evidence and clinical profiles. Eur Neuropsychopharmacol 2023 Jan;66:78-91. [doi: 10.1016/j.euroneuro.2022.11.010] [Medline: 36516511]
38. Marmet S, Studer J, Wicki M, Bertholet N, Khazaal Y, Gmel G. Unique versus shared associations between self-reported behavioral addictions and substance use disorders and mental health problems: a commonality analysis in a large sample of young Swiss men. J Behav Addict 2019 Dec 1;8(4):664-677. [doi: 10.1556/2006.8.2019.70] [Medline: 31891314]
39. Walther B, Morgenstern M, Hanewinkel R. Co-occurrence of addictive behaviours: personality factors related to substance use, gambling and computer gaming. Eur Addict Res 2012;18(4):167-174. [doi: 10.1159/000335662] [Medline: 22398819]

XSL•FO
RenderX

40. Mills DJ, Marchica L, Keough MT, Derevensky JL. Exploring differences in substance use among emerging adults at-risk for problem gambling, and/or problem video gaming. Int Gambl Stud 2020 Sep 1;20(3):539-555. [doi: 10.1080/14459795.2020.1752768]
41. Schluter MG, Hodgins DC, Wolfe J, Wild TC. Can one simple questionnaire assess substance-related and behavioural addiction problems? Results of a proposed new screener for community epidemiology. Addiction 2018 Aug;113(8):1528-1537. [doi: 10.1111/add.14166] [Medline: 29357188]
42. Brooks GA, Clark L. Associations between loot box use, problematic gaming and gambling, and gambling-related cognitions. Addict Behav 2019 Sep;96:26-34. [doi: 10.1016/j.addbeh.2019.04.009] [Medline: 31030176]
43. Greenberg JL, Lewis SE, Dodd DK. Overlapping addictions and self-esteem among college men and women. Addict Behav 1999;24(4):565-571. [doi: 10.1016/s0306-4603(98)00080-x] [Medline: 10466852]
44. Macey J, Abarbanel B, Hamari J. What predicts esports betting? A study on consumption of video games, esports, gambling and demographic factors. New Media Soc 2021 Jun;23(6):1481-1505. [doi: 10.1177/1461444820908510]
45. Spicer SG, Nicklin LL, Uther M, Lloyd J, Lloyd H, Close J. Loot boxes, problem gambling and problem video gaming: a systematic review and meta-synthesis. New Media Soc 2022 Apr;24(4):1001-1022. [doi: 10.1177/14614448211027175]
46. Xiao LY, Fraser TC, Nielsen RKL, Newall PWS. Loot boxes, gambling-related risk factors, and mental health in mainland China: a large-scale survey. Addict Behav 2024 Jan;148:107860. [doi: 10.1016/j.addbeh.2023.107860] [Medline: 37729858]
47. Spicer SG, Close J, Nicklin LL, et al. Exploring the relationships between psychological variables and loot box engagement, part 2: exploratory analyses of complex relationships. R Soc Open Sci 2024 Jan;11(1):231046. [doi: 10.1098/rsos.231046] [Medline: 38179078]
48. D'Amico NJ, Drummond A, de Salas K, et al. No effect of short term exposure to gambling like reward systems on post game risk taking. Sci Rep 2022 Oct 6;12(1):16751. [doi: 10.1038/s41598-022-21222-3] [Medline: 36202911]
49. Scholten OJ, Hughes NGJ, Deterding S, Drachen A, Walker JA, Zendle D. Ethereum crypto-games: mechanics, prevalence, and gambling similarities. Presented at: Computer-Human Interaction in Play; Oct 22-25, 2019; Barcelona, Spain p. 379-389. [doi: 10.1145/3311350.3347178]
50. Serada A. Why is cryptokitties (not) gambling? Presented at: FDG '20; Sep 15-18, 2020; Bugibba, Malta p. 1-4 URL: https://dl.acm.org/doi/proceedings/10.1145/3402942 [accessed 2024-12-18] [doi: 10.1145/3402942.3402985]
51. Johnson TE, Dixon MR. Influencing children's pregambling game playing via conditional discrimination training. J Appl Behav Anal 2009;42(1):73-81. [doi: 10.1901/jaba.2009.42-73] [Medline: 19721731]
52. Hayer T, Kalke J, Meyer G, Brosowski T. Do simulated gambling activities predict gambling with real money during adolescence? Empirical findings from a longitudinal study. J Gambl Stud 2018 Sep;34(3):929-947. [doi: 10.1007/s10899-018-9755-1] [Medline: 29435822]
53. Vadlin S, Åslund C, Nilsson KW. A longitudinal study of the individual- and group-level problematic gaming and associations with problem gambling among Swedish adolescents. Brain Behav 2018 Apr;8(4):e00949. [doi: 10.1002/brb3.949] [Medline: 29670829]
54. Brosowski T, Turowski T, Hayer T. Simulated gambling consumption mediation model (SGCMM): disentangling convergence with parallel mediation models. Int Gambl Stud 2020 Sep 1;20(3):466-486. [doi: 10.1080/14459795.2020.1799426]
55. González-Cabrera J, Basterra-González A, Ortega-Barón J, et al. Loot box purchases and their relationship with internet gaming disorder and online gambling disorder in adolescents: a prospective study. Comput Hum Behav 2023 Jun;143:107685. [doi: 10.1016/j.chb.2023.107685]
56. Tárrega S, Castro-Carreras L, Fernández-Aranda F, et al. A serious videogame as an additional therapy tool for training emotional regulation and impulsivity control in severe gambling disorder. Front Psychol 2015;6:1721. [doi: 10.3389/fpsyg.2015.01721] [Medline: 26617550]
57. Jouhki H, Savolainen I, Sirola A, Oksanen A. Escapism and excessive online behaviors: a three-wave longitudinal study in Finland during the COVID-19 pandemic. Int J Environ Res Public Health 2022 Sep 30;19(19):12491. [doi: 10.3390/ijerph191912491] [Medline: 36231799]
58. Mohamed MS, Rukh G, Schiöth HB, et al. Worsened anxiety and loneliness influenced gaming and gambling during the COVID-19 pandemic. JCM 2023;12(1):249. [doi: 10.3390/jcm12010249]
59. Wardle H, Tipping S. The relationship between problematic gambling severity and engagement with gambling products: longitudinal analysis of the Emerging Adults Gambling Survey. Addiction 2023 Jun;118(6):1127-1139. [doi: 10.1111/add.16125] [Medline: 36606732]
60. Brooks GA, Clark L. The gamblers of the future? Migration from loot boxes to gambling in a longitudinal study of young adults. Comput Human Behav 2023 Apr;141:107605. [doi: 10.1016/j.chb.2022.107605]
61. Armstrong T, Rockloff M, Browne M, Li E. An exploration of how simulated gambling games may promote gambling with money. J Gambl Stud 2018 Dec;34(4):1165-1184. [doi: 10.1007/s10899-018-9742-6] [Medline: 29322288]
62. Greer N, Rockloff MJ, Russell AMT, Lole L. Are esports bettors a new generation of harmed gamblers? A comparison with sports bettors on gambling involvement, problems, and harm. J Behav Addict 2021;10(3):435-446. [doi: 10.1556/2006.2021.00039]
63. Kristiansen S, Severin MC. Loot box engagement and problem gambling among adolescent gamers: findings from a national survey. Addict Behav 2020 Apr;103:106254. [doi: 10.1016/j.addbeh.2019.106254] [Medline: 31881408]

XSL•FO
RenderX

64. Gainsbury SM, Abarbanel B, Blaszczynski A. Game on: comparison of demographic profiles, consumption behaviors, and gambling site selection criteria of esports and sports bettors. Gaming Law Rev 2017 Oct;21(8):575-587. [doi: 10.1089/glr2.2017.21813]
65. DeCamp W. Loot boxes and gambling: similarities and dissimilarities in risk and protective factors. J Gambl Stud 2021 Mar;37(1):189-201. [doi: 10.1007/s10899-020-09957-y] [Medline: 32514653]
66. Ide S, Nakanishi M, Yamasaki S, et al. Adolescent problem gaming and loot box purchasing in video games: cross-sectional observational study using population-based cohort data. JMIR Serious Games 2021 Feb 9;9(1):e23886. [doi: 10.2196/23886] [Medline: 33560241]
67. DeCamp W, Daly K. Loot box consumption by adolescents pre- and post- pandemic lockdown. PeerJ 2023;11:e15287. [doi: 10.7717/peerj.15287] [Medline: 37151299]
68. Hing N, Lole L, Russell AMT, et al. Adolescent betting on esports using cash and skins: links with gaming, monetary gambling, and problematic gambling. PLoS One 2022;17(5):e0266571. [doi: 10.1371/journal.pone.0266571] [Medline: 35511808]
69. Hing N, Rockloff M, Russell AMT, et al. Loot box purchasing is linked to problem gambling in adolescents when controlling for monetary gambling participation. J Behav Addict 2022;11(2):396-405. [doi: 10.1556/2006.2022.00015]
70. Muela I, Navas JF, Barrada JR, et al. Operationalization and measurement of compulsivity across video gaming and gambling behavioral domains. BMC Psychol 2023 Nov 21;11(1). [doi: 10.1186/s40359-023-01439-1]
71. King DL, Russell AMT, Delfabbro PH, Polisena D. Fortnite microtransaction spending was associated with peers' purchasing behaviors but not gaming disorder symptoms. Addict Behav 2020 May;104:106311. [doi: 10.1016/j.addbeh.2020.106311] [Medline: 31978759]
72. Close J, Spicer SG, Nicklin LL, Uther M, Lloyd J, Lloyd H. Secondary analysis of loot box data: are high-spending "whales" wealthy gamers or problem gamblers? Addict Behav 2021 Jun;117:106851. [doi: 10.1016/j.addbeh.2021.106851] [Medline: 33578105]
73. Close J, Spicer SG, Nicklin LL, Lloyd J, Lloyd H. Loot box engagement: relationships with educational attainment, employment status and earnings in a cohort of 16 000 United Kingdom gamers. Addiction 2022 Aug;117(8):2338-2345. [doi: 10.1111/add.15837] [Medline: 35129238]
74. von Meduna M, Steinmetz F, Ante L, Reynolds J, Fiedler I. Loot boxes are gambling-like elements in video games with harmful potential: results from a large-scale population survey. Technol Soc 2020 Nov;63:101395. [doi: 10.1016/j.techsoc.2020.101395]
75. Wardle H. The same or different? Convergence of skin gambling and other gambling among children. J Gambl Stud 2019 Dec;35(4):1109-1125. [doi: 10.1007/s10899-019-09840-5] [Medline: 30852730]
76. King A, Wong-Padoongpatt G, Barrita A, Phung DT, Tong T. Risk factors of problem gaming and gambling in us emerging adult non-students: the role of loot boxes, microtransactions, and risk-taking. Issues Ment Health Nurs 2020 Dec;41(12):1063-1075. [doi: 10.1080/01612840.2020.1803461] [Medline: 32822251]
77. Teichert T, Gainsbury SM, Mühlbach C. Positioning of online gambling and gaming products from a consumer perspective: a blurring of perceived boundaries. Comput Hum Behav 2017 Oct;75:757-765. [doi: 10.1016/j.chb.2017.06.025]
78. Imataka G, Izumi S, Miyamoto Y, Maehashi A. Gaming disorders: navigating the fine line between entertainment and addiction-gaming history, health risks, social consequences, and pathways to prevention. J Clin Med 2024 Aug 29;13(17):5122. [doi: 10.3390/jcm13175122] [Medline: 39274335]
79. Saini N, Hodgins DC. Investigating gaming structural features associated with gaming disorder and proposing a revised taxonomical model: a scoping review. J Behav Addict 2023 Jun 29;12(2):352-374. [doi: 10.1556/2006.2023.00019] [Medline: 37256740]
80. Macey J, Hamari J, Adam M. A conceptual framework for understanding and identifying gamblified experiences. Comput Human Behav 2024 Mar;152:108087. [doi: 10.1016/j.chb.2023.108087]
81. Kim HS, Leslie RD, Stewart SH, et al. A scoping review of the association between loot boxes, esports, skin betting, and token wagering with gambling and video gaming behaviors. J Behav Addict 2023 Jun 29;12(2):309-351. [doi: 10.1556/2006.2023.00013] [Medline: 37159337]
82. Labrador FJ, Bernaldo-de-Quirós M, Sánchez-Iglesias I, et al. Advertising games of chance in adolescents and young adults in Spain. J Gambl Stud 2021 Sep;37(3):765-778. [doi: 10.1007/s10899-020-09988-5] [Medline: 33169222]
83. Russell AMT, Armstrong TA, Rockloff MJ, Greer NM, Hing N, Browne M. Exploring the changing landscape of gambling in childhood, adolescence and young adulthood. : CQUniversity; 2020. [doi: 10.25946/5F2335F6D50D1]
84. Abarbanel B, Phung D. Exploring gamers' perceptions of esports betting advertising. Gaming Law Rev 2019 Nov 1;23(9):640-644. [doi: 10.1089/glr2.2019.2394]
85. Rossi R, Nairn A, Smith J, Inskip C. "Get a £10 free bet every week!"—gambling advertising on Twitter: volume, content, followers, engagement, and regulatory compliance. J Public Policy & Marketing 2021 Oct;40(4):487-504. [doi: 10.1177/0743915621999674]
86. Abarbanel B, Johnson MR. Gambling engagement mechanisms in Twitch live streaming. Int Gambl Stud 2020 Sep 1;20(3):393-413. [doi: 10.1080/14459795.2020.1766097]

XSL•FO
RenderX

87. Denoo M, Bibert N, Zaman B. Disentangling the motivational pathways of recreational esports gamblers: a laddering study. Presented at: Computer-Human Interaction in Play; Oct 18-21, 2021; Yokohama Japan p. 1-15 URL: https://dl.acm.org/doi/proceedings/10.1145/3411764 [accessed 2024-09-20] [doi: 10.1145/3411764.3445287]
88. Lakić N, Bernik A, Čep A. Addiction and spending in gacha games. Information 2023;14(7):399. [doi: 10.3390/info14070399]
89. Bujňáková E. Consumer behavior of league of legends players [Bachelor's thesis]. : Satakunta University of Applied Sciences; 2023 URL: http://www.theseus.fi/handle/10024/812974 [accessed 2024-12-13]
90. Wieczorek Ł, Bujalski M, Dąbrowska K. "I can tell you it's a bit of a gamble": a qualitative analysis of how people who engage in gaming and gambling understand a link between these two behaviours. J Gambl Stud 2024 Jun;40(2):859-871. [doi: 10.1007/s10899-023-10275-2] [Medline: 38100034]
91. Calado F, Alexandre J, Griffiths MD. Mom, dad it's only a game! perceived gambling and gaming behaviors among adolescents and young adults: an exploratory study. Int J Ment Health Addict 2014 Dec;12(6):772-794. [doi: 10.1007/s11469-014-9509-y]
92. Rockloff M, Russell AMT, Greer N, Lole L, Hing N, Browne M. Young people who purchase loot boxes are more likely to have gambling problems: an online survey of adolescents and young adults living in NSW Australia. J Behav Addict 2021 Feb 24;10(1):35-41. [doi: 10.1556/2006.2021.00007] [Medline: 33625382]
93. Duffy L. Gen bet: a plain english summary of research into gambling and young people. Victorian Responsible Gambling Foundation. 2021. URL: https://responsiblegambling.vic.gov.au/resources/publications/gen-bet-a-plain-english-summary-of-research-into-gambling-and-young-people-990/ [accessed 2022-08-18]
94. Greer N, Rockloff M, Hing N, Browne M, King DL. Skin gambling contributes to gambling problems and harm after controlling for other forms of traditional gambling. J Gambl Stud 2023 Mar;39(1):225-247. [doi: 10.1007/s10899-022-10111-z] [Medline: 35217933]
95. Sidloski B, Brooks GA, Zhang K, Clark L. Exploring the association between loot boxes and problem gambling: are video gamers referring to loot boxes when they complete gambling screening tools? Addict Behav 2022 Aug;131:107318. [doi: 10.1016/j.addbeh.2022.107318] [Medline: 35381433]
96. Greer N. Experiences with esports betting and skin gambling: exposure, access, motivations and impacts [Thesis]. : CQUniversity; 2023 URL: https://acquire.cqu.edu.au/articles/thesis/Experiences_with_esports_betting_and_skin_gambling_Exposure_access_motivations_and_impacts/23897322/1 [accessed 2024-09-20]
97. Drummond A, Hall LC, Sauer JD. Surprisingly high prevalence rates of severe psychological distress among consumers who purchase loot boxes in video games. Sci Rep 2022 Sep 27;12(1):16128. [doi: 10.1038/s41598-022-20549-1] [Medline: 36167899]
98. Cole JD. Understanding school counselors' perceptions of esports and igaming as a career choice [Doctoral dissertation]. : Oregon State University; 2023 URL: https://ir.library.oregonstate.edu/concern/graduate_thesis_or_dissertations/j9602803m [accessed 2024-09-20]
99. Declerck P, Feci N. Mapping and analysis of the current regulatory framework on gambling(- like) elements in video games: a report in the framework of the 'gam(e)(a)ble' research project. : Universiteit Gent; 2022 Report No: 8768539 URL: http://hdl.handle.net/1854/LU-8768539 [accessed 2024-12-13]
100. Laato S. Is simulating casino environments in video games worse than gambling with loot boxes? The case of the removed pokémon game corner. Presented at: Conference on Technology Ethics; Oct 21, 2020; Turku, Finland p. 100-109 URL: http://ceur-ws.org/Vol-2737/SP_1.pdf [accessed 2024-09-20]
101. Xiao LY. Which implementations of loot boxes constitute gambling? A UK legal perspective on the potential harms of random reward mechanisms. Int J Ment Health Addict 2022 Feb;20(1):437-454. [doi: 10.1007/s11469-020-00372-3]
102. Kolandai-Matchett K, Wenden Abbott M. Gaming-gambling convergence: trends, emerging risks, and legislative responses. Int J Ment Health Addict 2022 Aug;20(4):2024-2056. [doi: 10.1007/s11469-021-00498-y]
103. Whitson J, French M. Productive play: the shift from responsible consumption to responsible production. J Consumer Cult 2021 Feb;21(1):14-33. [doi: 10.1177/1469540521993922]
104. Schwiddessen S, Karius P. Watch your loot boxes! – Recent developments and legal assessment in selected key jurisdictions from a gambling law perspective. IELR 2018 Jun 20;1(1):17-43 [FREE Full text] [doi: 10.4337/ielr.2018.01.02]
105. McCaffrey M. A cautious approach to public policy and loot box regulation. Addict Behav 2020 Mar;102:106136. [doi: 10.1016/j.addbeh.2019.106136] [Medline: 31704434]
106. Amadieu T. Addictive technologies? The moral and normative dynamics shaping the Chinese gaming culture. In: Chrétien-Ichikawa S, Pawlik K, editors. Creative Industries and Digital Transformation in China: Springer Nature Singapore; 2022:59-82. [doi: 10.1007/978-981-19-3049-2_4]
107. Delfabbro P, King DL. Gaming-gambling convergence: evaluating evidence for the 'gateway' hypothesis. Int Gambl Stud 2020 Sep 1;20(3):380-392. [doi: 10.1080/14459795.2020.1768430]
108. Garea SS, Drummond A, Sauer JD, Hall LC, Williams MN. Meta-analysis of the relationship between problem gambling, excessive gaming and loot box spending. Int Gambl Stud 2021 Sep 2;21(3):460-479. [doi: 10.1080/14459795.2021.1914705]

XSL•FO
RenderX

109. Biegun J, Edgerton JD, Roberts LW. Measuring problem online video gaming and its association with problem gambling and suspected motivational, mental health, and behavioral risk factors in a sample of university students. Games Cult 2021 Jun;16(4):434-456. [doi: 10.1177/1555412019897524]
110. Delfabbro P, King DL, Lambos C, Puglies S. Is video-game playing a risk factor for pathological gambling in Australian adolescents? J Gambl Stud 2009 Sep;25(3):391-405. [doi: 10.1007/s10899-009-9138-8] [Medline: 19578983]
111. Castrén S, Järvinen-Tassopoulos J, Raitasalo K. Money used in gaming is associated with problem gambling: results of the ESPAD 2019 Finland. J Behav Addict 2021 Nov 17;10(4):932-940. [doi: 10.1556/2006.2021.00076] [Medline: 34797777]
112. Rodda S. A rapid review and research gap analysis: a 2020 update. Commissioned by the NSW Responsible Gambling Fund. GambleAware. 2020. URL: https://www.gambleaware.nsw.gov.au/resources-and-education/check-out-our-research/published-research/a-rapid-review-and-research-gap-analysis---a-2020-update [accessed 2022-08-21]
113. Stevens MW, Dorstyn D, Delfabbro PH, King DL. Global prevalence of gaming disorder: a systematic review and meta-analysis. Aust N Z J Psychiatry 2021 Jun;55(6):553-568. [doi: 10.1177/0004867420962851] [Medline: 33028074]
114. De Leon A. Cutting losses: Southeast Asia's crackdown on online gambling. The Diplomat. 2024. URL: https://thediplomat.com/2024/08/cutting-losses-southeast-asias-crackdown-on-online-gambling/ [accessed 2024-12-13]
115. Rattaphong S. Governance and strategies of gambling business in Southeast Asia. CBSR 2022;3(2):55-64. [doi: 10.22495/cbsrv3i2art5]
116. Wardle H, Degenhardt L, Marionneau V, et al. The Lancet Public Health Commission on gambling. Lancet Public Health 2024 Nov;9(11):e950-e994. [doi: 10.1016/S2468-2667(24)00167-1]
117. Burleigh TL, Griffiths MD, Sumich A, et al. Co-occurrence of gaming disorder and other potentially addictive behaviours between Australia, New Zealand, and the United Kingdom. Int J Environ Res Public Health 2022 Dec 1;19(23):16078. [doi: 10.3390/ijerph192316078] [Medline: 36498151]
118. King DL, Delfabbro PH. Adolescents' perceptions of parental influences on commercial and simulated gambling activities. Int Gambl Stud 2016 Sep;16(3):424-441. [doi: 10.1080/14459795.2016.1220611]
119. Lia DAZ, Natswa SL. Buy-now-pay-later (BNPL): generation Z's dilemma on impulsive buying and overconsumption intention. Presented at: Proceedings of the BISTIC Business Innovation Sustainability and Technology International Conference East Java; Jul 27-28, 2021; East Java, Indonesia p. 130-137.
120. Brand M, Wegmann E, Stark R, et al. The Interaction of Person-Affect-Cognition-Execution (I-PACE) model for addictive behaviors: update, generalization to addictive behaviors beyond internet-use disorders, and specification of the process character of addictive behaviors. Neurosci Biobehav Rev 2019 Sep;104:1-10. [doi: 10.1016/j.neubiorev.2019.06.032] [Medline: 31247240]
121. Håkansson A, Widinghoff C. Television gambling advertisements: extent and content of gambling advertisements with a focus on potential high-risk commercial messages. Addict Behav Rep 2019 Jun;9:100182. [doi: 10.1016/j.abrep.2019.100182] [Medline: 31193826]
122. Bae S, Han DH, Jung J, Nam KC, Renshaw PF. Comparison of brain connectivity between internet gambling disorder and Internet gaming disorder: a preliminary study. J Behav Addict 2017 Dec 1;6(4):505-515. [doi: 10.1556/2006.6.2017.061] [Medline: 29039224]
123. Close J, Spicer SG, Nicklin LL, Lloyd J, Whalley B, Lloyd H. Gambling and gaming in the United Kingdom during the COVID-19 lockdown. COVID 2021;2(2):87-101. [doi: 10.3390/covid2020007]
124. Zendle D, Cairns P. Video game loot boxes are linked to problem gambling: results of a large-scale survey. PLoS One 2018;13(11):e0206767. [doi: 10.1371/journal.pone.0206767] [Medline: 30462669]
125. Zendle D, Cairns P. Loot boxes are again linked to problem gambling: results of a replication study. PLoS One 2019;14(3):e0213194. [doi: 10.1371/journal.pone.0213194] [Medline: 30845155]
126. Drummond A, Sauer JD, Ferguson CJ, Hall LC. The relationship between problem gambling, excessive gaming, psychological distress and spending on loot boxes in Aotearoa New Zealand, Australia, and the United States—a cross-national survey. PLoS One 2020;15(3):e0230378. [doi: 10.1371/journal.pone.0230378] [Medline: 32203522]
127. Liu K. A global analysis into loot boxes: is it "virtually" gambling? Washington. Int Law J 2019;28(3):7-8.
128. Guerrin CGJ, Doorduin J, Sommer IE, de Vries EFJ. The dual hit hypothesis of schizophrenia: evidence from animal models. Neurosci Biobehav Rev 2021 Dec;131:1150-1168. [doi: 10.1016/j.neubiorev.2021.10.025] [Medline: 34715148]
129. Bayer TA, Falkai P, Maier W. Genetic and non-genetic vulnerability factors in schizophrenia: the basis of the "two hit hypothesis". J Psychiatr Res 1999;33(6):543-548. [doi: 10.1016/s0022-3956(99)00039-4] [Medline: 10628531]
130. Shi J, Colder Carras M, Potenza MN, Turner NE. A perspective on age restrictions and other harm reduction approaches targeting youth online gambling, considering convergences of gambling and videogaming. Front Psychiatry 2020;11:601712. [doi: 10.3389/fpsyt.2020.601712] [Medline: 33584369]
131. Haskell JV. More than just skin(s) in the game: how one digital video game item is being used for unregulated gambling purposes online. J High Technol Law 2015;18(1):125-160.
132. McCaffrey M. The macro problem of microtransactions: the self-regulatory challenges of video game loot boxes. Bus Horiz 2019 Jul;62(4):483-495. [doi: 10.1016/j.bushor.2019.03.001]
133. Sanders J, Williams R. The relationship between video gaming, gambling, and problematic levels of video gaming and gambling. J Gambl Stud 2019 Jun;35(2):559-569. [doi: 10.1007/s10899-018-9798-3] [Medline: 30121839]

XSL•FO
RenderX

Abbreviations

GLF: gambling-like feature
***ICD-11*:** *International Classification of Diseases, 11th Revision*
JBI: Joanna Briggs Institute
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews

Edited by A Coristine, T Leung; submitted 24.04.24; peer-reviewed by I Gunardi, J Machimbarrena, M Huot-Lavoie; revised version received 24.11.24; accepted 25.11.24; published 15.01.25.

Please cite as:
Siste K, King DL, Hanafi E, Sen LT, Adrian A, Murtani BJ
Scrutinizing the Gateway Relationship Between Gaming and Gambling Disorder: Scoping Review With a Focus on the Southeast Asian Region
JMIR Serious Games 2025;13:e59740
URL: https://games.jmir.org/2025/1/e59740
doi:10.2196/59740

© Kristiana Siste, Daniel L King, Enjeline Hanafi, Lee Thung Sen, Adrian Adrian, Belinda Julivia Murtani. Originally published in JMIR Serious Games (https://games.jmir.org), 15.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Pervasive Games for Sexual Health Promotion: Scoping Literature Review

Claudio Rubio[1], BEng; Felipe Besoain[2], PhD

[1]Faculty of Engineering, Universidad de Talca, Curicó, Chile

[2]Department of Interactive Visualization and Virtual Reality, Faculty of Engineering, University of Talca, Talca, Chile

Corresponding Author:
Felipe Besoain, PhD
Department of Interactive Visualization and Virtual Reality
Faculty of Engineering
University of Talca
2 norte #685
Talca, 3460000
Chile
Phone: 56 2201687 ext 2687
Email: fbesoain@utalca.cl

Abstract

Background: Serious games play a fundamental role in promoting safe sexual behaviors. This medium has great potential for promoting healthy behaviors that prevent potential risk factors, such as sexually transmitted infections, and promote adherence to sexual health treatments, such as antiretroviral therapy. The ubiquity of mobile devices enhances access to such tools, increasing the effectiveness of video games as agents of change.

Objective: In this scoping review, we aimed to (1) identify the extent to which pervasive games have been used in the field of sexual health, (2) determine the theories used in the design and evaluation of pervasive games for sexual health, (3) identify the methods used to evaluate pervasive games for sexual health, and (4) explore the reported benefits of using pervasive games for sexual health.

Methods: Following the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) methodology, we conducted a comprehensive literature search in the Web of Science, Scopus, IEEE Xplore, and ACM databases for articles published between January 1, 2000, and August 4, 2024. Included articles were published in English between 2000 and 2024 and involved the design, implementation, or evaluation of a ubiquitous video game focused on promoting safe sexual behaviors, with qualitative and/or quantitative results based on theory-based techniques and ubiquitous technologies. Review articles, conference papers, or books without available data or quantitative or qualitative results were excluded.

Results: We screened 521 of 612 articles (85.1%) after removing duplicates. After the title and abstract review, 51 (9.8%) articles were assessed for eligibility, and 30 (5.8%) articles meeting the criteria were studied and evaluated in depth. The results suggested that the use of pervasive video games has a positive impact on promoting safe sexual behaviors. This is enhanced by the effectiveness of theory-based techniques and the use of mobile technologies as developmental factors that drive the gaming experience. The results indicated that this domain is a growing field that should not be ignored.

Conclusions: The literature showed that pervasive video games have been effective in promoting safe sexual behaviors. Substantial growth has been seen in scientific community interest in researching this domain; nevertheless, there is still much to work on. In this context, we advocate for the standardization of design, implementation, and experimentation as essential phases in creating video game experiences. These 3 fundamental aspects are critical in the development of video game–based studies to ensure the reproducibility of experiments.

(JMIR Serious Games 2025;13:e58912) doi:10.2196/58912

KEYWORDS

serious games; promotion; ubiquitous technologies; healthy behaviors; HIV; sexually transmitted infection; STI; scoping review; mobile phone

XSL•FO
RenderX

Introduction

Background

The global effort to combat sexually transmitted infections (STIs), including HIV and viral hepatitis, faces significant hurdles in promoting sexual health safety. These infections account for 2.3 million deaths and 1.2 million cancer cases annually. Daily, there are >1 million new STI cases, with 4.5 million individuals contracting HIV and hepatitis B and C annually [1,2]. Despite progress in HIV treatment, 2021 witnessed 1.5 million new HIV infections and 650,000 AIDS-related deaths, underscoring the persistent challenge HIV presents [3]. By the end of 2022, the population that tested positive for HIV reached approximately 39 million, with a notable 38% reduction in new infections since 2010; yet, HIV infection remains a major health issue [4]. Statistics also reveal alarming rates of STI transmission, with >1 million new infections daily, 374 million new cases of curable STIs per year, >500 million people living with herpes simplex virus, and human papillomavirus (HPV) causing >311,000 cervical cancer deaths annually [5]. These figures highlight the critical need for intensified global action aimed at achieving the diagnosis, treatment, and suppression goals for STIs by 2030 [6].

One current focus of these strategies is to promote sexual education and sexual well-being through access to reliable and detailed information covering topics related to sexuality and the risk of conditions such as HIV, STIs, and HPV [2]. However, despite significant progress in raising awareness and promoting safe sexual practices, effectively implementing these strategies presents notable challenges. Cultural taboos and sensitivities associated with traditions can complicate conversations about sexuality in some communities [7]. In addition, conservative roles and gender expectations among teenagers create barriers that limit access to sexual education, especially in areas where health professionals and parents are the primary providers of information [8]. As a result, there is a lack of comprehensive sexual education that provides teenagers with the necessary knowledge for a safe and healthy sexual life [9].

In a world where rapid and efficient access to information has become indispensable, ubiquitous and pervasive technologies have become a necessity. Ubiquitous computing focuses on the seamless and natural integration of technology into the environment, aiming to be omnipresent but noninvasive [10]. In contrast, pervasive computing focuses on the immediate use of processed information through the use of mobile devices and wireless communication [11,12]. Both concepts share the goal of making technology accessible and useful in everyday life but differ in their approach and application [13].

With the proliferation of mobile devices, the use of pervasive apps has become a standard, primarily due to the portability they offer to the users, enabling efficient access to knowledge through any device, at any time, and anywhere [14]. This advancement has created a unique space for innovative software development in the health sector [15]. Apps in this area have been vital for health promotion, and due to their omnipresent nature, they have integrated into people's daily lives [16].

For example, studies have been conducted on the use of mobile devices for the continuous monitoring, diagnosis, and treatment of mental health conditions [17,18] as well as to improve adherence to antiretroviral therapy (ART) in young people with HIV [19]. A review of serious video games for the prevention of HIV concluded that these games can be innovative and effective, as they provide interactive experiences with realistic scenarios, establish an agency in a safe virtual environment, and allow individuals to practice and develop skills without social stigma [20]. Another review highlighted the effectiveness of serious games on mobile platforms and the application of gamification techniques as innovative strategies in the prevention and care of HIV or AIDS, emphasizing their cost efficiency compared to long-term treatment options [21]. A third review in the field emphasized the use of persuasive games as tools for promoting healthy behaviors in domains such as physical activity, disease control, nutrition, and environmental sustainability, among others. Besides highlighting the importance of basing these games on behavioral theory, it suggested the need for longitudinal studies and more detailed evaluations of persuasive strategies to optimize the design of future persuasive games [22].

These studies underscored the value of serious games as effective tools for promoting adherence to healthy behaviors. Through simulations, these games provided exceptional opportunities for learning and skill development, becoming optimal means for awareness, education, and treatment of diseases. They also offered a pleasant and motivating way to interact with complex topics [23], including learning [24], behavior modification [25], and the promotion of healthy habits [26].

However, in health promotion through pervasive apps, it is crucial to distinguish between serious games and gamified apps. Serious games naturally present themselves as real games, are focused on improving performance or learning, and are designed to drive behavioral and attitudinal changes, offering an immersive experience where the act of playing itself is central [22,27]. In contrast, gamified apps incorporate ludic elements into traditionally nongaming environments to create an experience similar to a video game [28], with the aim of motivating and engaging users in the process of attitudinal change [29]. While they apply ludic aspects, they are not games themselves but tools that enhance the user experience by integrating game elements to enrich conventional activities [21,28]. Although both methods can be effective in shaping attitudes, this review focuses on serious games due to the significant difference in the gaming experience they offer compared to gamified systems.

This Study

This scoping literature review focused on exploring the intersection between theory-based techniques (persuasive, cognitive, or behavioral) applied in serious games and the use of pervasive media as integral components of the gaming experience. Exploring this relationship is crucial to identify the strengths, weaknesses, gaps, and opportunities that these media offer in promoting safe sexual behaviors. This study aimed to provide a deeper and more detailed perspective on how

XSL•FO
RenderX

pervasive mobile games can effectively contribute to sexual health promotion, attitude change, and sexual education in our society. Building on gaps observed in previous works that studied the use of video games for promoting safe sexual behaviors [20-22], this study sought to answer 4 research questions (RQs): RQ1—How have pervasive games been used in the field of sexual health? RQ2—What methods and technologies have been used to design pervasive video games for sexual health? RQ3—What theories are used in the design and evaluation of pervasive games for sexual health? and RQ4—What methods, along with their benefits, have been used to evaluate pervasive games for sexual health?

Methods

Overview

This scoping review followed the established methodological framework proposed by Arksey and O'Malley [30]. We adopted the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines [31]. The completed checklist is provided in Multimedia Appendix 1. We did not preregister for this scoping review because of its descriptive and exploratory nature.

Search Strategy

We explored 4 databases: Web of Science, Scopus, IEEE Xplore, and ACM. Web of Science and Scopus were chosen due to their ability to provide a comprehensive view of the evolution of serious games focused on promoting healthy behaviors and their relationship with pervasive games. In addition, IEEE Xplore and ACM were included due to their specialized focus on engineering and technology. The time frame analyzed spans from January 1, 2000, to August 4, 2024. In addition, we used the population, intervention, comparison, and outcomes (PICO) structure to define a suitable search string [32]. Textbox 1 presents the keywords of the PICO structure.

Textbox 1. population, intervention, comparison, and outcomes analysis with search terms.

Population

- Video games focused on promoting safe sexual health
- Terms: ((videogame OR games) AND (Health* OR well* OR self* OR care*) AND ((study OR intervention) AND (sex* OR STI)))

Intervention

- Design, implementation, or evaluation of pervasive video games and use of theory-based techniques
- Terms: ((videogame OR games) AND (persuasive OR persuasion OR theories OR strategies OR behav* OR change OR changing OR promote OR promotion OR promoting OR attitude OR prevent OR prevention OR preventing OR improve OR improving))

Comparison

- No specific direct comparison specified
- Terms: not applicable

Outcomes

- Qualitative and quantitative analysis of obtained data
- Terms: ((videogame OR games) AND ((study OR intervention) AND (sex* OR STI)))

We connected these concepts using the "AND" and "OR" operators, resulting in the following search string: ((videogame OR games) AND (persuasive OR persuasion OR theories OR strategies OR behav* OR change OR changing OR promote OR promotion OR promoting OR attitude OR prevent OR prevention OR preventing OR improve OR improving) AND (pervasive OR ubiquitous OR mobile OR smartphone OR omnipresent OR interactive OR presence) AND (design OR implement OR evaluation OR development OR technology OR technologies) AND (Health* OR well* OR self* OR care*)) AND ((study OR intervention) AND (sex* OR STI)). The complete search strategy, including a full list of search terms, is provided in Multimedia Appendix 2.

Eligibility Criteria

To systematically discern between eligible and ineligible articles for review, we defined a set of inclusion and exclusion criteria. Textbox 2 summarizes the criteria used.

XSL•FO
RenderX

Textbox 2. Inclusion and exclusion criteria.

Inclusion criteria

- Language: articles must be written in English
- Date: articles published between January 1, 2000, and August 4, 2024
- Database: articles must belong to the databases Web of Science, Scopus, IEEE Xplore, or ACM
- Document type: article or protocol
- Study type: studies based on the design, implementation, or evaluation of pervasive video games
- Health area: studies focusing on promoting safe sexual health
- Result type: articles with qualitative and quantitative data analysis.
- Device: specifically, the studies involving mobile devices in the context of promoting safe sexual health through pervasive video games

Exclusion Criteria

- Intervention: articles presenting gamified apps not specifically related to the use of pervasive video games for promoting safe sexual behaviors
- Data: studies with unavailable data
- Document type: nonempirical studies such as reviews, conference papers, or books that do not present empirical data
- Result type: articles that lack qualitative or quantitative data analysis
- Empirical: articles that do not focus on the use of video games for promoting safe sexual behaviors and present content relating to topics other than sexual health promotion (eg, nutrition, physical activity, smoking cessation, and mental health)

Study Selection

We used the Hubmeta platform to facilitate the process of duplicate removal, title screening, and full-text screening [33]. Once the queries were made in the selected databases, the results were imported into the platform and the reviewers did the screening process independently. At the end of the process (title screening and full-text screening), if reviewers disagreed about inclusion, a meeting was held to discuss and review the points from the PICO structure. All disagreements were resolved without the need for a third reviewer. Finally, the included articles were studied thoroughly in search of elements that answer the RQs.

Data Charting

The final list of data charting elements is presented in Textbox 3, in which we outline the proposed organization for storing the key data from each article reviewed. Data charting for all elements was conducted using a Microsoft Excel sheet. All the categories for different data charting options were initially created based on brainstorming relevant information and then refined with the information of a small subset of articles. Next, data were registered and discussed among the authors involved in the charting process.

XSL•FO
RenderX

Textbox 3. Data charting categorization.

Article data

- Name
- Year of publication
- Database
- Journal
- Type (protocol or article)

Game characteristics

- Name
- Topics
- Backstory or objective
- Genres
- Tasks

Development design

- Design methodology
- Target audience
- Device
- Inputs
- Behavioral change theories

Intervention characteristics

- Research design
- Sample population
- Country
- Mean age and SD
- Duration of dose
- Follow-up

Experimental data

- Results (preliminary or significative)
- Evaluation (qualitative, quantitative, or mixed)
- Outcomes (positive, negative, or neutral)

The information is categorized into five main groups: (1) article data, which include general information about each article; (2) game characteristics, which detail the key aspects of the video game; (3) development design, which focuses on the critical aspects of the video game's development phase; (4) intervention characteristics, which capture data related to the experimental stage of the various studies; and (5) experimental data, which present the results obtained during the experimental process.

Article data include the name of the article, the year of publication, the database from which it was retrieved, the journal, and the type of document (article or protocol). Although protocols only present the experimental design of future research, they were included because they helped identify important studies that were not captured in the initial queries.

Game characteristics capture the name of the video game, the topics addressed, the backstory or objective the character must achieve, the game's genres, and the tasks the user must complete to finish the experience. It is important to note that a video game can cover multiple topics and belong to more than one genre, so both categories allow for multiple selections.

In development design, the design methodologies guiding the experience were documented, along with the target audience, following the categorization guidelines from the World Health Organization [34,35] and the Centers for Disease Control and Prevention [36]. In addition, the devices for which the experience was developed, the inputs used, and the behavioral change theories that guided the attitudinal change process within the game were recorded. It is worth noting that a video game

XSL•FO
RenderX

can incorporate multiple behavioral change theories, so an article might present >1 theory.

Intervention characteristics such as the research design guiding the experience were identified, along with the sample population and its characteristics. The locations of the participants were recorded based on the countries from which they were recruited (if applicable). If an article did not describe participant recruitment, the authors' countries were recorded based on their affiliations. The mean (SD) ages of participants were obtained from the articles, and when unavailable, they were estimated following the recommendations by Hozo et al [37]. Alternatively, a weighted mean and weighted SD calculation was performed. If it was not possible to obtain the value by any of these methods, the box was marked as not available. The duration of the intervention, including the number of sessions and their length, and the follow-up period were also recorded.

Experimental data store the experimental results obtained from the articles. Since protocols are experimental designs for future studies, they do not provide experimental results and thus do not include this type of information. Articles were first categorized into 2 types: preliminary (those presenting proof of concept and their results) and significant (those conducting statistically significant studies), following the recommendations by Jitmun et al [21]. In terms of evaluation, articles were divided into three categories: qualitative, quantitative, and mixed methods. Finally, outcomes were categorized based on the contributions indicated by the authors as positive, negative, or neutral.

Results

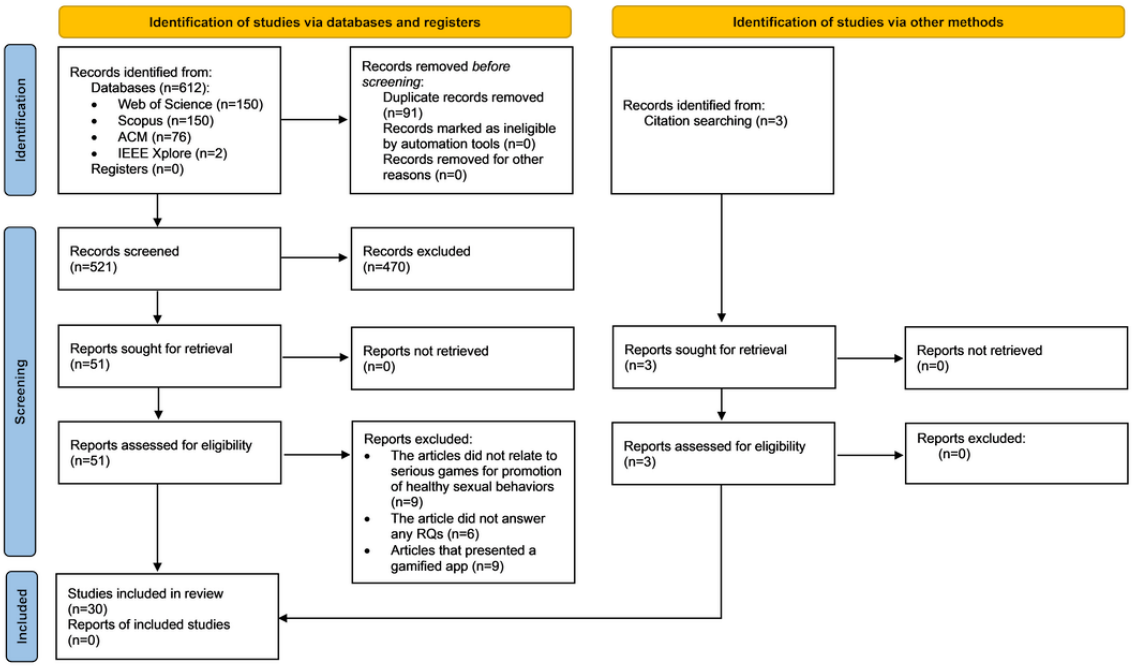
Search Results

Our search returned 612 articles (Web of Science: n=150, 24.5%; Scopus: n=384, 62.7%; ACM: n=76, 12.4%; and IEEE Xplore: n=2, 0.3%). Of 612 articles, we found and discarded 91 (14.9%) duplicate articles. Results of the 521 articles reviewed after remove duplicates can be founded in Multimedia Appendix 3. In the title screening, we evaluated the title and abstract of each article. From this step, we discarded 470 (90.2%) of the 521 documents for not meeting the inclusion and exclusion criteria.

The remaining 51 articles were included for full-text screening. After checking the full text of these 51 articles, 24 (47%) were excluded for presenting gamified apps or the central theme of the article not being related to the use of serious games for promoting safe sexual behaviors.

Finally, we identified 27 articles after this process, including 22 (82%) research articles and 5 (18%) protocols. Following the 5 protocols retrieved, we identified 3 additional research articles related to the protocols that were not retrieved by the original queries. Accordingly, the final number of included studies was 30. Figure 1 [38] presents the PRISMA-ScR flow diagram. The details of the papers' findings are provided in Multimedia Appendix 4.

Figure 1. PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) extension for scoping reviews flow diagram. RQ: research question.



General Screening

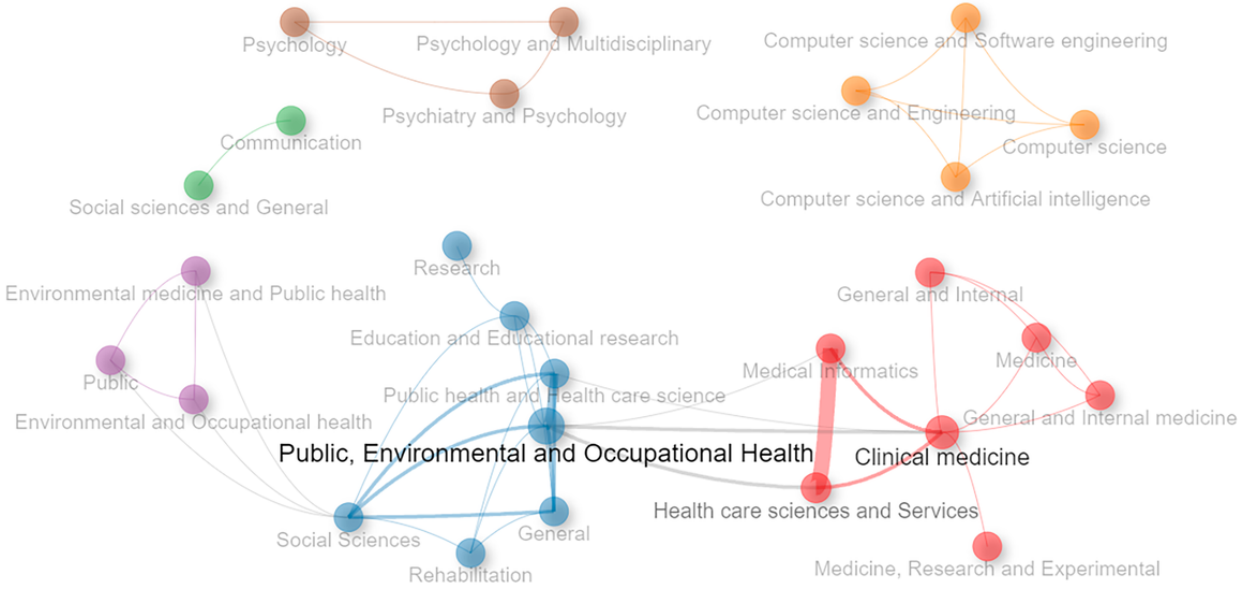
We observed that the most common journal where articles were published was *JMIR Serious Games* (8/30, 27%), followed by journals *JMIR Research Protocols* (3/30, 10%) and *Games for Health Journal* (3/30, 10%). The third position was shared by journals *JMIR mHealth and uHealth* (2/30, 7%), *BMC Public Health* (2/30, 7%), and *JMIR Formative Research* (2/30, 7%).

XSL•FO
RenderX

The remaining articles (10/30, 33%) were published in multiple journals, each with only 1 article. However, we did not find any studies published from 2000 to 2009, 2011 to 2014, or in 2017.

Figure 2 presents a visualization of the association network between journal categories from the master journal list of Web of Science. The connections in the network indicate how the different disciplines are thematically related based on their indexing (core collection, essential science indicators, and current contents). This network is shown using a network design [39] with a Walktrap clustering algorithm and association normalization. The network includes a maximum of 50 nodes and a repulsion strength set to 0.1. As a result, 6 clusters representing different subject areas are identified: 3 of them are highly interconnected (clusters 1, 2, and 3) and 3 are isolated groups (clusters 4, 5, and 6).

Figure 2. The visualization of the association network between journal categories from the master journal list of Web of Science. The connections in the network indicate how the different disciplines are thematically related based on their indexing.



Cluster 1 (blue) groups disciplines related to public, environmental, and occupational health; social sciences; rehabilitation; general research; and education and educational studies. The density of connections within this cluster suggests a strong interdisciplinary interrelation, reflecting the convergence of studies that combine public and environmental health with social and educational approaches.

Cluster 2 (red) concentrates on disciplines focused on medical informatics, health care sciences and services, and several subcategories of clinical medicine and research. The proximity and density of the interconnections suggest a strong relationship between medical research and the use of IT in health, such as medical informatics and clinical services, as well as studies in internal medicine.

Cluster 3 (purple) addresses categories related to occupational health, education, and public health. Although it is interconnected with other clusters, it maintains its distinct identity, possibly focused on issues that combine prevention and education in occupational health, as well as related public policies.

Cluster 4 (brown) is composed of categories related to psychology, psychiatry, and multidisciplinary psychology. The isolated nature of this cluster suggests that, although it is an internally interconnected field, its influence and relationship with other areas of research is limited within the analyzed network.

Cluster 5 (orange) includes categories of computer science, software engineering, artificial intelligence, and computer science and engineering. Its isolation reflects the specificity of the field and its high degree of specialization, suggesting that research topics in these areas intersect less with broader disciplines in the context of the network.

Cluster 6 (green) contains categories related to communication and social sciences in general. The isolation of this cluster could indicate that communication research, although relevant, is not strongly linked to the core areas of health and social sciences in the other larger clusters.

In the following subsections, results for each of the 4 RQs are presented.

RQ1: How Have Pervasive Games Been Used in the Field of Sexual Health?

Geographic Distribution of Target Populations in Sexual Health Promotion Research

The analysis of the geographic origins of the target populations in the reviewed research articles reveals a clear concentration of studies in specific regions. The United States stands out prominently, with 50% (15/30) of the articles focusing on populations from this country, indicating a strong emphasis on American contexts and possibly reflecting the substantial resources and infrastructure available for research in sexual health promotion there. Kenya follows with 17% (5/30) of the articles, suggesting an interest in addressing sexual health issues within this East African country. India and China each contribute

XSL•FO
RenderX

to the research landscape with 7% (2/30) of the articles from each country, highlighting emerging interests in sexual health promotion in South Asia and East Asia. Mexico, the Philippines, Singapore, Chile, Norway, and the United Kingdom each contributed 3% (1/30) of the articles, indicating a more limited but nonetheless notable representation from diverse geographic regions. This distribution points to a predominance of research focused on North American and African contexts, with growing attention to other global regions.

Audience Targeting in Sexual Health Promotion Research

The examination of the research articles revealed a varied focus on different target audiences within the realm of sexual health promotion. Adolescents, with a mean age of 14.58 (SD 0.71) years, were the most frequently addressed group, with 37% (11/30) of the articles focusing on this demographic, reflecting its significant representation and the urgent need for tailored sexual health interventions in this age group. Men who have sex with men, with a mean age of 22.29 (SD 2.64) years, were the focus of 30% (9/30) of the articles, highlighting the importance of addressing the specific health needs and risks faced by this population. Adolescents and young adults, with a mean age of 21.94 (SD 3.63) years, were featured in 20% (6/30) of the articles, indicating a recognition of the unique challenges faced by individuals in this transitional life stage. Preadolescents, with a mean age of 12.1 (SD 0.63) years, and university students, with a mean age of 20.35 (SD 1.87) years, appeared in 13% (4/30) of the articles, with 2 (7%) articles focusing on each group separately, suggesting a targeted approach to sexual health promotion in these distinct developmental and educational contexts. Although less frequently addressed, women, with a mean age of 47.5 (SD 6.64) years, and lesbian, gay, bisexual, transgender, queer, intersex, and asexual individuals were the focus of 3% (1/30) of the articles, pointing to the need for more inclusive research in these areas. This distribution underscores the diverse yet specific targeting of sexual health promotion efforts across various populations.

Research Dimensions for Sexual Health Promotion Through Pervasive Video Games

We observed that video games were used in 2 main areas and covered 7 topics. The first area was before the exposure, or before acquiring an STI, where video games have been developed as prevention tools. The second area was after the exposure, where video games were developed as treatment tools.

In pre-exposure category (video games as a prevention tool), we identified 6 recurring topics focusing on various aspects of STI prevention. First, we identified topics that aimed to provide general information about the impact of STIs on sexual health and the importance of protective methods, such as correct and consistent condom use and vaccination against hepatitis B (11/30, 37%). Second, we identified topics that are built as sexual health education tools, teaching about the risks of STIs; understanding one's body, gender identity, and sexuality; building and maintaining healthy relationships; and engaging in healthy communication and decision-making about sex (10/30, 33%). Third, topics that focused specifically on promoting healthy behaviors to reduce the risk of HIV or AIDS, including decisions about whether to have sex (eg, delaying sex, abstinence, and reducing the frequency of sex), having long-term mutually monogamous relationships (eg, reducing the number of partners, avoiding concurrent partners, and increasing the time interval between sex with different partners), and consistent and correct condom use (13/30, 43%) were identified. Fourth, topics that are developed to improve understanding and uptake of pre-exposure prophylaxis medication, emphasizing the importance of preventing STIs and early diagnosis and treatment to reduce the risk of HIV transmission (3/30, 10%) were identified. Fifth, we identified topics that aimed to increase awareness about the importance of HPV vaccination and reducing the incidence of STIs, including those caused by HPV (2/30, 7%). Finally, we identified topics that focused on reducing the stigma associated with living with HIV, aiming to foster empathy in people who do not have STIs by highlighting the challenges faced by those who do (2/30, 7%).

In the postexposure category (video games as a treatment tool), we identified ART as a recurring topic, in which topics were aimed at improving adherence to ART, focusing on informing users about the benefits of ART, and including strategies for preventing HIV transmission by suppressing viral replication in people living with HIV (4/30, 13%).

Evolution in Time

We observed a gradual increase in scientific interest in pervasive video games focused on promoting safe sexual behaviors between the years 2015 and 2024, although with fluctuations from 1 year to another. Before 2015, only 1 study was found in this domain, with the subsequent years showing a mode of 2 papers per year. Between 2018 and 2020, significant fluctuation was observed compared to other years of research in this domain. This could suggest changes in academic priorities or the absorption and maturation of ideas introduced by these studies. The stability observed between 2020 and 2023 indicates a major interest and academic production in this domain. These results align with findings in other literature reviews, such as the study by Ndulue and Orji [22], where the trend was similar while studying multiple domains of healthy behaviors.

RQ2: What Methods and Technologies Have Been Used to Design Pervasive Video Games for Sexual Health?

Overview

To our knowledge, there are no classifications or specific attributes to consider in the design, implementation, and evaluation of video games for promoting sexual health. Therefore, we subsequently propose a type of classification to organize the results in this section.

In this scenario, a general perspective on good game design is that it should specify the rules defining the world (game mechanics), user interaction with the environment (game dynamics), and the visual aspects guiding the experience (game aesthetics) [40]. Complementing this point, multiple authors affirmed that the design of the game narrative, or how events are told, directly influences the quality of the experience. In

XSL•FO
RenderX

consequence, designing good storytelling guidelines and integrating them cohesively can promote a sense of personal risk or improve self-efficacy to overcome communication barriers [41-44]. They also emphasized that including game elements, such as badges, scores, or communication tables, in the design of the experience can motivate users to engage in activities they would otherwise be reluctant to perform [45]. In addition, they suggested that the incorporation of these elements in the design of the experience could improve retention and reach, promoting sustained changes over time [46]. In contrast, evaluating the effectiveness of a video game requires time and multiple studies to validate not only the robustness of the implementation but also the usability and acceptability of the experience in the population. Using the previously established definitions as a reference for the effective design, implementation, and evaluation of development, we made the classification presented in Table 1.

Table 1. Phases of game development.

Aspect	Description
Design phase	
World rules	Describes the rules governing the game world
User interaction	Describes how users should interact with the game world
Aesthetics	Describes the esthetic aspects guiding the game experience
Implementation phase	
Device requirements	Specifies the devices for which the experience is intended and its requirements
Input methods	Describes the input methods used and their relationship to the game experience
System testing	Executes the system and tests the video game with an isolated population
Evaluation phase	
Population	Indicates the analyzed population and its characteristics
Dose duration	Indicates the number of sessions and playtime per session
Follow-up	Conducts interviews at different periods
Measurement instruments	Describes the measurement instruments used

Considering the proposed classification listed in Table 1, the articles that were found were categorized to present a structured data format. Table 2 provides a summary of these findings. We observe that 33% (10/30) of the articles provided a detailed exposition of design, implementation, and evaluation within subcategories of the video game. On the other hand, 67% (20/30) of the studies provided partial information in these categories. For the design and its subcategories, 80% (24/30) presented all information (3/3 subcategories), 13% (4/30) partial information (2/3 subcategories), and 7% (2/30) limited information (1/3 subcategories). Then, for the implementation category and its subcategories, 57% (17/30) presented all information (3/3 subcategories), 30% (9/30) partial information (2/3 subcategories), and 10% (3/30) limited information (1/3 subcategories); also, 3% (1/30) did not present any information associated with these subcategories. Next, for the evaluation category and its subcategories, 57% (17/30) presented all information (4/4 subcategories), 37% (11/30) partial information (3/4 subcategories), and 7% (2/30) limited information (2/4 or less subcategories). Finally, it is noteworthy that all studies presented information associated with at least 2 of the 3 main categories, though the emphasis, and how deeply the topic was explored, depended on the main objective of each study. All studies incorporated some form of experimental evaluation. These findings allowed us to recognize that studies in the area of pervasive games for sexual health promotion could provide more detailed descriptions of the design and implementation processes.

XSL•FO
RenderX

Table 2. Classification of articles based on the presence or absence of elements of experimental design, implementation, and evaluation.

Articles	Design			Implementation			Evaluation			
	World rules	User interaction	Aesthetics	Device requirement	Inputs methods	System testing	Population	Dose duration	Follow-up	Measurement instruments
[47-51]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
[43,52,53]	✓	✓	✓	✓	✓[a]	✓	✓	✓	✓	✓
[54]	✓	✓[a]	✓[a]	✓	✓	✓[a]	✓	✓	✓	✓
[55]	✓	✓[a]	✓	✓	✓[a]	✓[a]	✓	✓	✓	✓
[56,57]	✓	✓	✓	✓	✓	✓	✓	✓		✓
[19,58]	✓	✓	✓	✓	✓[a]		✓	✓	✓	✓
[45,59]	✓	✓	✓	✓	✓[a]	✓	✓	✓		✓
[8]	✓	✓		✓	✓	✓	✓	✓		✓
[60]	✓	✓	✓	✓	✓		✓	✓		✓
[46]	✓	✓		✓	✓		✓	✓	✓	✓
[61]	✓	✓	✓	✓			✓	✓	✓	✓
[62,63]	✓	✓	✓	✓	✓[a]		✓	✓		✓
[64]	✓	✓	✓		✓[a]	✓	✓		✓	✓
[65]	✓			✓	✓[a]	✓	✓	✓	✓	✓
[66]	✓	✓		✓[a]	✓[a]	✓	✓	✓		✓
[67]	✓			✓	✓[a]		✓	✓	✓	✓
[44]	✓	✓	✓[a]	✓[a]			✓		✓	✓
[68]	✓	✓	✓	✓	✓[a]					✓
[7]	✓	✓					✓	✓	✓	✓
[69]	✓[a]	✓[a]	✓[a]	✓			✓			✓

[a]These articles referenced these elements in a related article.

We identified 9 different methodological approaches used during the design and implementation stage. The most common approach observed was iterative development (9/30, 30%), which is a cyclical process where developers prototype an idea, conduct laboratory and user testing, and refine the game based on user feedback. The next most frequently used method was user-centered design (3/30, 10%), which is an iterative process that focuses on users and their needs at each phase of the design. This was followed by agile software development (3/30, 10%) and design thinking (2/30, 7%) methodologies, both emphasizing iterative development. Agile uses collaboration among self-organizing, cross-functional teams, and design thinking uses a problem-solving process that prioritizes user needs; it is based on empathy and hands-on iteration. Several methods were used in single instances, such as the participatory serious game development model (1/30, 3%), human-centered design (1/30, 3%), design and creation (1/30, 3%), participatory design (1/30, 3%), and embedded design (1/30, 3%). In addition, 11 video games did not specify the design method (11/30, 37%), focusing instead on presenting the research's methodological approach. This may suggest that the design stage methodologies were considered less critical in the reporting of the development process.

Principal Video Game Genres

We identified and classified the video games according to their genres, listing them in descending order. The most common genre was role-playing games, which featured in 70% (21/30) of the articles. The next most popular genres were action and dating simulation, each representing 10% (3/30) of the articles. The puzzle, minigames, and education genres each appeared in 7% (2/30) of the articles. Finally, the genres casual, simulation, strategy, platformer, deck builder, and party games were each mentioned only once, accounting for 3% (1/30) of the articles in each case. One article, representing 3% (1/30) of the total articles, did not specify the game genre.

Physical Dimensions in Video Games: An Analysis

We identified a trend toward using 2D and 3D visuals as multimedia representations for developers and researchers. The most common dimension was 2D, appearing in 67% (20/30) of the articles. The prevalence of 2D visuals indicated a continued preference for this visual style, likely due to its straightforward graphical requirements. In contrast, 3D visuals were featured

XSL•FO
RenderX

in 33% (10/30) of the articles, highlighting a significant but less frequent emphasis on more complex, immersive environments. 2.5D visual were used in 3% (1/30) of the articles, reflecting a less common but innovative approach that combines 2D and 3D graphics to create a layered visual experience. In addition, 3% (1/30) of the articles did not present the game's physical dimensions.

Devices Used in Interventions

While examining the range of devices used in the interventions reported across various studies, we observed a clear predominance of smartphones. Smartphones were the most frequently used device, appearing in 80% (24/30) of the articles. This widespread use of smartphones highlights their accessibility and convenience as a platform for delivering interventions, given their ubiquity and the ease with which users can engage with content on the go. Tablets were the second most common device, featuring in 30% (9/30) of the articles, often used in conjunction with smartphones to enhance the flexibility of the intervention delivery. Computers appeared in 13% (4/30) of the articles, reflecting a preference for more traditional, larger screens in certain contexts where detailed interaction or extensive data processing might be required. In addition, web-based platforms were mentioned in 23% (7/30) of the articles, indicating a growing trend toward leveraging web environments for broader reach and interactive experiences. Finally, in 7% (2/30) of the articles, the device was not specified.

Analysis of Inputs and Supporting Technologies in Interventions

Analysis of inputs used across the interventions in the reviewed articles reveals a strong emphasis on multimedia and interactive elements designed to enhance user engagement and experience. A dominant input category across studies was touch screens (26/30, 87%), reflecting the versatility and the increasing reliance on touch-based interfaces for intuitive user interactions. Game controllers (game pads and joysticks) were also frequently used (21/30, 70%), highlighting their role in providing precise and interactive gameplay experiences. In addition to these common inputs, internet connection (Wi-Fi, 4G LTE, and 5G) were noted (8/30, 27%), suggesting a significant role in supporting real-time interactions and online connectivity for multiplayer or cloud-based elements. Another prevalent input was audio (7/30, 23%). This widespread use underscores the importance of immersive soundscapes in creating engaging and realistic environments for users. Artificial intelligence for nonplayable character behavior (5/30, 17%), GPS (3/30, 3%), sharing on social networks (2/30; 7%) and chatbots (2/30; 7%) were less frequent, highlighting their specialized use in certain contexts, such as location-based gaming or simulation-heavy environments. Finally, 13% (4/30) of the articles did not detail the specific inputs used.

RQ3: What Theories are Used in the Design and Evaluation of Pervasive Games for Sexual Health?

Identification of Theories in the Design Stage

We were able identify 22 unique theory-based techniques used in the design stage of game development. A compilation of the observed theories and their implementation is provided in Table 3, where a breakdown of each theory, related articles, and the total number of documents using that theory are presented. It is important to note that a video game can encompass multiple theories, implying that a single document can be linked to several theories simultaneously. From the theory-based techniques observed, we noted that 90% (27/30) of the articles indicated the theory used and its relation to the research objective. We found that the social cognitive theory (SCT; 11/30, 37%) is the most recurring theory presented in the literature, followed by social learning theory (6/30, 20%), the health belief model (5/30, 17%), the theory of planned behavior (5/30, 17%), the theory of possible selves (3/30, 10%), entertainment-education literature (3/30, 10%), games for health literature (3/30, 10%) and information-motivation-behavioral skills (3/30, 10%), among others with less presence. In total, 10% (3/30) of the studies did not specify the theory upon which the video game was developed; they only stated that they were effective for behavior modification. In this case, they were categorized as "not specified" for not presenting the theory used or any support to validate its use in promoting safe sexual behaviors.

XSL•FO
RenderX

Table 3. Summary of theories used in the studied video games (N=30).

Theory	References	Games, n (%)
Not specified	[53,57,65]	3 (10)
Social cognitive theory	[44,45,48,49,52,55,58,64,67-69]	11 (37)
Social learning theory	[8,19,46,51,54,63]	6 (20)
Health belief model	[49,52,56,66,68]	5 (17)
Theory of planned behavior	[44,49,61,68,69]	5 (17)
Theory of possible selves	[55,58,67]	3 (10)
Entertainment-education literature	[55,58,67]	3 (10)
Games for health literature	[55,58,67]	3 (10)
Information-motivation-behavioral skills	[19,59,63]	3 (10)
Self-determination theory of motivation	[8,56]	2 (7)
Elaboration likelihood model	[60]	1 (3)
Entertainment overcoming resistance model	[62]	1 (3)
Extended parallel process model	[7]	1 (3)
Behavioral model by Fogg	[50]	1 (3)
Game-based learning and goal setting	[8]	1 (3)
Integrated model of behavior change	[50]	1 (3)
Integrative model of predictive behavior and predicted game influence	[46]	1 (3)
Motivation theory of role modeling	[8]	1 (3)
Framework of learner engagement by Prensky	[8]	1 (3)
Problem-solving theory	[48]	1 (3)
Self-efficacy message framing delay discounting	[46]	1 (3)
Yale 4Ps framework for behavior change	[8]	1 (3)
Proteus effect	[8]	1 (3)

Some examples of articles that presented the attitudinal theory used in the development of the video game and its relationship with the research objective presented are provided in Textbox 4.

Textbox 4. Examples of articles that include attitudinal theory in video game development and research objective.

- Life-simulation game: based on the health belief model theory and social cognitive theory, it seeks to evaluate the effectiveness of the intervention in reducing HIV risk behaviors in an adolescent and young adult population in the United States [52].
- MyPEEPS: based on social learning theory, it aims to test the efficiency of the intervention in reducing sexual risk for HIV and promoting healthy behaviors among young men aged between 13 and 18 years [54].
- Battle Viro: based on the information-motivation-behavioral skills theory and social learning theory, it is an iPhone intervention that seeks to improve adherence to medication and antiretroviral therapy [19].
- ViralCombat: based on the social cognitive theory, theory of possible selves, entertainment-education literature, and the games for health literature, it is a second version of the Battle Viro video game that aims to improve the technical aspects of the original experience to enhance adherence to antiretroviral therapy [63].
- Tumaini: grounded in social cognitive theory and theories of narrative and applied communication, it uses a role-playing gaming experience and minigames to help young adolescents acquire the information, skills, and motivation to avoid and reduce sexual risks. The game also incorporates personal storytelling and a reward system to maintain engagement and link the gameplay to players' real-life goals [55,58,67,68].

Impact of Theory-Based Techniques in the Evaluation Stage

Multiple studies have observed that teenagers and young adults are more receptive to messages related to sexuality when the information is presented in an entertaining way and comes from peers or social networks [54,57,63]. Other authors pointed out that if games are based appropriately and effectively on behavior theories, they can become valuable tools for reducing the spread of STIs [45,65] or promoting adherence to medical treatments in at-risk populations [47,59]. In addition, by building experiences that simulate risky conditions and their potential

XSL•FO
RenderX

consequences, users can observe in a safe environment how STIs could influence their lives. Consequently, these experiences can help influence the determinant factors of behavior [43,68]. In contrast, authors in the field reinforced that the widespread appeal of video games among young people creates a unique opportunity to convey educational sexual messages in the users' free time [7]. This, combined with the fact that video games are crucial for attracting and maintaining attention, makes them key components for effective behavioral change [53]. They also pointed out that this medium is ideal for promoting safe sexual behaviors outside the laboratory context, in a cost-effective and scalable manner, ensuring that the promotion is effective and not localized [19,46,56].

In more traditional contexts, where group dynamics are characterized by being focused on the family group rather than being individualistic, video games emerge as tools that provide key opportunities to extend the effects of intervention to multiple generational levels [53,67]. This interaction facilitates attitude modification, especially in contexts where gaming plays a fundamental role in the incorporation of skills related to the prevention of STIs and effective communication between parents and children [49,58]. At the same time, these experiences reduce resistance to discussing controversial topics due to age gaps, especially in contexts where support and social norms are critical for sustained behavioral change [8,50,55,67].

RQ4: What Methods, Along With Their Benefits, Have Been Used to Evaluate Pervasive Games for Sexual Health?

Research Dimensions to Evaluate Sexual Health Promotion Through Pervasive Video Games

We observed that the research articles used several main methods and combined different approaches. The first and most prevalent method was the randomized controlled trial, which was used in various studies to rigorously test interventions and their effectiveness in promoting sexual health. Within this method, we identified the following frequencies: randomized controlled trial (11/30, 37%) and random clinical trial (1/30, 3%), indicating a strong emphasis on controlled experimental designs in this research area. Mixed methods research was also frequently used (11/30, 37%); this combines qualitative and quantitative approaches to provide a comprehensive understanding of the RQs. This method allows researchers to validate findings across different types of data and address complex issues from multiple perspectives. The third most common method was focus groups (9/30, 30%). This method is used to gather qualitative data, enabling researchers to explore participant experiences, perceptions, and attitudes in depth. Formative research (6/30, 20%) was used to develop and refine interventions by understanding the needs, preferences, and contexts of the target populations before full-scale implementation. Less frequently used methods included correlational studies (2/30, 7%), which examine the relationships between variables without implying causation; grounded theory (2/30, 7%), which guides the development of research based on existing theories; surveys (2/30, 7%), which collect quantitative data on participant behaviors and attitudes; and interviews (2/30, 7%), which provide in-depth qualitative insights. Several methods were used in single instances, including multimix (1/30, 3%), community-based participatory research (1/30, 3%), think aloud (1/30, 3%), descriptive research (1/30, 3%), evaluative research (1/30, 3%), participatory research (1/30, 3%), and preliminary research (1/30, 3%). Finally, 7% (2/30) of the articles did not specify the methods that were used in the evaluation stage.

This diversity of research methods highlighted the multifaceted approach to investigating sexual health promotion, using both quantitative and qualitative techniques to address the complexity of the topic across different study designs.

Period for Measuring the Impact of Sexual Health Interventions in Pervasive Video Games

The research on sexual health promotion through pervasive video games has used various study durations to assess the effectiveness of interventions. Short term follow-ups, from 0 to 3 months, were the most common (14/30, 47%); this allows researchers to capture immediate effects. Medium-term assessments, over 3 to 6 months (4/30, 13%), provide insights into the sustainability of intervention impacts. Long-term follow-ups, conducted over 6 to 9 months (3/30, 10%) and 10 months to ≥1 year (5/30, 17%), allow exploration of the durability of the effects. In some cases, the exact measurement periods were not specified (11/30, 37%), indicating variability in reporting practices. Pre- and postintervention outcomes (8/30, 27%) highlight the importance of understanding changes from baseline to the conclusion of the intervention.

Reported Benefits of Experimental Articles

In this section, only articles containing experimental data were considered; in other words, this analysis excluded articles in "protocol" format for not presenting these data. In this context, the reported benefits presented in this section only consider the 25 articles in research article format. Among the observed studies, all authors reported obtaining positive results in their experiments. We categorized the articles into 2 types: preliminary (those presenting proofs of concept and their results) and significant (those conducting statistically significant studies). Of the 25 articles, we observed that 13 (52%) were classified as preliminary, while 12 (48%) were significant. This indicates a slightly lower focus on studies with statistically significant outcomes. In addition, we found that 7 articles presented quantitative results (7/25, 28%), 10 presented qualitative data (10/25, 40%), and 8 offered mixed methods analyses (8/25, 32%). This balance suggests a robust exploration of both numerical data and in-depth insights, providing a comprehensive view of the research outcomes.

Discussion

Principal Findings

We reviewed 24 years of research on pervasive games aimed at promoting safe sexual behaviors. The reviewed literature suggests a positive relationship in the use of this medium for the prevention and treatment of STIs. Predominantly, research has been done in the United States (15/30, 50%) and Kenya (5/30, 17%), with growing attention across the globe. The main target population for the topic focused primarily on adolescents

XSL•FO
RenderX

(11/30, 37%), men who have sex with men (9/30, 30%), and adolescents and young adults (5/30, 17%), which is in line with the priority populations for STIs from the World Health Organization [35] and Centers for Disease Control and Prevention [36].

The association network between journal categories from the master journal list of Web of Science demonstrates a strong interdisciplinary interrelation. Within the first group (highly interconnected clusters), we can appreciate a strong interdisciplinary interrelation among the 3 main clusters related to occupational health, medical informatics and education, and public health (Figure 1), which is in line with the main purpose of prevention and promotion of healthy behaviors. The second group (isolated clusters) is related mainly to psychology, computer science, and communication.

Although the focus of the first group (highly interconnected clusters) of studies is on health, in general, they are multidisciplinary in nature, reflecting how video game development for health promotion is a multidisciplinary endeavor. Furthermore, they integrate the disciplines that are isolated in the other group. However, there is room for development and research in technical terms (computer science) and psychological theories. The challenge is to generate more unified work between all disciplines to establish strong foundations in the technical, theoretical, and applied fields.

Results have shown that games for promoting safe sexual health have been used both as prevention (26/30, 87%) and treatment tools (4/30, 13%), with several approaches for prevention. In contrast, ART is a recurrent topic in this classification, showing that efforts are primarily focused on its use as a prevention tool.

Digital video games have been observed to be unique in their ability to change attitudes and behaviors in an engaging manner [61,62]. A well-designed serious game based on real-life situations serves not only to observe but also to induce behavioral changes that promote safe sexual behaviors [44].

The increasing accessibility of mobile technology provides new opportunities to implement health interventions targeted at teenagers [58,60]. This is particularly relevant in areas where discussing sex education is challenging [8,50]. The ubiquity of mobile devices allows for widespread and cost-effective prevention and skills interventions [55,67,68], and combining these technologies with persuasive approaches could improve receptivity and reduce resistance to change [7]. In this context, mobile games and web platforms have become the preferred tools for young people for sex education and behavior change [43,46,59]. The information on these platforms has the potential to influence how young people see and act [57]. Moreover, their ability to capture attention has been crucial for effective behavior change [63].

The studies presented in this review demonstrated that pervasive video games can be agents of change. However, there is a lack of work in standardizing and presenting the procedures carried out in experimental design, implementation, and evaluation. Table 2 presents a classification for each development dimension to examine what previous research has emphasized. We believe that the great variety seen in this table is natural because the field that has been developed is interdisciplinary. For example, some studies accurately determined that the use of pervasive games helps promote safe sexual behaviors, but they did not present or allow other researchers to interact with the gaming experience or understand the design and implementation, making it complex to generate new experiences using those studies as a reference, which could be interesting from a technical point of view.

These findings are consistent with other research in a similar field, where it was highlighted that the lack of standardization in mobile health apps could lead to inconsistencies in user experience and the overall effectiveness of health interventions [70]. Similarly, it was shown that the use of metadata in serious games for health not only better organizes and structures information but also increases the effectiveness of interventions [71]. Both studies highlight the importance of standardized frameworks and reinforce the need to formalize procedures in interdisciplinary fields, such as pervasive video games, emphasizing the relevance of implementing clear standards in experimental design, implementation, and evaluation stages.

In this scenario, we found that a significant number of the video games analyzed could be more easily reproduced by providing more details and characteristics about the gaming experience. Some studies could add details about the technology used, such as device specifications and input methods. This could be helped by standardizing reporting practices in a way that encourages including technological aspects rather than only focusing on content. Including technological aspects helps in the task of producing derivative versions that would enable researchers to replicate the experiment in varied environments. Consequently, adding detailed information would contribute to the ability to replicate the game experience effectively, making it easier to compare the results with findings from new studies.

Standardization is key because recognizing the vital components that affect user perception and immersion can significantly impact our comprehension of the factors influencing the experimentation process. Such insights could clarify whether issues in research findings stem from the experimental procedures or the design of the gaming experience itself.

In that sense, the need for consistency in both the design of gaming experiences and the data collection and evaluation processes becomes evident. As emphasized by Kinross [72], by implementing standardized methods, researchers can ensure that variations in user outcomes are attributable to the gaming intervention rather than inconsistencies in experimental design. This approach is essential for improving the reliability of findings in gaming research, where clear and reproducible procedures are crucial for validating the impact of game-based interventions on user perception and immersion.

We observed that 90% (27/30) of the reviewed articles used behavioral or attitudinal theories that allowed researchers to predict the outcome. This is relevant because the research and results are associated with strong theories and research from social sciences in behavioral health and not just the production of the video game.

XSL•FO
RenderX

From the theories that were found, it is notable that research has mainly concentrated on social learning theory or SCT. The most used theory was SCT, which highlights the concept of reciprocal determinism, emphasizing the interaction between individuals and their environments [73]. While many behavioral and social theories concentrate on individual, social, and environmental factors shaping behavior, SCT proposes that human behavior results from the dynamic interplay of personal, behavioral, and environmental influences [74]. Conversely, other theories focus on attitudes and their change and formation through a cognitive process.

In this context, there is a need to continue researching behavioral or attitudinal theories in other video game experiences with different designs, genres, mechanics, and so on to advance the use of mixed theories in encouraging long-term behavior. For evaluating and validating serious games for health, some guidelines are presented in the study by Kato [75], with a focus on experimental design and considerations for reporting results. Furthermore, a critical review of the literature on games for health by Verschueren et al [76] proposed a more specific framework with 5 stages (scientific, design, development, validation, and implementation) and considered all the dimensions and aspects of the development process. By covering scientific, technological, and design aspects, the review serves as an excellent guide for developing serious games based on theory for health.

We observed that the interventions reviewed mostly focused their objectives on the short and medium terms. By taking advantage of various models of behavior change, these interventions focused on reducing risky sexual behaviors, improving HIV prevention, and increasing adherence to treatments, such as pre-exposure prophylaxis and ART. For example, interventions aimed at improving sexual health knowledge, reducing stigma, and promoting health-seeking behaviors among high-risk groups have proven valuable in creating educational and engaging experiences. In contrast, little work has been done to measure the impact of long-term interventions.

These findings aligned with other research, including studies such as the study by Pouls et al [77], which emphasized the need for long-term engagement in interactive eHealth interventions to improve medication adherence over time effectively. Furthermore, as highlighted in the study by Sharifzadeh et al [78], serious games have demonstrated their capacity to influence behaviors among diverse populations, suggesting that their design can be adapted to promote sustained health outcomes. By focusing on long-term objectives, future interventions can take advantage of the engaging qualities of video games to promote continuous health-seeking behaviors while improving the effectiveness of these tools in achieving enduring behavioral changes in sexual health.

This opens space for research to demonstrate not only the capacity of video games as tools for change in the short term but also their effectiveness over time. To exemplify, the Tumaini game has undergone different stages of development and evaluation, ranging from laboratory tests to field tests with other users in various periods [55,58,67,68]. Findings from the Tumaini game consistently show positive outcomes regarding feasibility, acceptability, and effectiveness, highlighting the potential of these game-based approaches to effectively reach populations at risk. This variety of periods underlines the comprehensive approach taken to assess the effectiveness of sexual health interventions, from immediate outcomes to sustained impacts over time.

Finally, the multidisciplinary methodologies used ensure that the games are not only engaging but also grounded in sound theoretical frameworks, making them effective tools for influencing sexual health–related attitudes, intentions, and behaviors. These results reinforce the importance of innovative and accessible platforms in public health, offering a promising avenue for the continued promotion of safe sexual practices among young people. However, the studies in this review already showed promising results, with overall well-functioning technical implementation of the game elements.

Limitations

This work might have missed some relevant articles because we made a distinction between serious games and gamified apps. Given that this study is limited to pervasive video games, several ubiquitous experiences categorized as gamified apps were excluded. This selection criterion could have omitted the analysis of certain video games in cases where the author of the original study categorized the experience as a gamified app rather than a pervasive video game and vice versa. Similarly, when categorizing articles to present a structured data format in Table 2, the definition of search terms may have omitted articles where synonyms were used or elements were described in a different way.

Future Work and Recommendations

With the aim of driving development in the field of pervasive games, we suggest 2 directions for future research. First direction is expanding the studies to other healthy behaviors, such as mental health, tobacco cessation, and nutrition, evaluating their applicability and effectiveness in a broader spectrum. Second direction deals with working on standardizing and documenting the processes of design, implementation, and evaluation of pervasive games to improve reproducibility and facilitate access to these educational resources over time.

We recommend that researchers include a trial version or “sandbox” of their video games in their contributions. This way, other researchers can better understand how the gaming experience affects results. It would also facilitate research sustainability by providing continuous access to these resources without the need to reimplement the experience for future studies. This not only saves time but also adds long-term value to the field of study. In this context, we propose the standardization of these elements, as outlined in Table 1. This table details these aspects, emphasizing the importance of standardizing the process to advance the understanding of developments within this field.

Conclusions

The reviewed articles suggest that pervasive video games are effective tools in promoting safe sexual behaviors, showing

XSL•FO
RenderX

promising results and increasing relevance as educational tools. Their design, implementation, and evaluation are crucial factors (although not all research emphasizes these 3 dimensions), merging entertainment with education and facilitating behavior change and knowledge acquisition, especially among young people.

Pervasive video games incorporate a wide range of techniques that align effectively with theory-based techniques. Although still emerging, the results indicate that this is a growing field with considerable potential that should not be underestimated. These findings suggest that the use of pervasive video games with interactive approaches based on theories can be an effective and attractive methodology to promote education and the adoption of healthy behaviors, representing a research area with a promising future.

Finally, we encourage multidisciplinary research that integrates areas such as psychology, technology, and education to enrich both the design and effectiveness of pervasive games. We believe that the proposed approaches will not only expand knowledge in this field but also optimize the quality and effectiveness of game-based interventions.

Acknowledgments

The authors would like to thank Andrea Olivos for her support. This research was funded by the Agencia Nacional de Investigación y Desarrollo, ANID; Subdirección del Capital Humano; Doctorado Nacional; 2023-21232404.

Data Availability

The data are available in Multimedia Appendices 1-4.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA-ScR (Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews) checklist.
[DOCX File , 109 KB - games_v13i1e58912_app1.docx]

Multimedia Appendix 2
Search strategy for each database.
[DOCX File , 25 KB - games_v13i1e58912_app2.docx]

Multimedia Appendix 3
Results from queries.
[ZIP File (Zip Archive), 481 KB - games_v13i1e58912_app3.zip]

Multimedia Appendix 4
Full data charting.
[XLSX File (Microsoft Excel File), 299 KB - games_v13i1e58912_app4.xlsx]

References

1. New report highlights global progress on reducing HIV, viral hepatitis and sexually transmitted infections and signals need for renewed efforts to reach 2030 targets. World Health Organization (WHO). URL: https://tinyurl.com/3zp2kb8v
2. Global health sector strategies on, respectively, HIV, viral hepatitis and sexually transmitted infections for the period 2022-2030. World Health Organization (WHO). URL: https://www.who.int/teams/global-hiv-hepatitis-and-stis-programmes/strategies/global-health-sector-strategies [accessed 2024-11-21]
3. UNAIDS calls for urgent global action as progress against HIV falters. United Nations. URL: https://news.un.org/en/story/2022/07/1123332 [accessed 2024-11-21]
4. HIV data and statistics. World Health Organization (WHO). URL: https://www.who.int/teams/global-hiv-hepatitis-and-stis-programmes/hiv/strategic-information/hiv-data-and-statistics
5. Sexually transmitted infections (STIs). World Health Organization (WHO). 2024 May 21. URL: https://www.who.int/news-room/fact-sheets/detail/sexually-transmitted-infections-(stis) [accessed 2024-11-21]
6. HIV and AIDS. World Health Organization (WHO). 2024 Jul 22. URL: https://www.who.int/news-room/fact-sheets/detail/hiv-aids

XSL•FO
RenderX

7. Chi AI, Lwin MO, Lee Z, Ng VW, Wong PH. Learning AIDS in Singapore: examining the effectiveness of HIV/AIDS efficacy messages for adolescents using ICTs. Knowl Manag E-Learn 2010;2(2):169-187 [FREE Full text] [doi: 10.34105/j.kmel.2010.02.013]
8. Shankar L, Dixit A, Howard S. A design-led theory of change for a mobile game app (Go Nisha Go) for adolescent girls in India: multimix methodology study. JMIR Form Res 2023 Apr 18;7:1-11 [FREE Full text] [doi: 10.2196/43085] [Medline: 37071463]
9. Kirby DB. The impact of abstinence and comprehensive sex and STD/HIV education programs on adolescent sexual behavior. Sex Res Soc Policy 2008 Sep;5(3):18-27. [doi: 10.1525/srsp.2008.5.3.18]
10. Weiser M. The computer for the 21st century. AMC SIGMOBILE Mob Comput Commun Rev 1999 Jul 01;3(3):3-11 [FREE Full text] [doi: 10.1145/329124.329126]
11. Saha D, Mukherjee A, Bandyopadhyay S. Networking Infrastructure for Pervasive Computing: Enabling Technologies and Systems. Boston, MA: Springer; 2002.
12. Hansmann U, Merk L, Nicklous MS, Stober T. Pervasive Computing Handbook. Berlin, Germany: Springer; 2001.
13. Greenfield A. Everyware: The Dawning Age of Ubiquitous Computing. Washington DC: Pearson; Nov 01, 2006.
14. Shuib L, Shamshirband S, Ismail MH. A review of mobile pervasive learning: applications and issues. Comput Human Behav 2015 May;46:239-244. [doi: 10.1016/j.chb.2015.01.002]
15. Shakya S. Computational enhancements of wearable healthcare devices on pervasive computing system. J Ubiquitous Comput Commun Technol 2020 May 27;2(2):98-108. [doi: 10.36548/jucct.2020.2.005]
16. Dewi DA, Thinakan R, Batumalay M, Kurniawan TB. A model for pervasive computing and wearable devices for sustainable healthcare applications. Int J Adv Comput Sci Appl 2023;14(10):1-4 [FREE Full text] [doi: 10.14569/ijacsa.2023.0141056]
17. Doraiswamy PM, Narayan VA, Manji HK. Mobile and pervasive computing technologies and the future of Alzheimer's clinical trials. NPJ Digit Med 2018 Jan 25;1(1):1 [FREE Full text] [doi: 10.1038/s41746-017-0008-y] [Medline: 31304287]
18. Bardram JE, Matic A. A decade of ubiquitous computing research in mental health. IEEE Pervasive Comput 2020 Jan;19(1):62-72. [doi: 10.1109/mprv.2019.2925338]
19. Whiteley L, Brown L, Lally M, Heck N, van den Berg JJ. A mobile gaming intervention to increase adherence to antiretroviral treatment for youth living with HIV: development guided by the information, motivation, and behavioral skills model. JMIR Mhealth Uhealth 2018 Apr 23;6(4):1-14 [FREE Full text] [doi: 10.2196/mhealth.8155] [Medline: 29685863]
20. Smith AU, Khawly GM, Jann J, Zetina AP, Padilla J, Schnall R. A review of serious gaming as an intervention for HIV prevention. Curr HIV/AIDS Rep 2023 Aug 22;20(4):181-205 [FREE Full text] [doi: 10.1007/s11904-023-00659-9] [Medline: 37213045]
21. Jitmun W, Palee P, Choosri N, Surapunt T. The success of serious games and gamified systems in HIV prevention and care: scoping review. JMIR Serious Games 2023 Sep 05;11:1-16 [FREE Full text] [doi: 10.2196/39915] [Medline: 37669098]
22. Ndulue C, Orji R. Games for change—a comparative systematic review of persuasive strategies in games for behavior change. IEEE Trans Games 2023 Jun;15(2):121-133. [doi: 10.1109/tg.2022.3159090]
23. Sandí Delgado JC, Sanz CV. Juegos serios para potenciar la adquisición de competencias digitales en la formación del profesorado. Rev Educ 2019 Oct 28:34 [FREE Full text] [doi: 10.15517/revedu.v44i1.37228]
24. Drummond D, Hadchouel A, Tesnière A. Serious games for health: three steps forwards. Adv Simul (Lond) 2017 Feb 4;2(1):3 [FREE Full text] [doi: 10.1186/s41077-017-0036-3] [Medline: 29450004]
25. Antle AN, Tanenbaum TJ, Macaranas A, Robinson J. Games for change: looking at models of persuasion through the lens of design. In: Nijholt A, editor. Playful User Interfaces: Interfaces that Invite Social and Physical Interaction. Cham, Switzerland: Springer; 2014:163-184.
26. Fleming TM, Bavin L, Stasiak K, Hermansson-Webb E, Merry SN, Cheek C, et al. Serious games and gamification for mental health: current status and promising directions. Front Psychiatry 2016;7:215 [FREE Full text] [doi: 10.3389/fpsyt.2016.00215] [Medline: 28119636]
27. Ndulue C, Orji R. STD Pong: changing risky sexual behaviour in Africa through persuasive games. In: Proceedings of the 2nd African Conference for Human Computer Interaction on Thriving Communities. 2018 Presented at: AfriCHI '18; December 3-7, 2018; Windhoek, Namibia p. 1-5 URL: https://dl.acm.org/doi/10.1145/3283458.3283463 [doi: 10.1145/3283458.3283463]
28. De Croon R, Wildemeersch D, Wille J, Verbert K, Vanden V. Gamification and serious games in a healthcare informatics context. In: Proceedings of the 2018 IEEE International Conference on Healthcare Informatics. 2018 Presented at: ICHI ,18; June 4-7, 2018; New York, NY p. 53-63 URL: https://ieeexplore.ieee.org/document/8419347 [doi: 10.1109/ichi.2018.00014]
29. Hamari J. Gamification. In: Ritzer G, editor. The Blackwell Encyclopedia of Sociology. Hoboken, NJ: John Wiley & Sons; 2019:1-3.
30. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
31. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]

XSL•FO
RenderX

32. Petersen K, Vakkalanka S, Kuzniarz L. Guidelines for conducting systematic mapping studies in software engineering: an update. Inf Softw Technol 2015 Aug;64:1-18. [doi: 10.1016/j.infsof.2015.03.007]
33. Steel P, Fariborzi H, Hendijani R. An Application of Modern Literature Review Methodology: Finding Needles in Ever-Growing Haystacks. Thousand Oaks, CA: Sage Publications; 2023.
34. Consolidated guidelines on HIV prevention, diagnosis, treatment and care for key populations. World Health Organization (WHO). 2016 Jul 01. URL: https://www.who.int/publications/i/item/9789241511124 [accessed 2024-04-29]
35. Global HIV, Hepatitis and STIs programmes. World Health Organization (WHO). URL: https://www.who.int/teams/global-hiv-hepatitis-and-stis-programmes/populations [accessed 2024-11-21]
36. Screening recommendations and considerations referenced in treatment guidelines and original sources. Center for Disease Control and Prevention (CDC). 2021. URL: https://www.cdc.gov/std/treatment-guidelines/screening-recommendations.htm [accessed 2024-11-21]
37. Hozo SP, Djulbegovic B, Hozo I. Estimating the mean and variance from the median, range, and the size of a sample. BMC Med Res Methodol 2005 Apr 20;5:13 [FREE Full text] [doi: 10.1186/1471-2288-5-13] [Medline: 15840177]
38. Haddaway NR, Page MJ, Pritchard CC, McGuinness LA. PRISMA2020: an R package and shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency and open synthesis. Campbell Syst Rev 2022 Jun 27;18(2):1-12 [FREE Full text] [doi: 10.1002/cl2.1230] [Medline: 36911350]
39. Fruchterman TM, Reingold EM. Graph drawing by force‐directed placement. Softw Pract Exp 2006 Oct 30;21(11):149-160. [doi: 10.1002/spe.4380211102]
40. Hunicke R, LeBlanc M, Zubek R. MDA: a formal approach to game design and game research. In: Proceedings of the 2004 Conference on Association for the Advancement of Artificial Intelligence. 2004 Presented at: AAAI '04; July 25-29, 2004; San Jose, CA p. 1-5 URL: https://users.cs.northwestern.edu/~hunicke/MDA.pdf
41. Heussner T, Finley TK, Hepler JB, Lemay A. The Game Narrative Toolbox. New York, NY: Routledge; 2015.
42. Shen F, Han JA. Effectiveness of entertainment education in communicating health information: a systematic review. Asian J. Commun 2014 Jul 15;24(6):605-616. [doi: 10.1080/01292986.2014.927895]
43. Hemingway C, Baja ES, Dalmacion GV, Medina PM, Guevara EG, Sy TR, et al. Development of a mobile game to influence behavior determinants of HIV service uptake among key populations in the Philippines: user-centered design process. JMIR Serious Games 2019 Dec 20;7(4):e13695 [FREE Full text] [doi: 10.2196/13695] [Medline: 31859673]
44. Wang L, Christensen JL, Smith BJ, Gillig TK, Jeong DC, Liu M, et al. User-agent bond in generalizable environments: long-term risk-reduction via nudged virtual choices. Front Psychol 2021 Aug 26;12:695389 [FREE Full text] [doi: 10.3389/fpsyg.2021.695389] [Medline: 34512452]
45. Andrade-Romo Z, Chavira-Razo L, Buzdugan R, Bertozzi E, Bautista-Arredondo S. Hot, horny and healthy-online intervention to incentivize HIV and sexually transmitted infections (STI) testing among young Mexican MSM: a feasibility study. Mhealth 2020 Jul;6:28 [FREE Full text] [doi: 10.21037/mhealth.2020.03.01] [Medline: 32632366]
46. Fiellin LE, Kyriakides TC, Hieftje KD, Pendergrass TM, Duncan LR, Dziura JD, et al. The design and implementation of a randomized controlled trial of a risk reduction and human immunodeficiency virus prevention videogame intervention in minority adolescents: PlayForward: Elm City stories. Clin Trials 2016 Aug;13(4):400-408 [FREE Full text] [doi: 10.1177/1740774516637871] [Medline: 27013483]
47. Gilliam M, Jagoda P, Jones IB, Rowley J, Hill B. Embedded game design as a method for addressing social determinants of health. Am J Sex Educ 2018 May 25;13(3):378-398 [FREE Full text] [doi: 10.1080/15546128.2018.1445572] [Medline: 34140841]
48. Patchen L, Ellis L, Ma TX, Ott C, Chang KH, Araya B, et al. Engaging African American youth in the development of a serious mobile game for sexual health education: mixed methods study. JMIR Serious Games 2020 Jan 31;8(1):e16254 [FREE Full text] [doi: 10.2196/16254] [Medline: 32012041]
49. Gariepy AM, Hieftje K, Pendergrass T, Miller E, Dziura JD, Fiellin LE. Development and feasibility testing of a videogame intervention to reduce high-risk sexual behavior in black and Hispanic adolescents. Games Health J 2018 Dec;7(6):393-400 [FREE Full text] [doi: 10.1089/g4h.2017.0142] [Medline: 30299980]
50. Saha A, Dixit A, Shankar L, Battala M, Khan N, Saggurti N, et al. Study protocol for an individually randomized control trial for India's first roleplay-based mobile game for reproductive health for adolescent girls. Reprod Health 2023 Sep 14;20(1):140 [FREE Full text] [doi: 10.1186/s12978-023-01665-1] [Medline: 37710347]
51. Schnall R, Kuhns LM, Pearson C, Batey DS, Bruce J, Hidalgo MA, et al. Efficacy of MyPEEPS mobile, an HIV prevention intervention using mobile technology, on reducing sexual risk among same-sex attracted adolescent males: a randomized clinical trial. JAMA Netw Open 2022 Sep 01;5(9):e2231853 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.31853] [Medline: 36129712]
52. Castel AD, Wilbourn B, Trexler C, D'Angelo LD, Greenberg D. A digital gaming intervention to improve HIV testing for adolescents and young adults: protocol for development and a pilot randomized controlled trial. JMIR Res Protoc 2021 Jun 24;10(6):e29792 [FREE Full text] [doi: 10.2196/29792] [Medline: 34185022]
53. Chu SK, Kwan AC, Reynolds R, Mellecker RR, Tam F, Lee G, et al. Promoting sex education among teenagers through an interactive game: reasons for success and implications. Games Health J 2015 Jun;4(3):168-174. [doi: 10.1089/g4h.2014.0059] [Medline: 26182060]

XSL•FO
RenderX

54. Kuhns LM, Garofalo R, Hidalgo M, Hirshfield S, Pearson C, Bruce J, et al. A randomized controlled efficacy trial of an mHealth HIV prevention intervention for sexual minority young men: MyPEEPS mobile study protocol. BMC Public Health 2020 Jan 15;20(1):65 [FREE Full text] [doi: 10.1186/s12889-020-8180-4] [Medline: 31941475]
55. Sabben G, Mudhune V, Ondeng'e K, Odero I, Ndivo R, Akelo V, et al. A smartphone game to prevent HIV among young Africans (Tumaini): assessing intervention and study acceptability among adolescents and their parents in a randomized controlled trial. JMIR Mhealth Uhealth 2019 May 21;7(5):e13049 [FREE Full text] [doi: 10.2196/13049] [Medline: 31115348]
56. Cates JR, Fuemmeler BF, Diehl SJ, Stockton LL, Porter J, Ihekweazu C, et al. Developing a serious videogame for preteens to motivate HPV vaccination decision making: land of secret gardens. Games Health J 2018 Feb;7(1):51-66 [FREE Full text] [doi: 10.1089/g4h.2017.0002] [Medline: 29161529]
57. Ruiz-López T, Sen S, Jakobsen E, Tropé A, Castle PE, Hansen BT, et al. FightHPV: design and evaluation of a mobile game to raise awareness about human papillomavirus and nudge people to take action against cervical cancer. JMIR Serious Games 2019 Apr 08;7(2):e8540 [FREE Full text] [doi: 10.2196/games.8540] [Medline: 30958271]
58. Sabben G, Akelo V, Mudhune V, Ondeng'e K, Ndivo R, Stephenson R, et al. A smartphone game to prevent HIV among young Africans: protocol for a randomized pilot study of a mobile intervention. JMIR Res Protoc 2019 Mar 27;8(3):e11209 [FREE Full text] [doi: 10.2196/11209] [Medline: 30916661]
59. LeGrand S, Muessig KE, McNulty T, Soni K, Knudtson K, Lemann A, et al. Epic allies: development of a gaming app to improve antiretroviral therapy adherence among young HIV-positive men who have sex with men. JMIR Serious Games 2016 May 13;4(1):e6 [FREE Full text] [doi: 10.2196/games.5687] [Medline: 27178752]
60. Rubio C, Gallardo I, Besoain F. Promoting healthy behaviors through an RPG videogame: an elaboration likelihood model based approach. IEEE Trans Games 2024 Mar;16(1):241-246. [doi: 10.1109/tg.2023.3258840]
61. Bayley JE, Brown KE. Translating group programmes into online formats: establishing the acceptability of a parents' sex and relationships communication serious game. BMC Public Health 2015 Dec 09;15(1):1225 [FREE Full text] [doi: 10.1186/s12889-015-2545-0] [Medline: 26651616]
62. Zhang X, Lai E. A web-based gaming approach to decrease HIV-related stigma: game development and mixed methods evaluation. JMIR Serious Games 2022 Dec 15;10(4):e37219 [FREE Full text] [doi: 10.2196/37219] [Medline: 36520508]
63. Whiteley L, Mena L, Craker LK, Healy MG, Brown LK. Creating a theoretically grounded gaming app to increase adherence to pre-exposure prophylaxis: lessons from the development of the viral combat mobile phone game. JMIR Serious Games 2019 Mar 27;7(1):e11861 [FREE Full text] [doi: 10.2196/11861] [Medline: 30916652]
64. Castel AD, Qasmieh S, Greenberg D, Ellenberger N, Howell TH, Griffith C, et al. Digital gaming to improve adherence among adolescents and young adults living with HIV: mixed-methods study to test feasibility and acceptability. JMIR Serious Games 2018 Oct 15;6(4):e10213 [FREE Full text] [doi: 10.2196/10213] [Medline: 30322838]
65. Mudhune V, Sabben G, Ondenge K, Mbeda C, Morales M, Lyles RH, et al. The efficacy of a smartphone game to prevent HIV among young Africans: protocol for a randomized controlled trial in the context of COVID-19. JMIR Res Protoc 2022 Mar 03;11(3):e35117 [FREE Full text] [doi: 10.2196/35117] [Medline: 35030090]
66. Wilbourn B, Howard-Howell T, Castel A, D'Angelo L, Trexler C, Carr R, et al. Barriers and facilitators to HIV testing among adolescents and young adults in Washington, district of Columbia: formative research to inform the development of an mHealth intervention. JMIR Form Res 2022 Mar 11;6(3):e29196 [FREE Full text] [doi: 10.2196/29196] [Medline: 35275083]
67. Winskell K, Sabben G, Ondeng'e K, Odero I, Akelo V, Mudhune V. A smartphone game to prevent HIV among young Kenyans: household dynamics of gameplay in a feasibility study. Health Educ J 2019 Aug 01;78(5):595-606 [FREE Full text] [doi: 10.1177/0017896919832344] [Medline: 34219796]
68. Winskell K, Sabben G, Obong'o C. Interactive narrative in a mobile health behavioral intervention (Tumaini): theoretical grounding and structure of a smartphone game to prevent HIV among young Africans. JMIR Serious Games 2019 May 08;7(2):e13037 [FREE Full text] [doi: 10.2196/13037] [Medline: 31066692]
69. Wang LY, Brown A, Miller LC. Who you are can predict what you say on a virtual date: traits as predictors of communication patterns of young men who have sex with men. Int J Commun 2022:1-23 [FREE Full text]
70. Llorens-Vernet P, Miró J. Standards for mobile health-related apps: systematic review and development of a guide. JMIR Mhealth Uhealth 2020 Mar 03;8(3):1-9 [FREE Full text] [doi: 10.2196/13057] [Medline: 32130169]
71. Göbel S, Maddison R. Serious games for health: the potential of metadata. Games Health J 2017 Feb;6(1):49-56. [doi: 10.1089/g4h.2016.0034] [Medline: 28051899]
72. Kinross JM. Precision gaming for health: computer games as digital medicine. Methods 2018 Dec 01;151:28-33. [doi: 10.1016/j.ymeth.2018.09.009] [Medline: 30273711]
73. Bandura A. Social cognitive theory of mass communication. Media Psychol 2001 Aug;3(3):265-299. [doi: 10.1207/S1532785XMEP0303_03]
74. Glanz K, Lewis FM, Rimer BK. Health Behavior and Health Education: Theory, Research, and Practice. Hoboken, NJ: Jossey-Bass/Wiley; 1990.
75. Kato PM. Evaluating efficacy and validating games for health. Games Health J 2012 Feb;1(1):74. [doi: 10.1089/g4h.2012.1017] [Medline: 26196436]

XSL•FO
RenderX

76. Verschueren S, Buffel C, Vander Stichele G. Developing theory-driven, evidence-based serious games for health: framework based on research community insights. JMIR Serious Games 2019 May 02;7(2):1-16 [FREE Full text] [doi: 10.2196/11565] [Medline: 31045496]
77. Pouls BP, Vriezekolk JE, Bekker CL, Linn AJ, van Onzenoort HA, Vervloet M, et al. Effect of interactive eHealth interventions on improving medication adherence in adults with long-term medication: systematic review. J Med Internet Res 2021 Jan 08;23(1):1-16 [FREE Full text] [doi: 10.2196/18901] [Medline: 33416501]
78. Sharifzadeh N, Kharrazi H, Nazari E, Tabesh H, Edalati Khodabandeh M, Heidari S, et al. Health education serious games targeting health care providers, patients, and public health users: scoping review. JMIR Serious Games 2020 Mar 05;8(1):1-16 [FREE Full text] [doi: 10.2196/13459] [Medline: 32134391]

Abbreviations

ART: antiretroviral therapy
HPV: human papillomavirus
PICO: population, intervention, comparison, and outcomes
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Review
RQ: research question
SCT: social cognitive theory
STI: sexually transmitted infection

Edited by A Coristine, T Leung; submitted 02.04.24; peer-reviewed by C Hemingway, F Xu; comments to author 12.07.24; revised version received 06.09.24; accepted 02.10.24; published 15.01.25.

Please cite as:
Rubio C, Besoain F
Pervasive Games for Sexual Health Promotion: Scoping Literature Review
JMIR Serious Games 2025;13:e58912
URL: https://games.jmir.org/2025/1/e58912
doi:10.2196/58912
PMID:

©Claudio Rubio, Felipe Besoain. Originally published in JMIR Serious Games (https://games.jmir.org), 15.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Effectiveness of Mobile Health–Based Gamification Interventions for Improving Physical Activity in Individuals With Cardiovascular Diseases: Systematic Review and Meta-Analysis of Randomized Controlled Trials

Tianzhuo Yu[1,2], BS; Monica Parry[2], PhD; Tianyue Yu[1], MSc; Linqi Xu[3], PhD; Yuejin Wu[1], MSc; Ting Zeng[1], MSc; Xin Leng[1], MSc; Qian Tong[4], PhD; Feng Li[1], PhD

[1]School of Nursing, Jilin University, Changchun, China

[2]Lawrence Bloomberg Faculty of Nursing, University of Toronto, Toronto, ON, Canada

[3]Faculty of Medicine and Life Sciences, University of Hasselt, Diepenbeek, Belgium

[4]Department of Cardiovascular Medicine, Bethune First Hospital of Jilin University, Changchun, China

Corresponding Author:
Feng Li, PhD
School of Nursing
Jilin University
School of Nursing Office Building, 4th Fl.
965 Xinjiang St
Changchun, 130021
China
Phone: 86 0431 85619377
Email: fli@jlu.edu.cn

Abstract

Background: Gamification refers to using game design elements in nongame contexts. Promoting physical activity (PA) through gamification is a novel and promising avenue for improving lifestyles and mitigating the advancement of cardiovascular diseases (CVDs). However, evidence of its effectiveness remains mixed.

Objective: This systematic review and meta-analysis aimed to evaluate the efficacy of gamification interventions in promoting PA during short-term and follow-up periods in individuals with CVDs and to explore the most effective game design elements.

Methods: A comprehensive search of 7 electronic databases was conducted for randomized controlled trials published in English from January 1, 2010, to February 3, 2024. Eligible studies used mobile health–based gamification interventions to promote PA or reduce sedentary behavior in individuals with CVDs. In total, 2 independent reviewers screened the retrieved records, extracted data, and evaluated the risk of bias using the RoB 2 tool. Discrepancies were resolved by a third reviewer. Meta-analyses were performed using a random-effects model with the Sidik-Jonkman method adjusted by the Knapp-Hartung method. Sensitivity analysis and influence analysis examined the robustness of results, while prediction intervals indicated heterogeneity. A meta-regression using a multimodel inference approach explored the most important game design elements. Statistical analyses were conducted using R (version 4.3.2; R Foundation for Statistical Computing).

Results: In total, 6 randomized controlled trials were included. Meta-analysis of 5 studies revealed a small effect of gamification interventions on short-term PA (after sensitivity analysis: Hedges g=0.32, 95% CI 0.19-0.45, 95% prediction interval [PI] 0.02-0.62). Meta-analysis of 3 studies found the maintenance effect (measured with follow-up averaging 2.5 months after the end of the intervention) was small (Hedges g=0.20, 95% CI 0.12-0.29, 95% PI –0.01 to 0.41). A meta-analysis of 3 studies found participants taking 696.96 more steps per day than the control group (95% CI 327.80 to 1066.12, 95% PI –121.39 to 1515.31). “Feedback” was the most important game design element, followed by “Avatar.”

Conclusions: This meta-analysis demonstrates that gamification interventions effectively promote PA in individuals with CVD, with effects persisting beyond the intervention period, indicating they are not merely novel effects caused by the game nature of gamification. The 95% PI suggests that implementing gamification interventions in similar populations in the future will lead to actual effects in promoting PA in the vast majority of cases. However, the limited number of included studies underscores the urgent need for more high-quality research in this emerging field.

XSL•FO
RenderX

Trial Registration: PROSPERO CRD42024518795; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=518795

(JMIR Serious Games 2025;13:e64410) doi:10.2196/64410

KEYWORDS

cardiovascular diseases; digital health; mobile health; gamification; exercise; physical activity; systematic review; meta-analysis

Introduction

Background

Cardiovascular diseases (CVDs) are a group of disorders affecting the heart and blood vessels that significantly contribute to premature mortality and escalating health care expenses [1,2]. According to the Global Burden of Disease 2021 Study, the global prevalence of CVD cases has reached 612 million [3]. Regular physical activity (PA) is widely recognized as one of the most effective lifestyle interventions for managing CVD risk factors, slowing disease progression, and reducing CVD-related mortality [4]. The World Health Organization guidelines on PA recommend that individuals with CVD engage in 150-300 minutes of moderate-intensity PA, 75-150 minutes of vigorous-intensity PA, or an equivalent combination of moderate-to-vigorous intensity PA per week [5]. Moreover, as self-reported sedentary behavior (SB) is independently associated with an increased risk of CVD, irrespective of PA levels [6,7], the guidelines emphasize minimizing sedentary time and replacing it with any intensity of PA to achieve health benefits [5].

Adherence in the health domain is defined as "the extent to which a person's behavior—taking medication, following a diet, and/or executing lifestyle changes—corresponds with agreed recommendations from a health care provider" [8]. Despite the well-documented benefits of PA, studies indicate that many individuals with CVD demonstrate poor adherence, often failing to meet the PA levels recommended by guidelines [9,10]. Promoting health behavior change and improving PA adherence in individuals with CVD is challenging due to various factors, including lack of motivation or self-efficacy, cognitive or physical limitations, and other barriers [11]. In response, there is a growing call to shift the focus of PA from its health-promoting utility to emphasizing the personal experience [12]. Since emotion plays an essential role in driving behavior, creating opportunities for individuals with CVD to "feel good" while engaging in PA and building a connection between pleasurable feelings and the activity may provide an effective strategy to enhance PA adherence.

Gamification, defined as the use of game design elements in nongame contexts [13], has shown promise in promoting adherence to PA among adults [14], offering a novel approach to modifying health behaviors [15]. Notably, studies have demonstrated that gamification is not only enjoyable for individuals with CVD but also increases the pleasure of engaging in PA [16] and supports maintaining a high level of adherence [17]. In recent years, gamification has been increasingly integrated into mobile health (mHealth), which leverages mobile computing and communication technologies—such as mobile phones and wearable sensors—to deliver medical services and information [18]. With its powerful capabilities for sensing, processing, storing, and displaying data, mHealth enables continuous tracking and collecting of PA-related information, extending gamification into daily health behaviors [15,19]. The combination of gamification and mHealth is mutually reinforcing, providing opportunities to enhance the quality and experience of CVD care [20].

Although gamification has been studied as a strategy to improve PA- and health-related outcomes in individuals with CVD [16,17,21], no systematic review on this specific topic has been published to date. Existing systematic reviews have primarily examined the application of gamification in populations with conditions such as hypertension, excess weight, and diabetes [22,23]. However, these reviews did not exclusively focus on secondary prevention or PA outcomes in individuals with established CVD. Notably, these reviews highlighted the necessity for high-quality studies [23]. They emphasized the need for randomized controlled trials (RCTs) to identify effective and acceptable gamification interventions for the self-management of CVD [22]. Furthermore, evidence regarding the effectiveness of mHealth-based gamification interventions in improving PA among individuals with CVD remains inconclusive. Given the growing body of RCTs in this area, a systematic review and meta-analysis focused on secondary prevention of CVD appears timely and warranted to address these gaps.

Objectives

This study aims to quantify the effects of gamification interventions on PA in individuals with established CVD. Effective gamification outcomes should extend beyond short-term novelty effects measured immediately after the intervention to include lasting impacts assessed at the end of a predefined follow-up period [24]. Therefore, we evaluated the effects of PA both postintervention and at the end of follow-up periods as defined by the included studies. Additionally, prediction intervals (PIs) were reported alongside CIs in the meta-analyses. PIs provide an estimate of the range within which future individual observations will likely fall, offering insights often overlooked [25]. IntHout et al [26] demonstrated that implementing PIs in over 400 published meta-analyses led to completely opposite effects in more than 20% of cases. Last, a fundamental limitation of current gamification research is the inability to determine which game design elements contribute most to its efficacy [16,22]. This highlights the need for further research to identify and isolate the most active and effective gamification elements [27].

This systematic review and meta-analysis aim to address key research gaps by (1) assessing the impact of gamification interventions on PA in individuals with CVD, (2) evaluating the follow-up effects of these interventions, (3) reporting the PIs to estimate the potential intervention effects in future studies,

XSL•FO
RenderX

and (4) determining the most effective game design element for influencing PA behaviors.

Methods

This review follows the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 statement [28]. The methods were preregistered with PROSPERO (International Prospective Register of Systematic Reviews, registration number CRD42024518795, registration on February 29, 2024).

Eligibility Criteria

Eligible studies met the following criteria:

1. Participants: Adults aged 18 years or older with CVD, such as coronary heart disease, peripheral artery disease (PAD), as well as heart attacks and strokes.
2. Interventions: Included studies focused on gamification interventions delivered via mHealth, aiming to improve PA. These interventions included game design elements such as points, levels, challenges, progress bars, leaderboards, rewards, collaboration, social support, and avatars. Gamification and serious games should be clearly distinguished. Therefore, interventions involving virtual reality, active video games, or motion-sensing technologies (eg, Xbox 360, Kinect, Wii) were excluded. mHealth incorporated at least one of the following components: wearable devices, portal websites, smartphone applications, or messaging services.
3. Comparators: When available, control groups were included for between-group comparisons using meta-analysis.
4. Outcomes: Studies assessing changes in PA-related outcomes (eg, steps) or SB were included. These outcomes were required to be continuous data, obtained either through device-based measurements or a 6-minute walking test or subjectively via self-reported questionnaires.
5. Study designs: Only RCTs were eligible.
6. Publication status: Full-text research papers were eligible, while conference proceedings, dissertations, and grey literature were excluded.
7. Language: Only English-language studies were included due to the researchers' language proficiency.

Information Sources

As the term "gamification" gained widespread adoption in 2010 [13], we set this year as the starting point of our search. We systematically searched 7 electronic databases—Ovid MEDLINE, PubMed, Web of Science, Embase, Scopus, Cochrane, and CINAHL—for studies published between January 1, 2010, and February 3, 2024. To ensure comprehensive coverage, we manually screened reference lists and studies included in relevant systematic reviews and meta-analyses for additional eligible studies.

Search Strategy

The search strategy, informed by previous systematic reviews [19], targeted 6 key topics: CVD, gamification, mHealth, PA, SB, and RCTs. It was initially formulated for Ovid MEDLINE and subsequently adapted for use in the other databases.

Selection Process

The search results were exported into EndNote (version 20.6) for document management, and duplicates were removed both automatically and manually. In total, 2 independent reviewers (Tianzhuo Y and Tianyue Y) screened the retrieved records by title, abstract, and full text to identify potentially relevant studies. A third reviewer (FL) resolved any disagreements between the reviewers.

Data Extraction

In total, 2 independent reviewers (Tianzhuo Y and Tianyue Y) extracted and verified the data, with FL arbitrating disagreements. Data were organized into preestablished Microsoft Excel sheets, capturing study characteristics (author, year, and country), participant details (sample size, age, sex, ethnicity, and diagnosis), and outcomes (measurement tools, methods, units, data at different time points). The intervention details were described using the Template for Intervention Description and Replication (TIDieR) checklist [29], covering aspects such as Why (theoretical framework), What (materials, procedures, and game design elements of gamification intervention), Who provided (intervention provider), How (mode of delivery), Where (location), When and How much (duration and frequency), Tailoring (eg, individualized goal setting), Modifications, and How well (adherence and attrition). Game design elements were categorized based on the gamification persuasion architecture and its 7 persuasion strategies [24]. Additionally, the checklist was used to assess the reporting completeness of each intervention, with items rated as "present," "absent," or "unclear."

Risk of Bias

The risk of bias was assessed using the Revised Cochrane risk-of-bias tool for randomized trials (RoB 2), which evaluates 5 aspects: randomization process, deviations from intended interventions, missing outcome data, measurement of the outcomes, and selection of the reported results [30]. In total, 2 reviewers (Tianzhuo Y and Tianyue Y) independently performed the assessments, and any discrepancies were resolved through discussion with a third reviewer (FL).

Effect Measures

Studies reporting PA data as mean (SD) or mean differences with SD of the differences were eligible for inclusion in the meta-analysis. Data reported as medians and interquartile ranges were also included after transformation into means and SD [31].

Synthesis Methods

Characteristics and TIDieR findings were synthesized narratively, with frequencies and percentages summarized in tables. This qualitative review included all studies meeting the eligibility criteria, even if data could not be obtained for quantitative analysis. Statistical analyses were conducted using R (version 4.3.2). An overall meta-analysis was performed to determine the summary effect, with additional meta-analyses for follow-up effects and specific outcomes (eg, daily steps) when sufficient data were available. The standardized mean difference was applied to calculate effect sizes for continuous variables measured with different instruments, while the mean

XSL•FO
RenderX

difference was used for those measured with similar instruments. To address small-sample bias in the included studies, we calculated the effect size Hedges *g* [32], which adjusts Cohen *d* for small-sample bias. A Hedges g of 0.20 indicates a small effect, 0.50 a moderate effect, and 0.80 a large effect [33]. Given the expected between-study heterogeneity, a random-effects model was used to pool effect sizes. To mitigate potential biases associated with the DerSimonian-Laird method, particularly in cases of few studies and high heterogeneity [34], the Sidik-Jonkman method with Knapp-Hartung adjustments was used for more robust variance estimates [35]. According to the Cochrane Handbook [36], when a study included multiple intervention groups and aimed to compare the effects of different interventions, each intervention group was treated as a separate study. The "shared" control group was split into 2 or more groups with reduced sample sizes for comparison. Only the total number of participants in the control group was divided for continuous outcomes, while the mean and SD remained unchanged.

Statistical Heterogeneity

To investigate sources of heterogeneity, we first identified potential outliers, defined as studies with effect sizes that were extreme and significantly deviated from the overall effect [37]. An influence analysis was then performed using the leave-one-out method to determine which studies exerted the greatest impact on the pooled estimate and to assess whether this influence distorted the overall results [38]. Additionally, we used a Baujat plot, a diagnostic tool designed to identify studies that contribute disproportionately to heterogeneity in a meta-analysis [39]. Finally, a sensitivity analysis was conducted to evaluate the robustness of the findings by excluding studies with a high risk of bias, significant heterogeneity, or identified outliers.

The I^2 statistic reflects the proportion of variance in observed effects attributable to variance in true effects but does not indicate the extent to which effect sizes vary across studies. Consequently, categorizing heterogeneity as low, moderate, or high based solely on the I^2 statistic is not recommended [40]. In this meta-analysis, PIs were used to quantify heterogeneity [38]. PIs represent the range within which the effect size of a new study, randomly chosen from the same population as those included in the meta-analysis, is likely to fall [41]. Unlike CI, PIs provide valuable clinical decision-making insights by estimating the likely intervention effect in future studies [25]. PIs also use the same scale as the effect size, illustrating both the interval's width and limits, which helps determine whether the intervention consistently produces beneficial effects or has the potential for harm. PIs can be calculated when a meta-analysis includes at least 3 studies [38] and are most suitable when the included studies have a low risk of bias [36]. Therefore, we calculated 95% PIs for each meta-analysis.

Meta-Regression

To evaluate the most effective game design elements, we performed a meta-regression using a multimodel inference approach. Game design elements were treated as predictors to identify the best combination and the most influential predictor overall [42]. We used the Sidik-Jonkman random-effect model to pool effect sizes and the Knapp-Hartung adjustment method to calculate the test statistic and CI. The small sample-correction Akaike's information criterion (AICc) was applied as the evaluation criterion for the fitted model.

Certainty Assessment

We assessed the quality of the evidence using the web-based version GRADEpro GDT (Grading of Recommendations Assessment, Development and Evaluation professional guideline development tool). Although RCT evidence typically starts at a high-quality rating, we evaluated the following domains to determine any necessary downgrades: (1) risk of bias, (2) inconsistency, (3) indirectness, (4) imprecision, and (5) publication bias [43].

Protocol Deviations

We made the following protocol deviations to facilitate a more comprehensive synthesis of gamification interventions for individuals with CVD. First, the TIDieR checklist was used for narrative synthesis. Second, the Revised Cochrane risk-of-bias tool was used for bias assessment. Third, effect sizes were calculated using Hedges *g*, with CI adjusted via the Sidik-Jonkman method and the Knapp-Hartung adjustment. Fourth, all analyses were conducted using R (version 4.3.2). Last, outlier detection and influential analysis were performed to explore sources of heterogeneity instead of subgroup analyses, and 95% PIs were plotted to measure interstudy heterogeneity in the meta-analysis. No additional protocol deviations from the PROSPERO (International Prospective Register of Systematic Reviews) registration were made.

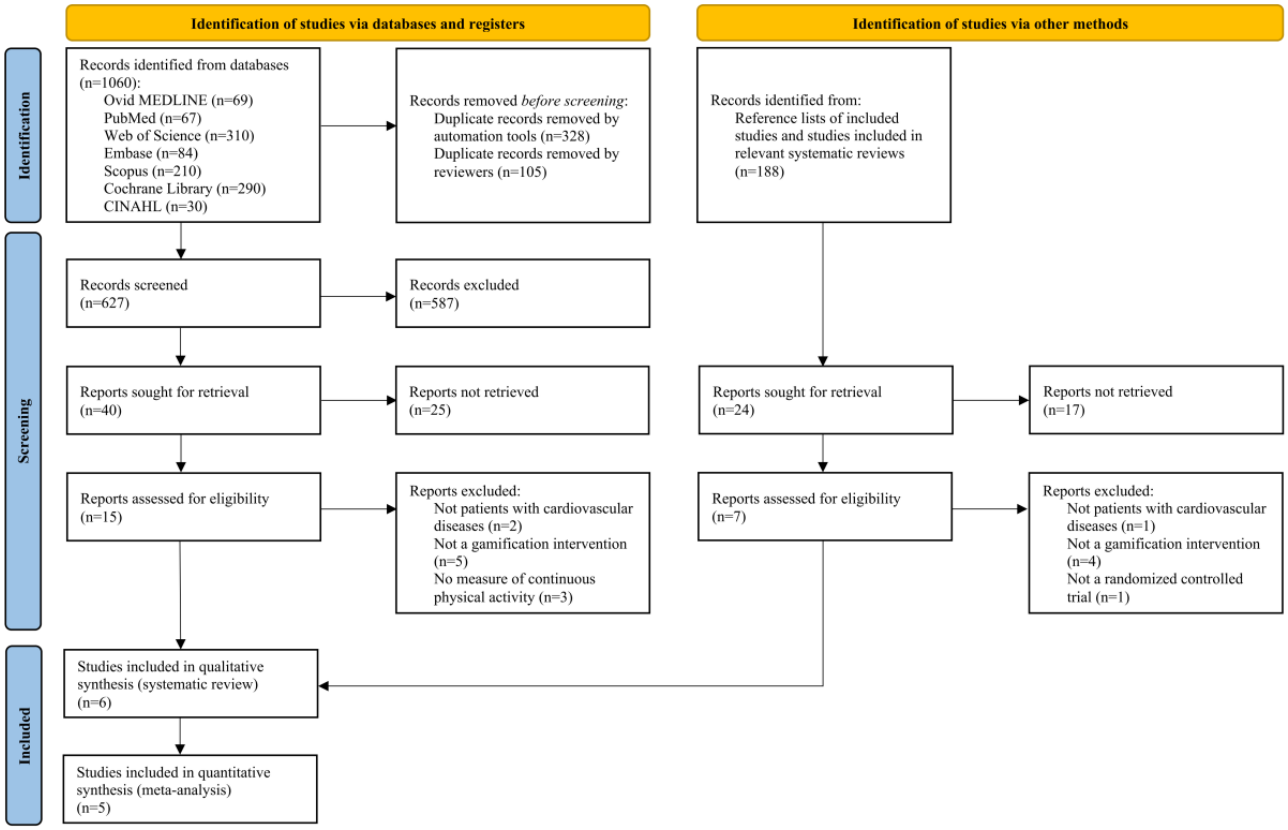
Results

Study Selection

The initial search yielded 1060 records. The detailed search strategies for each database are provided in Table S1 in Multimedia Appendix 1. Additionally, 188 records were identified through manual searches of reference lists and studies included in relevant systematic reviews. After removing duplicates and screening titles and abstracts, 22 full-text papers were assessed for eligibility. In total, 6 studies met the inclusion criteria and were included in the narrative synthesis. Among them, 5 studies were eligible for inclusion in the meta-analysis. The study selection process is illustrated in Figure 1.

XSL•FO
RenderX

Figure 1. Flowchart for study identification, screening, eligibility, and inclusion.



Study Characteristics

Overview

Table S2 in Multimedia Appendix 1 summarizes the characteristics of the included studies. All studies were published within the past 3 years. The majority were conducted in the United States (n=3, 50%), with the remainder conducted in China, Australia, and Germany. A total of 1109 participants were included across the studies, with an average age of 59 years or older and sample sizes ranging from 34 [21] to 500 [44]. Males comprised 55.3% (n=613) of the total sample. Participants were primarily diagnosed with coronary heart disease, heart failure, stroke, atherosclerotic CVD, or PAD.

Intervention Characteristics According to the TIDieR Checklist

A summary of the intervention characteristics, including "Why," "How," "How long," "Tailoring," and "How well," is provided in Table S3 in Multimedia Appendix 1. Details related to "What" are briefly described below and further summarized in Table S2 in Multimedia Appendix 1.

Theoretical Framework (Why)

All but one study [17] used a theoretical framework to guide their interventions, with 2 studies incorporating 2 distinct theories [16,44]. The most frequently applied framework was behavioral economics principles (BEPs) (n=3), followed by self-determination theory (SDT) [16], social cognitive theory (SCT) [27], goal-setting theory (GST) [44], and the Fogg behavioral model (FBM) [45].

Intervention Content (What)

All included studies incorporated game design elements, ranging from 4 [44] to 7 elements [16,45]. Among the 11 identified elements, "Goals" and "Feedback" were the most frequently used (n=5, 83.3%), while "Social support" and "Collaboration" were the least used (n=1, 16.7%). None of the studies applied all 7 gamification persuasion strategies. Detailed descriptions of the gamification interventions are provided in Table S4 in Multimedia Appendix 1.

In the control groups, participants in 3 studies were asked only to use wearable devices for passive PA monitoring without receiving additional interventions [21,44,45]. One study required control group participants to use the same application as the intervention group **but without the active engagement of the game design elements** [16]. In 2 studies, control group participants received standard care only [17,27].

Mode of Delivery (How)

Applications were the most commonly used delivery method, with 4 studies using independently designed and developed applications to deliver gamification interventions [16,17,27,45]. In total, 2 studies used the "Way to Health" research technology platform [21,44]. Additionally, 3 studies provided participants with PA feedback via SMS text messages and emails [21,27,44]. Wearable devices used in the 3 studies not only measured PA but also synchronized the data to applications and websites [21,44,45].

XSL•FO
RenderX

Intervention and Follow-Up Duration (When and How Much)

The intervention periods ranged from 2 months [21] to 6 months [27]. More than half of the interventions (n=4) were considered medium in duration (≥3 months), one was short (<3 months), and one was long (≥6 months). In total, 2 studies included predefined follow-up periods ranging from 2 months [44] to 3 months [16].

Adherence and Attrition (How Well)

Attrition rates varied among the studies: 3 reported low dropout rates (<13%) [16,21,44], 2 had medium rates (13%-26%) [17,45], and 1 reported a high dropout rate (>26%) [27]. Intervention adherence was evaluated through task completion [27] and application usage [17,45].

TIDieR Coding

The completeness of reporting across TIDieR items varied, ranging from 41.7% (n=5) [44] to 83.3% (n=10) [45], with an average of 7 out of 12 items adequately reported. The most consistently reported items were brief name (item 1), rationale or theory (item 2), intervention content (item 4), mode of delivery (item 6), and duration of intervention and follow-up (item 8), all covered in all studies. In total, 4 studies (66.7%) tailored participants' weekly PA goals based on baseline data (item 9) [16,17,21,45]. Only one study reported intervention modifications (item 10) [22]. Similarly, just one study explicitly stated that the intervention was self-delivered by participants (item 5) [27], and another specified that the intervention occurred at the participant's home or convenient locations (item 7) [45]. Other studies' descriptions of both items lack clarity. However, since all interventions involve self-managed PA, participants and their homes or a convenient location were expected to serve as intervention providers and locations in other studies. For intervention materials (item 3), 2 of the 3 studies provided descriptions, but the level of detail was insufficient for replication [17,27]. Adherence or fidelity (item 11 and item 12) was reported in 3 studies (50%) [17,27,45]. A detailed summary of TIDieR items' completeness is presented in Table S5 in Multimedia Appendix 1.

Outcomes Characteristics

In total, 4 studies focused on steps [16,21,44,45], while others measured moderate-to-vigorous PA [44], quantity in metabolic equivalent of task [27], and walking test distance [17]. Except for one study [27], which used the self-reported Global Physical Activity Questionnaire to assess total PA, all other studies used objective measurement tools, including wearable devices (n=3) [21,44,45], smartphone accelerometers [16], or the 6-minute walking test [17]. A detailed summary of PA-related outcomes is presented in Table S6 in Multimedia Appendix 1.

Risk of Bias

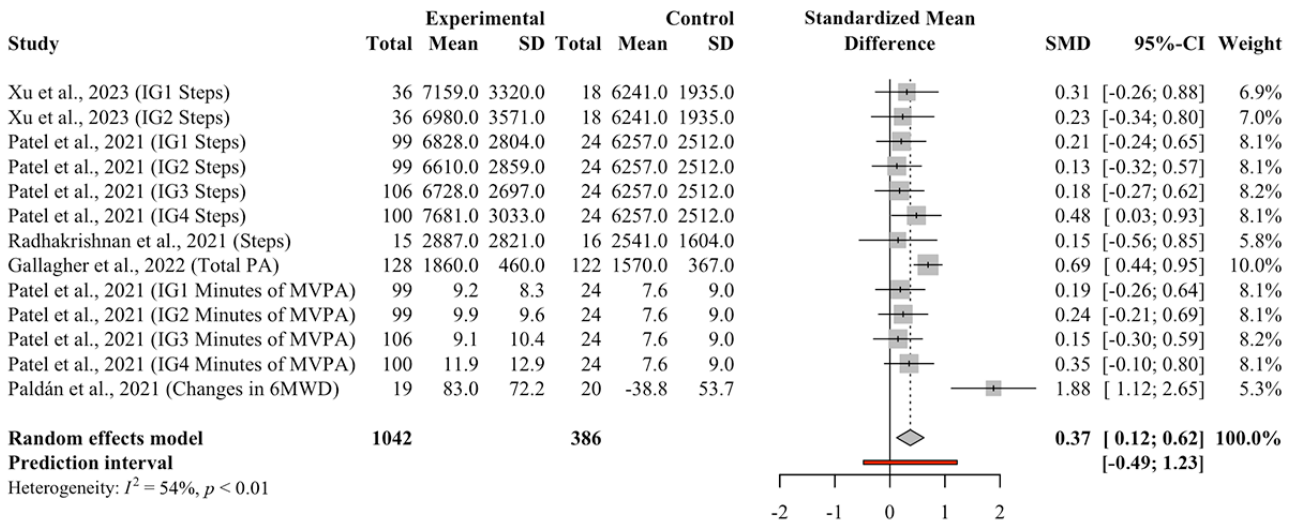
Among the 6 included studies, 5 had a low overall risk of bias, and one raised some concerns [17] due to not using appropriate analyses to estimate intervention effects (see Figure S1 in Multimedia Appendix 1). All studies were at low risk of bias in domains of the randomization process, missing outcome data, outcome measurement, and selection of the reported results. Detailed risk of bias assessments are provided in Table S7 in Multimedia Appendix 1.

Results of Syntheses

Summary Effect

At postintervention, the overall effect size for PA-related outcomes was Hedges *g* of 0.37 (95% CI 0.12-0.62), indicating a statistically significant small-to-moderate effect (Figure 2). Assuming a normal distribution of effects, the PI ranged from –0.49 to 1.23, suggesting that the true effect size for any single population would usually fall within this range. Based on the formula [26], the estimated probability that the true effect of gamification compared to nongamification on PA in individuals with CVD exceeds zero in the new study is 82.1%.

Figure 2. Forest plot for the effect of gamification on PA at postintervention. 6MWD: 6-minute walking distance; IG: intervention group; PA: physical activity; MVPA: moderate-to-vigorous intensity physical activity; SMD: standardized mean difference.

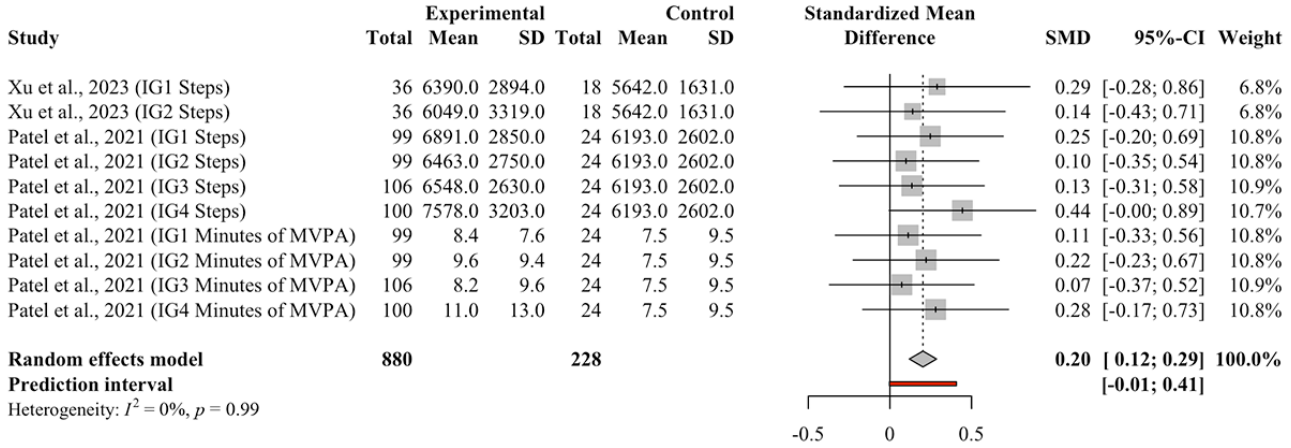


XSL•FO
RenderX

Predefined Follow-Up

At the end of the predefined follow-up, gamification interventions resulted in a statistically significant small increase in PA (Hedges *g*=0.20, 95% CI 0.12-0.29; Figure 3). The 95% PI ranged from –0.01 to 0.41, indicating that in approximately 95% of similar studies, the actual effect size for PA would usually fall within this range. Based on the formula [26], the probability that the true effect of a gamification intervention compared to a nongamification intervention on PA exceeds zero in the new study is estimated at 99.7%.

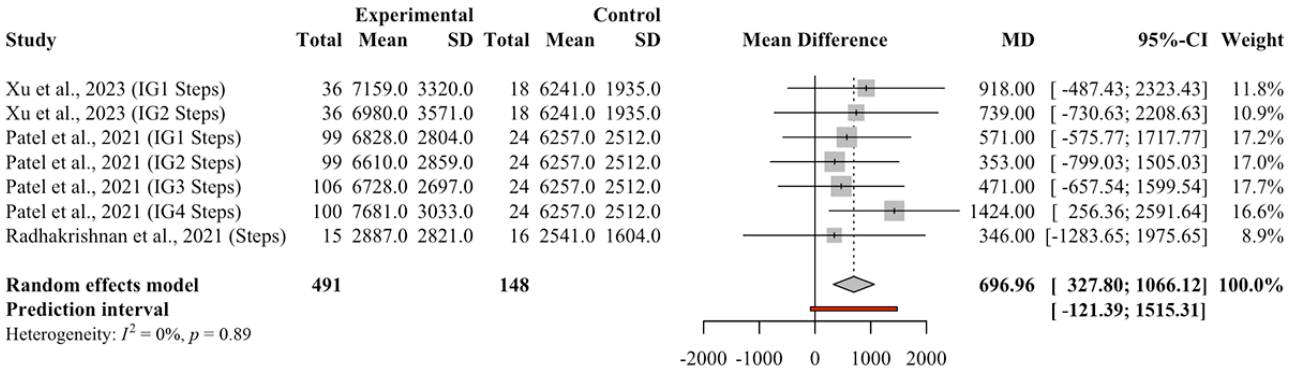
Figure 3. Forest plot for the effect of gamification on PA at the end of the follow-up period. IG: intervention group; PA: physical activity; MVPA: moderate-to-vigorous intensity physical activity; SMD: standardized mean difference.



Daily Steps

A statistically significant increase in daily steps was observed, with a mean difference of 696.96 daily steps (Figure 4). The 95% PI indicated that the change in daily steps for participants using gamification ranged from a decrease of 121.39 steps to an increase of 1515.31 steps. Based on the formula [26], the probability that the true effect of gamification will lead to an increase in daily steps compared to nongamification in the new study is estimated at 80.9%.

Figure 4. Forest plot for the effect of gamification on daily steps. IG: intervention group; MD: mean difference.



Outliers and Influence Analyses

In the short-term PA analysis, substantial statistical heterogeneity was observed. To this, we first identified outliers, with one study by Paldán et al [17] flagged as an outlier. Leave-one-out analyses revealed that sequential removal of each study did not significantly affect the overall effect size, which ranged from a Hedges *g* of 0.32 (95% CI 0.19-0.45) to 0.39 (95% CI 0.12-0.66; see Figure S2 in Multimedia Appendix 1). The Baujat plot also identified the study by Paldán et al [17] as contributing most to the heterogeneity (influence value is 14.61; see Figure S3 in Multimedia Appendix 1). Given its outlier status and concerns about the risk of bias, we excluded this study from the sensitivity analysis. After exclusion, we still obtained a statistically significant effect of a Hedges g of 0.32 (95% CI 0.19-0.45), representing a small effect with minor heterogeneity (95% PI 0.02-0.62; see Figure S4 in Multimedia Appendix 1).

Meta-Regression

Meta-regression using the multimodel inference method identified the top 5 models, ranked by increasing AICc values, with “Feedback plus Goals” having the lowest AICc (4.0) and thus showing the best fit. However, other predictor combinations had similar AICc values, making it difficult to determine a definitive “best” model. Notably, all top 5 models included the predictor “Feedback,” suggesting it may be particularly important. The top 5 models are summarized in Table S8 in Multimedia Appendix 1. As shown in Figure S5 in Multimedia Appendix 1, only 10 of the 11 game design elements were included in the predictor importance plot, as none of the studies in the meta-analysis used the element “Social support.” The

XSL•FO
RenderX

plot illustrates the average importance of each predictor across all models, highlighting "Feedback" as the most important predictor (importance value is 0.71), followed by "Avatars" (importance value is 0.59).

Certainty Assessment

After conducting the sensitivity analysis, the quality of evidence for short-term PA, predefined follow-up PA, and daily steps was rated as high (see Table S9 in Multimedia Appendix 1).

Discussion

Summary Effect

In total, 6 studies were included in this review, all published within the past 3 years, reflecting the growing interest in gamification interventions in the field of CVD. This meta-analysis of 1075 participants from 5 RCTs demonstrated that mHealth-based gamification interventions had a statistically significant effect on overall PA over a mean intervention duration of 3.8 months (Hedges *g*=0.32, 95% CI 0.19-0.45 after sensitivity analysis), indicating that gamification interventions effectively promote PA in individuals with CVD. This effect remained robust even after conducting different influence analyses, ie, the leave-one-out method.

These findings align with those of Mazeas et al [19], who assessed the impact of gamification interventions on PA in a broader population. Their study, which included healthy individuals and those with chronic diseases, demonstrated that a 12-week gamification intervention statistically significantly improved PA levels (Hedges *g*=0.42, 95% CI 0.14-0.69 after sensitivity analysis). However, the mean age of participants in Mazeas et al's study was 35.7 years [19], raising questions about the generalizability of results to older individuals [46]. In contrast, the present meta-analysis focuses on participants with a mean age exceeding 59 years, addressing, to some extent, the research gap regarding the effects of gamification interventions on PA in older individuals. Nonetheless, future research is needed to comprehensively evaluate the impact of gamification interventions on older adults [47].

Follow-Up Effect

When analyzing the predefined follow-up effects of gamification interventions on PA, we observed a statistically significant effect size (Hedges *g*=0.20, 95% CI 0.12-0.29) at a mean follow-up time of 2.5 months after the intervention ended. These findings suggest that the effects of gamification persist beyond the intervention period, although the maintenance effect diminishes over time. Similarly, Mazeas et al [19] reported a statistically significant effect at a longer follow-up period (mean 3.6 months) with a weaker effect size (Hedges *g*=0.15, 95% CI 0.07-0.23 after sensitivity analysis). This consistency between studies indicates that gamification is not merely a short-term novelty effect for increasing PA. However, additional research is warranted to explore the longer-term sustainability of gamification interventions, particularly beyond 6 months.

Effect on Daily Steps

The impact of the gamification intervention on daily steps was similarly promising. The meta-analysis demonstrated a statistically significant effect of mHealth-based gamification interventions, with participants in the intervention group increasing their daily steps by 696.96 compared to the control group (Hedges *g*=696.96, 95% CI 327.80-1066.12). In contrast, Mazeas et al [19] reported a larger increase of 1609.56 steps per day among participants in their gamification group. This discrepancy could be attributed to differences in participant characteristics; this review exclusively included individuals with CVD, who are generally less physically active than those without CVD [48]. Nevertheless, individuals with CVD often derive greater health benefits from the same level of PA [49], which underscores the potential meaning of our findings.

Furthermore, while current guidelines emphasize the intensity of PA [5], Oja et al [50] concluded that even moderate walking intervention yields cardiovascular health benefits. Paluch et al [51] demonstrated that higher daily steps were associated with a progressively lower risk of CVD among older adults. Banach et al [52] found that an increase of just 500 steps per day was linked to a 7% reduction in cardiovascular mortality. Based on the PI, in a new study, the probability that gamification interventions would increase daily steps by at least 500 in individuals with CVD is estimated to be 59.8%. This highlights the potential of gamification interventions to statistically and clinically reduce CVD risk by promoting higher daily steps.

Statistical Heterogeneity

The meta-analyses conducted in this review all demonstrated statistically significant effects, with 95% CI consistently located on the same side of the null. However, due to between-study heterogeneity, the corresponding 95% PI spanned both sides of the null. While most PIs were primarily to the right of the null, indicating that gamification interventions are generally effective, the overlap with the null suggests that their effectiveness may vary in certain contexts. The uncertainty reflected by the PI is tied to how closely future studies align with the characteristics of the completed studies [26]. This indicates that optimizing gamification interventions based on existing studies may help minimize null results in future research.

Theoretical Foundations of Gamification

Overview

A growing body of research has examined how various theoretical foundations can inform the design and effectiveness of gamification interventions [53]. A systematic review of mHealth-based gamification interventions for promoting PA participation found that interventions with theoretical guidance were more effective than those without [54]. Drawing on the theories applied in the included studies, we discuss how different theoretical foundations, each with distinct focuses, can elucidate the underlying mechanisms of gamification. Currently, the theoretical foundations guiding the development, implementation, and evaluation of gamification interventions can be broadly categorized into 3 main groups [55].

Focus on Affect and Motivation

The first category of theoretical foundations emphasizes affect and motivation. SDT is a central representative of this category, evolving over decades into an organic and dialectical

XSL•FO
RenderX

meta-theory of human motivation [56]. As the ubiquitous theoretical framework in gamification, SDT has been extensively applied to guide intervention design [55]. For example, Xu et al used game design elements that support SDT's 3 basic psychological needs—autonomy, competence, and relatedness—to develop a gamification intervention [16]. Satisfying these needs fosters participants' autonomous motivation, enhancing their sustained participation in PA [57].

GST also belongs to this category, primarily serving to refine and optimize gamification interventions [55]. Patel et al [44] integrated gamification interventions with various goal-setting methods grounded in GST. Their findings revealed that behavioral design gamification interventions led to increased PA, with more significant improvements observed when interventions included self-chosen and immediate goals.

Focus on Behavior

The second category of theoretical foundations emphasizes behavior. An example is the FBM, which Radhakrishnan et al used to guide interventions and select digital game elements [45]. FBM is built on 3 core concepts: motivation, ability, and triggers. According to FBM, individuals can perform the target healthy behavior only when they have sufficient motivation, the ability to perform the behavior, and are prompted by appropriate triggers [58].

Focus on Learning

The third category of theoretical foundation emphasizes learning and is primarily derived from social psychology, notably SCT [55]. Within the SCT framework, self-efficacy—an individual's belief in their ability to perform a specific behavior—is regarded as a core determinant of task-oriented behavior [59]. An example of incorporating SCT into gamification is the study by Gallagher et al [27], which used strategies designed to enhance participants' self-efficacy, such as providing incremental challenges, monitoring and tracking activity performance, and a "coin" reward system.

Behavioral Economics Principles

In addition to the 3 main theoretical foundations of gamification, BEP emerged as the most frequently applied theory in the included studies [16,21,44]. Principles such as the "Fresh start effect," "Prospect theory/Loss aversion," and "Goal gradient" were ingeniously integrated with game design elements to establish effective "gamification rules" to drive behavior change. A detailed summary of the application and implications of BEP in the included studies is provided in Table S10 in Multimedia Appendix 1. BEP has gained increasing traction in the primary and secondary prevention of CVD. By enabling more targeted and strategic application of these principles, behavioral economics has the potential to serve as a powerful tool for promoting positive clinical outcomes in individuals with CVD [60].

Game Design Elements

Overview

The most frequently used game design elements in the included studies were "Goals" and "Feedback," followed by "Rewards" and "Progress bars." These findings align with previous reviews. For example, Davis et al [22] summarized the gamification strategies included in health mobile applications for older adults at high risk for CVD, reporting universal use of "Goal setting" (n=7, 100%) and frequent use of "Rewards (virtual and tangible)" and "Track/show progress" (n=6, 85.7%). Similarly, Xu et al [54] reviewed gamification interventions targeting PA participation and identified "Goal Setting" as the most commonly used game design element (n=30, 60%), followed by "Progress bars" (n=26, 52%), "Rewards" (n=25, 50%), and "Feedback" (n=21, 42%). These findings suggest that goal setting, performance feedback, progress visualization, and rewards may be among the most appealing components of gamification interventions.

Using meta-regression multimodel inference, we identified "Feedback" as the most important predictor, followed by "Avatars." While multimodel inference helps to provide a comprehensive overview of predictors influencing effect sizes, it remains an exploratory method. Moreover, the small number of included studies limits the generalizability of these findings. To offer a broader perspective, we discussed all 11 game design elements included in the systematic review, providing insights and references for designing future gamification interventions.

Goals

Setting goals facilitates behavioral change by focusing attention and effort while enhancing perseverance toward achieving specific proficiency levels [61,62]. Most studies in this review implemented "gradual goals" that were updated weekly [16,17,44]. Evidence suggests that adjusting goals weekly or bi-weekly, when necessary, better supports sustained PA behavior change [63]. Combining goal-setting attributes such as feedback, rewards, and task strategies appears to be beneficial in maximizing the effectiveness of interventions on PA behaviors [63]. This validates why goals, feedback, and rewards are among the most frequently used game design elements.

Challenges

Paul et al [64] propose that layered or incremental challenges are among the most effective gamification strategies, particularly when progress can be tracked through wearable devices. For example, the "MyHeartMate" application used incremental challenges but lacked an objective method to track their completion [27]. Challenges can also serve as an effective way to drive other gamification mechanisms [65]. For instance, challenges can be nested within badges, as demonstrated in the "TrackPAD" application, where completing a 7-consecutive-day training challenge unlocks the corresponding badge [17]. Similarly, challenges can be linked to a points system, as seen in the Samsung Health application, where participants earn points for completing challenges and progress to higher levels [66]. Through ongoing challenges, participants may be motivated to continue using an application, especially when these challenges validate their understanding of its goals [65]. Miller et al [67] emphasize the importance of regularly updating challenges to ensure participants have sufficient choices, which helps maintain their motivation to use the application.

XSL•FO
RenderX

Points and Levels

"Points" play an essential place in gamification environments [68]. They serve as a unit of measurement for game scores and can be assigned metaphorical icons or titles, commonly called "Levels," to represent participants' progress [67]. Points are also considered a form of gamification reward [69]. Using points and levels can incentivize participants to consistently engage with mHealth-based gamification interventions for PA and to acquire knowledge about relevant CVD. Additionally, points can serve as a feedback mechanism; for instance, they can stay or be deducted to indicate whether a step goal for the day has been achieved [16,21,44]. By accumulating points and progressing through point-based levels, participants may experience an enhanced sense of competence [70,71].

Feedback

"Feedback" is a crucial component of remotely delivered PA interventions [72], and PA interventions that incorporate performance feedback tend to achieve greater success [73]. Our meta-regression results identified "Feedback" as the most important predictor, reinforcing this perspective. Zuckerman and Gal-Oz [74] developed the "StepByStep" application to encourage participants to walk more. However, participants reported a lack of statistics or graph representations of their daily steps progress and suggested including their personal activity history to facilitate postactivity reflection. This underscores the value of summarized feedback. Moreover, there is a correlation between the amount of PA and the frequency of receiving summary feedback. Research has shown that participants who received daily feedback demonstrated more significant increases in daily steps than those receiving weekly feedback [72].

Rewards

Gamification focuses on motivating participants, often by leveraging extrinsic and intrinsic motivation [75]. While extrinsic motivation may drive short-term behavioral changes, intrinsic motivation is more strongly linked to sustained behavior changes over time [76,77]. Rewards in gamification are thought to stimulate intrinsic motivation by fulfilling the basic psychological needs outlined in SDT [78]. Ideally, rewards should primarily promote autonomy and competence while providing enjoyment and fun without being perceived as overly controlling [79]. However, practical research indicates that not all types of rewards effectively maintain or enhance intrinsic motivation. Lewis et al recommend prioritizing verbal, task-noncontingent, and glory rewards while modifying tangible and task-contingent rewards to reduce the perception of control and foster intrinsic motivation [69]. Future research should focus on identifying specific rewards that evoke feelings of volition, willingness, and enjoyment while avoiding those associated with tension, unwillingness, or coercion [80].

Progress Bars

"Feedback" provides daily or weekly summaries of PA, while "Progress bars" offer real-time feedback. Health gamification operates at the intersection of persuasive technologies, serious games, and personal informatics. Like personal informatics, gamification often centers on tracking individual behaviors [15], aligning with the quantified self-movement concept [81]. Using accelerometers built into wearable devices or mobile phones to track participants' PA and visualize the progress through progress bars represents a straightforward yet quantifiable system that can effectively prompt behavior change [74]. One possible reason for its effectiveness is its ability to support reflection in action by delivering real-time feedback (ie, "Progress bars") and reflection after action by summarizing prior activities (ie, "Feedback"). Both modes of reflection can potentially motivate participants to change their current level of PA [82].

Leaderboards

"Progress bars" monitor individual progress, while "Leaderboards" enable social comparison. Half of the studies in this review used "Leaderboards," highlighting their popularity and effectiveness as a gamification strategy [83]. According to Zichermann and Cunningham [65], applications should offer switchable leaderboards, allowing users to compare rankings within their social network or against a global user base, offering a broader context to measure their progress. Additional columns and filters can be incorporated to provide other relevant elements aligned with the application's design and purpose. For instance, the "TrackPAD" application implemented 4 different types of leaderboards (ie, number of steps in single training sessions, number of completed training sessions, total minutes of PA, and percent increase of PA) [17]. Leaderboards that facilitate social comparison help participants assess their performance relative to others. The dynamic comparison process can elevate individuals' expectations for goal achievement, ultimately leading to increased PA levels [84].

Social Support

Social support is critical in sustaining engagement with health behavior change interventions and enhancing outcomes [85]. Among the included studies, social support was often facilitated through participants' existing social networks, involving significant others such as spouses, friends, or family members. These individuals provided emotional support, positively influenced participants' health behaviors [21]. Research indicates that higher levels of social support, especially from family, are associated with increased PA levels in older adults [86]. However, future studies should prioritize assessing participants' levels of social engagement and exploring strategies to enhance social support for populations at high risk of social isolation, such as older adults [87]. An RCT by Greysen et al [87] investigated the impact of gamification combined with a social support partner intervention on daily steps among hospitalized adults after discharge. The study found that gamification-based social incentives improved mobility only in participants with higher levels of social engagement.

Collaboration and Competition

One study found that adherence rates were 66% higher when participants used a team-based application than exercise alone [85]. Increased adherence rates may be associated with improved health behaviors. However, the only study in this review that incorporated a "Collaborative" game design element reported "poor" results: while there was a statistically significant increase

XSL•FO
RenderX

in mean daily steps in the individual group compared to the control group, no statistically significant difference was observed in the team group [16]. This outcome may be attributed to negative participants reducing their team's overall motivation and the benefits of social bonding being more fully realized when team members are equally engaged [22]. Additionally, Xu et al [16] suggested that collaboration may not be effective without established social relationships among participants. Supporting this view, a previous RCT demonstrated that gamification collaboration statistically significantly increased daily steps among household members [88], highlighting the potential effectiveness of behavior change programs when participants are engaged together and socially connected [89]. However, collaborating with family members may not always be effective for participants with specific diseases, such as CVD, particularly if family members lack equal motivation to engage in health behavior change. Future research should focus on identifying best practices for organizing effective collaborative teams, considering factors such as team size, member relationships, and incentives to balance peer support, team responsibility reinforcement, and individual achievement rewards.

Patel et al [90] introduced a novel perspective for interventions involving participants without preexisting social relationships. In a 24-week PA intervention with overweight or obese adults unfamiliar with each other, they found that the gamification elements supporting collaboration and competition all statistically significantly increased PA compared to the control group, with competition showing the strongest effect. This suggests that competition may be more effective than collaboration in the absence of prior social relationships. Competing with teammates or other teams stimulates a powerful innate human drive [91]. However, the game design element "Competition" was not used in the included studies, likely due to concerns that excessive PA could be counterproductive for individuals with CVD [16]. Future studies could explore competition among individuals with CVD based on achieving daily PA goals rather than total activity levels. This approach may reduce the risk of overexercising while using the game design element of competition safely and effectively.

Avatars

Researchers now widely agree on the central theoretical architecture of gamification, emphasizing intrinsic and extrinsic motivation as defined by SDT [92]. Avatars provide significant visual feedback with endogenous value. Changes in avatars are linked to participants' PA behaviors. This connection fosters a sense of participation and engagement, which can sustain and enhance motivation and willingness to engage in PA [93]. In our meta-regression, "Avatar" emerged as the second most important predictor, confirming this connection. According to SDT, sociocontextual events (eg, feedback, communication, or rewards) that generate a sense of competence during actions can enhance intrinsic motivation. Several applications demonstrate this principle. For instance, the "MyHeartMate" application uses a cartoon heart avatar to virtually represent the health status of individuals with coronary artery disease [27]. The "Heart Health Mountain" application for individuals with heart failure incorporates communication through avatars [45]. The "STARFISH" application for individuals with stroke translates participants' active PA into rewards that customize their colorful fish avatars in a virtual fish tank [64]. Future research should focus on designing attractive avatars tailored to specific populations and crafting interactions with avatars to enhance intrinsic motivation for behavior and subsequently improve PA.

Limitations

This review has 4 main limitations. First, the small number of included studies, some of which were feasibility or pilot studies, resulted in small sample sizes and underpowered analyses. However, this limitation reflects the emerging momentum of gamification interventions as secondary prevention for CVD. Second, all studies incorporated multiple gamification elements, making it challenging to isolate the effects of individual elements. While we conducted meta-regression multimodel inference to explore this, the results remain exploratory. Further research focusing on the independent effects of specific gamification elements is needed. Third, although we had an interest in exploring whether gamification can reduce SB, none of the included studies evaluated that outcome. Designing gamification interventions targeting SB may be a valuable direction for future research. Finally, none of the studies conducted cost-effectiveness analyses, which are crucial for evaluating the public health impact of gamification interventions and guiding resource allocation. Incorporating cost-effectiveness analyses in future studies could provide evidence to support the adoption of more efficient interventions and replace fewer effective ones [94].

Conclusions

In conclusion, gamification interventions show promise in promoting PA, particularly in increasing daily steps among individuals with CVD. While the effects may diminish over time, their persistence during follow-up suggests that gamification is not merely a novelty effect. However, attributing effects to individual game design elements remains challenging, as no studies have independently tested their impacts. Future studies should include larger sample sizes, longer durations, and more rigorous designs to further explore the effectiveness and persistence of gamification interventions and the impact of individual game design elements.

Acknowledgments

This work was supported by the National Key Research and Development Plan Project Sub-topic of China (grant number 2022YFC3601305) and the China Scholarship Council scholarship program (grant number 202306170197).

XSL•FO
RenderX

Data Availability

All data generated or analyzed during this study are included in the supplementary files accompanying this published article.

Authors' Contributions

Tianzhuo Y was involved in conceptualization, methodology, software, formal analysis, investigation, data curation, writing – original draft, writing – review and editing, and visualization. MP was involved in conceptualization, formal analysis, resources, writing – review and editing, and supervision. Tianyue Y was involved in validation, formal analysis, investigation, and writing – review and editing. LX was involved in formal analysis and writing – review and editing. YW was involved in writing – review and editing. TZ was involved in writing – review and editing. XL was involved in writing – review and editing. QT was involved in writing – review and editing, project administration, and funding acquisition. FL was involved in conceptualization, methodology, validation, writing – review and editing, supervision, project administration, and funding acquisition.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Related multimedia appendix tables and figures.
[PDF File (Adobe PDF File), 948 KB - games_v13i1e64410_app1.pdf]

Multimedia Appendix 2
PRISMA 2020 Checklist.
[DOCX File , 32 KB - games_v13i1e64410_app2.docx]

References

1. GBD 2019 Diseases and Injuries Collaborators. Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global burden of disease study 2019. Lancet 2020;396(10258):1204-1222 [FREE Full text] [doi: 10.1016/S0140-6736(20)30925-9] [Medline: 33069326]
2. Mensah GA, Roth GA, Fuster V. The global burden of cardiovascular diseases and risk factors: 2020 and beyond. J Am Coll Cardiol 2019;74(20):2529-2532 [FREE Full text] [doi: 10.1016/j.jacc.2019.10.009] [Medline: 31727292]
3. Institute for Health Metrics and Evaluation (IHME). GBD 2021 Cause and Risk Summary: Cardiovascular Diseases-Level 2 Cause. Seattle, USA: IHME, University of Washington URL: https://www.healthdata.org/research-analysis/diseases-injuries-risks/factsheets/2021-cardiovascular-diseases-level-2-disease [accessed 2024-10-14]
4. Piercy KL, Troiano RP. Physical activity guidelines for Americans from the US Department of Health and Human Services. Circ Cardiovasc Qual Outcomes 2018;11(11):e005263 [FREE Full text] [doi: 10.1161/CIRCOUTCOMES.118.005263] [Medline: 30571339]
5. Bull FC, Al-Ansari SS, Biddle S, Borodulin K, Buman MP, Cardon G, et al. World Health Organization 2020 guidelines on physical activity and sedentary behaviour. Br J Sports Med 2020;54(24):1451-1462 [FREE Full text] [doi: 10.1136/bjsports-2020-102955] [Medline: 33239350]
6. Shiroma EJ, Lee IM. Physical activity and cardiovascular health: lessons learned from epidemiological studies across age, gender, and race/ethnicity. Circulation 2010;122(7):743-752. [doi: 10.1161/CIRCULATIONAHA.109.914721] [Medline: 20713909]
7. Ekelund U, Tarp J, Steene-Johannessen J, Hansen BH, Jefferis B, Fagerland MW, et al. Dose-response associations between accelerometry measured physical activity and sedentary time and all cause mortality: systematic review and harmonised meta-analysis. Br Med J 2019;366:l4570 [FREE Full text] [doi: 10.1136/bmj.l4570] [Medline: 31434697]
8. World Health Organization. Adherence to Long-Term Therapies: Evidence for Action. Geneva, Switzerland: World Health Organization; 2003.
9. Dibben GO, Dalal HM, Taylor RS, Doherty P, Tang LH, Hillsdon M. Cardiac rehabilitation and physical activity: systematic review and meta-analysis. Heart 2018;104(17):1394-1402 [FREE Full text] [doi: 10.1136/heartjnl-2017-312832] [Medline: 29654095]
10. Kronish IM, Diaz KM, Goldsmith J, Moise N, Schwartz JE. Objectively measured adherence to physical activity guidelines after acute coronary syndrome. J Am Coll Cardiol 2017;69(9):1205-1207 [FREE Full text] [doi: 10.1016/j.jacc.2016.10.087] [Medline: 28254185]
11. Shrank WH, Choudhry NK. Therapy. Affect and affirmations--a 'basic' approach to promote adherence. Nat Rev Cardiol 2012;9(5):263-265. [doi: 10.1038/nrcardio.2012.35] [Medline: 22411288]
12. Barreto Pde S. Why are we failing to promote physical activity globally? Bull World Health Organ 2013;91(6):390 [FREE Full text] [doi: 10.2471/BLT.13.120790] [Medline: 24052671]

XSL•FO
RenderX

13. Deterding S, Dixon D, Khaled R, Nacke L. From game design elements to gamefulness: defining 'gamification'. 2011 Presented at: Proceedings of 15th International Academic MindTrek Conference: Envisioning Future Media Environments; September 30, 2011; New York. [doi: 10.1145/2181037.2181040]
14. Gremaud AL, Carr LJ, Simmering JE, Evans NJ, Cremer JF, Segre AM, et al. Gamifying accelerometer use increases physical activity levels of sedentary office workers. J Am Heart Assoc 2018;7(13):e007735 [FREE Full text] [doi: 10.1161/JAHA.117.007735] [Medline: 29967221]
15. Johnson D, Deterding S, Kuhn KA, Staneva A, Stoyanov S, Hides L. Gamification for health and wellbeing: a systematic review of the literature. Internet Interv 2016;6:89-106 [FREE Full text] [doi: 10.1016/j.invent.2016.10.002] [Medline: 30135818]
16. Xu L, Tong Q, Zhang X, Yu T, Lian X, Yu T, et al. Smartphone-based gamification intervention to increase physical activity participation among patients with coronary heart disease: a randomized controlled trial. J Telemed Telecare 2024;30(9):1425-1436. [doi: 10.1177/1357633X221150943] [Medline: 36794484]
17. Paldán K, Steinmetz M, Simanovski J, Rammos C, Ullrich G, Jánosi RA, et al. Supervised exercise therapy using mobile health technology in patients with peripheral arterial disease: pilot randomized controlled trial. JMIR mHealth uHealth 2021;9(8):e24214 [FREE Full text] [doi: 10.2196/24214] [Medline: 34398800]
18. Burke LE, Ma J, Azar KM, Bennett GG, Peterson ED, Zheng Y, American Heart Association Publications Committee of the Council on Epidemiology and Prevention, Behavior Change Committee of the Council on Cardiometabolic Health, Council on Cardiovascular and Stroke Nursing, Council on Functional Genomics and Translational Biology, Council on Quality of Care and Outcomes Research, Stroke Council. Current science on consumer use of mobile health for cardiovascular disease prevention: a scientific statement from the American Heart Association. Circulation 2015;132(12):1157-1213 [FREE Full text] [doi: 10.1161/CIR.0000000000000232] [Medline: 26271892]
19. Mazeas A, Duclos M, Pereira B, Chalabaev A. Evaluating the effectiveness of gamification on physical activity: systematic review and meta-analysis of randomized controlled trials. J Med Internet Res 2022;24(1):e26779 [FREE Full text] [doi: 10.2196/26779] [Medline: 34982715]
20. Berglund A, Jaarsma T, Berglund E, Strömberg A, Klompstra L. Understanding and assessing gamification in digital healthcare interventions for patients with cardiovascular disease. Eur J Cardiovasc Nurs 2022;21(6):630-638 [FREE Full text] [doi: 10.1093/eurjcn/zvac048] [Medline: 35709297]
21. Waddell KJ, Patel MS, Clark K, Harrington TO, Greysen SR. Effect of gamification with social incentives on daily steps after stroke: a randomized clinical trial. JAMA Neurol 2022;79(5):528-530 [FREE Full text] [doi: 10.1001/jamaneurol.2022.0231] [Medline: 35344027]
22. Davis AJ, Parker HM, Gallagher R. Gamified applications for secondary prevention in patients with high cardiovascular disease risk: a systematic review of effectiveness and acceptability. J Clin Nurs 2021;30(19-20):3001-3010. [doi: 10.1111/jocn.15808] [Medline: 33872436]
23. Mitra S, Kroeger CM, Wang T, Masedunskas A, Cassidy SA, Huang R, et al. Gamified smartphone-app interventions on behaviour and metabolic profile in patients at risk of cardiovascular disease. Stud Health Technol Inform 2024;310:1542-1543. [doi: 10.3233/SHTI231284] [Medline: 38269736]
24. Cugelman B. Gamification: what it is and why it matters to digital health behavior change developers. JMIR Serious Games 2013;1(1):e3 [FREE Full text] [doi: 10.2196/games.3139] [Medline: 25658754]
25. Christaina E, Trypsianis G, Lazarides MK. Prediction interval in meta-analyses. Int J Low Extrem Wounds 2023. [doi: 10.1177/15347346231185615] [Medline: 37545295]
26. IntHout J, Ioannidis JP, Rovers MM, Goeman JJ. Plea for routinely presenting prediction intervals in meta-analysis. BMJ Open 2016;6(7):e010247 [FREE Full text] [doi: 10.1136/bmjopen-2015-010247] [Medline: 27406637]
27. Gallagher R, Chow CK, Parker H, Neubeck L, Celermajer DS, Redfern J, et al. The effect of a game-based mobile app 'MyHeartMate' to promote lifestyle change in coronary disease patients: a randomized controlled trial. Eur Heart J Digit Health 2023;4(1):33-42 [FREE Full text] [doi: 10.1093/ehjdh/ztac069] [Medline: 36743873]
28. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Br Med J 2021;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
29. Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. Br Med J 2014;348:g1687 [FREE Full text] [doi: 10.1136/bmj.g1687] [Medline: 24609605]
30. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. Br Med J 2019;366:l4898 [FREE Full text] [doi: 10.1136/bmj.l4898] [Medline: 31462531]
31. Hozo SP, Djulbegovic B, Hozo I. Estimating the mean and variance from the median, range, and the size of a sample. BMC Med Res Methodol 2005;5:13 [FREE Full text] [doi: 10.1186/1471-2288-5-13] [Medline: 15840177]
32. Hedges LV. Distribution theory for Glass's estimator of effect size and related estimators. J Educ Stat 1981;6(2):107-128. [doi: 10.3102/10769986006002107]
33. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd ed. New York, NY: Routledge; 2013.

XSL•FO
RenderX

34. Hartung J. An alternative method for meta-analysis. Biom J 1999;41(8):901-916. [doi: 10.1002/(sici)1521-4036(199912)41:8<901::aid-bimj901>3.3.co;2-n]
35. IntHout J, Ioannidis JP, Borm GF. The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method. BMC Med Res Methodol 2014;14:25 [FREE Full text] [doi: 10.1186/1471-2288-14-25] [Medline: 24548571]
36. Higgins JPT, Thomas J, Chandler J, Li T, Page MJ, Welch VA. Cochrane Handbook for Systematic Reviews of Interventions. 2nd ed. Chichester, UK: John Wiley & Sons; 2019.
37. Viechtbauer W, Cheung MW. Outlier and influence diagnostics for meta-analysis. Res Synth Methods 2010;1(2):112-125. [doi: 10.1002/jrsm.11] [Medline: 26061377]
38. Riley RD, Higgins JPT, Deeks JJ. Interpretation of random effects meta-analyses. Br Med J 2011;342:d549. [doi: 10.1136/bmj.d549] [Medline: 21310794]
39. Baujat B, Mahé C, Pignon JP, Hill C. A graphical method for exploring heterogeneity in meta-analyses: application to a meta-analysis of 65 trials. Stat Med 2002;21(18):2641-2652. [doi: 10.1002/sim.1221] [Medline: 12228882]
40. Borenstein M. Research Note: in a meta-analysis, the I index does not tell us how much the effect size varies across studies. J Physiother 2020;66(2):135-139 [FREE Full text] [doi: 10.1016/j.jphys.2020.02.011] [Medline: 32307309]
41. Ades AE, Lu G, Higgins JPT. The interpretation of random-effects meta-analysis in decision models. Med Decis Making 2005;25(6):646-654. [doi: 10.1177/0272989X05282643] [Medline: 16282215]
42. Harrer M, Cuijpers P, Furukawa TA, Ebert DD. Doing Meta-Analysis With R: A Hands-On Guide. Boca Raton, FL: Chapman & Hall/CRC Press; 2021.
43. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, GRADE Working Group. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. Br Med J 2008;336(7650):924-926 [FREE Full text] [doi: 10.1136/bmj.39489.470347.AD] [Medline: 18436948]
44. Patel MS, Bachireddy C, Small DS, Harrison JD, Harrington TO, Oon AL, et al. Effect of goal-setting approaches within a gamification intervention to increase physical activity among economically disadvantaged adults at elevated risk for major adverse cardiovascular events: the ENGAGE randomized clinical trial. JAMA Cardiol 2021;6(12):1387-1396 [FREE Full text] [doi: 10.1001/jamacardio.2021.3176] [Medline: 34468691]
45. Radhakrishnan K, Julien C, Baranowski T, O'Hair M, Lee G, Sagna De Main A, et al. Feasibility of a sensor-controlled digital game for heart failure self-management: randomized controlled trial. JMIR Serious Games 2021;9(4):e29044 [FREE Full text] [doi: 10.2196/29044] [Medline: 34747701]
46. Hung CH, Kao YS. Learning more about the effects of gamification on physical activity. Comment on 'Evaluating the Effectiveness of Gamification on Physical Activity: Systematic Review and Meta-analysis of Randomized Controlled Trials'. J Med Internet Res 2022;24(5):e36396 [FREE Full text] [doi: 10.2196/36396] [Medline: 35503654]
47. Mazeas A, Duclos M, Pereira B, Chalabaev A. Authors' reply to: learning more about the effects of gamification on physical activity. Comment on 'Evaluating the Effectiveness of Gamification on Physical Activity: Systematic Review and Meta-analysis of Randomized Controlled Trials'. J Med Internet Res 2022;24(5):e38212 [FREE Full text] [doi: 10.2196/38212] [Medline: 35503414]
48. Vasankari V, Husu P, Vähä-Ypyä H, Suni JH, Tokola K, Borodulin K, et al. Subjects with cardiovascular disease or high disease risk are more sedentary and less active than their healthy peers. BMJ Open Sport Exerc Med 2018;4(1):e000363 [FREE Full text] [doi: 10.1136/bmjsem-2018-000363] [Medline: 29765701]
49. Jeong SW, Kim SH, Kang SH, Kim HJ, Yoon CH, Youn TJ, et al. Mortality reduction with physical activity in patients with and without cardiovascular disease. Eur Heart J 2019;40(43):3547-3555 [FREE Full text] [doi: 10.1093/eurheartj/ehz564] [Medline: 31504416]
50. Oja P, Kelly P, Murtagh EM, Murphy MH, Foster C, Titze S. Effects of frequency, intensity, duration and volume of walking interventions on CVD risk factors: a systematic review and meta-regression analysis of randomised controlled trials among inactive healthy adults. Br J Sports Med 2018;52(12):769-775. [doi: 10.1136/bjsports-2017-098558] [Medline: 29858464]
51. Paluch AE, Bajpai S, Ballin M, Bassett DR, Buford TW, Carnethon MR, Steps for Health Collaborative. Prospective association of daily steps with cardiovascular disease: a harmonized meta-analysis. Circulation 2023;147(2):122-131 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.122.061288] [Medline: 36537288]
52. Banach M, Lewek J, Surma S, Penson PE, Sahebkar A, Martin SS, et al. The association between daily step count and all-cause and cardiovascular mortality: a meta-analysis. Eur J Prev Cardiol 2023;30(18):1975-1985. [doi: 10.1093/eurjpc/zwad229] [Medline: 37555441]
53. Nacke LE, Deterding S. The maturing of gamification research. Comput Hum Behav 2017;71:450-454. [doi: 10.1016/j.chb.2016.11.062]
54. Xu L, Shi H, Shen M, Ni Y, Zhang X, Pang Y, et al. The effects of mHealth-based gamification interventions on participation in physical activity: systematic review. JMIR Mhealth Uhealth 2022;10(2):e27794 [FREE Full text] [doi: 10.2196/27794] [Medline: 35113034]

XSL•FO
RenderX

55. Krath J, Schürmann L, von Korflesch HFO. Revealing the theoretical basis of gamification: a systematic review and analysis of theory in research on gamification, serious games and game-based learning. Comput Hum Behav 2021;125:106963. [doi: 10.1016/j.chb.2021.106963]
56. Ryan RM, Deci EL. Intrinsic and extrinsic motivation from a self-determination theory perspective: definitions, theory, practices, and future directions. Contemp Educ Psychol 2020;61:101860. [doi: 10.1016/j.cedpsych.2020.101860]
57. Ryan RM, Deci EL. Self-Determination Theory: Basic Psychological Needs in Motivation, Development, and Wellness. New York: The Guilford Press; 2017.
58. Fogg BJ. Behavior model for persuasive design. 2009 Presented at: Proceedings of the 4th International Conference on Persuasive Technology; 2009 April 29; USA. [doi: 10.1145/1541948.1541999]
59. Beauchamp MR, Crawford K, Jackson B. Social cognitive theory and physical activity: mechanisms of behavior change, critique, and legacy. Psychol Sport Exerc 2019;42:110-117.
60. Adusumalli S, Aragam G, Patel M. A nudge towards cardiovascular health: applications of behavioral economics for primary and secondary cardiovascular prevention. Curr Treat Options Cardiovasc Med 2020;22(9). [doi: 10.1007/s11936-020-00824-y]
61. Locke EA, Latham GP. Building a practically useful theory of goal setting and task motivation. A 35-year odyssey. Am Psychol 2002;57(9):705-717. [doi: 10.1037//0003-066x.57.9.705] [Medline: 12237980]
62. Locke EA, Latham GP. New directions in goal-setting theory. Curr Dir Psychol Sci 2006;15(5):265-268. [doi: 10.1111/j.1467-8721.2006.00449.x]
63. McEwan D, Harden SM, Zumbo BD, Sylvester BD, Kaulius M, Ruissen GR, et al. The effectiveness of multi-component goal setting interventions for changing physical activity behaviour: a systematic review and meta-analysis. Health Psychol Rev 2016;10(1):67-88. [doi: 10.1080/17437199.2015.1104258] [Medline: 26445201]
64. Paul L, Wyke S, Brewster S, Sattar N, Gill JMR, Alexander G, et al. Increasing physical activity in stroke survivors using STARFISH, an interactive mobile phone application: a pilot study. Top Stroke Rehabil 2016;23(3):170-177 [FREE Full text] [doi: 10.1080/10749357.2015.1122266] [Medline: 27077973]
65. Zichermann G, Cunningham C. Gamification by Design: Implementing Game Mechanics in Web and Mobile Apps. Sebastopol, CA: O'Reilly Media, Inc; 2011.
66. Saleh ZT, Elshatarat RA, Elhefnawy KA, Helmi Elneblawi N, Abu Raddaha AH, Al-Za'areer MS, et al. Effect of a home-based mobile health app intervention on physical activity levels in patients with heart failure: a randomized controlled trial. J Cardiovasc Nurs 2023;38(2):128-139. [doi: 10.1097/JCN.0000000000000911] [Medline: 35389920]
67. Miller AS, Cafazzo JA, Seto E. A game plan: gamification design principles in mHealth applications for chronic disease management. Health Informatics J 2016;22(2):184-193 [FREE Full text] [doi: 10.1177/1460458214537511] [Medline: 24986104]
68. Alsawaier RS. The effect of gamification on motivation and engagement. Int J Inf Learn Technol 2018;35(1):56-79. [doi: 10.1108/ijilt-02-2017-0009]
69. Lewis ZH, Swartz MC, Lyons EJ. What's the point?: a review of reward systems implemented in gamification interventions. Games Health J 2016;5(2):93-99. [doi: 10.1089/g4h.2015.0078] [Medline: 26812253]
70. Lyons EJ. Cultivating engagement and enjoyment in exergames using feedback, challenge, and rewards. Games Health J 2015;4(1):12-18 [FREE Full text] [doi: 10.1089/g4h.2014.0072] [Medline: 26181675]
71. Cruz C, Hanus MD, Fox J. The need to achieve: players' perceptions and uses of extrinsic meta-game reward systems for video game consoles. Comput Hum Behav 2017;71:516-524. [doi: 10.1016/j.chb.2015.08.017]
72. Kerrigan SG, Forman EM, Williams D, Patel M, Loyka C, Zhang F, et al. Project step: a randomized controlled trial investigating the effects of frequent feedback and contingent incentives on physical activity. J Phys Act Health 2021;18(3):247-253. [doi: 10.1123/jpah.2020-0198] [Medline: 33601334]
73. Bravata DM, Smith-Spangler C, Sundaram V, Gienger AL, Lin N, Lewis R, et al. Using pedometers to increase physical activity and improve health: a systematic review. J Am Med Assoc 2007;298(19):2296-2304. [doi: 10.1001/jama.298.19.2296] [Medline: 18029834]
74. Zuckerman O, Gal-Oz A. Deconstructing gamification: evaluating the effectiveness of continuous measurement, virtual rewards, and social comparison for promoting physical activity. Pers Ubiquit Comput 2014;18(7):1705-1719. [doi: 10.1007/s00779-014-0783-2]
75. Hamari J, Koivisto J. Social motivations to use gamification: an empirical study of gamifying exercise. 2013 Presented at: European Conference on Information Systems; Utrecht; 2013 June 06.
76. Deci EL, Ryan RM. Intrinsic Motivation and Self-Determination in Human Behavior. New York: Springer; 1985.
77. Curry S, Wagner EH, Grothaus LC. Intrinsic and extrinsic motivation for smoking cessation. J Consult Clin Psychol 1990;58(3):310-316. [doi: 10.1037//0022-006x.58.3.310] [Medline: 2195084]
78. Robson K, Plangger K, Kietzmann JH, McCarthy I, Pitt L. Is it all a game? Understanding the principles of gamification. Bus Horiz 2015;58(4):411-420. [doi: 10.1016/j.bushor.2015.03.006]
79. Rigby C. Findings the right rewards to sustain player engagement. 2009 Presented at: Game Developers Conference; 2009 Sep 18; USA.

XSL•FO
RenderX

80. Deci EL, Ryan RM. Motivation, personality, and development within embedded social contexts: an overview of self-determination theory. In: The Oxford Handbook of Human Motivation. United Kingdom: Oxford University Press; 2012.
81. Li I, Dey A, Forlizzi J. A stage-based model of personal informatics systems. 2010 Presented at: Proceedings of the SIGCHI Conference on Human Factors in Computing System; 2010 April 10; USA. [doi: 10.1145/1753326.1753409]
82. Ploderer B, Reitberger W, Oinas-Kukkonen H, van Gemert-Pijnen J. Social interaction and reflection for behaviour change. Pers Ubiquit Comput 2014;18(7):1667-1676. [doi: 10.1007/s00779-014-0779-y]
83. De Croon R, Geuens J, Verbert K, Vanden Abeele V. A systematic review of the effect of gamification on adherence across disciplines. In: HCI in Games: Experience Design and Game Mechanics. Cham: Springer International Publishing; 2021.
84. Zhang J, Brackbill D, Yang S, Becker J, Herbert N, Centola D. Support or competition? How online social networks increase physical activity: a randomized controlled trial. Prev Med Rep 2016;4:453-458 [FREE Full text] [doi: 10.1016/j.pmedr.2016.08.008] [Medline: 27617191]
85. Du H, Venkatakrishnan A, Youngblood GM, Ram A, Pirolli P. A group-based mobile application to increase adherence in exercise and nutrition programs: a factorial design feasibility study. JMIR mHealth uHealth 2016;4(1):e4 [FREE Full text] [doi: 10.2196/mhealth.4900] [Medline: 26772910]
86. Lindsay Smith G, Banting L, Eime R, O'Sullivan G, van Uffelen JGZ. The association between social support and physical activity in older adults: a systematic review. Int J Behav Nutr Phys Act 2017;14(1):56 [FREE Full text] [doi: 10.1186/s12966-017-0509-8] [Medline: 28449673]
87. Greysen SR, Changolkar S, Small DS, Reale C, Rareshide CAL, Mercede A, et al. Effect of behaviorally designed gamification with a social support partner to increase mobility after hospital discharge: a randomized clinical trial. JAMA Netw Open 2021;4(3):e210952 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.0952] [Medline: 33760089]
88. Patel MS, Benjamin EJ, Volpp KG, Fox CS, Small DS, Massaro JM, et al. Effect of a game-based intervention designed to enhance social incentives to increase physical activity among families: the BE FIT randomized clinical trial. JAMA Intern Med 2017;177(11):1586-1593 [FREE Full text] [doi: 10.1001/jamainternmed.2017.3458] [Medline: 28973115]
89. Patel MS, Asch DA, Rosin R, Small DS, Bellamy SL, Eberbach K, et al. Individual versus team-based financial incentives to increase physical activity: a randomized, controlled trial. J Gen Intern Med 2016;31(7):746-754 [FREE Full text] [doi: 10.1007/s11606-016-3627-0] [Medline: 26976287]
90. Patel MS, Small DS, Harrison JD, Fortunato MP, Oon AL, Rareshide CAL, et al. Effectiveness of behaviorally designed gamification interventions with social incentives for increasing physical activity among overweight and obese adults across the United States: the STEP UP randomized clinical trial. JAMA Intern Med 2019;179(12):1624-1632 [FREE Full text] [doi: 10.1001/jamainternmed.2019.3505] [Medline: 31498375]
91. Visch VT, Vegt NJH, Anderiesen H, Van der Kooij K. Persuasive game design: a model and its definitions. 2013 Presented at: CHI; 2013 May 2; Paris.
92. Seaborn K, Fels DI. Gamification in theory and action: a survey. Int J Hum Comput Stud 2015;74:14-31. [doi: 10.1016/j.ijhcs.2014.09.006]
93. White JS, Salem MK, Toussaert S, Westmaas JL, Raiff BR, Crane D, et al. Developing a game (Inner Dragon) within a leading smartphone app for smoking cessation: design and feasibility evaluation study. JMIR Serious Games 2023;11:e46602 [FREE Full text] [doi: 10.2196/46602] [Medline: 37566442]
94. Sanders GD, Neumann PJ, Basu A, Brock DW, Feeny D, Krahn M, et al. Recommendations for conduct, methodological practices, and reporting of cost-effectiveness analyses: second panel on cost-effectiveness in health and medicine. J Am Med Assoc 2016;316(10):1093-1103. [doi: 10.1001/jama.2016.12195] [Medline: 27623463]

Abbreviations

AICc: sample-correction Akaike's information criterion
BEP: behavioral economics principles
CVD: cardiovascular disease
FBM: Fogg behavioral model
GST: goal-setting theory
mHealth: mobile health
PA: physical activity
PAD: peripheral artery disease
PI: prediction intervals
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROSPERO: International Prospective Register of Systematic Reviews
RCT: randomized controlled trial
SB: sedentary behavior
SCT: social cognitive theory
SDT: self-determination theory

XSL•FO
RenderX

TIDieR: Template for Intervention Description and Replication

Edited by A Coristine; submitted 16.07.24; peer-reviewed by C Monroe, T Baranowski; comments to author 11.10.24; revised version received 28.11.24; accepted 06.12.24; published 24.01.25.

Please cite as:
Yu T, Parry M, Yu T, Xu L, Wu Y, Zeng T, Leng X, Tong Q, Li F
Effectiveness of Mobile Health–Based Gamification Interventions for Improving Physical Activity in Individuals With Cardiovascular Diseases: Systematic Review and Meta-Analysis of Randomized Controlled Trials
JMIR Serious Games 2025;13:e64410
URL: https://games.jmir.org/2025/1/e64410
doi:10.2196/64410
PMID:

©Tianzhuo Yu, Monica Parry, Tianyue Yu, Linqi Xu, Yuejin Wu, Ting Zeng, Xin Leng, Qian Tong, Feng Li. Originally published in JMIR Serious Games (https://games.jmir.org), 24.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Electrophysiological Insights in Exergaming—Electroencephalography Data Recording and Movement Artifact Detection: Systematic Review

Carolina Rico-Olarte[1,2], MSc; Diego M Lopez[1], PhD; Bjoern M Eskofier[2], PhD; Linda Becker[3], PhD

[1]Telematics Department, Universidad del Cauca, Popayán, Colombia

[2]Machine Learning and Data Analytics Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany

[3]Department of Psychology, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany

Corresponding Author:
Carolina Rico-Olarte, MSc
Machine Learning and Data Analytics Lab
Friedrich-Alexander-Universität Erlangen-Nürnberg
Carl-Thiersch-Straße 2b
Erlangen, 91052
Germany
Phone: 49 9131 85 28990
Email: carolinarico@unicauca.edu.co

Abstract

Background: Exergames are interactive solutions that require physical activity and are commonly used in learning or rehabilitation settings. For cognitive rehabilitation with exergames, the assessment of the intervention progress can be conducted by verifying the changes in brain activity. Electroencephalography (EEG) is a well-known method for this evaluation. However, motion artifacts due to large body movements can impede signal quality. No comprehensive guide on the artifact removal methods in the context of exergaming has been found.

Objective: This paper aimed to identify studies that have assessed EEG signals while a user interacts with an exergame and the applied methods for data handling and analysis with a focus on dealing with movement artifacts.

Methods: This review included studies on human participants while engaging in exergames, where the primary outcome was brain activity measured by EEG. A total of 5 databases were searched at 3 time points: March 2021, October 2022, and February 2024. The Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies assessed methodological quality, rating studies as “good,” “fair,” or “poor.” Data were synthesized quantitatively to identify characteristics across studies, including sample demographics and intervention details, and basic statistics (mean [SD]) were calculated.

Results: A total of 494 papers were screened, resulting in 17 studies having been included. All studies carried out EEG recordings during exergame interactions, primarily assessing attention and concentration, with the alpha wave being the most analyzed EEG band. Common motion artifact removal methods included visual inspection and independent component analysis. The review identified significant risks of bias, with 2 studies rated as “good,” 7 as “fair,” and 8 as “poor.” Due to the small number of studies and their heterogeneity, a meta-analysis was not feasible.

Conclusions: The study successfully identifies the feasibility of recording electrophysiological brain activity during exergaming and provides insights into EEG devices, analysis methods, and exergaming systems used in previous studies. However, limitations, such as the lack of sufficient detail on motion artifact removal and a focus on short-term effects, underscore the need for improved methodologies and reporting standards, with recommendations for enhancing reliability in cognitive rehabilitation with exergames.

(JMIR Serious Games 2025;13:e50992) doi:10.2196/50992

KEYWORDS

exergaming; EEG; brain activity; motion artifact; artifact removal

XSL•FO
RenderX

Introduction

Rationale

Over the last 15 years, research around the concept of exergaming has expanded in the scientific community as well as in the commercial industry, such as the development of motion controllers for gaming consoles and virtual reality (VR) headsets [1-4]. The closest to a formal definition is found in [5], where the term exergame refers to "digital games that require bodily movements to play, stimulating an active gaming experience to function as a form of physical activity." Based on this, the scope of exergames can also extend to include VR, augmented reality, and similar technologies that provide immersive gaming experiences. Due to their nature, exergames are used across interdisciplinary fields, including application areas such as psychology, sports science, neuroscience, and computer science. Indeed, the largest application area is in public health and rehabilitation [6-8].

The main populations targeted for the use of exergames are children and adolescents due to their interest in video games and their potential benefits [9]. Some advantages of exergames are their ability to increase motivation and engagement in physical activity [5,10,11]. They also present features such as individualization, adaptivity (by providing feedback for adjustments), and specificity (task-centered and outcome-specific training) [12]. Consequently, exergaming can have a positive impact on physical, cognitive, and psychosocial variables in users [13].

Given their potential to improve health, exergames have evolved and have been evaluated in several scenarios as promising rehabilitation tools [14-16], especially for cognitive rehabilitation [17,18]. The constant training generated through the interaction with exergames may result in the acquisition or improvement of cognitive abilities such as executive functions (eg, attention and memory) [19]. These cognitive skills are necessary for effective performance of everyday tasks, especially in the case of older adults, as well as for school performance in children [20,21].

One specific exergame, HapHop-Physio, was designed and developed by our research group as part of an ongoing research project. This exergame targets children and adolescents with specific learning disorders, aiming to improve their memory and attention through physical activity. Initial findings are promising in supporting children's cognitive development and thus, improving their daily routines [22]. In previous studies, electrodermal activity has been recorded during interactions with HapHop-Physio, providing valuable insights into the cognitive processes influenced by exergaming [23].

However, to fully assess cognitive-specific outcomes, exergames like HapHop-Physio require the additional measurement of neurophysiological changes related to brain activity. Our research focuses on evaluating HapHop-Physio's effectiveness through electroencephalography (EEG) [22]. EEG is one of the most commonly used methods for inspecting electrical activity in the brain by recording brain signals with electrodes placed on the scalp [24,25]. EEG is increasingly used in studies on infant development [23]. EEG captures voltage fluctuations resulting from ionic currents within neurons, which produces signals that represent various brain waves. EEG analysis allows an assessment of the underlying processes of cognitive functions [26]. However, EEG signals are highly sensitive to external noise and interference, such as muscle movement, eye blinks, or electrical noise, which introduces artifacts in the EEG signal (ie, unwanted additional components that distort the raw signal) [27]. These artifacts pose a challenge to the measurement of brain activity during the interaction with an exergame, since movements cannot be avoided during exergaming. The movement artifacts complicate or impede the accurate interpretation of the EEG signals. Thus, removing or minimizing artifacts is crucial in ensuring valid EEG readings [28].

Consequently, some researchers have addressed this issue from different approaches. Artifact removal methods vary from manual visual inspection or the application of band filters to complex procedures such as independent component analysis (ICA) or regression-based techniques. ICA aids in detecting facial muscle movements, eye blinks, or other movements. Regression-based techniques predict and subtract the contribution of artifacts to the signal (particularly from eye movements) using mathematical models [29-31]. The specific choice of artifact removal method depends on the specific study design, the types of artifacts present, and the trade-offs between preserving brain signals and removing noise [32].

Nevertheless, the body movements generated by the interaction with an exergame are a piece of motion artifact that are just beginning to be explored. Accurately quantifying data loss due to artifacts is important because large portions of EEG data may be discarded, leading to reduced sample sizes or biased results. When reporting data loss, the research community benefits from better interpretation and an effective comparison across studies. Evaluating signal quality becomes essential to determine whether the brain activity signals remain useful for interpretation after cleaning [33]. Similarly, the associations between any specific cognitive domain induced by the interaction with the exergame and its content are not yet generalizable and depend on the type of exergame as well as the type and strength of body movements.

Objectives

The primary objective of this systematic review was to identify studies that have measured EEG signals during exergame interaction with a focus on examining the methods used for artifact detection and removal addressing the "large" movement artifact generated. Thus, this review aimed to answer the following questions:

1. Main review questions: (1) Can electrophysiological brain activity be recorded during exergaming? (2) Which EEG devices, methods for data analysis with a focus on methods for artifact detection and removal, and exergaming systems have been used in previous studies?
2. Further review questions: (1) Which population groups have been studied so far (eg, with respect to health status, age, gender, ethnicity, and education)? (2) Are there differences in brain activity during exergaming across the

XSL•FO
RenderX

lifespan (eg, between children, adolescents, younger adults, and older adults)?

Methods

Eligibility Criteria

The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement [34] was followed to report the process and results obtained in the revision (Multimedia Appendix 1). In this systematic review, we established explicit inclusion and exclusion criteria to determine the eligibility of studies. Detailed specifications of these criteria are provided in Textbox 1.

Textbox 1. Details on eligibility criteria.

Inclusion criteria

- Human participants, all age groups.
- All types of exergames and cognitive-motor games.
- Original studies, longitudinal or experimental studies, (controlled) cross-sectional studies, or case-control studies.
- Electrophysiological activity (via electroencephalography [EEG]), participants' characteristics (eg, age, gender, and health status), characteristics of the game and the EEG device, and the preprocessing pipeline.
- Brain activity measured by EEG during the exergame activity.

Exclusion criteria

- Animal studies.
- Single cognitive or motor tasks (eg, classical video games).
- Brain-computer interface games.
- Literature reviews, case reports, and qualitative studies.
- Other neuroscientific methods (eg, functional near-infrared spectroscopy, magnetoencephalography, and functional magnetic resonance imaging).

Information Sources

We used 5 bibliographic databases to identify peer-reviewed studies for their inclusion in this systematic review: PubMed, IEEE Xplore, ACM Digital Library, Scopus, and Web of Science. Also, we searched the reference list from retrieved full-text articles for additional studies. To keep the results up to date, we conducted searches based on 2 search strings at 3 separate times. The first search date was March 2021, the second search date was October 2022; and finally, the third search was performed in February 2024.

Search Strategy

For the search strategy, we considered 2 main concepts to retrieve the studies: "exergaming" and "electroencephalography." We adjusted these concepts according to the respective databases' Thesaurus and MeSH (Medical Subject Headings) terms. Also, a second search string was built to replace the concept of exergaming with commercial devices that meet the concept. Thus, for the first string, we searched 10 synonyms within the exergaming concept and 36 synonyms within the EEG concept covering the obtained measures when recording brain activity through EEG; as for the second string, we searched for eleven devices (mostly VR headsets) along with the electroencephalography concept. Multimedia Appendix 2 describes the strings, filters, and limits used in each information source. A further source of information was the references to the papers that supplied the following selection process.

Selection Process

After retrieving the studies from the information sources, we uploaded the files to the Rayyan platform [35] to remove the duplicates and start the screening and selection process. We established a 3-stage approach. First, 2 authors (CRO and LB) independently screened all the titles and abstracts to identify articles that potentially met the inclusion criteria. Second, the introduction and conclusions of the remaining articles were screened to further refine the selection. Finally, full texts of the articles were independently assessed for eligibility and data extraction. Any disagreements between the 2 raters were solved through discussion with a third author (DML). The Rayyan platform enabled blinded screening and selection of papers.

Data Collection Process

We extracted data from the selected papers generating a standardized prepiloted spreadsheet. As in the selection process, 2 authors (CRO and LB) collected the data independently. The spreadsheet comprised study identification, methodological design, sample characteristics, details of the exergame and the EEG device, analysis methods, outcomes, study limitations, and some additional information.

Data Items

The most relevant conditions for the papers to be selected were (1) that there was an exergame, that is, a game that stimulates the participants both physically and cognitively (or generates entertainment); (2) that authors recorded brain (EEG) activity during the game, that is, while the participants were stimulated

XSL•FO
RenderX

and were moving; (3) that authors provided a detailed description of the EEG analysis methods.

Other variables considered included whether the application was for clinical patients, the data collection setting (laboratory, clinic, or at home), characteristics of the intervention, results in terms of cognition and brain activity, and study limitations; and covariates.

Study Risk of Bias Assessment

The 2 authors (CRO and LB) independently evaluated the methodological quality of the selected studies. The Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies [36] from the National Heart, Lung, and Blood Institute (NHLBI) [37] contains fourteen questions designed to assess the internal validity of a study. Any disagreement in the evaluation was solved by the third author (DML).

The evaluation process consisted of 2 steps. First, we assigned importance levels to the 14 yes or no questions based on the elements queried by each question, such as research question, sample size justification, etc. Four questions were classified as very important, 1 as important, and 9 as less important. The importance level determined the weight of each question in the assessment process.

Subsequently, we evaluated each study based on its responses to the yes or no questions. A “good” rating was assigned when none or just 1 answer was “NO,” indicating minimal risk of bias. A “fair” rating was given when 2 or 3 answers were “NO,” suggesting some potential for bias. Conversely, a “poor” rating was assigned when there was a “NO” answer in 1 of the most important questions, or at least 4 “NO” answers in the less important questions, indicating a significant risk of bias. The final rating for each study was determined by tallying the number of “NO” answers across all questions.

Synthesis Methods

From the data collected and recorded in the matrix, we synthesized evidence quantitatively to determine the sets of characteristics that could be associated with each study, such as the population sample, the countries that conducted the research, or the gender distribution of the sample. In the same way, we calculated basic statistics (mean [SD]) for characteristics such as the ages reported in the studies and the intervention performed with the exergame and measured with the EEG device. A meta-analysis should be considered depending upon the availability of appropriate data (ie, sufficient number and homogeneity of studies).

The protocol for this systematic review was prospectively registered in the PROSPERO (International prospective register of systematic reviews) database (registration no CRD42020208131). No deviations from the registered protocol occurred during the review process.

Results

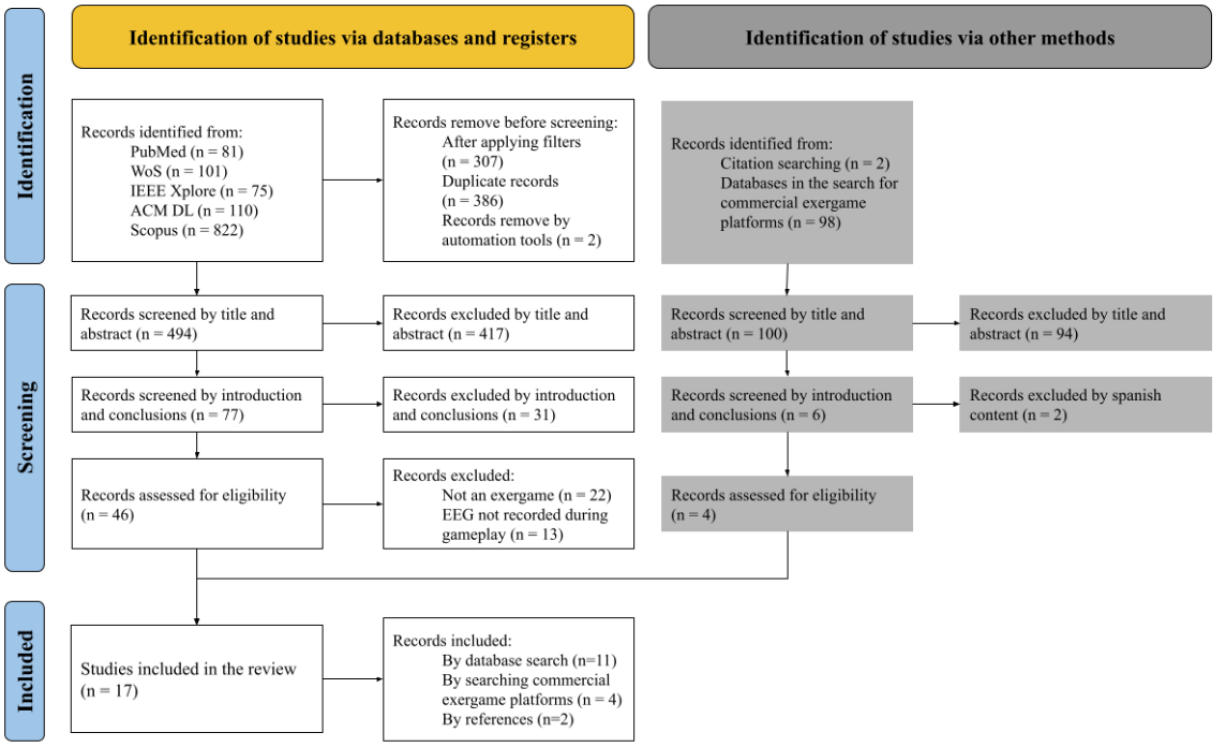
Study Selection

After removing the duplicates, we first screened 494 papers by title and abstract. Second, we screened 77 papers by introduction and conclusions. Third, we reviewed the full text of 46 documents, excluding 35 studies from this set, either because they did not refer to an exergame (n=22), or did not record EEG activity during gameplay (n=13; Figure 1). Finally, only 11 papers were included in the review. Additional papers fulfilling the inclusion criteria were found in other searches, such as the study’s references (n=2) and the search for commercial exergames platforms (Nintendo Wii and Kinect) that were not included in the searches results since they were referenced by their commercial name and not as exergames (n=4).

A considerable number of papers containing the keywords were excluded because they were within the brain-computer interface (BCI) paradigm. This paradigm uses EEG as an input device to control the game from the brain activity detected in the player. Although papers reported EEG recordings, there was no relationship reported between the physical activity generated by the game (not classified as exergame) and the brain signals.

XSL•FO
RenderX

Figure 1. Flowchart of the results from the review process.



Study Characteristics

The 17 included studies were conducted in diverse continent settings including Europe (5 studies: Norway, Germany [n=3] Austria, and Italy); Asia (5 studies: Korea, China [n=2], Turkey, and Iran); North America (3 studies: the United States and Canada [n=2]), Central America (1 study from Mexico), South America (2 studies from Brazil) and 1 in Oceania (New Zealand). Sample sizes across these studies were generally small, with 6 studies including between 10 and 19 participants, and 4 studies reporting more than 30 participants. Authors of 11 of the identified studies labelled the design as “experimental.” The remaining studies were reported to be pilot studies (n=2), a feasibility study (n=1), a preliminary study (n=1), a comparative study (n=1), and a case study (n=1).

In terms of population characteristics, 13 studies reported gender distribution. The female participation rate was 42% against a male participation rate with 58% (mean values from studies). In total, 16 studies reported the age of the participants. The mean age of the study’s participants within the young adult category (n=12) was 25.4 years. The population in 14 of the studies was nonclinical; 2 studies reported data from a clinical population (upper limb injuries and stroke). The children’s population was mixed between clinical (diagnosed with Autism) and a nonclinical sample.

Multimedia Appendix 3 [38-54] shows the characteristics of the included studies. All 17 studies used strategies to induce physical activity: 7 of them classified their games as exergames, 3 studies used the Kinect device, a motion-sensing input device developed by Microsoft for the Xbox gaming console that detects physical movements and translates them into game controls, to detect movement, 3 studies used the Nintendo Wii device, a gaming console with motion-sensing controllers that track physical movements to control game actions, to induce movement, 2 studies use a VR-based approach (MetaQuest2 and HTC Vive), 1 study used the SilverFit system, a rehabilitation device designed for motor and cognitive therapy that combines interactive gaming with physical exercises, and the remaining study used a static bike to induce the physical activity while playing the video games.

Multimedia Appendix 4 [38-60] details the information of EEG recordings in the studies. A total of 12 studies reported methods for the detection and removal of motion artifacts in the EEG signal. There were 2 studies for which the process is not transparent since it is made internally by the devices. The EEG activity was measured by its band waves (alpha, beta, theta, gamma, and delta) in 14 studies, 2 studies reported the brain activity in terms of the signal components (such as the N1 component of the Event-Related Potential) and finally, 1 study reported the processed signal from the device’s software. A narrative description of the results of individual studies is reported in Multimedia Appendix 5.

Synthesis of Results

According to the results obtained and from the data extracted from the reviewed papers, the following aspects can be synthesized.

Exergame Characteristics

Regarding the exergames used in the studies, 11 out of 17 exergames are commercially available; only 6 exergames were developed to assess the research objectives outlined in each paper. The additional tools that authors used to induce physical activity within the exergame correspond first to platforms for users to stand with different modalities (to apply pressure, to

XSL•FO
RenderX

control balance, or to practice specific ski movements), followed by, the motion detection of body gestures and positions through the Kinect and Wii, and finally, VR headset devices.

Cognitive Domains and EEG Analysis

The cognitive domains stimulated by the exergames and identified through the analysis of the EEG recordings include attention and concentration (n=4), followed by engagement (n=2), cognitive workload (n=2), stress (n=2), motor learning (n=2), working memory (n=1), and awareness (n=1). Only 1 study evaluated processing speed and language. Two studies did not report the cognitive domains.

In the studies using the analysis of the EEG recordings through band waves, the alpha wave was the most often analyzed (n=10), followed by the beta and theta waves (n=7). The least used waves were gamma and delta. Authors used various methods to measure their results according to the EEG band waves. Among them, the absolute power and the relative power were obtained, in addition to the asymmetrical comparisons between channels (in dB), changes in the bands (in terms of percentages), and applying equations for different indexes based on the values of the bands.

The most common electrode placement positions for EEG recording were the frontal and parietal lobes (n=6), followed by the prefrontal lobe (n=5), then, the central and precentral lobes (n=3). The least used positions were located occipitally (n=2) and temporally (n=1).

Motion Artifact Handling

Finally, in line with the aim of this systematic review, the methods for detecting and removing motion artifacts were summarized. The most frequently reported methods were filtering the signal (n=6), using the ICA method (n=4), visual and manual inspection (n=2), and applying the automated subspace reconstruction method (n=2). Other methods used by the authors were based on previous contributions such as plug-ins, pipelines, and metrics, including the use of several types of band-pass filters. Some authors used recording devices that automatically filtered motion artifacts by deleting them before delivering the data to the user.

Risk of Bias Assessment in Studies

The risk of bias was evaluated across 2 dimensions: internal validity (risk of bias within individual studies) and external validity (reporting biases affecting the overall synthesis).

The methodological quality of the studies included was assessed based on essential factors like study population, exposure measures, and statistical analysis. Among the studies, only 2 [40,50] were rated as "good," indicating the least risk of bias and validating their results. Seven studies [38,39,45,47,49,51,53] were rated as "fair," suggesting some potential for bias, though not enough to invalidate their findings. However, 8 studies [41-44,46,48,52,54] received a rating of "poor," indicating a significant risk of bias. These studies were included in the review due to the limited availability of alternative evidence. Full details of the risk of bias assessment are provided in Multimedia Appendix 6.

In the assessment of reporting biases, 2 papers received a "good" rating, indicating a minimal risk of bias. These papers demonstrated clear objectives, transparent recruitment process, and consistent exposure and outcome measure implementations. Seven papers were rated as "fair" suggesting a moderate risk of bias, due to shortcomings such as insufficient sample size justification or limited exposure assessment at only 1 time point. In contrast, the remaining 8 papers were rated as "poor" indicating a high risk of bias. These papers lacked clear definition of the study population, exhibited inconsistent implementation of outcome measures, and had an unclear recruitment process. In addition, none of the papers reported blinding of assessors to participant exposure status, and none accounted for confounding variables. Furthermore, some papers included unnecessary information in their reporting.

Certainty of Evidence

Unfortunately, the number of papers that met the inclusion criteria to fulfill the aim of this systematic review was low (n=17). This low number of studies did not allow for a meta-analysis of the information extracted from each paper due to the heterogeneity of the studies. Likewise, it is not possible to generalize the results, since the risk-of-bias assessment resulted in inferior quality in 47% of the papers analyzed.

Two papers [40], [50] mentioned the need to detect and remove movement artifacts from the EEG signal recorded during the participants' interaction with the exergame. One important piece of information from [40] was the decision to remove around 50% of the signal. In contrast, for the EEG signal epochs in [50], only between 5% and 7% were removed. In the remaining papers, although noise detection and removal methods were applied, no specific reference was made to the motion artifacts generated by physical activity during playing the exergame or the amount of signal that was removed for further analysis.

Discussion

Principal Findings

Our review revealed several key findings regarding the recording of electrophysiological brain activity during exergaming. First, we demonstrated the feasibility of this approach, highlighting the potential for capturing neural signals during interactive physical activity. In addition, we identified a variety of EEG devices, analysis methods, and exergaming systems used in previous studies, providing valuable insights into the technological landscape of the field. The analysis also covered significant challenges, including issues related to data loss and motion artifact removal, which impact statistical power and introduces bias in estimation of EEG outcomes. We observed a lack of reporting of detailed information on motion artifact identification, particularly in relation to the motion artifacts generated by large, whole-body movements inherent to exergame interaction, removal methods in the included studies, traditional techniques were primarily developed to address smaller localized artifacts and may not effectively account for the more "substantial" artifacts generated by whole-body movements. This lack of detail highlights the need for greater methodological transparency in future research, especially in

XSL•FO
RenderX

how larger motion artifacts are managed to ensure valid EEG findings.

Our review highlighted a notable gap in literature, with few studies conducting long-term measurements of the cognitive and physical effects of exergaming through EEG recordings. There is a predominant focus on young adult populations in existing research, with limited exploration of differences across age groups and ethnicities. These findings collectively contribute to a deeper understanding of the challenges and opportunities in the field of electrophysiological research in exergaming.

The main review questions (Can electrophysiological brain activity be recorded during exergaming? Which EEG devices, methods for data analysis, and exergaming systems have been used in previous studies?) were answered: at least 1 of the included studies carried out a feasibility study determining that, although approximately 50% of the EEG signal data were lost, it was possible to record and subsequently analyze the information from the signals that were obtained while the players interacted with the exergame. This finding, along with the much lower amount reported in [50], is contradictory and highlights the importance of defining comparable quality metrics for future studies. However, data loss negatively impacts the results in several ways. First, it reduces the statistical power of the study. Second, data loss can introduce bias in the estimation of parameters in classification or prediction models, leading to inaccurate results. Finally, it may diminish the representativeness of the sample, which may limit the generalization of the findings [61].

Comparison to Previous Work

In addition, information on EEG devices, analysis methods, and exergaming systems was obtained and reported in this review. Of the 16 included studies, although all recorded EEG signals, the aims were different in terms of determining the feasibility or quality of the signal in the presence of movements (physical activity) made by the participants. Of these studies, 12 papers attempted to detail the process of identifying and removing motion artifacts from the signal. However, the information provided is not sufficient for replicating the procedures.

Regarding the devices found in the included studies, we classified them according to the type of recording hardware. Four of the EEG devices can be classified as complete systems; that is, they have a signal amplifier and a cap that facilitates the easy placement of EEG electrodes on the participant's head. Three of the studies simply used a signal amplifier, without mentioning the type of electrodes used for signal capturing or using generic EEG electrodes. Ten studies used headset-type devices. These devices are wireless and store the signal in their memory, which is then downloaded to a software for reading and analyzing. These devices are wearables [62], so their use is much easier for nonexperts in the placement of conventional electrodes according to the 10-20 reference system [63]. In terms of usability in environments other than clinical, wearable devices are convenient to validate hypotheses before conducting more formal studies [64].

Another aspect that was evaluated within this review was the analysis method for processing the EEG data. Of the studies included, 14 used analysis methods in the frequency domain, obtaining the absolute power of the band waves [65]; this is the most commonly used approach in the literature. Only 1 study included time-frequency maps showing increasing and decreasing power events (event-related desynchronization/event-related synchronization). Two studies analyzed the signal in the time domain, examining the N1 component and using its amplitude metrics. Defining the analysis methods is an important task to determine the effects of movement on the EEG signal and, at the same time, obtain information on the changes or cognitive processes that are activated during the interaction with an exergame [66].

Exergames are still being tested to corroborate the benefits attributed to them, such as the increase in motivation and enjoyment, the reach of populations, and their capacity for individualization, adaptivity, specificity, and scalability [5]. Even though 9 of the studies included used game systems that were not categorized as exergames by the study's authors (using a Kinect, a Nintendo Wii, stationary bicycle, a balance board, and a VR headset), the authors attributed improvement of cognitive processes to the physical activity during the games [67,68].

In addition to our systematic review, a recent literature review on motion artifact reduction in BCI systems [69], highlights the challenges associated with EEG data collection during physical activities, which closely mirrors our findings in exergaming studies. Both reviews highlight the limitations of widely used artifact removal techniques, such as ICA and regression-based methods, which are often insufficient when addressing the larger artifacts caused by whole-body movements in exergaming. This aligns with our observation that many exergaming studies lack sufficient detail for replicating effective motion artifact removal techniques.

Furthermore, the studies reviewed in [69] revealed a tendency to focus on introducing new methods for artifact reduction, rather than improving or thoroughly comparing established ones. Both our review and the BCI literature emphasize the difficulty of generalizing findings due to small sample sizes and the variations in experimental paradigms, hardware setups, and preprocessing steps. Furthermore, the "ground truth" problem (distinguishing between true brain signals and noise) identified in [69] is especially relevant in exergaming contexts, where participant movement is integral to the activity. By framing our findings within this broader context, we emphasize the need for future research to improve artifact removal methods, enhance study transparency, and adopt advanced technologies to refine EEG signal quality in exergaming.

Limitations

The greatest limitation found in the included studies is the lack of detail in the processes of identification and removal of motion artifacts in the EEG signal. This limitation hinders a comprehensive understanding of data quality and may impact the accuracy of findings. Another limitation is the formality of the studies conducted, as their objectives only go as far as measuring the immediate effect of 1 game session on cognitive processes. While other studies have data on long-term outcomes, they did not measure EEG during gameplay [70-73]. To date,

XSL•FO
RenderX

no study includes long-term measurement of the cognitive effect or the effect on the person's physical state by the periodic use of an exergame with EEG recordings. This narrow scope limits the depth of insight into the sustained benefits or potential drawbacks of exergaming interventions.

Several limitations were identified in our systematic review. Many of the included studies were formative in nature and had small sample sizes, which restrict the generalization of the findings. However, we opted for a systematic review to evaluate the feasibility of recording EEG during exergaming and to assess methods for managing motion artifacts. Despite the heterogeneity across, including the variation in EEG devices, data analysis methods, and exergames used, we aimed to provide structured insight into the field.

Third, the absence of information on differences in brain activity across age groups and ethnicities restricts the generalizability of our findings. Without this crucial demographic context, the applicability of our results to diverse populations is compromised, highlighting the need for more inclusive research approaches in future investigations. While we used a comprehensive search strategy and followed the PRISMA guidelines, some studies may have been missed, particularly unpublished or grey literature.

Heterogeneity across studies, the small sample sizes, and the formative nature of many studies made synthesis challenging, both qualitatively and quantitatively. It also limited our ability to draw firm conclusions or identify trends across the studies included.

Future Directions

The results of this systematic review provide answers to further review questions (Which population groups have been studied so far, eg, with respect to health status, age, gender, ethnicity, education? Are there differences in brain activity during exergaming across the lifespan, eg, between children, adolescents, younger adults, and older adults?). As stated in the results section, most of the included studies targeted young adults and no differences were examined between age groups even when some exergames were intended for children but evaluated in young adults. Unfortunately, none of the studies specified information about differences in brain activity during exergaming across the lifespan or the ethnicity of the population. This lack of detail represents a limitation when applying the results of published investigations to other contexts, such as comparing age groups [74,75] or targeting multiethnic populations [76].

Future research should focus on more standardized methods, larger and more diverse sample sizes, and long-term evaluations of cognitive and physical outcomes through EEG recordings during exergaming interventions.

Conclusions

The review demonstrates the feasibility of recording electrophysiological brain activity during exergaming and offers valuable insights into the EEG devices, analysis methods, and exergaming systems used in previous research. Our findings indicate that EEG signals can be successfully measured in the midst of physical activity induced by player interaction with exergames, which may be associated with cognitive domains and processes. However, the presence of motion artifacts may lead to the loss of up to 50% of the EEG signal, posing challenges for data analysis and interpretation. Despite these limitations, our study underscores the potential for conducting meaningful analyses with the available data. Moving forward, research should address these limitations by enhancing reporting standards and methodological practices. Specifically, researchers should provide a rationale for EEG device selection, determine the type and intensity of movements during recording, assess appropriate motion artifact detection and removal methods, report details of data loss due to artifacts, evaluate signal quality post artifact removal, and conduct more formal longitudinal experiments to explore the medium and long-term effects of exergaming on cognitive processes and physical well-being.

Furthermore, our review has practical implications for the development of therapeutic interventions, as exemplified by our design of the HapHop-Physio exergame targeting children with specific learning disorders [77]. This exergame has the potential to assist in cognitive rehabilitation for this population [17]. By incorporating EEG signal recording during gameplay, our ongoing study aims to characterize electrical signals corresponding to typical player movements [78]. More information on the protocol for the study can be found in the Open Science Framework [79].

Acknowledgments

The authors gratefully acknowledge the support received for this research project. Funding for this review was provided by the project titled "Electrophysiological brain activity during cognitive rehabilitation training for children with learning disabilities." In addition, support was received from MinCiencias (Colombian government) under call 785 from 2017. The authors (LB, BME, and DML) received support from the BAYLAT (Bavarian University Center for Latin America) project entitled "Electrophysiological brain activity during cognitive rehabilitation training for children with learning disabilities."

Data Availability

The precise sources (extraction matrix, tables, and figures) for each estimate are described in a data spreadsheet. Available in Open Science Framework [79].

XSL•FO
RenderX

Authors' Contributions

CRO contributed to the search, collection, synthesis, and analysis of the results from the review process, along with the building of the first draft and subsequent corrections to the manuscript. DML contributed to the analysis of the results from the search and the reviews of the manuscript. BME contributed to the review and approval of the manuscript for publication. Finally, LB contributed to building the search protocol, to analyze the results of the search and to correct the different versions of the draft and final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 checklist.
[PDF File (Adobe PDF File), 127 KB - games_v13i1e50992_app1.pdf]

Multimedia Appendix 2
Table A. Filters and limits in information sources from search strings.
[PDF File (Adobe PDF File), 35 KB - games_v13i1e50992_app2.pdf]

Multimedia Appendix 3
Table B. Details of intervention or exposure in the studies.
[PDF File (Adobe PDF File), 78 KB - games_v13i1e50992_app3.pdf]

Multimedia Appendix 4
Table C. Details of EEG recordings in the studies.
[PDF File (Adobe PDF File), 155 KB - games_v13i1e50992_app4.pdf]

Multimedia Appendix 5
Details of the results of individual studies.
[PDF File (Adobe PDF File), 176 KB - games_v13i1e50992_app5.pdf]

Multimedia Appendix 6
Quality assessment for papers.
[PDF File (Adobe PDF File), 79 KB - games_v13i1e50992_app6.pdf]

References

1. Best JR. Exergaming in youth: effects on physical and cognitive health. Z Psychol 2013;221(2):72-78 [FREE Full text] [doi: 10.1027/2151-2604/a000137] [Medline: 25097828]
2. Street TD, Lacey SJ, Langdon RR. Gaming your way to health: a systematic review of exergaming programs to increase health and exercise behaviors in adults. Games Health J 2017;6(3):136-146. [doi: 10.1089/g4h.2016.0102] [Medline: 28448175]
3. Williams WM, Ayres CG. Can active video games improve physical activity in adolescents? A review of RCT. Int J Environ Res Public Health 2020;17(2):669 [FREE Full text] [doi: 10.3390/ijerph17020669] [Medline: 31968642]
4. O'Loughlin EK, Dutczak H, Kakinami L, Consalvo M, McGrath JJ, Barnett TA. Exergaming in youth and young adults: a narrative overview. Games Health J 2020;9(5):314-338. [doi: 10.1089/g4h.2019.0008] [Medline: 32017864]
5. Benzing V, Schmidt M. Exergaming for children and adolescents: strengths, weaknesses, opportunities and threats. J Clin Med 2018;7(11):422 [FREE Full text] [doi: 10.3390/jcm7110422] [Medline: 30413016]
6. Gao Z, Chen S. Are field-based exergames useful in preventing childhood obesity? A systematic review. Obes Rev 2014;15(8):676-691. [doi: 10.1111/obr.12164] [Medline: 24602072]
7. Barry G, Galna B, Rochester L. The role of exergaming in parkinson's disease rehabilitation: a systematic review of the evidence. J Neuroeng Rehabil 2014;11(1):33 [FREE Full text] [doi: 10.1186/1743-0003-11-33] [Medline: 24602325]
8. Taylor MJD, Griffin M. The use of gaming technology for rehabilitation in people with multiple sclerosis. Mult Scler 2015;21(4):355-371. [doi: 10.1177/1352458514563593] [Medline: 25533296]
9. Baranowski T. Exergaming: hope for future physical activity? or blight on mankind? J Sport Health Sci 2017;6(1):44-46 [FREE Full text] [doi: 10.1016/j.jshs.2016.11.006] [Medline: 30356568]
10. Osorio G, Moffat DC, Sykes J. Exergaming, exercise, and gaming: sharing motivations. Games Health J 2012;1(3):205-210. [doi: 10.1089/g4h.2011.0025] [Medline: 26193438]

XSL•FO
RenderX

11. Cacciata MC, Stromberg A, Klompstra L, Jaarsma T, Kuriakose M, Lee J, et al. Facilitators and challenges to exergaming. J Cardiovasc Nurs 2021;37(3):281-288. [doi: 10.1097/jcn.0000000000000825]
12. Smits-Engelsman B, Vinçon S, Blank R, Quadrado VH, Polatajko H, Wilson PH. Evaluating the evidence for motor-based interventions in developmental coordination disorder: a systematic review and meta-analysis. Res Dev Disabil 2018;74:72-102. [doi: 10.1016/j.ridd.2018.01.002] [Medline: 29413431]
13. Gao Z. Fight fire with fire? Promoting physical activity and health through active video games. J Sport Health Sci 2017;6(1):1-3 [FREE Full text] [doi: 10.1016/j.jshs.2016.11.009] [Medline: 30356540]
14. Mat Rosly M, Mat Rosly H, Davis Oam GM, Husain R, Hasnan N. Exergaming for individuals with neurological disability: a systematic review. Disabil Rehabil 2017;39(8):727-735. [doi: 10.3109/09638288.2016.1161086] [Medline: 27108475]
15. Mura G, Carta MG, Sancassiani F, Machado S, Prosperini L. Active exergames to improve cognitive functioning in neurological disabilities: a systematic review and meta-analysis. Eur J Phys Rehabil Med 2018;54(3):450-462. [doi: 10.23736/s1973-9087.17.04680-9]
16. Chan KGF, Jiang Y, Choo WT, Ramachandran HJ, Lin Y, Wang W. Effects of exergaming on functional outcomes in people with chronic stroke: a systematic review and meta-analysis. J Adv Nurs 2022;78(4):929-946. [doi: 10.1111/jan.15125] [Medline: 34877698]
17. Rico-Olarte C, Narváez-Muñoz N, López DM, Becker L, Tovar-Ruiz L. Assessing HapHop-Physio: an exer-learning game to support therapies for children with specific learning disorders. Applied Sciences 2022;12(16):8281. [doi: 10.3390/app12168281]
18. Li K, Wang Y, Wu Z, Yao X, Fan Y. Effectiveness of active exergames for improving cognitive function in patients with neurological disabilities: a systematic review and meta-analysis. Games Health J 2023;12(3):198-210. [doi: 10.1089/g4h.2022.0134] [Medline: 36445748]
19. Hötting K, Röder B. Beneficial effects of physical exercise on neuroplasticity and cognition. Neurosci Biobehav Rev 2013;37(9 Pt B):2243-2257. [doi: 10.1016/j.neubiorev.2013.04.005] [Medline: 23623982]
20. Gunzenhauser C, Nückles M. Training executive functions to improve academic achievement: tackling avenues to far transfer. Front Psychol 2021;12:624008 [FREE Full text] [doi: 10.3389/fpsyg.2021.624008] [Medline: 34108906]
21. Corbo I, Casagrande M. Higher-level executive functions in healthy elderly and mild cognitive impairment: a systematic review. J Clin Med 2022;11(5):1204 [FREE Full text] [doi: 10.3390/jcm11051204] [Medline: 35268294]
22. Tost A, Migliorelli C, Bachiller A, Medina-Rivera I, Romero S, García-Cazorla Á, et al. Choosing strategies to deal with artifactual EEG data in children with cognitive impairment. Entropy (Basel) 2021;23(8):1030 [FREE Full text] [doi: 10.3390/e23081030] [Medline: 34441170]
23. Noreika V, Georgieva S, Wass S, Leong V. 14 challenges and their solutions for conducting social neuroscience and longitudinal EEG research with infants. Infant Behav Dev 2020;58:101393. [doi: 10.1016/j.infbeh.2019.101393] [Medline: 31830682]
24. Müller-Putz GR. Electroencephalography. Handb Clin Neurol 2020;168:249-262. [doi: 10.1016/B978-0-444-63934-9.00018-4] [Medline: 32164856]
25. Troller-Renfree SV, Morales S, Leach SC, Bowers ME, Debnath R, Fifer WP, et al. Feasibility of assessing brain activity using mobile, in-home collection of electroencephalography: methods and analysis. Dev Psychobiol 2021;63(6):e22128 [FREE Full text] [doi: 10.1002/dev.22128] [Medline: 34087950]
26. Lau-Zhu A, Lau MP, McLoughlin G. Mobile EEG in research on neurodevelopmental disorders: opportunities and challenges. Dev Cogn Neurosci 2019;36:100635 [FREE Full text] [doi: 10.1016/j.dcn.2019.100635] [Medline: 30877927]
27. Reis PMR, Hebenstreit F, Gabsteiger F, von Tscharner V, Lochmann M. Methodological aspects of EEG and body dynamics measurements during motion. Front Hum Neurosci 2014;8:156 [FREE Full text] [doi: 10.3389/fnhum.2014.00156] [Medline: 24715858]
28. Mumtaz W, Rasheed S, Irfan A. Review of challenges associated with the EEG artifact removal methods. Biomedical Signal Processing and Control 2021;68:102741. [doi: 10.1016/j.bspc.2021.102741]
29. Independent Component Analysis for artifact removal. EEGLAB Wiki. URL: https://eeglab.org/tutorials/06_RejectArtifacts/RunICA.html [accessed 2025-02-27]
30. Islam MK, Rastegarnia A, Yang Z. Methods for artifact detection and removal from scalp EEG: a review. Neurophysiol Clin 2016;46(4-5):287-305. [doi: 10.1016/j.neucli.2016.07.002] [Medline: 27751622]
31. Gorjan D, Gramann K, De Pauw K, Marusic U. Removal of movement-induced EEG artifacts: current state of the art and guidelines. J Neural Eng 2022. [doi: 10.1088/1741-2552/ac52d2] [Medline: 35133298]
32. Mannan MMN, Kamran MA, Jeong MY. Identification and removal of physiological artifacts from electroencephalogram signals: a review. IEEE Access 2018;6:30630-30652. [doi: 10.1109/access.2018.2842082]
33. Jiang X, Bian GB, Tian Z. Removal of artifacts from EEG signals: a review. Sensors (Basel) 2019;19(5):987 [FREE Full text] [doi: 10.3390/s19050987] [Medline: 30813520]
34. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
35. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016;5(1):210 [FREE Full text] [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]

XSL•FO
RenderX

36. Study quality assessment tools. National Institutes of Health (NIH). URL: https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools [accessed 2025-02-27]
37. Background: development and use of study quality assessment tools. National Institutes of Health (NIH). URL: https://www.nhlbi.nih.gov/node/80102 [accessed 2025-02-27]
38. Anders P, Lehmann T, Müller H, Grønvik KB, Skjæret-Maroni N, Baumeister J, et al. Exergames inherently contain cognitive elements as indicated by cortical processing. Front Behav Neurosci 2018;12:102 [FREE Full text] [doi: 10.3389/fnbeh.2018.00102] [Medline: 29867400]
39. Scherer R, Moitzi G, Daly I, Muller-Putz GR. On the use of games for noninvasive EEG-based functional brain mapping. IEEE Trans. Comput. Intell. AI Games 2013;5(2):155-163. [doi: 10.1109/tciaig.2013.2250287]
40. Ghani U, Signal N, Niazi IK, Taylor D. Efficacy of a single-task ERP measure to evaluate cognitive workload during a novel exergame. Front Hum Neurosci 2021;15:742384 [FREE Full text] [doi: 10.3389/fnhum.2021.742384] [Medline: 34566610]
41. Parent M, Albuquerque I, Tiwari A, Cassani R, Gagnon J, Lafond D, et al. PASS: a multimodal database of physical activity and stress for mobile passive body/ brain-computer interface research. Front Neurosci 2020;14:542934 [FREE Full text] [doi: 10.3389/fnins.2020.542934] [Medline: 33363449]
42. Xu W, Liang HN, Zhang Z, Baghaei N. Studying the effect of display type and viewing perspective on user experience in virtual reality exergames. Games Health J 2020;9(6):405-414. [doi: 10.1089/g4h.2019.0102] [Medline: 32074463]
43. Ko J, Jang SW, Lee HT, Yun H, Kim YS. Effects of virtual reality and non-virtual reality exercises on the exercise capacity and concentration of users in a ski exergame: comparative study. JMIR Serious Games 2020;8(4):e16693 [FREE Full text] [doi: 10.2196/16693] [Medline: 33112240]
44. Elor A, Powell M, Mahmoodi E, Teodorescu M, Kurniawan S. Gaming beyond the novelty effect of immersive virtual reality for physical rehabilitation. IEEE Trans. Games 2022;14(1):107-115. [doi: 10.1109/tg.2021.3069445]
45. Baumeister J, Reinecke K, Cordes M, Lerch C, Weiss M. Brain activity in goal-directed movements in a real compared to a virtual environment using the nintendo wii. Neurosci Lett 2010;481(1):47-50. [doi: 10.1016/j.neulet.2010.06.051] [Medline: 20600604]
46. Kandemir H, Kose H. Development of adaptive human–computer interaction games to evaluate attention. Robotica 2021;40(1):56-76. [doi: 10.1017/s0263574721000370]
47. Pacheco TBF, Oliveira Rego IA, Campos TF, Cavalcanti FADC. Brain activity during a lower limb functional task in a real and virtual environment: a comparative study. NeuroRehabilitation: An International, Interdisciplinary Journal 2017;40(3):391-400. [doi: 10.3233/nre-161426]
48. Dang X, Wei R, Li G. An efficient movement and mental classification for children with autism based on motion and EEG features. J Ambient Intell Human Comput 2016;8(6):907-912. [doi: 10.1007/s12652-016-0424-x]
49. Fernandes ABGS, de Melo JCP, de Oliveira DC, Cavalcanti FADC, Postolache OA, Passos PJM, et al. Is motor learning of stroke patients in non-immersive virtual environment influenced by laterality of injury? A preliminary study. J Bodyw Mov Ther 2021;25:53-60. [doi: 10.1016/j.jbmt.2020.10.015] [Medline: 33714511]
50. Olyaei G, Khanmohammadi R, Talebian S, Hadian MR, Bagheri H, Najafi M. The effect of exergaming on cognition and brain activity in older adults: a motor- related cortical potential study. Physiol Behav 2022;255:113941. [doi: 10.1016/j.physbeh.2022.113941] [Medline: 35963295]
51. Müller H, Baumeister J, Bardal EM, Vereijken B, Skjæret-Maroni N. Exergaming in older adults: the effects of game characteristics on brain activity and physical activity. Front Aging Neurosci 2023;15:1143859 [FREE Full text] [doi: 10.3389/fnagi.2023.1143859] [Medline: 37213536]
52. GomezRomero-Borquez J, Del Puerto-Flores JA, Del-Valle-Soto C. Mapping EEG alpha activity: assessing concentration levels during player experience in virtual reality video games. Future Internet 2023;15(8):264. [doi: 10.3390/fi15080264]
53. Amprimo G, Rechichi I, Ferraris C, Olmo G. Measuring brain activation patterns from raw single-channel EEG during exergaming: a pilot study. Electronics 2023;12(3):623. [doi: 10.3390/electronics12030623]
54. Moinnereau MA, Oliveira AA, Falk TH. Instrumenting a virtual reality headset for at-home gamer experience monitoring and behavioural assessment. Front. Virtual Real 2022;3. [doi: 10.3389/frvir.2022.971054]
55. Mullen T. CleanLine. NeuroImaging Tools and Resources Collaboratory (NITRC). URL: https://www.nitrc.org/projects/cleanline [accessed 2025-02-27]
56. Ian Daly I, Pichiorri F, Faller J, Kaiser V, Kreilinger A, Scherer R. What does clean EEG look like. 2012 Presented at: Annual International Conference of the IEEE Engineering in Medicine and Biology Society; 2012 August 28; San Diego, CA, USA. [doi: 10.1109/embc.2012.6346834]
57. Bigdely-Shamlo N, Mullen T, Kothe C, Su K, Robbins KA. The PREP pipeline: standardized preprocessing for large-scale EEG analysis. Front Neuroinform 2015;9:16. [doi: 10.3389/fninf.2015.00016] [Medline: 26150785]
58. Pion-Tonachini L, Kreutz-Delgado K, Makeig S. The ICLabel dataset of electroencephalographic (EEG) independent component (IC) features. Data Brief 2019;25:104101. [doi: 10.1016/j.dib.2019.104101] [Medline: 31294058]
59. Casiez G, Roussel N, Vogel D. 1 € filter: a simple speed-based low-pass filter for noisy input in interactive systems. 2012 Presented at: CHI '12: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; 2012 May 05; Austin Texas USA p. 2527-2530. [doi: 10.1145/2207676.2208639]

XSL•FO
RenderX

60. Li R, Principe JC. Principe, Blinking artifact removal in cognitive EEG data using ICA, Conf. 2006 Presented at: International Conference of the IEEE Engineering in Medicine and Biology Society; 2006 August 30; New York, NY, USA p. 5273-5276. [doi: 10.1109/iembs.2006.260605]
61. Kang H. The prevention and handling of the missing data. Korean J Anesthesiol 2013;64(5):402-406 [FREE Full text] [doi: 10.4097/kjae.2013.64.5.402] [Medline: 23741561]
62. Mihajlovic V, Grundlehner B, Vullers R, Penders J. Wearable, wireless EEG solutions in daily life applications: what are we missing? IEEE J. Biomed. Health Inform 2015;19(1):6-21. [doi: 10.1109/jbhi.2014.2328317]
63. Morley A, Hill L, Kaditis A. Welcome to the ERS respiratory channel. ERS Respiratory channel. URL: http://www.ers-education.org/lrmedia/2016/pdf/298830.pdf [accessed 2025-02-27]
64. Casson AJ. Wearable EEG and beyond. Biomed Eng Lett 2019;9(1):53-71 [FREE Full text] [doi: 10.1007/s13534-018-00093-6] [Medline: 30956880]
65. Michel CM. High-resolution EEG. Handb. Clin. Neurol 2019;160:185-201. [doi: 10.1016/b978-0-444-64032-1.00012-6]
66. Chikhi S, Matton N, Blanchet S. EEG power spectral measures of cognitive workload: a meta-analysis. Psychophysiology 2022;59(6):e14009. [doi: 10.1111/psyp.14009] [Medline: 35128686]
67. Stanmore E, Stubbs B, Vancampfort D, de Bruin ED, Firth J. The effect of active video games on cognitive functioning in clinical and non-clinical populations: a meta-analysis of randomized controlled trials. Neurosci Biobehav Rev 2017;78:34-43 [FREE Full text] [doi: 10.1016/j.neubiorev.2017.04.011] [Medline: 28442405]
68. Comeras-Chueca C, Marin-Puyalto J, Matute-Llorente A, Vicente-Rodriguez G, Casajus JA, Gonzalez-Aguero A. The effects of active video games on health-related physical fitness and motor competence in children and adolescents with healthy weight: a systematic review and meta-analysis. Int J Environ Res Public Health 2021;18(13):6965 [FREE Full text] [doi: 10.3390/ijerph18136965] [Medline: 34209767]
69. Schmoigl-Tonis M, Schranz C, Müller-Putz GR. Methods for motion artifact reduction in online brain-computer interface experiments: a systematic review. Front Hum Neurosci 2023;17:1251690. [doi: 10.3389/fnhum.2023.1251690] [Medline: 37920561]
70. Amjad I, Toor H, Niazi IK, Pervaiz S, Jochumsen M, Shafique M, et al. Xbox 360 kinect cognitive games improve slowness, complexity of EEG, and cognitive functions in subjects with mild cognitive impairment: a randomized control trial. Games Health J 2019;8(2):144-152. [doi: 10.1089/g4h.2018.0029] [Medline: 30239213]
71. Schättin A, Baier C, Mai D, Klamroth-Marganska V, Herter-Aeberli I, de Bruin ED. Effects of exergame training combined with omega-3 fatty acids on the elderly brain: a randomized double-blind placebo-controlled trial. BMC Geriatr 2019;19(1):81 [FREE Full text] [doi: 10.1186/s12877-019-1084-4] [Medline: 30866834]
72. Jirayucharoensak S, Israsena P, Pan-ngum S, Hemrungrojn S, Maes M. A game-based neurofeedback training system to enhance cognitive performance in healthy elderly subjects and in patients with amnestic mild cognitive impairment. CIA 2019;Volume 14:347-360. [doi: 10.2147/cia.s189047]
73. Schättin A, Arner R, Gennaro F, de Bruin ED. Adaptations of prefrontal brain activity, executive functions, and gait in healthy elderly following exergame and balance training: a randomized-controlled study. Front Aging Neurosci 2016;8:278 [FREE Full text] [doi: 10.3389/fnagi.2016.00278] [Medline: 27932975]
74. Dimitrova J, Hogan M, Khader P, O'Hora D, Kilmartin L, Walsh JC, et al. Comparing the effects of an acute bout of physical exercise with an acute bout of interactive mental and physical exercise on electrophysiology and executive functioning in younger and older adults. Aging Clin Exp Res 2017;29(5):959-967. [doi: 10.1007/s40520-016-0683-6] [Medline: 27866346]
75. Zhang H, Wang D, Wang Y, Chi Y, Miao C. Development and validation of a practical instrument for evaluating players' familiarity with exergames. International Journal of Human-Computer Studies 2021;145:102521. [doi: 10.1016/j.ijhcs.2020.102521]
76. Dotson VM, Duarte A. The importance of diversity in cognitive neuroscience. Ann N Y Acad Sci 2020;1464(1):181-191. [doi: 10.1111/nyas.14268] [Medline: 31663150]
77. Rico-Olarte C, Lopez D, Narváez S, Farinango CD, Pharow PS. HapHop-Physio: a computer game to support cognitive therapies in children. PRBM 2017;Volume 10:209-217. [doi: 10.2147/prbm.s130998]
78. Rico-Olarte C, Lopez DM, Becker L, Eskofier B. Towards classifying cognitive performance by sensing electrodermal activity in children with specific learning disorders. IEEE Access 2020;8:196187-196196. [doi: 10.1109/access.2020.3033769]
79. Electrophysiological brain activity during a cognitive rehabilitation training for children with learning disorders. BayLat. 2023. URL: https://osf.io/abrfx/ [accessed 2024-07-25]

Abbreviations

BCI: brain-computer interface
EEG: electroencephalography
ICA: independent component analysis
MeSH: Medical Subject Headings
NHLBI: National Heart, Lung, and Blood Institute
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

XSL•FO
RenderX

VR: virtual reality

Edited by T Leung, N Cahill; submitted 18.07.23; peer-reviewed by B Somon, S Xiong; comments to author 07.11.23; revised version received 30.03.24; accepted 06.11.24; published 07.04.25.

Please cite as:
Rico-Olarte C, Lopez DM, Eskofier BM, Becker L
Electrophysiological Insights in Exergaming—Electroencephalography Data Recording and Movement Artifact Detection: Systematic Review
JMIR Serious Games 2025;13:e50992
URL: https://games.jmir.org/2025/1/e50992
doi:10.2196/50992
PMID:

©Carolina Rico-Olarte, Diego M Lopez, Bjoern M Eskofier, Linda Becker. Originally published in JMIR Serious Games (https://games.jmir.org), 07.04.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Effectiveness of Serious Games as Digital Therapeutics for Enhancing the Abilities of Children With Attention-Deficit/Hyperactivity Disorder (ADHD): Systematic Literature Review

Jing Lin[1], PhD; Woo-Rin Chang[1], Prof Dr

Department of Digital Contents, College of Art and Design, Kyung Hee University, Yongin-si, Republic of Korea

Corresponding Author:
Woo-Rin Chang, Prof Dr
Department of Digital Contents
College of Art and Design
Kyung Hee University
1732 Deogyeong-daero, Giheung-gu
Yongin-si, 17104
Republic of Korea
Phone: 82 01088818489
Email: lastanimator01@gmail.com

Abstract

Background: Attention-deficit/hyperactivity disorder (ADHD) is a neurodevelopmental disorder that often begins in childhood and requires long-term treatment and management. Given the potential adverse effects of pharmacological interventions in children, interest in alternative treatments has increased. Among alternative therapies, serious games have emerged as a promising digital therapeutic approach and are increasingly recognized as an important intervention for children with ADHD.

Objective: This systematic review aims to evaluate the effectiveness of serious games as digital therapeutics for children with ADHD. It focuses on assessing therapeutic outcomes, including improvements in attention, hyperactivity-impulsivity, social skills, motor skills, executive functions, and enjoyment.

Methods: The review was conducted following PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. A comprehensive literature search was performed across 5 databases: PubMed, Web of Science, Scopus, IEEE Xplore, and ACM Digital Library, covering English studies published from January 2010 to January 2024. Eligibility criteria were established based on the PICOS (Participants, Intervention, Comparison, Outcomes, Study design) framework, with digital therapeutics guidelines pragmatically applied to inform inclusion criteria, exclusion criteria, and quality assessment. Standardized tools including the Cochrane Risk of Bias Tool for randomized controlled trials, the Cochrane Risk of Bias Tool for Non-Randomized Studies of Interventions (ROBINS-I) for nonrandomized controlled trial studies, and the Critical Appraisal Skills Program checklists were used to evaluate risk of bias. Data on study design, targeted abilities, game software and hardware, and intervention parameters (duration, frequency, and length) were extracted and synthesized descriptively.

Results: Of the 35 studies identified (1408 participants), gender data were available for 22 studies (888 participants), comprising 660 male and 228 female participants. Analysis revealed multiple abilities focused across many studies: 80% (28/35) assessed attention, 29% (10/35) addressed hyperactivity-impulsivity, 17% (6/35) explored improvements in social skills, 20% (7/35) evaluated motor skills, and 43% (15/35) investigated executive functions. Furthermore, in 89% (31/35) of the trials, children exhibited a positive attitude toward game interventions. Evidence suggests that serious games may contribute to improvements in attention, hyperactivity-impulsivity, social skills, and executive functions in children with ADHD. Although findings on motor skills were inconclusive, interventions incorporating somatosensory inputs demonstrated benefits for hand-eye coordination.

Conclusions: The findings support the potential of serious games as a digital therapeutic modality for children with ADHD, offering benefits in core symptoms and associated impairments while promoting engagement.

Trial Registration: PROSPERO CRD420250509693; https://www.crd.york.ac.uk/PROSPERO/view/CRD420250509693

(JMIR Serious Games 2025;13:e60937) doi:10.2196/60937

XSL•FO
RenderX

KEYWORDS

serious games; ADHD; attention deficit disorder with hyperactivity; neurodevelopmental disorders; digital therapeutics; DTx; systematic review; pediatric; children

Introduction

Background

Attention-deficit/hyperactivity disorder (ADHD) is a neurodevelopmental disorder [1], often diagnosed in childhood and persisting into adolescence and adulthood [2]. It is associated with a high rate of comorbidity, accidents, and mortality, among others [3,4]. Although ADHD medications exhibit significant therapeutic effects [5], some guardians are reluctant to medicate their children due to potential long-term adverse developmental effects and varying effectiveness across different ADHD subtypes [6,7]. Furthermore, the core symptoms of ADHD, which include inattention and hyperactivity-impulsivity, typically necessitate multiple therapeutic interventions to ensure effective management [8].

Serious games, which have emerged as a promising alternative therapy, are defined as "games that do not have entertainment, enjoyment or fun as their primary purpose" [9]. Serious games are used in medical diagnostics, therapy, prevention, health promotion, and medical or patient education [10]. Serious games have been proposed for neurodevelopmental disorder interventions primarily due to their appeal, engagement, and effectiveness [11,12]. Meanwhile, music therapy [13], exercise therapy [14], and chess therapy [15,16], shown to be effective for managing ADHD symptoms in children, can achieve better therapeutic effects when combined with serious games.

In 2020, EndeavorRx became the first US Food and Drug Administration–authorized prescription digital therapeutic (DTx) specifically designed for pediatric ADHD, leveraging a video game experience to enhance attention in children with this condition [17]. Subsequently, in June 2023, the International Organization for Standardization (ISO) introduced ISO/TR 11147, which formally defined "Digital Therapeutics" as evidence-based, software-driven interventions to prevent, manage, or treat a medical disorder or disease [18]. This standardization by ISO facilitates industry and global alignment regarding the scope and implementation of DTx.

To date, only 2 systematic reviews have specifically examined the effects of games on children with ADHD [19,20], revealing the potential of such interventions to address core ADHD symptoms and underscoring the need for greater collaboration between developers and health care professionals. However, children with ADHD manifest a range of symptoms beyond the core symptoms of inattention and hyperactivity-impulsivity [21], including challenges in social interaction, impaired motor skills, and deficits in executive functioning [22,23]. In addition, further research on nonpharmacological treatments targeting specific abilities in pediatric patients with ADHD indicates that interventions designed to enhance social, motor, and executive functioning skills may potentially benefit overall ADHD symptom management [24]. The role of environmental and familial factors, such as maternal depression and emotional attitudes, has also been convincingly demonstrated to impact the development of executive functions in children [25]. Consequently, an UpToDate literature review is imperative to address existing gaps in research on capability enhancement, and examining player enjoyment remains critical.

The current evidence is insufficient to demonstrate the specific impact of serious games on children with ADHD, mainly because of several limitations: the limited number of studies included, the insufficient variety of games, and the missing and inconsistent reporting of results. Therefore, this review aimed to include more appropriate studies, expand the coverage of serious games, and diversify the outcome assessments, so that a more comprehensive review could be conducted to further systematically assess the available evidence related to the impact of serious games on the various abilities of children with ADHD.

Objectives

Compared with other systematic reviews, this study offers 3 unique contributions: First, it provides a comprehensive review of the most recent studies over the past 14 years, using the basic definition of DTx as part of the inclusion and exclusion criteria, thereby addressing the gap in recent studies that adhere to industry standards. Second, it offers a broader evaluation of the impact of serious games as interventions on children's competencies, including attention, hyperactivity-impulsivity, social skills, motor skills, and executive functions. Third, it discusses outcomes related to participation, like enjoyment and adverse effects.

Methods

Overview

A total of 2 literature search phases were conducted in January 2024 and February 2025. The exploratory search phase identified 4420 records. Using Zotero Reference Management software, we automatically removed 1022 duplicate records. Subsequently, we manually deleted an additional 7 duplicates and 1 retracted article, resulting in a final set of 3390 records. Based on the exploration, the second search yielded 5109 records. After removing 1492 duplicate records and 2 retracted articles, and cross-validating with the exploratory search, 5510 records were retained for evaluation. The PICOS framework (Participants, Intervention, Comparison, Outcomes, Study design) informed both search strategy development and eligibility criteria formulation. This systematic review adhered to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines (Multimedia Appendix 1) [26]. In addition, we followed the recommendations of the Cochrane Consortium for conducting systematic reviews and used the RefHunter website for guidance [27]. For the final selection of 35 articles, established assessment tools were used, including the Cochrane Risk of Bias Tool for randomized controlled trials (RCTs), the Cochrane Risk of Bias Tool for Non-Randomized Studies of Interventions (ROBINS-I), and the CASP (Critical

XSL•FO
RenderX

Appraisal Skills Programme) checklists to evaluate risk of bias of each study.

Identification: Literature Search

The PICOS criterion [28] was used to guide the literature search strategy and identify inclusion and exclusion criteria. During the development of the literature search strategy, to ensure that the search covered a wider range of related studies, the search strategy was based on the broader P (Participants, patients with ADHD) and I (Intervention, games) without strictly limiting C (Comparison), O (Outcomes) and S (Study design). This is because existing studies cover a wide range of experimental designs, including exploratory and qualitative studies with no control group. A strict search strategy may limit the scope of studies, resulting in potentially valuable literature being missed. Also, the search strategy did not limit the age, but we limited the inclusion criteria to children with ADHD. Considering the interdisciplinary nature of this review, the selected databases should include relevant research literature from diverse disciplines, extending beyond psychology, game design, computer science, and medicine. Consequently, databases such as PubMed, Web of Science, Scopus, IEEE Xplore, and ACM Digital Library were chosen to address the needs of interdisciplinary research.

Close synonyms for “Serious Games” include “gamification,” “gamified,” “game mechanics,” “game dynamics,” “game design,” “game based,” and “gaming.” [29]. To minimize limitations, “DTx” and “DHI” (digital health interventions) were added to the search terms. DTx is considered a subset of “DHI,” a broader term that encompasses the use of digital technologies to support wellness and health care practices, particularly in the management of neurodevelopmental disorders [30-32].

A 2-phase literature search strategy was used. The initial exploration phase spanned January to March 2024. Relevant search terms for “ADHD” and “Serious Games” were identified by systematically mapping domain-specific subject headings and keywords informed by expert knowledge. The exploratory phase identified 4420 publications, with deduplication yielding 3390 records. Index terms, titles, and abstracts of retrieved articles were systematically analyzed to extract search terms informing subsequent query optimization. Relevant articles from the exploratory phase were archived for cross-verification during the second search screening.

The second search implementation occurred in February 2025, incorporating emergent terminology from exploratory phase findings, including educational gaming, exergaming, cognitive training, telemedicine, and mobile health. Consultations with a medical research librarian and pediatric ADHD specialist informed MeSH (Medical Subject Headings) term selection and Boolean operator construction (Multimedia Appendix 2).

Screening: Title and Abstract Screening

PICOS criterion guided inclusion and exclusion criteria and added other related labels. Duplicates were automatically identified using Zotero software (version 6.0.29, Corporation for Digital Scholarship, Roy Rosenzweig Center for History and New Media). Following iterative deduplication, publications underwent manual screening. Initially, 2 authors independently screened the titles and abstracts of relevant studies. The full texts of eligible studies were subsequently assessed independently by the same authors using inclusion criteria, with any arising conflicts resolved through conference discussions. Each study was evaluated against these criteria, resulting in 5210 exclusions for various reasons, including inappropriate target population, interventions not involving a game, unreasonable experimental design, nonjournal publication, and language not being English (Table 1).

XSL•FO
RenderX

Table 1. Title and abstract screening Participants, Intervention, Comparison, Outcomes, Study design (PICOS) diagram for children with attention-deficit/hyperactivity disorder (ADHD) game interventions (2010-2024). The systematic review screening process for ADHD children game studies is based on the PICOS framework.

	Inclusion criteria	Exclusion criteria[a]
Patient	• Children diagnosed with ADHD.	• Studies that did not only include children with ADHD and/or their caregivers (guardians, teachers, and clinicians). • Nonhuman studies.
Intervention	• Research focused on serious games designed specifically for targeting ADHD: cognitive training games, neurofeedback-based games, exergames, and social skills training games.	• Studies focused on other interventions (eg, pharmacological interventions, Pele medicine, mobile health and sensors, web-based intervention, augmented or virtual reality, and robot assistants).
Comparison	• The comparative arms could include various interventions or usual care.	• __[b]
Outcome	• Focus on outcomes in the following developmental domains: attention, hyperactivity-impulsivity, social skills, motor skills or physical activities, executive functions, enjoyment, and intervention adherence.	• Studies lacking concrete examples or evidence. • Examples are not games or game-based platforms or software. • Studies where the original text cannot be retrieved. • Repeated trials.
Study Design	• Any experimental or quasi-experimental evaluative design, including pilot and feasibility studies. • Nonrandomized studies (eg, pre-post study with no control group). • Cohort or longitudinal studies that had pre-post outcome measures. • Case series or case studies.	• Theoretical designs or frameworks without example data. • Studies addressing conditions other than ADHD (eg, autism spectrum disorder, anxiety disorders, bipolar disorder, and learning disorders). • Studies not aimed at treating or alleviating ADHD symptoms.
Publication Type	• Peer-reviewed article. • Full paper proceedings.	• Conference abstracts, study protocols, books, websites, reviews, theses or dissertations, short conference paper proceedings, posters, and demos.
Publication period	• From 1 January 2010 to 1 January 2024	• Before 2010, after 1 January 2024
Setting	• Any country or region	• Not applicable
Language	• English	• Any other language
Aggregate	• Included (300)	• Excluded (5210)

[a]Exclusion criteria include nonrelevant study populations, interventions lacking a game component, inappropriate study designs, nonjournal publications, and non–English-language articles.

[b]Not applicable.

Eligibility: Content Screening

Following the preliminary screening based on inclusion criteria, 300 studies were analyzed using the full-text exclusion assessment for children with ADHD game interventions as a benchmark for further content screening. Each study was evaluated against these criteria, resulting in 261 exclusions for various reasons such as lack of relevance to ADHD-only populations, excluding studies not focused on ADHD treatment, nonchild populations, nongame platforms, or lacking empirical data (Table 2). Subsequently, a quality assessment was conducted, where each criterion was rated as “yes,” “partially,” or “no,” and scored, respectively, with values of 1, 0.5, and 0. The final score, requiring a minimum of 4 out of 7 (about 60%) marks, was calculated as an average of the grades. The quality criteria and scoring methodology were based on research by Coelho et al [33]. This step led to excluding 4 articles that did not meet the minimum threshold of 0.6, while 35 other articles were included (Multimedia Appendix 3) [14,34-67].

In our literature screening process, it is crucial to ensure that the research selected develops a game specifically designed for treating children with ADHD, possesses a fundamental game structure, and is sufficiently comprehensive to allow for an analysis of its usefulness [68]. However, as most games evaluated are not yet commercially available, the 8 DTx guidelines were pragmatically applied to the design of the exclusion criteria and quality assessment, drawing upon the concept of the minimum viable product (MVP). Some guidelines were temporarily deprioritized (Table 3). Despite the limited availability of MVP-related content in digital contexts, the MVP concept continues to offer the potential for enhancing product development efficiency, quality, and innovation [69].

XSL•FO
RenderX

Table 2. Full-text exclusion assessment for children with attention-deficit/hyperactivity disorder (ADHD) game interventions. Details the full-text review phase, excluding studies not focused on attention deficit hyperactivity disorder treatment, nonchild populations, nongame platforms, or lacking empirical data.

Exclusion criteria	Excluded, n
Full-text articles excluded	
Studies addressing conditions other than ADHD	62
Studies that included not only children with ADHD and/or their caregivers	54
Studies not aimed at treating or alleviating ADHD symptoms	41
Studies lack concrete examples or evidence	35
Examples are not games/game-based platforms or software	39
Studies focused on other interventions	8
Studies where the original text cannot be retrieved	4
Repeated trials	18
Quality assessment	
Substandard quality	4
Aggregate	
Included: n=35	265

Table 3. Design of minimum viable product quality evaluation in children with attention-deficit/hyperactivity disorder game interventions as digital therapeutics. The 8 digital therapeutics guidelines were pragmatically integrated with minimum viable product principles to balance game design feasibility and therapeutic relevance.

DTx[a] core principles	Reasons for deprioritization	Prioritization recovery
Incorporate design, manufacturing, and quality best practices.	—[b]	—
Engage end users in product development and usability processes.	—	—
Incorporate patient privacy and security protections.	—	—
Apply product deployment, management, and maintenance best practices.	During initial development and prototyping, the primary focus might be on validating the therapeutic concept and functionality rather than on the complexities of deployment and ongoing maintenance.	Applying product deployment, management, and maintenance becomes critical as the product approaches a state ready for broader testing or market launch, to ensure scalability, reliability, and user support.
Publish trial results inclusive of clinically meaningful outcomes in peer-reviewed journals.	—	—
Be reviewed and cleared or certified by regulatory bodies as required to support product claims of risk, efficacy, and intended use.	Early-stage products, particularly those in the research, discovery, or pre-clinical phases, may not yet have undergone regulatory review.	The process of obtaining regulatory clearance or certification often comes after proving the product's efficacy and safety, which are established through clinical trials.
Make claims appropriate to clinical evaluation and regulatory status.	In the exploratory stages of development, a DTx product might not yet have undergone extensive clinical evaluation, and thus cannot make definitive claims about its efficacy, safety, or intended use.	Formal claims become relevant and necessary as the product undergoes clinical trials and seeks regulatory approval.
Collect, analyze, and apply real-world evidence and/or product performance data.	—	—

[a]DTx: digital therapeutics.

[b]Not available.

Included: Data Abstraction and Analysis

To screen the final record set for inclusion criteria, after manually removing duplicates (identical game instances with different paper titles), 2 authors used Zotero to further screen the publications manually. Together, the 2 authors determined keywords and selected databases in step 1. In the screening step, the first author screened the titles and abstracts of all records,

XSL•FO
RenderX

applying inclusion criteria to identify obvious exclusions. Subsequently, in the eligibility step, the 2 authors independently reviewed the remaining papers and resolved any conflicts through discussion, arriving at the final decision.

Significant heterogeneity in study types, comparisons, intervention programs, and outcomes indicated that the data from this review were not suitable for meta-analysis. Therefore, descriptive analysis was used to summarize the impact of serious games on the rehabilitation of children with ADHD. The process of conducting descriptive analysis involved tabulating features of the original studies from 2 perspectives: experimental design and game design. This included software, description of gameplay, duration, length and frequency, sample size and characteristics, and sex of children participants. Jing Lin and Woo-Rin Chang independently extracted this information using predefined Microsoft Excel (Table 4).

Subsequently, the authors thoroughly reviewed and validated the collected data to ensure accuracy and reliability. In cases of disagreement, consensus was reached through discussion. Finally, the data were organized and categorized for each specific outcome to enable a comprehensive assessment of the findings.

XSL•FO
RenderX

Table 4. Characteristics of included studies.

Studies	Software, sample size, and characteristics	Description and sex of the child participants	Duration, length, and frequency
Giannaraki et al [34]	• ADDventurous Rhythmical Planet • 4 children (including 2 with ADHD[a]) and 4 expert educators	• Players use the tin drum to produce rhythms and collaborate in multiuser mode to achieve common goals. • 3 male and 1 female	• —[b]
Rodrigo-Yanguas et al [35]	• The Secret Trail of Moon • 37 children and adolescents	• Including 5 VR[c] minigames, each aimed at enhancing specific cognitive skills. • 25 male and 12 female	• Single sessions for 10-40 min
Aghdam and Alavi [36]	• MIND PRO Working Memory Game • 6 children with ADHD, ages 6 to 11 years	• In the auditory section of the game, players identify and select boxes that produce distinct sounds from multiple options • The visual section requires players to select boxes that contain matching images • 4 male and 2 female	• 30 min per session, 12 sessions total
Batista et al [37]	• Taboo! • 184 children with ADHD, ages 7-14 years	• The game uses an endless runner format where the player avoids obstacles and solves math problems • Predominantly male	• —
Wiguna et al [38]	• Indonesian computer-based game prototype • 10 children with drug-naïve ADHD, aged 7-12 years	• A role-playing game features children acting as fruit car drivers, tasked with delivering color-specific fruits to corresponding houses • 9 male and 1 female	• 30 min per session, 20 sessions total over 4 weeks
Ou et al [39]	• Fishing Master/Fruit Train/Ocean Manager • 3 children aged 8-12 years old diagnosed with ADHD	• The games enhance hand-eye coordination and balance through activities like catching fish, picking fruit on a moving train, and feeding fish while dodging obstacles • 1 male and 2 female	• 40 min per session, 3 times a week for 3 months
Capelo et al [40]	• Multisensory Virtual Game • 20 children aged 7-12 years, including both children with and without ADHD	• In a multisensory game, children use the Leap Motion controller to match virtual objects to containers based on their color and shape • Mixed (10 children with ADHD and including 4 female and 6 male; 10 children without ADHD and including 4 female and 6 male)	• —
García-Baos et al [41]	• RECOGNeyes • 28 children aged 8-15 years diagnosed with ADHD	• In the word recognition task, participants need to identify the correct words displayed among scrambled words at the screen's center • 18 male and 10 female	• 30 min per session, 3 times per week for 3 weeks
Avila-Pesantez et al [42]	• ATHYNOS • 11 children diagnosed with ADHD, aged 7-10 years	• AR game, activities in the ATHYNOS prototype include “Drag and Drop” and “Shapes,” focusing on improving hand-eye coordination • 9 male and 2 female	• 20 min per session, 2 times per week during a month
Kanellos et al [43]	• REEFOCUS • 75 children diagnosed with ADHD aged 8-14 years	• A management simulation game set in an underwater environment where players engage in tasks	• 45 min per session for 8 weeks
Kollins et al [44]	• AKL-T01(An early version of Endeavor-Rx) • 206 children. 130 children continued stimulant medication during the study. 76 children, no ADHD medication.	• Players control a character navigating a colorful, animated world on a fixed path • 154 male and 52 female	• 12 weeks, split into two 4-week treatment phases with a 4-week pause in between • Daily sessions of ~25 minutes, consisting of five 5-minute missions, 5 days a week

XSL•FO
RenderX

Studies	Software, sample size, and characteristics	Description and sex of the child participants	Duration, length, and frequency
Machado et al [45]	• 3 children aged 7-12 years are diagnosed with ADHD	• A dry a wireless electroencephalography headset captures the player's brain activity, gameplay involves tasks like collecting stars or managing fuel levels • 3 male	• 20-60 min per session
De la Guía et al [46]	• StiCap • 12 children with ADHD, ages 5-16 years	• The games use tangible user interfaces with RFID technology for engaging learning activities • 8 male and 4 female	• —
Avila-Pesantez et al [47]	• CIUDAD PUZZLE • 20 children diagnosed with ADHD	• A puzzle game that incorporates NFC tags to interact with physical game pieces • 14 male and 6 female	• 15 min per session, 2 times per week for 3 months
Schena et al [48]	• IAmHero • 60 children with ADHD, age 5-12 years	• The application features 3 VR mini-games to enhance various abilities • 33 male and 27 female	• 30 min weekly for 6 months
Blandón et al [49]	• Harvest Challenge • 9 children diagnosed with ADHD, ages 5 to 12 years	• Gameplay is controlled by monitoring attention levels via a Brain-Computer Interface (BCI)	• 30 min per session, 2 sessions total
Vita and Mennitto [50]	• NEUROBOT • A 10-year-old child diagnosed with ADHD	• Participants control the speed of a Lego robot by concentrating and using brain wave activity to compete in races, aiming to increase the robot's speed	• 30 min weekly
Chen et al [51]	• 10 children diagnosed with ADHD	• The game features tasks that require players to respond in real time to changes in their brainwave status	• Twice a week, 30-60 min per session for 14 sessions
Soysal et al [52]	• Tetris • 6 children, including 2 diagnosed with ADHD	• The study explored how different types of background music affect the attention levels of children while playing Tetris	• —
Weerdmeester et al [53]	• Adventurous Dreaming Highflying Dragon • 73 school-aged children with elevated ADHD symptoms	• Players control avatars through full-body movements to solve cognitively challenging tasks in the game • 58 male and 15 female	• 15 min per session, 2 times per week for 3 weeks
Jácome et al [54]	• DIVIDI2 • 5 children diagnosed with ADHD	• Set on Mars, the gameplay transforms each child into an astronaut on various adventures • 3 male and 2 female	• 2 weekly sessions for a month
Kim et al [55]	• Eye-contact game • 40 children diagnosed with ADHD by DSM-5[d] criteria	• An eye contact game encourages children to recognize faces and improve attention and social interaction skills	• 30 min per session, 15 sessions total over 6 weeks
Martínez et al [56]	• KAPEAN • 15 children with ADHD	• Children can interact with digital games using either a traditional mouse or hand gestures detected by a Leap Motion device	• 10 min for 4 months
Crepaldi et al [57]	• Antonyms • 16 boys aged 8–11, including 8 with ADHD and 8 controls	• Players roleplay as superheroes in a realm where usual rules are inverted, needing to restrain impulsive actions to advance • 16 male	• 45 min per session
Bul et al [58]	• Plan-It Commander • 170 children with ADHD diagnosis, aged 8-12 years, and their parents and teachers	• The game engages players in a 10-mission online adventure, requiring specific skills to overcome challenges guided by a narrative • 137 male and 33 female	• 65 min per session, 3 times per week, over 20 weeks

XSL•FO
RenderX

Studies	Software, sample size, and characteristics	Description and sex of the child participants	Duration, length, and frequency
Castro and Huaman-chahua [59]	• Casa de Spots • 35 children (19 with ADHD, 16 without ADHD)	• The math learning game covers basic mathematics, including even and odd numbers, fractions, and geometric shapes	• —
Celis et al [60]	• Dilud • 15 children with ADHD and their parents	• Interactive and dynamic games designed to reinforce rote learning in children with ADHD employ the Tajima Cognitive Test method	• 10 min daily
Retalis et al [61]	• Kinems • 11 children diagnosed with ADHD	• Touchless motion-based games use the player's body movements and gestures for interaction • 10 male and 1 female	• 30 min per session, 2-3 times per week for a month
Dovis et al [62]	• Braingame Brian • 89 children with a clinical diagnosis of ADHD, aged 8-12 years	• Three minigames are designed to target distinct cognitive skills: working memory, inhibition, and cognitive flexibility • 25 male and 6 female in full active; 22 male and 6 female in partially active; 24 male and 6 female in placebo	• 35-50 min per session, 5 times per week for 5 weeks
Gizatdinova et al [63]	• PigScape • 10 Children with ADHD	• The game alternates between active and still phases to enhance impulse control, challenging children to mimic and hold postures	• —
Ji et al [14]	• Alchemist's Treasure • 30 children with ADHD (8-12 years).	• Combines physical activity (running/jumping on a sensor-equipped board) with cognitive challenges (avoiding obstacles, collecting items) • 26 male and 4 female	• 3 sessions per week, 50 min per day for 4 weeks
Ahmadi et al [64]	• TARLAN • 60 children (8-12 years): 40 with ADHD and 20 non-ADHD with social skills deficits	• Simulation game with 40 scenarios using the SOCCSS model across four social contexts	• 8 sessions, 2 sessions per week, 20-30 minutes each, 3 phases
García-Redondo et al [65]	• "Boogies Academy" and "Culbrain" • 44 children with ADHD and SLD (6-16 years).	• Each includes 10 subgames targeting specific intelligence (eg, logical-mathematical, spatial) • 27 male and 17 female	• 28 sessions, 2 sessions per week, 10 min per session
Barba et al [66]	• BRAVO • 60 children (3-12 years old)	• Three minigames, including spatial reasoning in virtual environments, obstacle avoidance with rule adherence and teamwork in a spaceship scenario	• 27 weeks
Kim et al [67]	• NeuroWorld DTx[e] • 30 children (6-13 years old), 15 experimental, 15 control	• Glass bridge recognition, spatial obstacle avoidance • Remembering animal sounds and counts • 23 male and 7 female	• 20 sessions, 30 minutes per day for 4 weeks

[a]ADHD: attention-deficit/hyperactivity disorder.

[b]Not available.

[c]VR: virtual reality.

[d]DSM-5: Diagnostic and Statistical Manual of Mental Disorders (Fifth Edition).

[e]DTx: digital therapeutics.

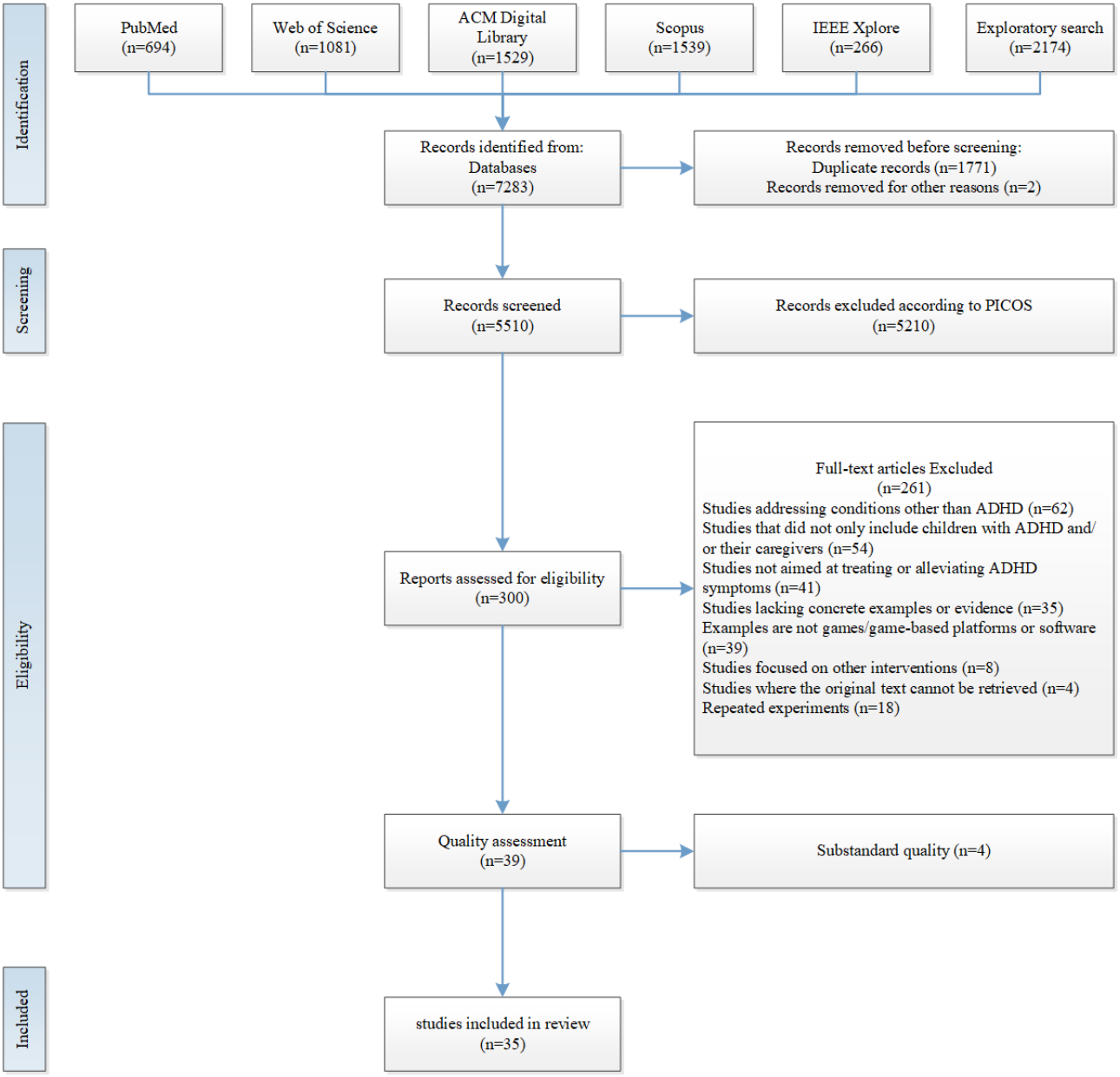
Results

Overview

The optimized search strategy identified 7283 additional records. After cross-validation with the exploration phase, with 1771 duplicates and 2 retracted articles excluded, yielding 5510 articles were aggregated using Zotero (version 6.0.36). After screening titles and abstracts, 300 articles met the inclusion criteria, and upon a complete reading of each article, 261 articles met the exclusion criteria. Finally, after a quality assessment excluded 4 articles, 35 publications were screened to fulfill the basic DTx criteria (Figure 1).

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram. ADHD: attention deficit hyperactivity disorder; PICOS: Participants, Intervention, Comparison, Outcomes, and Study.



The 35 papers in our review were published in journals or conference proceedings across a variety of fields (based on the journal descriptions), as described in the journals. Of these, 14 were from interdisciplinary publications, 8 from computer science, 7 from psychology and medicine, 3 from education, and 3 from design. Among the game examples, wearables and sensors were used in 37% of papers (13/35), 37% (13/35) being 3D games and 63% (22/35) being 2D, and most were single player (31/35, 89%; Multimedia Appendix 4).

The selected papers encompassed 1408 participants, with sample sizes ranging from 1 to 206. Gender was recorded for 22 of the 35 trials, totaling 888 children, 660 male and 228 female. These data confirm the established sex disparity in ADHD prevalence, with male participants demonstrating significantly higher diagnosis rates than female participants [70]. Notable methodological limitations were identified, particularly regarding sample size constraints (median 20). Among the included studies, 28 provided intervention duration, and 64% (18) implemented interventions lasting fewer than 8 weeks, coupled with heterogeneous outcome measurement approaches. As some of the trials were quantitative research or qualitative research, there were issues of heterogeneity and bias or the results of competence enhancement were not rigorously quantified in the conclusions, we used a uniform use of descriptive analyses to elaborate on the findings (Table 4).

Enjoyment

In total, 89% (31/35) of the studies investigated participants' enjoyment of these serious games. The results on enjoyment were generally positive (Multimedia Appendix 5) [14,34-67], with 19% (6/31) of the trials including negative results on enjoyment. Children as the primary target group not only

XSL•FO
RenderX

expressed their enjoyment through interviews and questionnaires, but some also directly expressed a desire to continue playing the game [35,54]. Although the games initially appeared very appealing, interest varied over time and could wane once children mastered them, leading to reduced enthusiasm [39,42]. However, children demonstrated consistently high satisfaction with the game intervention in most trials. In 45% (14/31) of the studies, player enjoyment was maintained by customizing the difficulty. Most measures of enjoyment were obtained through pre- and postintervention observations and assessments, satisfaction surveys, and electroencephalography (EEG) signal analyses. In particular, the game StiCap measures player enjoyment by testing the user experience with Smileyometer tests [46].

Attention

Attention, identified as the most competency in need of improvement in children with ADHD, received focus in 80% (28/35) of the studies [14,34-56,64-67], with 2 studies emphasizing visual attention [41,65]. Among these games, 46% (13/28) of the studies focused on using carefully crafted game rules to promote attention [34-43,64,65,67]. Furthermore, 36% (10/28) of the studies concentrated on attentional training using various technologies [14,45-52,66]. Specifically, 7 studies used EEG headsets to capture brainwave data for observing attention levels [43,49-52,56,66], and 2 of these used attention levels, as measured by the brain-computer interface, to control gameplay [49,52]. 21% (6/28) of the studies focused on using in-game interaction design to motivate player focus, with 83% (5/6) using a Kinect sensor to recognize players' movements or gestures and input this data into the virtual world to generate interactions [14,42,53,56,66], while 17% (1/6) applied design principles from the EMOINAD guide for mobile interfaces [54]. Notably, KAPEAN and BRAVO integrated wearable sensors, leveraging Kinect and electrocardiography to adapt therapy [56,66].

The studies consistently demonstrate significant improvements in attention spans among ADHD children across various gaming interventions (Multimedia Appendix 5) [14,34-67]. Changes in attention before and after the trials were assessed using various methods, the most common being standardized tests such as the Conners-3 scales, the Italian Battery for ADHD (BIA), the Attention Test for Elementary School Children (ATESC), and Swanson, Nolan, and Pelham Rating Scale, Version IV ratings, along with behavioral observations and electroencephalography signal analyses. Specifically, the game CIUDAD PUZZLE uses the perception of differences test (Faces-R) to measure improvements in attention performance [47].

Hyperactivity-Impulsivity

In total, 29% (10/35) of studies focused on research related to hyperactivity-impulsivity; 30% (3/10) of studies focused on the interaction design and user experience [14,43,53], and 30% (3/10) of studies used virtual reality to create an immersive therapeutic environment for relief of hyperactivity-impulsivity [35,39,48]. Another 30% (3/10) of the studies focused on improving hyperactivity-impulsivity symptoms with counterintuitive game rules designed to stimulate impulsive responses, requiring players to inhibit these responses to succeed, and the game's difficulty can be adjusted based on the child's progress [57,63,67]. Hyperactivity-impulsivity symptom improvement was measured primarily by direct observation and decreased hyperactivity-impulsivity scores. Although some of the studies were not statistically emphasized, and 1 study reported no significant changes in hyperactivity-impulsivity symptoms (Evaluation of the Deficit of Attention and Hyperactivity scale) following a gaming intervention [65], it was possible to see improvements in inhibitory control in most studies.

Social Skills

In total, 17% (6/35) of the studies focused on research related to social skills. In 3 of these studies, the single-player game Plan-It Commander, TARLAN, and BRAVO provide an in-game interactive environment where children learn and practice social skills through task challenges [58,64,66]. The eye-contact game requires participants to match and maintain eye contact with a virtual character [55]. The multiplayer game PigScape supports colocated gameplay, meaning that children play in pairs. This setup aims to create a learning environment in which children can support each other and enhance their social interactions and communication skills [63]. In particular, ADDventurous Rhythmical Planet offers the option to play alone or in collaboration with others, using a multiuser mode that fosters teamwork and communication. In addition to observational assessments, social skills can be scientifically assessed using the social skills rating system [34]. Although some of the trials were not statistically significant though not statistically significant, the results were generally positive.

Motor Skills

In total, 7 (7/35, 20%) studies focused on motor skills. 57% (4/7) of these studies were designed to engage children in physical activities, using their body movements to interact with the game through a Kinect camera [42,53,61,63]. The remaining 3 studies used VR technology to train body coordination, specifically focusing on hand-eye and hand-foot coordination [39,43,66]. Children were observed to have enhanced hand-eye coordination because of interacting with the interface, but no significant changes in gross motor skills were noted, suggesting no targeted training or measurement for substantial motor skill development [39,42,53].

Executive Functions

In total, 43% (15/35) of the studies focused on enhancing executive functioning, comprising 3 core functions: working memory, cognitive flexibility, and inhibitory control. Inhibitory control is discussed in the hyperactivity-impulsivity section. A total of 87% (13/15) of studies targeted working memory, primarily through repetitive practice in the games, with 2 studies focused on enhancing math skills [37,59]. Around 13% (2/15) of the studies aimed at enhancing children's emotion regulation [38,56] were viewed as integrating cognitive flexibility and inhibitory control. Four of these studies enhanced children's cognitive flexibility and problem-solving through virtual reality environments [35,39,48,66]. A total of 5 studies reported synthesizing improvements in executive function through multiple mini-games [39,48,58,64,66].

XSL•FO
RenderX

Most of the trials concluded that serious games had positive effects on executive functions, except for one report where enhanced cognitive flexibility and problem-solving abilities were observed in 1 participant through improved Wisconsin Card Sorting Test (WCST) scores, although the results for the other participants were not favorable [39]. Improvements in executive functioning are evaluated primarily through observational assessments by parents and teachers, and through scientific tests such as the Wechsler Memory Scale, Tower of London task (TOL), Behavioral Rating Inventory of Executive Function (BRIEF), and WCST. In addition, it is assessed by the scores achieved by the children during the game [60,62], such as behavioral tests like the Frankfurt Attention Inventory, Go/No-go [14], and through electroencephalography technology to monitor the children's emotional and cognitive states during gameplay [56].

Discussion

Principal Findings

To our knowledge, this is not the inaugural systematic review in this field. However, it incorporates DTx guidelines to provide current and industry-standard insights, critically examining many original studies, serious games of various types, and outcomes. In addition, it addresses the multisymptom nature of ADHD, focusing on enhancements in social interactions, motor skills, and executive functions, extending beyond the core symptoms of inattention, and hyperactivity-impulsivity, to include children's enjoyment of games. The review suggests that serious games hold great promise in the field of treatment for children with ADHD. However, findings were interpreted cautiously using descriptive analyses due to the wide variation in study design, sample size, targeting ability and duration, outcomes, and associated risks.

The results indicate that serious games may be beneficial for improving inattention, potentially serving as an effective alternative or supplement to traditional rehabilitation methods for key ADHD symptoms. Changes in EEG patterns, especially in the alpha and beta bands, support the effectiveness of serious games for enhancing attention [49]. Serious games provide a more engaging and digital format, enhancing enjoyment and promoting adherence among children [35,62]. This factor has been recognized as crucial for improving adherence to interventions [54]. The findings reveal that serious games use feedback mechanisms and engaging digital experience through game rules, technology, and interaction design to maximize participant motivation and minimize boredom [67,71,72]. Serious games can also be tailored to individual patient needs, resulting in more personalized and effective treatment. However, further research is necessary to examine the long-term effects of serious games on attention.

The integration of serious games into ADHD interventions provides a promising avenue for addressing hyperactivity-impulsivity symptoms in children. The aspect of impulsivity in ADHD, characterized by rash decision-making and difficulty in delaying gratification, can be addressed through games that require players to make calculated decisions to progress. Games that incorporate delay mechanisms or require the player to strategize can help in cultivating patience and strategic thinking, thus directly engaging with impulsivity control mechanisms. The games which focus on inhibitory control and cognitive flexibility, are reported to enhance the ability to suppress inappropriate motor responses and reduce impulsive behaviors in ADHD children [57,63]. Specifically, games that involve physical activity and require controlled responses have shown potential in managing these symptoms by improving motor skills and behavioral responses [39,48,53]. These findings are aligned with our observations where engagement in game-based tasks that demanded physical interaction not only helped in controlling impulsivity but also appeared to improve overall motor coordination [53] However, while digital games show promise, their effectiveness can vary based on the individual's specific characteristics and the nature of the game mechanics. Games that require consistent engagement and provide immediate feedback, such as those with a narrative or competitive elements, may sustain attention better and lead to more significant improvements in hyperactive and impulsive behaviors [48,57].

The results of the study suggest that serious games may enhance the social skills of children with ADHD. Multiplayer gaming is not a requirement, games are primarily set in virtual environments that facilitate interaction or collaboration among peers, thereby aiding children with ADHD in developing their social skills. Most studies report consistent results, showing improvement in children's social skills after using serious games, compared with the pre-experimental period. Children found the games engaging and expressed a desire to share their experiences with friends. However, when playing games with peers, children may experience frustration from not keeping up with friends for various reasons, leading to a loss of interest in continuing the game [34].

The current study does not find evidence to suggest a positive effect of serious games on the motor abilities of children with ADHD. However, the use of a Kinect sensor for full-body interaction during serious games is reported to be more engaging than traditional forms of exercise, enhancing fine motor skills, but there are no significant changes in gross motor skills [53]. In addition, limitations in motor skills enhancement results may stem from variations in measurement methods and study quality. Most studies concentrate on improving attention and inhibitory control in children with ADHD through exercise or physical activity, rather than sports instruction. Further research examining motor practice in serious games could shed more light on potential benefits for individuals with ADHD. This review also indicates that somatosensory digital systems, which use body movements and gestures as inputs, may achieve higher adherence rates than traditional games [63]. This underscores the importance of incorporating digital systems in serious gaming interventions, which could significantly affect their effectiveness. However, adherence rates may vary by population and specific intervention [73], necessitating further research to fully assess the impact of digital systems on adherence in serious gaming interventions.

Findings suggest that serious games may be beneficial and may improve executive functioning in children with ADHD. The studies in this review focused on working memory for executive

XSL•FO
RenderX

functions, such as math skills, inhibitory control-related impulsivity symptoms, and emotion regulation. It should be emphasized that improvements in executive functioning may be influenced by a variety of factors and therefore may result in a lack of observed improvements [74]. Improvements in working memory can be more intuitively reflected in scores on games, and inhibitory control-related abilities are improved to some extent primarily by completing impulse-suppressing game challenges. Beyond this, it is not clear whether long-term interventions lead to significant improvements, which is consistent with the results of previous reviews [75].

The results of the current research on the enjoyment of serious games are generally positive, some of the studies that reported adverse outcomes involved difficult, reproducible, and device-related adverse events [44]. As the aim of these serious games is to improve various symptoms of ADHD, this often results in a relatively repetitive gameplay process. While games are initially appealing, participant interest varies over time, making compliance with the program one of the most significant challenges. [38,39,45]. Therefore, controlling the game's difficulty is crucial, a game that is too difficult can frustrate children, leading them to give up [35,53,56]. However, games that are too easy may become uninteresting to children once mastered, leading to reduced enthusiasm unless new games are introduced. Of the 35 studies reviewed, 19 discussed customizing the game: 4 were customized for appearance only, while the remaining 15 involved adjusting levels and tasks according to the player's performance. This suggests that game designers, to prevent boredom during gameplay, use progressively increasing levels of difficulty to stimulate children's desire for a challenge [76]. For example, in "Alchemist's Treasure," exercise intensity was adjusted to 60%-80% of heart rate reserve based on individual metrics [14]. The difficulty levels in "Boogies Academy" and "Culbrain" were adapted based on age [65]. At the same time, the game uses gamification elements such as points, rewards, rankings, and levels as feedback mechanisms to minimize tedium [59,60]. Higher interactivity enhances player acceptance, and appropriate difficulty control can gain more favor among players. For example, in the game "The Secret Trail of the Moon," among the 5 minigames, Teka Teki is highly interactive but also the most difficult, with a strict difficulty curve that impedes progress and may lead to a decline in initial motivation. In contrast, Kuburi, another reasonably difficult and interactive mini-game, received the highest usability score [35]. While serious games aim to address these issues through interesting experiences, they may still struggle to fully engage children who have severe attention deficits and impulsiveness, traits that are common in ADHD [49,50,56]. It appears that a suitable solution for this issue has not yet been identified and requires further examination by future research.

Recent advancements in serious games for ADHD incorporate diverse interventions. For example, cognitive training games strengthen attention and executive functions through adaptive memory and game tasks [35,48], whereas neurofeedback integrates real-time electrocardiography data to promote self-regulation [45,49,66]. Exergaming combines physical activity with gaming [14,53], while social simulation games use role-playing scenarios to enhance emotional awareness [58,64]. By blending cognitive, physical, and emotional strategies, serious games provide scalable, complementary tools that may amplify traditional ADHD interventions.

Although the probability of having more than 1 comorbid condition in people with ADHD is high [3], no in-depth or categorical studies of comorbid conditions were found in the included studies. Instead, researchers chose to prioritize a more representative sample of children with ADHD over a "pure" ADHD sample, which would be less generalizable [44]. Meanwhile, only one study categorized and discussed the subtypes of study participants to target treatment more effectively [35]. This oversight of comorbidities and subtypes may increase the difficulty of integrating the game into daily patient use.

Most studies in this review did not report safety outcomes associated with serious games. Of the 5 studies using VR technology, only 1 reported adverse effects, such as dizziness and virtual reality motion sickness [35]. Some children found the EEG headsets uncomfortable and encountered challenges in setting up and using this equipment effectively during gameplay [43]. Some children exhibited decreased performance in certain cognitive tasks after participating in game-based interventions, indicating that these games may not enhance all cognitive functions equally [39]. The common device-related adverse events were decreased frustration tolerance, headache, and irritability [44]. This indicates that potential adverse effects have not been fully identified, necessitating future studies to assess the safety outcomes of serious games and identify any possible adverse effects.

Limitations

This review covers only original research in English from PubMed, Web of Science, Scopus, IEEE Xplore, and ACM Digital Library databases, published from January 2010 to January 2024. Although these publications furnish a robust foundation for academic research, they do not encompass the entirety of available literature, particularly that from disciplines outside our primary focus. Crucial studies published in other languages or indexed in databases predominant in other disciplines or geographical regions might not have been included. This limitation potentially omits crucial interdisciplinary insights and international perspectives that could influence the global understanding and treatment of ADHD. Future reviews could derive benefit from adopting a more inclusive approach, extending beyond English language publications to include a broader array of databases. Such inclusivity would enhance the global relevance and applicability of systematic reviews, ensuring a comprehensive synthesis that incorporates diverse methodological approaches and cultural perspectives. To the best of our ability, this limitation has been minimized by combining search strings and including more broadly defined terms in the keywords. Although the review incorporated DTx guidelines into the inclusion and exclusion criteria to comply with industry standards, the temporary deprioritization of certain fundamental DTx principles that do not currently apply to the research has also contributed to the limitations of the study.

XSL•FO
RenderX

In the review of 35 articles, bias was analyzed using 3 distinct methods, depending on the trial type: RCTs were assessed using the Cochrane Risk of Bias Tool 2.0 [77], nonrandomized quantitative studies were evaluated using the ROBINS-I [78], and qualitative studies were appraised using CASP checklists (Multimedia Appendix 6) [14,34-67,79]. RCTs typically showed low bias but were affected by deviations from intended interventions and subjective outcome assessments, raising concerns about generalizability. Quantitative studies faced biases from confounding factors and selective participant pools, lacking robust randomization and control. Qualitative research was compromised by nonrepresentative samples and insufficient methodological rigor, with potential skew from researcher and observer biases. These findings underscore the necessity for more rigorously designed studies with robust randomization, clear reporting, and diverse participant pools to ensure broader generalizability and more dependable conclusions. Despite limiting the study population to children with ADHD, the included studies were highly heterogeneous in terms of experiment type, game type, duration, sample size, targeting ability, assessment methods, and findings. Bias and heterogeneity necessitated descriptive rather than meta-analyses in this review. However, these studies initially demonstrate the potential for using serious games in ADHD treatment, providing data and insights for future research. They are expected to serve as a basis for obtaining clinical trial data and conducting more studies.

Although these limitations remain unresolved for the time being, the review still provides important implications for further research to enhance the evidence supporting serious games as digital therapeutics for enhancing ADHD treatment effectiveness in children.

Conclusion

This review synthesized evidence from 35 studies on serious games as DTx in children with ADHD. The results indicated that the potential of serious games as a digital therapeutic modality for children with ADHD offers benefits in attention, hyperactivity-impulsivity symptoms, social skills, and executive functions. And interventions incorporating somatosensory input enhanced hand-eye coordination. While these findings align with growing interest in nonpharmacological alternative therapies, limitations like small sample sizes, short intervention durations, risk of bias, and heterogeneity of outcome measures temper their generalizability. Due to the highly heterogeneous nature of the included studies, the review opted for descriptive analysis over meta-analysis; yet, it still offers significant reference value for future research. To advance the field, future research should combine with DTx guidelines, and prioritize large-scale trials with standardized ADHD assessment protocols, rigorous experimental designs, and long-term efficacy evaluations. By bridging these gaps, serious games could become scalable, personalized interventions within children's ADHD care frameworks.

Acknowledgments

The publication of this paper is funded by the Kyung Hee University SCI Paper Support Fund. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Both authors JL and WRC collaboratively designed the study framework and methodology, including keyword selection and database identification. JL conducted the initial screening of titles and abstracts using inclusion criteria, while both authors independently performed full-text eligibility assessments. Discrepancies in study inclusion were resolved through iterative discussions to reach a consensus. JL led data extraction and formal analysis, with WRC overseeing methodological rigor and validation. The manuscript was drafted by JL and critically revised by WRC. Both authors approved the final version and assume full responsibility for its content.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA checklist.
[DOCX File , 32 KB - games_v13i1e60937_app1.docx]

Multimedia Appendix 2
Search strategies.
[DOCX File , 22 KB - games_v13i1e60937_app2.docx]

Multimedia Appendix 3

XSL•FO
RenderX

Descriptive analysis.
[XLSX File (Microsoft Excel File), 13 KB - games_v13i1e60937_app3.xlsx]

Multimedia Appendix 4
Quality assessment.
[DOCX File , 30 KB - games_v13i1e60937_app4.docx]

Multimedia Appendix 5
Results.
[DOCX File , 41 KB - games_v13i1e60937_app5.docx]

Multimedia Appendix 6
The risk of bias.
[DOCX File , 31 KB - games_v13i1e60937_app6.docx]

References

1. Carmona S, Vilarroya O, Bielsa A, Trèmols V, Soliva JC, Rovira M, et al. Global and regional gray matter reductions in ADHD: a voxel-based morphometric study. Neurosci Lett 2005;389(2):88-93. [doi: 10.1016/j.neulet.2005.07.020] [Medline: 16129560]
2. Dessie M, Techane MA, Tesfaye B, Gebeyehu DA. Elementary school teachers knowledge and attitude towards attention deficit-hyperactivity disorder in Gondar, Ethiopia: a multi-institutional study. Child Adolesc Psychiatry Ment Health 2021;15(1):16 [FREE Full text] [doi: 10.1186/s13034-021-00371-9] [Medline: 33827642]
3. Buitelaar J, Medori R. Treating attention-deficit/hyperactivity disorder beyond symptom control alone in children and adolescents: a review of the potential benefits of long-acting stimulants. Eur Child Adolesc Psychiatry 2010;19(4):325-340 [FREE Full text] [doi: 10.1007/s00787-009-0056-1] [Medline: 19823900]
4. Dalsgaard S, Østergaard SD, Leckman JF, Mortensen PB, Pedersen MG. Mortality in children, adolescents, and adults with attention deficit hyperactivity disorder: a nationwide cohort study. Lancet 2015;385(9983):2190-2196. [doi: 10.1016/S0140-6736(14)61684-6] [Medline: 25726514]
5. Johnston C, Hommersen P, Seipp C. Acceptability of behavioral and pharmacological treatments for attention-deficit/hyperactivity disorder: relations to child and parent characteristics. Behav Ther 2008;39(1):22-32. [doi: 10.1016/j.beth.2007.04.002] [Medline: 18328867]
6. Walden EL, Thompson SA. A review of some alternative approaches to drug management of hyperactivity in children. J Learn Disabil 1981;14(4):213-7, 238. [doi: 10.1177/002221948101400408] [Medline: 7014748]
7. Pelham WE, Carlson C, Sams SE, Vallano G, Dixon MJ, Hoza B. Separate and combined effects of methylphenidate and behavior modification on boys with attention deficit-hyperactivity disorder in the classroom. J Consult Clin Psychol 1993;61(3):506-515. [doi: 10.1037/0022-006X.61.3.506] [Medline: 8326053]
8. Davis CC, Claudius M, Palinkas LA, Wong JB, Leslie LK. Putting families in the center: family perspectives on decision making and ADHD and implications for ADHD care. J Atten Disord 2012;16(8):675-684. [doi: 10.1177/1087054711413077] [Medline: 21976032]
9. Michael DR, Chen SL. Serious Games: Games That Educate, Train, and Inform. Hagerstown: Muska & Lipman/Premier-Trade; 2005.
10. Tolks D, Lampert C, Dadaczynski K, Maslon E, Paulus P, Sailer M. [Game-based approaches to prevention and health promotion: serious games and gamification]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2020;63(6):698-707. [doi: 10.1007/s00103-020-03156-1] [Medline: 32451596]
11. Chicchi Giglioli IA, de Juan Ripoll C, Parra E, Alcañiz Raya M. EXPANSE: A novel narrative serious game for the behavioral assessment of cognitive abilities. PLoS One 2018;13(11):e0206925 [FREE Full text] [doi: 10.1371/journal.pone.0206925] [Medline: 30412614]
12. Mittmann G, Zehetmayer S, Schrank B. Study protocol for a randomised controlled trial to evaluate the effectiveness of a serious game targeting interpersonal emotion regulation in early adolescents. Trials 2021;22(1):741 [FREE Full text] [doi: 10.1186/s13063-021-05706-7] [Medline: 34696816]
13. Martin-Moratinos M, Bella-Fernández M, Blasco-Fontecilla H. Effects of music on attention-deficit/hyperactivity disorder (ADHD) and potential application in serious video games: systematic review. J Med Internet Res 2023;25:e37742 [FREE Full text] [doi: 10.2196/37742] [Medline: 37171837]
14. Ji H, Wu S, Won J, Weng S, Lee S, Seo S, et al. The effects of exergaming on attention in children with attention deficit/hyperactivity disorder: randomized controlled trial. JMIR Serious Games 2023;11:e40438 [FREE Full text] [doi: 10.2196/40438] [Medline: 37159253]
15. Rodrigo-Yanguas M, Martin-Moratinos M, Menendez-Garcia A, Gonzalez-Tardon C, Sanchez-Sanchez F, Royuela A, et al. A virtual reality serious videogame versus online chess augmentation in patients with attention deficit hyperactivity

XSL•FO
RenderX

disorder: A randomized clinical trial. Games Health J 2021;10(4):283-292. [doi: 10.1089/g4h.2021.0073] [Medline: 34370610]
16. Rodrigo-Yanguas M, Martín-Moratinos M, González-Tardón C, Sanchez-Sanchez F, Royuela A, Bella-Fernández M, et al. Effectiveness of a personalized, chess-based training serious video game in the treatment of adolescents and young adults with attention-deficit/hyperactivity disorder: randomized controlled trial. JMIR Serious Games 2023;11:e39874 [FREE Full text] [doi: 10.2196/39874] [Medline: 37093628]
17. Kollins SH, DeLoss DJ, Cañadas E, Lutz J, Findling RL, Keefe RSE, et al. A novel digital intervention for actively reducing severity of paediatric ADHD (STARS-ADHD): a randomised controlled trial. Lancet Digit Health 2020;2(4):e168-e178 [FREE Full text] [doi: 10.1016/S2589-7500(20)30017-0] [Medline: 33334505]
18. ISO/TR 11147:2023 Health informatics—personalized digital health—digital therapeutics health software systems. ISO. 2023. URL: https://www.iso.org/standard/83767.html [accessed 2024-12-21]
19. Peñuelas-Calvo I, Jiang-Lin LK, Girela-Serrano B, Delgado-Gomez D, Navarro-Jimenez R, Baca-Garcia E, et al. Video games for the assessment and treatment of attention-deficit/hyperactivity disorder: a systematic review. Eur Child Adolesc Psychiatry 2022;31(1):5-20. [doi: 10.1007/s00787-020-01557-w] [Medline: 32424511]
20. Oh S, Choi J, Han DH, Kim E. Effects of game-based digital therapeutics on attention deficit hyperactivity disorder in children and adolescents as assessed by parents or teachers: a systematic review and meta-analysis. Eur Child Adolesc Psychiatry 2024;33(2):481-493. [doi: 10.1007/s00787-023-02174-z] [Medline: 36862162]
21. Vahia V. Diagnostic and statistical manual of mental disorders 5: A quick glance. Indian J Psychiatry 2013;55(3):220-223 [FREE Full text] [doi: 10.4103/0019-5545.117131] [Medline: 24082241]
22. Kuppili PP, Aneja J. Developmental delay, hyperactivity and mania: A perfect blend of confusion in a preschool child. Indian J Psychol Med 2019;41(3):286-288 [FREE Full text] [doi: 10.4103/IJPSYM.IJPSYM_379_18] [Medline: 31142934]
23. Guo C, Assumpcao L, Hu Z. Efficacy of non-pharmacological treatments on emotional symptoms of children and adults with attention-deficit/Hyperactivity Disorder: A meta-analysis. J Atten Disord 2022;26(4):508-524. [doi: 10.1177/10870547211001953] [Medline: 33759605]
24. Cibrian FL, Lakes KD, Schuck SE, Hayes GR. The potential for emerging technologies to support self-regulation in children with ADHD: A literature review. International Journal of Child-Computer Interaction 2022;31:1-15. [doi: 10.1016/j.ijcci.2021.100421]
25. Power J, van IJzendoorn M, Lewis AJ, Chen W, Galbally M. Maternal perinatal depression and child executive function: A systematic review and meta-analysis. J Affect Disord 2021;291:218-234. [doi: 10.1016/j.jad.2021.05.003] [Medline: 34049191]
26. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
27. Nordhausen T, Hirt J. Manual zur literaturrecherche in fachdatenbanken. RefHunter. 2020. URL: https://www.researchgate.net/publication/344337007_Manual_zur_Literaturrecherche_in_Fachdatenbanken [accessed 2025-04-03]
28. Eriksen MB, Frandsen TF. The impact of patient, intervention, comparison, outcome (PICO) as a search strategy tool on literature search quality: a systematic review. J Med Libr Assoc 2018;106(4):420-431 [FREE Full text] [doi: 10.5195/jmla.2018.345] [Medline: 30271283]
29. Susi T, Johannesson M, Backlund P. Serious Games—An Overview. Skövde: School of Humanities and Informatics, University of Skövde; 2007.
30. Vacca RA, Augello A, Gallo L, Caggianese G, Malizia V, La Grutta S, et al. Serious games in the new era of digital-health interventions: A narrative review of their therapeutic applications to manage neurobehavior in neurodevelopmental disorders. Neurosci Biobehav Rev 2023;149:105156 [FREE Full text] [doi: 10.1016/j.neubiorev.2023.105156] [Medline: 37019246]
31. Khan K, Hall CL, Davies EB, Hollis C, Glazebrook C. The effectiveness of web-based interventions delivered to children and young people with neurodevelopmental disorders: systematic review and meta-analysis. J Med Internet Res 2019;21(11):e13478 [FREE Full text] [doi: 10.2196/13478] [Medline: 31682573]
32. Vajawat B, Varshney P, Banerjee D. Digital gaming interventions in psychiatry: evidence, applications and challenges. Psychiatry Res 2021;295:113585. [doi: 10.1016/j.psychres.2020.113585] [Medline: 33303223]
33. Coelho T, Barbosa GAR, Silva IS, Coutinho FR, Silva FR. An overview of researches on digital accessibility before and after the great challenges of SBC 2006-2016: a systematic literature review. : Association for Computing Machinery; 2017 Presented at: IHC '17: Proceedings of the XVI Brazilian Symposium on Human Factors in Computing Systems; 2017 October 23 - 27; Joinville Brazil p. 1-10. [doi: 10.1145/3160504.3160537]
34. Giannaraki M, Moumoutzis N, Papatzanis Y, Kourkoutas E, Mania K. A 3D rhythm-based serious game for collaboration improvement of children with attention deficit hyperactivity disorder (ADHD). : IEEE; 2021 Presented at: IEEE Global Engineering Education Conference (EDUCON); 2021 April 21-23; Vienna, Austria p. 1217-1225. [doi: 10.1109/educon46332.2021.9453999]
35. Rodrigo-Yanguas M, Martin-Moratinos M, Menendez-Garcia A, Gonzalez-Tardon C, Royuela A, Blasco-Fontecilla H. A virtual reality game (The Secret Trail of Moon) for treating attention-deficit/hyperactivity disorder: development and usability study. JMIR Serious Games 2021;9(3):e26824 [FREE Full text] [doi: 10.2196/26824] [Medline: 34468332]

XSL•FO
RenderX

36. Aghdam KS, Alavi MH. Designing MIND PRO working memory game and evaluating its effectiveness on working memory in ADHD children. 2019 Presented at: International Serious Games Symposium (ISGS); 2019 December 26; Tehran, Iran p. 124-128. [doi: 10.1109/isgs49501.2019.9047038]
37. Batista BG, Rodrigues AFD, Miranda DM, Ishitani L, Nobre CN. Developing an edutainment game, taboo!, for children with ADHD based on socially aware design and VCIA model. 2022 Presented at: IHC '22: Proceedings of the 21st Brazilian Symposium on Human Factors in Computing Systems; 2022 October 17 - 21; Diamantina Brazil p. 1-11. [doi: 10.1145/3554364.3559121]
38. Wiguna T, Ismail RI, Kaligis F, Minayati K, Murtani BJ, Wigantara NA, et al. Developing and feasibility testing of the Indonesian computer-based game prototype for children with attention deficit/hyperactivity disorder. Heliyon 2021;7(7):e07571 [FREE Full text] [doi: 10.1016/j.heliyon.2021.e07571] [Medline: 34345741]
39. Ou Y, Wang Y, Chang H, Yen S, Zheng Y, Lee B. Development of virtual reality rehabilitation games for children with attention-deficit hyperactivity disorder. J Ambient Intell Human Comput 2020;11(11):5713-5720. [doi: 10.1007/s12652-020-01945-9]
40. Capelo DC, Sánchez ME, Hurtado JS, Chicaiza DB. Multisensory virtual game with use of the device leap motion to improve the lack of attention in children of 7–12 years with ADHD. 2018 Presented at: Proceedings of the International Conference on Information Technology & Systems (ICITS 2018); 2018 January 10-12; Ecuador p. 897-906. [doi: 10.1007/978-3-319-73450-7_85]
41. García-Baos A, D'Amelio T, Oliveira I, Collins P, Echevarria C, Zapata LP, et al. Novel interactive eye-tracking game for training attention in children with attention-deficit/hyperactivity disorder. Prim Care Companion CNS Disord 2019;21(4) [FREE Full text] [doi: 10.4088/PCC.19m02428] [Medline: 31274260]
42. Avila-Pesantez D, Rivera LA, Vaca-Cardenas L, Aguayo S, Zuñiga L. Towards the improvement of ADHD children through augmented reality serious games: preliminary results. : IEEE; 2018 Presented at: IEEE Global Engineering Education Conference (EDUCON); 2018 April 17-20; Santa Cruz de Tenerife, Spain p. 843-848. [doi: 10.1109/educon.2018.8363318]
43. Kanellos T, Doulgerakis A, Georgiou E, Bessa M, Thomopoulos S, Vatakis A. User experience evaluation of the REEFOCUS ADHD management gaming system. : IEEE; 2019 Presented at: 4th International Conference on Smart and Sustainable Technologies (SpliTech); 2019 June 18-21; Split, Croatia p. 1-6. [doi: 10.23919/splitech.2019.8783141]
44. Kollins SH, Childress A, Heusser AC, Lutz J. Effectiveness of a digital therapeutic as adjunct to treatment with medication in pediatric ADHD. NPJ Digit Med 2021;4(1):58 [FREE Full text] [doi: 10.1038/s41746-021-00429-0] [Medline: 33772095]
45. Machado FSV, Casagrande WD, Frizera A, Rocha FEM. Development of serious games for neurorehabilitation of children with attention-deficit/hyperactivity disorder through neurofeedback. 2019 Presented at: 18th Brazilian Symposium on Computer Games and Digital Entertainment (SBGames); 2019 October; Rio de Janeiro, Brazil p. 91-97. [doi: 10.1109/sbgames.2019.00022]
46. de la Guía E, Lozano MD, Penichet VMR. Educational games based on distributed and tangible user interfaces to stimulate cognitive abilities in children with ADHD. Brit J Educational Tech 2014;46(3):664-678. [doi: 10.1111/bjet.12165]
47. Avila-Pesantez D, Santillán GS, Padilla N, Miriam AL, Arellano-Aucancela A. Exploring learning in near-field communication-based serious games in children diagnosed with ADHD. 2021 Presented at: Advances in Emerging Trends and Technologies; 2021 May 29-31; Quito, Ecuador p. 314-324. [doi: 10.1007/978-3-030-63665-4_25]
48. Schena A, Garotti R, D'Alise D, Giugliano S, Polizzi M, Trabucco V, et al. IAmHero: preliminary findings of an experimental study to evaluate the statistical significance of an intervention for ADHD conducted through the use of serious games in virtual reality. Int J Environ Res Public Health 2023;20(4) [FREE Full text] [doi: 10.3390/ijerph20043414] [Medline: 36834109]
49. Blandon DZ, Munoz JE, Lopez DS, Gallo OH. Influence of a BCI neurofeedback videogame in children with ADHD. Quantifying the brain activity through an EEG signal processing dedicated toolbox. 2016 Presented at: IEEE 11th Colombian Computing Conference (CCC); 2016 September 27-30; Popayan, Colombia p. 1-8. [doi: 10.1109/columbiancc.2016.7750788]
50. Vita S, Mennitto A. Neurobot: a psycho-edutainment tool to perform neurofeedback training in children with ADHD. 2019 Presented at: CEUR Workshop Proceedings; 2019 November 25-26; Naples; Italy.
51. Chen CL, Tang YW, Zhang NQ, Shin J. Neurofeedback based attention training for children with ADHD. : IEEE; 2017 Presented at: IEEE 8th International Conference on Awareness Science and Technology (iCAST); 2017 November 08-10; Taichung, Taiwan p. 93-97. [doi: 10.1109/icawst.2017.8256530]
52. Soysal Ö, Kiran F, Chen J. Quantifying brain activity state: EEG analysis of background music in a serious game on attention of children. 2020 Presented at: 4th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT); 2020 October 22-24; Istanbul, Turkey p. 1-7. [doi: 10.1109/ismsit50672.2020.9255308]
53. Weerdmeester J, Cima M, Granic I, Hashemian Y, Gotsis M. A feasibility study on the effectiveness of a full-body videogame intervention for decreasing attention deficit hyperactivity disorder symptoms. Games Health J 2016;5(4):258-269. [doi: 10.1089/g4h.2015.0103] [Medline: 27304677]
54. Jácome V ID, Páez O JS, Cóllazos O CA, Fardoun HM. DIVIDI2: reinforcing divided attention in children with AD/HD through a mobile application. 2019 Presented at: REHAB '19: Proceedings of the 5th Workshop on ICTs for improving Patients Rehabilitation Research Techniques; 2019 September 11 - 13; Popayan Columbia p. 106-110. [doi: 10.1145/3364138.3364161]

XSL•FO
RenderX

55. Kim S, Ryu J, Choi Y, Kang Y, Li H, Kim K. Eye-contact game using mixed reality for the treatment of children with attention deficit hyperactivity disorder. IEEE Access 2020;8:45996-46006. [doi: 10.1109/access.2020.2977688]
56. Fernando M, Claudia B, Nimrod G, Juan G. KAPEAN: understanding affective states of children with ADHD. Journal of Educational Technology & Society 2016;19(2):18-28.
57. Crepaldi M, Colombo V, Mottura S, Baldassini D, Sacco M, Cancer A, et al. Antonyms: A computer game to improve inhibitory control of impulsivity in children with attention deficit/hyperactivity disorder (ADHD). Information 2020;11(4):230. [doi: 10.3390/info11040230]
58. Bul KCM, Kato PM, Van der Oord S, Danckaerts M, Vreeke LJ, Willems A, et al. Behavioral outcome effects of serious gaming as an adjunct to treatment for children with attention-deficit/hyperactivity disorder: A randomized controlled trial. J Med Internet Res 2016;18(2):e26 [FREE Full text] [doi: 10.2196/jmir.5173] [Medline: 26883052]
59. Castro R, Huamanchahua D. Development and validation of a gamified videogame for math learning in attention deficit hyperactivity disorder children (ADHD). 2021 Presented at: CEUR Workshop Proceedings; 2021 November 16-18; Chiclayo, Peru p. 17-25.
60. Celis G, Casas M, Mauricio D, Santisteban J. Dilud: A mobile application to reinforce rote learning in elementary school children with attention deficit hyperactivity disorder. Int. J. Interact. Mob. Technol 2023;17(06):62-80. [doi: 10.3991/ijim.v17i06.35259]
61. Retalis S, Korpa T, Skaloumpakas C, Boloudakis M, Kourakli M, Altanis G, et al. Empowering children with ADHD learning disabilities with the kinems kinect learning games. 2014 Presented at: 8th European Conference on Games Based Learning; 2014 October 9-10; Berlin, Germany p. 28-39. [doi: 10.1016/j.ijcci.2016.10.009]
62. Dovis S, Van der Oord S, Wiers RW, Prins PJM. Improving executive functioning in children with ADHD: training multiple executive functions within the context of a computer game. a randomized double-blind placebo controlled trial. PLoS One 2015;10(4):e0121651 [FREE Full text] [doi: 10.1371/journal.pone.0121651] [Medline: 25844638]
63. Gizatdinova Y, Remizova V, Sand A, Sharma S, Rantanen K, Helminen T, et al. PigScape: An embodied video game for cognitive peer-training of impulse and behavior control in children with ADHD. : Association for Computing Machinery; 2022 Presented at: ASSETS '22: Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility; 2022 October 23 - 26; Athens Greece p. 1-4. [doi: 10.1145/3517428.3550401]
64. Ahmadi A, Mitrovic A, Najmi B, Rucklidge J. TARLAN: a simulation game to improve social problem-solving skills of ADHD children. In: Artificial Intelligence in Education. Cham: Springer International Publishing; 2015:328-337.
65. García-Redondo P, García T, Areces D, Núñez JC, Rodríguez C. Serious games and their effect improving attention in students with learning disabilities. Int J Environ Res Public Health 2019;16(14):2480 [FREE Full text] [doi: 10.3390/ijerph16142480] [Medline: 31336804]
66. Barba MC, Covino A, De LV, DePaolis LT, D'Errico G, Di B. BRAVO: a gaming environment for the treatment of ADHD. In: Augmented Reality, Virtual Reality, and Computer Graphics. Cham: Springer International Publishing; 2019.
67. Kim S, Lee H, Lee H, Kim G, Song J. Adjuvant therapy for attention in children with ADHD using game-type digital therapy. Int J Environ Res Public Health 2022;19(22):14982 [FREE Full text] [doi: 10.3390/ijerph192214982] [Medline: 36429699]
68. Moogk DR. Minimum viable product and the importance of experimentation in technology startups. Technology Innovation Management Review 2012;2(3). [doi: 10.22215/timreview/535]
69. Fleming TM, de Beurs D, Khazaal Y, Gaggioli A, Riva G, Botella C, et al. Maximizing the impact of e-therapy and serious gaming: time for a paradigm shift. Front Psychiatry 2016;7:65 [FREE Full text] [doi: 10.3389/fpsyt.2016.00065] [Medline: 27148094]
70. Slobodin O, Davidovitch M. Gender differences in objective and subjective measures of ADHD among clinic-referred children. Front Hum Neurosci 2019;13:441 [FREE Full text] [doi: 10.3389/fnhum.2019.00441] [Medline: 31920599]
71. Elkin JA, McDowell M, Yau B, Machiri SV, Pal S, Briand S, et al. The Good Talk! A serious game to boost people's competence to have open conversations about COVID-19: protocol for a randomized controlled trial. JMIR Res Protoc 2023;12:e40753 [FREE Full text] [doi: 10.2196/40753] [Medline: 36884269]
72. Jacobs J, Heymann R, Greeff JJ. A telehealth system incorporating a serious game intervention to aid occupational therapists in identifying and Treating Children with difficulty crossing the body's midline: key informant interviews among occupational therapists. JMIR Serious Games 2021;9(4):e27761 [FREE Full text] [doi: 10.2196/27761] [Medline: 34723821]
73. Shinwari JMA, Al Yemni EAA, Alnaemi FM, Abebe D, Al-Abdulaziz BS, Al Mubarak BR, et al. Analysis of shared homozygosity regions in Saudi siblings with attention deficit hyperactivity disorder. Psychiatr Genet 2017;27(4):131-138 [FREE Full text] [doi: 10.1097/YPG.0000000000000173] [Medline: 28452824]
74. Wang Z, Cao H, Cao Y, Song H, Jiang X, Wei C, et al. Clinical characteristics and cognitive function in bipolar disorder patients with different onset symptom. Front Psychiatry 2023;14:1253088 [FREE Full text] [doi: 10.3389/fpsyt.2023.1253088] [Medline: 37840798]
75. Caselles-Pina L, Sújar A, Quesada-López A, Delgado-Gómez D. Adherence, frequency, and long-term follow-up of video game-based treatments in patients with attention-deficit/hyperactivity disorder: A systematic review. Brain Behav 2023;13(11):e3265 [FREE Full text] [doi: 10.1002/brb3.3265] [Medline: 37743605]

XSL•FO
RenderX

76. Ratra D, Rakshit A, Ratra V. Visual rehabilitation using video game stimulation for stargardt disease. Ther Adv Ophthalmol 2019;11:2515841419831158 [FREE Full text] [doi: 10.1177/2515841419831158] [Medline: 30886944]
77. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019;366:l4898 [FREE Full text] [doi: 10.1136/bmj.l4898] [Medline: 31462531]
78. Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ 2016;355:i4919 [FREE Full text] [doi: 10.1136/bmj.i4919] [Medline: 27733354]
79. CASP checklist: CASP qualitative studies checklis. Critical Appraisal Skills Programme. URL: https://casp-uk.net/casp-tools-checklists/qualitative-studies-checklist/ [accessed 2024-12-21]

Abbreviations

ADHD: attention-deficit/hyperactivity disorder
ATESC: Attention Test for Elementary School Children
BIA: Italian Battery for ADHD
BRIEF: Behavioral Rating Inventory of Executive Function
CASP: Critical Appraisal Skills Programme
DHI: digital health intervention
DTx: digital therapeutics
EEG: electroencephalography
ISO: International Organization for Standardization
MeSH: Medical Subject Headings
MVP: minimum viable product
PICOS: Participants, Intervention, Comparison, Outcomes, Study design
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analysis
RCT: randomized controlled trial
ROBINS-I: Cochrane Risk of Bias Tool for Non-Randomized Studies of Interventions
TOL: Tower of London task
WCST: Wisconsin Card Sorting Test

Edited by A Coristine, T Rashid Soron; submitted 28.05.24; peer-reviewed by E Kourkoutas, A Kononowicz; comments to author 04.11.24; revised version received 24.12.24; accepted 22.03.25; published 06.05.25.

Please cite as:
Lin J, Chang WR
Effectiveness of Serious Games as Digital Therapeutics for Enhancing the Abilities of Children With Attention-Deficit/Hyperactivity Disorder (ADHD): Systematic Literature Review
JMIR Serious Games 2025;13:e60937
URL: https://games.jmir.org/2025/1/e60937
doi:10.2196/60937
PMID:

©Jing Lin, Woo-Rin Chang. Originally published in JMIR Serious Games (https://games.jmir.org), 06.05.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

The Impact of Gamified Interventions on the Management of Chronic Obstructive Pulmonary Disease: Systematic Literature Review

Jinsong Chen[1,2,3], PhD; Tingzhong Yang[4], PhD; Qilian He[5], PhD; Mingli Pang[3], MPH; Ying Cao[6]; Zheng Liu[6]; Linfei Li[7]; Hsing-I Liu[7]; Christopher Bullen[3], PhD

[1]Department of Public Administration, Law School, Hangzhou City University, Hangzhou, China

[2]School of Public Affairs, Zhejiang University, Hangzhou, China

[3]National Institute for Health Innovation, University of Auckland, Auckland, New Zealand

[4]Research Center for Tobacco Control, Zhejiang University and Research Center for Digital Health Theory and Management, ZJU National Health Big Data Institute, Center for Tobacco Control Research, Zhejiang University School of Medicine, Zhejiang University, Hangzhou, China

[5]School of Nursing, College of Medicine, Dali University, Dali, China

[6]Tencent IEG Social Value Exploration Center, Shenzhen, China

[7]China Taiping Insurance Group Ltd., Shanghai, China

Corresponding Author:
Christopher Bullen, PhD
National Institute for Health Innovation
University of Auckland
28 Park Ave.
Grafton
Auckland, 1023
New Zealand
Phone: 64 93737599
Email: c.bullen@auckland.ac.nz

Abstract

Background: Chronic obstructive pulmonary disease (COPD) requires consistent sustained management, including regular physical activity, pulmonary rehabilitation, and self-care adherence. Despite strong clinical guidelines, patient engagement remains a major challenge, leading to suboptimal disease control and increased health care use. Gamified interventions have emerged as potential tools to improve adherence, motivation, and outcomes in chronic disease management. However, their effectiveness and implementation in COPD remain underexplored.

Objective: This review synthesizes current evidence on gamified interventions for COPD management to evaluate their effectiveness, focusing on patient engagement, physical outcomes, and quality of life.

Methods: We conducted a systematic search in PubMed, Scopus, Web of Science, Embase, IEEE Xplore, Cochrane Library, and China National Knowledge Infrastructure for studies published from January 2014 to October 2024. Only original studies involving trials (both randomized controlled trials [RCTs] and non-RCTs), intervention studies, feasibility studies, cross-sectional surveys, or qualitative studies were included.

Results: A total of 29 studies met the inclusion criteria: 11 (38%) RCTs; 7 (24%) pilot studies; 5 (17%) observational studies (including qualitative studies); and 6 (21%) other studies using gamified technologies such as virtual reality, exergames, and mobile apps. Compared to nongamified methods, gamified interventions provided an engaging, home-based alternative for COPD management, supporting long-term rehabilitation. Gamification features such as real-time feedback, adaptive challenges, and personalized goals increased patient adherence and motivation, with high engagement seen in virtual reality and exergame-based interventions, and showed notable improvements in COPD management, enhancing exercise tolerance, self-management, and symptom control. However, most of the studies (22/29, 76%) were of short duration, with small sample sizes.

Conclusions: Gamified COPD management tools offer flexibility and empower patients to self-manage their condition, potentially reducing the need for clinic visits. Gamified interventions show promise in COPD management, although current studies have methodological limitations. Future research should focus on conducting larger trials to assess the sustained impact of gamified

XSL•FO
RenderX

interventions on COPD outcomes; developing culturally relevant adaptations to enhance the global applicability of these interventions; and collaborating with patients, clinicians, and game developers to make the interventions more engaging and effective.

(JMIR Serious Games 2025;13:e69510) doi:10.2196/69510

KEYWORDS

gamified interventions; chronic obstructive pulmonary disease; COPD management; digital health; patient engagement; pulmonary rehabilitation

Introduction

Background

Chronic obstructive pulmonary disease (COPD) is a progressive, irreversible respiratory condition characterized by persistent airflow limitation, which collectively reduces patients' physical capabilities and impairs their quality of life (QoL). COPD affects >300 million individuals worldwide and ranks as the third leading cause of death, underscoring an urgent need for innovative, accessible management solutions [1]. COPD management typically focuses on symptom alleviation and preventing disease progression through pulmonary rehabilitation (PR; defined as a multidisciplinary, evidence-based intervention for chronic respiratory diseases such as COPD, combining exercise training, education, and behavioral therapy to reduce symptoms, improve physical and psychological well-being, and enhance long-term health behaviors), pharmacological interventions, and behavior modification (such as smoking cessation and dietary change) [1]. However, adherence to these programs is often low, due in part to a lack of patient engagement, access barriers, and the need for long-term ongoing use of the interventions [2]. Recent advances in digital health, particularly gamified interventions, show promise to enhance adherence and self-management in patients with COPD by incorporating engaging, game-like elements into COPD management programs.

Gamification in health care, defined as using game design elements in nongame contexts, is emerging as a valuable tool in managing chronic diseases, particularly in PR and home-based COPD management [3]. Gamified interventions leverage mechanisms such as rewards, competitive elements, real-time feedback, and immersive virtual environments to enhance motivation; foster active participation; and sustain adherence, particularly to exercise-focused regimens [4]. Recent systematic reviews indicate that gamified eHealth interventions can improve physical activity levels in adults with chronic diseases and significantly enhance physical outcomes, adherence, and QoL [1,5]. In COPD, where symptoms often lead to a sedentary lifestyle, these elements can provide a sense of accomplishment and motivation, addressing common barriers to physical activity and rehabilitation.

A systematic review by Chang et al [5] highlighted the potential of gamified interventions to enhance exercise endurance and QoL among patients with COPD, especially those incorporating telemonitoring, web-based tools, and virtual reality (VR). VR as a vehicle for PR has a growing evidence base; for example, Liu et al [2] conducted a meta-analysis to explore VR's role in COPD management and found that VR-assisted PR can improve lung function and exercise capacity. By making exercise engaging and immersive, delivery via VR potentially mitigates the monotony associated with traditional exercise–focused PR, encouraging regular participation and improving outcomes. In addition, a systematic review and meta-analysis by Obrero-Gaitán et al [3] found supportive evidence for the effectiveness of VR-based PR in COPD, reporting significant improvements in physical performance and mobility.

Such immersive, game-like platforms not only aid in physical rehabilitation but also enhance patient engagement. In a scoping review, Dalko et al [6] found supportive evidence for VR in rehabilitation for patients recovering from post–COVID-19 condition who had chronic respiratory problems [6]. These findings support the idea that VR's engaging nature may appeal to patients with COPD facing similar challenges of motivation and compliance.

Objectives

While these studies found promising results, most gamified interventions are in the early stages of development. As the recent systematic reviews highlight, there is a need for larger, long-term rigorous studies to validate their effectiveness, cost-effectiveness, acceptability, and sustained use for different populations. This review synthesizes existing literature to assess the impact of gamified interventions on COPD management and PR, focusing not only on physical outcomes and QoL improvements but also on engagement and adherence. By consolidating these findings, this study aims to provide a comprehensive understanding of gamification's potential role in enhancing care for patients with COPD.

Methods

Overview

The primary objective of this study was to systematically review and evaluate the impact of gamified interventions on the management of COPD, including PR. Specifically, this study aimed to assess how gamification influences key outcomes in COPD management, such as patient adherence to treatment regimens, physical activity, symptom management, and QoL. In addition, we sought to assess the accessibility, usability, and patient satisfaction associated with these interventions, considering whether gamification supports long-term behavior change and self-management. By synthesizing findings from recent studies, this study aimed to provide valuable insights into the current evidence, limitations, and the potential of gamified approaches in enhancing COPD care, as well as to offer recommendations for future research directions in this emerging area of digital health.

XSL•FO
RenderX

Search Strategy

The search strategy targeted studies on gamified interventions for COPD management, emphasizing recent advancements over the past 10 years (January 2014-October 2024) to maintain relevance in this rapidly evolving field. We searched 7 databases—PubMed, Scopus, Web of Science, Embase, IEEE Xplore, Cochrane Library, and China National Knowledge Infrastructure—capturing a broad range of research in biomedical and digital health.

The search terms in Table 1 were identified for 3 domains: "chronic obstructive pulmonary disease," "gamification," and "condition management" as well as "rehabilitation," with terms within each domain connected by the Boolean operator "OR" and domains linked by "AND." Detailed search strategies for each database are shown in Multimedia Appendix 1.

The searches included studies published in English and Chinese to capture a more global perspective than reviews relying solely on English-language databases. Boolean operators were used to ensure comprehensive coverage, and all retrieved studies were imported into EndNote 20 (Clarivate) to facilitate screening, deduplication, and data extraction.

Table 1. Search domains and terms.

Domains	Search terms
Chronic obstructive pulmonary disease	"Chronic Obstruct* Pulmon* Disease" OR "COPD" OR "Pulmon* Disease, Chronic Obstruct*" OR "Chronic Bronchit*" OR "Emphysem*" OR "Respirator* Disease" OR "Pulmon* Disorder"
Gamification	"Gamif*" OR "Game-based" OR "Serious Game*" OR "Gaming" OR "Digital Game*" OR "Health Game*" OR "Game Mechanic*" OR "Mobile Game*" OR "Interactive Game*" OR "Exergame*" OR "Behavioral Gamif*" OR "Virtual Reality" OR "Augmented Reality"
Condition management	"Self-Manage*" OR "Self-Care" OR "Disease Manage*" OR "Chronic Disease Manage*" OR "Symptom Monitor*" OR "Pulmon* Rehab*" OR "Pulmon* Rehabilitat*" OR "Exacerbat* Prevent*" OR "Patient Engage*" OR "Patient Adher*" OR "Treat* Adher*" OR "Health Monitor*" OR "Behavior* Change" OR "Behaviour* Change" OR "Condition Manage*" OR "Clinical Monitor*" OR "Intervent* Manage*"

Inclusion and Exclusion Criteria

Inclusion criteria were as follows: (1) languages—studies published in English and Chinese were included to capture a breadth of global research; (2) duration—publications from the last 10 years (January 2014-October 2024) were included to reflect recent advances in gamification and digital health; and (3) study types—only original studies involving trials (both randomized controlled trials [RCTs] and non-RCTs), intervention studies, feasibility studies, cross-sectional surveys, or qualitative studies were included. Studies published solely as protocols or conference proceedings were excluded due to limited information on outcomes and intervention efficacy. Studies that did not incorporate gamification elements and those not addressing COPD as the primary health condition were also excluded.

Data Extraction and Analysis

Data extraction and analysis were conducted in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 guidelines to ensure thorough and transparent reporting [7]. The PRISMA checklist is presented in Multimedia Appendix 2. Using a standardized data extraction form, 2 independent reviewers extracted relevant information from each study, reducing the risk of bias and enhancing data reliability. Key data collected included study design, sample size, population characteristics, intervention details (such as the type of gamification, platform, and duration), outcomes (eg, physical function, adherence, engagement metrics, and QoL), and any limitations reported. Multimedia Appendix 3 provides the template used for the data extraction and analysis form. Data on primary and secondary outcomes—such as physical activity levels, lung function, and QoL metrics—were specifically targeted to assess the effectiveness of gamified interventions in COPD management. A third reviewer addressed any discrepancies between the 2 primary reviewers. Once extracted, data were tabulated and summarized to facilitate cross-study comparisons. Given the diversity in gamification strategies and outcome measures, a narrative synthesis approach was used to identify common themes, potential correlations, and patterns across studies. Each study's methodological quality and risk of bias were assessed using criteria aligned with the PRISMA guidelines. The risk-of-bias assessment was conducted using a qualitative approach, given the diverse study designs (both RCTs and non-RCTs) included in this review. For each study, the review evaluated methodological limitations and potential sources of bias, including risk-of-bias assessment, risk-of-bias findings, and limitations. The assessment framework considered several key domains that could potentially influence study validity: randomization procedures (eg, control group presence and characteristics, as well as blinding methods), sample characteristics (eg, sample size adequacy, dropout rates, and selection procedures), and implementation quality (eg, study duration, follow-up assessment, and intervention adherence). On the basis of these criteria, studies were classified as low risk, moderate risk, or high risk.

Results

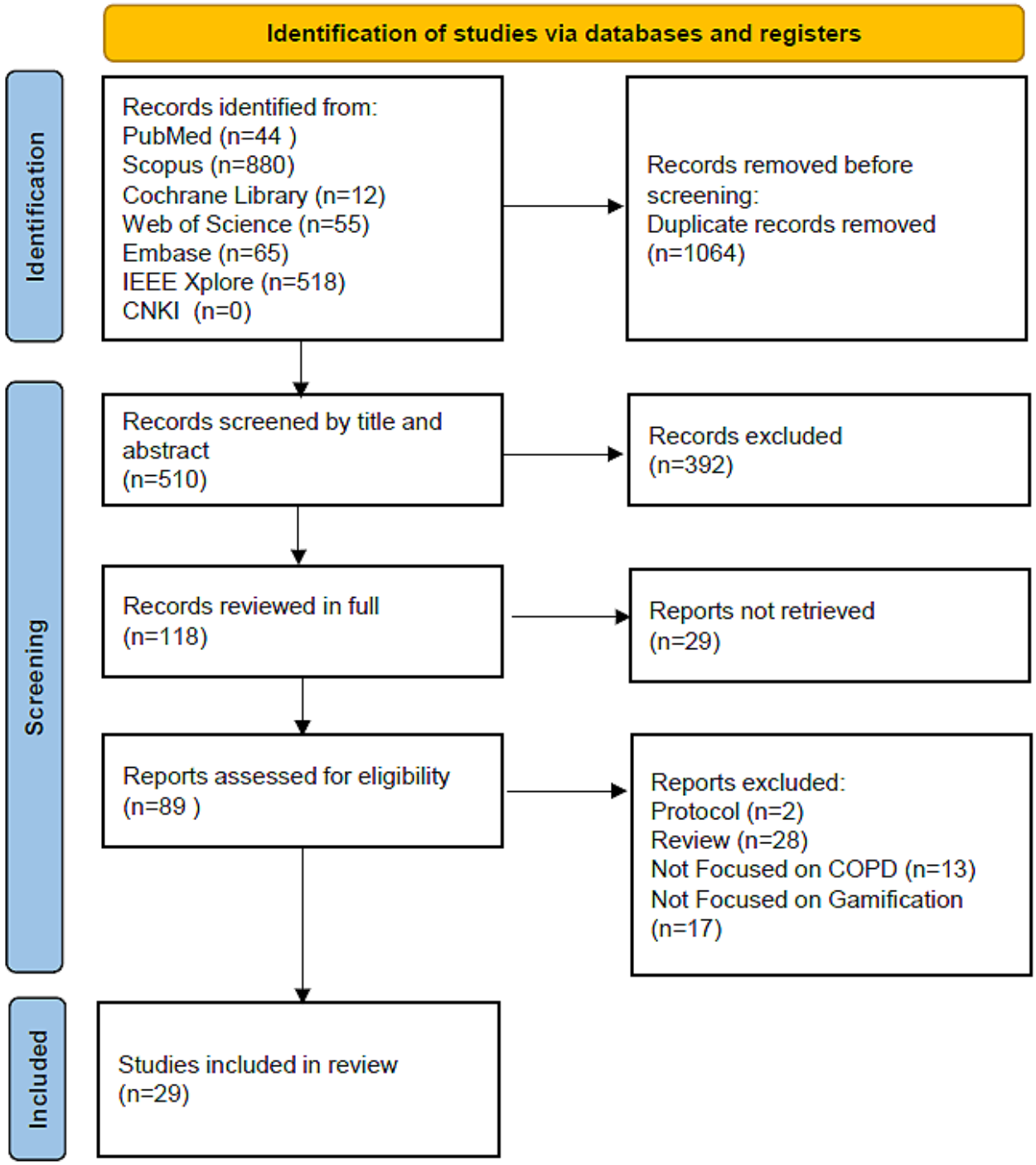
Summary of Included Studies

The study identified 29 studies published between January 2014 and October 2024 that investigated the impact of gamified interventions on COPD management. The studies examined a range of gamification approaches, including VR-based rehabilitation, mobile apps, and digital games designed to enhance patient engagement, physical activity, or symptom management. Multimedia Appendix 4 [8-36] provides a detailed summary of the data extraction and analysis for each included

XSL•FO
RenderX

study. The PRISMA flowchart (Figure 1) outlines the study selection process.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram showing the number of studies identified, screened, assessed for eligibility, and included in the final analysis. CNKI: China National Knowledge Infrastructure; COPD: chronic obstructive pulmonary disease.



Study Designs and Settings

Most of the studies (10/29, 35%) were RCTs, reflecting the focus on evaluating intervention effectiveness [8,9]. Sample sizes ranged from small-scale feasibility trials with as few as 4 participants to larger studies with up to 106 participants [10,11]. Intervention durations varied substantially from single sessions to long-term studies lasting more than a year. Settings varied widely, including clinical, home-based, and hybrid models, with home-based approaches being common for virtual and gamified interventions [12,13].

Table 2 provides a summary of the study design, setting, sample size, and intervention duration for each reviewed study, capturing the diversity in research approaches across this field.

XSL•FO
RenderX

Table 2. Overview of study designs and settings.

Study	Study design	Setting	Sample size, n	Intervention duration
Tabak et al [8], 2014	Pilot RCT[a]	Hospital and primary care physiotherapy practices, Enschede, Netherlands	29	9 mo
Mazzoleni et al [14], 2014	RCT	Auxilium Vitae Rehabilitation Center, Volterra, Italy	40	3 wk
Kotrach et al [15], 2015	Pilot RCT	Mount Sinai Hospital, Montreal, Canada	12	3-6 h of individualized training, with follow-up at home
Hoaas et al [12], 2016	Mixed methods pilot study	Home-based with telemonitoring, Norway	10	2 y
LeGear et al [16], 2016	Randomized, within-participant crossover trial	St. Paul's Hospital, Vancouver, Canada	10	Single session with two 15-min exercise interventions
Liu et al [17], 2016	Cross-sectional observational study	CIRO, Horn, Netherlands	61	—[b]
Bamidis et al [18], 2017	Multilevel intervention project	EU[c]-funded multisite study	Data not explicitly provided (ongoing multicenter trials)	3 y
Burkow et al [19], 2018	Feasibility study	Virtual group-based home setting	10	6 wk
De Las Heras et al [20], 2018	Qualitative study with focus groups and semistructured interviews	Aarhus University Hospital, Denmark; and Oulu University Hospital, Finland	13	—
Parent et al [21], 2018	Pilot feasibility study	Supervised hospital setting in Montreal, Canada	14	Single 30-min session
Rutkowski et al [22], 2019	RCT	Hospital-based pulmonary rehabilitation (stationary)	68	14 d
Sutanto et al [23], 2019	RCT	Outpatient clinic in Dr Moewardi Hospital, Surakarta, Indonesia	23	6 wk
Jung et al [13], 2020	Mixed methods study	Home-based rehabilitation, South and West Cumbria, United Kingdom	10	8 wk
Rutkowski et al [11], 2020	RCT	Specialist hospital in Głuchołazy, Poland	106	2 wk
Tu et al [24], 2020	Pilot feasibility study	Home-based (in-laboratory demonstration using smartwatch, smartphone, and VR[d] headset)	Proof-of-concept demonstration; no specific patient sample size indicated	Demonstration sessions lasted 2-5 min
Rutkowski et al [9], 2021	RCT	Specialist hospital in Głuchołazy, Poland	50	2 wk (10 VR sessions)
Simmich et al [25], 2021	Qualitative study with semistructured interviews	Pulmonary support groups in Brisbane, Australia	19	NR (single interview session per participant)
Simmich et al [26], 2021	Pilot RCT	Home-based intervention, Brisbane, Australia	18	3 wk
Baxter et al [27], 2022	Mixed methods randomized usability study	Urban locations in southeast Queensland, Australia	24	Single session (3 inspirations using each device)
Oberschmidt et al [28], 2022	Qualitative study using interviews	Physiotherapy office in the Netherlands	7	6 mo
Finkelstein et al [29], 2023	Mixed methods randomized usability study	Icahn School of Medicine at Mount Sinai, New York, United States	9	Single session of using the VR app
Gabriel et al [30], 2023	Mixed methods quantitative and qualitative assessments	Icahn School of Medicine at Mount Sinai, New York, United States	18	Single session using the VR app
Gabriel et al [31], 2023	Qualitative study with semistructured interviews	Home-based, conducted virtually	9	Single session
Pancini et al [32], 2023	RCT	IRCCS INRCA, Italy	Not reported (study is in planning phase)	2 wk (four 20-min sessions)

XSL•FO
RenderX

Study	Study design	Setting	Sample size, n	Intervention duration
Pardos et al [33], 2023	Methodology and pilot design	Remote care platforms for chronic disease management	Pilot system; no specific patient sample provided	—
Colombo et al [34], 2024	Single-group pilot study	In-hospital rehabilitation program, Italy	14	2 wk, with 20-min sessions, twice daily
Jin et al [35], 2024	RCT	The First Affiliated Hospital of Soochow University, China	80	6 wk
Kizmaz et al [36], 2024	RCT	Hospitalized patients at Pamukkale University Health Research and Application Centre, Turkey	50	Until discharge from hospital
McAnirlin et al [10], 2024	Proof of concept, mixed methods	Participants' homes in upstate South Carolina, United States	4	4 mo

[a]RCT: randomized controlled trial.

[b]Not applicable.

[c]EU: European Union.

[d]VR: virtual reality.

Types and Characteristics of Gamified Interventions

The reviewed studies applied various gamified interventions to support COPD management, ranging from VR programs to wearable activity trackers. Platforms included VR headsets, gaming consoles, and smartphone apps, each incorporating gamification elements such as motivational cues, real-time feedback, and rewards for progress [8,14,15]; for example, the Condition Coach COPD telehealth program used wearable devices and daily feedback tailored by physiotherapists [8], while BreathCoach provided real-time biofeedback for respiratory exercises [24].

Several interventions offered individualized tailoring to enhance relevance and usability, adapting exercise intensity or integrating patient preferences into their design. Integration with health care systems varied, with some interventions offering telemonitoring and health care provider supervision, such as the SmokeFreeBrain project [18]. Table 3 shows the details of the interventions, the platform, game elements, details of tailoring if included, and integration with health care.

XSL•FO
RenderX

Table 3. Summary of intervention characteristics.

Study	Intervention	Type	Platform or technology used	Game elements	Tailoring	Integration with health care
Tabak et al [8], 2014	Condition Coach COPD telehealth program	Motivational cues	Wearable device, smartphone app, and web-based portal	Motivational cues and daily feedback	Exercise schemes tailored by physiotherapists	Integrated with primary and secondary care; health care professionals monitored progress
Mazzoleni et al [14], 2014	Exergaming with Nintendo Wii	Exergames using Nintendo Wii Fit Plus	Nintendo Wii and Wii Balance Board	Real-time feedback and visual and auditory cues	Exercise intensity adjusted by physiotherapist	Supervised by health care professionals
Kotrach et al [15], 2015	Virtual game system for home exercise	Exergames using Nintendo Wii Fit U	Nintendo Wii Fit U	11 prevalidated games, with real-time feedback on performance	Individualized training sessions based on patient needs	Monitored by physiotherapists during in-hospital and home training
Hoaas et al [12], 2016	Telerehabilitation program	Telerehabilitation via self-management and telemonitoring	Treadmill, pulse oximeter, and iPad with a web page for telemonitoring	Regular feedback and remote support	Individually tailored exercise program	Weekly videoconferencing sessions with a physiotherapist
LeGear et al [16], 2016	Nintendo Wii-based exercise	Exergame using EA Sports Active on Nintendo Wii	Nintendo Wii with SenseWear Armband	Physical activities, including marching, dancing, and punching	Exercise intensity adjusted based on perceived exertion	NR[a]
Liu et al [17], 2016	GRAIL[b]	NR; focus on GRAIL technology and 6MWT[c]	GRAIL, 3D motion capture, and VR[d] treadmill	VR environment used to simulate walking conditions	Not personalized beyond adjusting walking pace	NR
Bamidis et al [18], 2017	SmokeFreeBrain	Smoking cessation intervention using gamified apps, mini-games, and social media	Gamification app, neurofeedback, social media, mobile SMS text messaging, and pharmacological interventions	Tasks, goals, achievements (eg, “Crushing cigarettes” game, breathing control, and exercise gamification)	Interventions customized by demographic and geographic needs	Public health campaigns, electronic health records, and open data integration
Burkow et al [19], 2018	Virtual group intervention for behavior change	Virtual group exercise with follow-along videos and exercise diaries	Tablet-based app	Virtual group status, rewards, and exercise diary	Self-chosen individual exercises	Peer group exercises combined with remote support
De Las Heras et al [20], 2018	AR[e] glasses	Use of AR glasses	Laster WAV AR glasses	Visual and audio guidance during exercises	Adjustable brightness, head fixation, and interface improvements	Remote communication and feedback
Parent et al [21], 2018	High-intensity active video game	High-intensity exergames	Kinect motion capture (Xbox One) and Shape Up game	Repetitive exercise mini-games	Adjustments to exercise difficulty	NR
Rutkowski et al [22], 2019	VR training program	Virtual rehabilitation using Kinect-based motion training	Xbox 360 Kinect system and Kinect Adventures! game	Avatar-based mini-games involving rafting, ball hitting, and roller coaster riding	VR exercises tailored for each patient	Integrated into a structured pulmonary rehabilitation program
Sutanto et al [23], 2019	Videogame-assisted exercise training	Wii Fit–based exercise training	Wii Fit balance board and television system	Yoga, strength training, aerobic exercises, feedback system, and virtual trainer	Individualized program with game duration, difficulty, and scores recorded	Integrated into a hospital-based outpatient exercise program
Jung et al [13], 2020	VR pulmonary rehabilitation program	VR-supported pulmonary rehabilitation	VR headset (Pico Goblin) and pulmonary rehabilitation in VR app	3D avatars, educational modules, and immersive experience	Exercises tailored to severity level (MRC[f] breathlessness scale level 4 or 5) of patient with COPD[g]	Real-time remote monitoring of heart rate and oxygen saturation

XSL•FO
RenderX

Study	Intervention	Type	Platform or technology used	Game elements	Tailoring	Integration with health care
Rutkowski et al [11], 2020	VR rehabilitation for COPD	VR-based rehabilitation using Xbox Kinect and Kinect Adventures! software	Xbox 360, Kinect, and Kinect Adventures!	Mini-games involving rafting, ball hitting, dynamic balance, and coordination	Exercises adapted to patients' abilities	Supervised by physiotherapists; heart rate monitored
Tu et al [24], 2020	BreathCoach	VR-assisted biofeedback breathing training using RSA-BT[h]	Smartphone, smartwatch (Empatica E4), and VR viewer (Google Cardboard)	Breathing control, interactive VR environments, avatars, and real-time biofeedback	Dynamic adjustment of breathing patterns based on real-time physiological data	NR
Rutkowski et al [9], 2021	Immersive VR therapy	Immersive VR therapy	VR TierOne device	Virtual therapeutic garden and metaphoric health recovery	Emotional balance recovery and mood improvement	Supervised by therapists
Simmich et al [25], 2021	Wearable technology and video games	AVGs[i]	Wearable activity trackers, smartphones, and AVGs such as Wii and Xbox Kinect	AVGs and wearable trackers	NR	Data shared with clinicians for feedback and improved clinical care
Simmich et al [26], 2021	AVG for physical activity	AVG focusing on physical activities	Smartphone app with Fitbit integration	Single-player and multiplayer modes, progress tracking, and rewards for completing exercises	Players selected difficulty levels for each exercise	Clinicians monitored progress via a web interface
Baxter et al [27], 2022	Virtual respiratory therapy	Virtual incentive spirometry via smartphone app	QUT Inspire app and smartphone	Visual rewards, breath timer, and inspiration counter	Adjustable microphone sensitivity and text or video instructions	NR
Oberschmidt et al [28], 2022	Exergame for COPD treatment	Exergame used as part of physiotherapy treatment	Television screen with motion-sensing camera	Audiovisual feedback during exercises and score tracking	Feedback based on exercise accuracy and adjustable difficulty levels	Integrated into routine physiotherapy treatment
Finkelstein et al [29], 2023	VR app for pulmonary rehabilitation	VR educational app	Oculus Quest 2	Interactive educational modules, MCQs[j], and visual feedback	Simplified user interface with preset controls for ease of use	NR
Gabriel et al [30], 2023	VR-based system for pulmonary rehabilitation	VR-based system for pulmonary rehabilitation	Oculus Quest 2	Educational modules, MCQs, and guided exercises	Simplified interface, single-button navigation, and custom instructions	NR
Gabriel et al [31], 2023	VR-based system for pulmonary rehabilitation	VR-based exercise app for pulmonary rehabilitation	VR headset and controllers	Interactive guided exercises and visual feedback	Simplified controls and visual guidance for exercises	Focused on self-management
Pancini et al [32], 2023	Overcoming COPD	VR-based relaxation combined with savoring strategies	VR headset with immersive natural scenarios	Narrated virtual walks, visual and audio feedback, and positive emotion amplification	Personalized savoring exercises	Incorporated into standard pulmonary rehabilitation
Pardos et al [33], 2023	Remote monitoring and gamification platform	Personalized coaching with exergames and mental health games	Smartphone app, smartwatches, and Bluetooth-enabled devices	Credit-based system, scores, rewards, and health recommendations	Customized recommendations	Integration with PHR[k] and third-party apps
Colombo et al [34], 2024	VR endurance training	Semi-immersive VR cycling in a virtual park environment	Cycle ergometer, pulse oximeter, and wide screen projection	Visual feedback, real-time cycling metrics, and first-person navigation	Exercise intensity based on baseline conditions	Continuous supervision by physiotherapists

XSL•FO
RenderX

Study	Intervention	Type	Platform or technology used	Game elements	Tailoring	Integration with health care
Jin et al [35], 2024	Somatosensory interactive game	Somatosensory interactive games involving arm movements for exercise	Motion-based games: Kitchen Sharp Knife, Swimming Master, and Table Tennis Master	Real-time visual feedback and engaging tasks	Patients adjusted exercise based on comfort level	Integrated with physiotherapist-supervised pulmonary rehabilitation programs
Kizmaz et al [36], 2024	VR for COPD exacerbation	VR cycling simulation in the forest combined with pulmonary rehabilitation	Oculus Quest 2	Immersive cycling simulation in a forest environment	NR	Integrated with pulmonary rehabilitation sessions, supervised by physiotherapists
McAnirlin et al [10], 2024	Nature-based VR experiences	Nature-based VR experiences	Oculus Quest 2	Cocreated 360-degree videos of personalized, nature-based scenes	Personalized VR based on participants' outdoor memories	NR

[a]NR: not reported.

[b]GRAIL: Gait Real-time Analysis Interactive Lab.

[c]6MWT: 6-minute walk test.

[d]VR: virtual reality.

[e]AR: augmented reality.

[f]MRC: Medical Research Council.

[g]COPD: chronic obstructive pulmonary disease.

[h]RSA-BT: respiratory sinus arrhythmia biofeedback-based breathing training.

[i]AVG: active video game.

[j]MCQ: multiple-choice question.

[k]PHR: personal health record.

Effectiveness of Gamified Interventions

Gamified interventions showed positive impacts on COPD management outcomes, particularly in improving exercise tolerance and physical fitness; for instance, Mazzoleni et al [14] found a significant increase in 6-minute walk test distance among participants using exergames, with the experimental group improving by 97.4 meters. Similarly, Rutkowski et al [11] observed increased exercise performance in the VR groups compared to the traditional exercise group, as shown in Table 4.

Regarding behavioral outcomes, gamified interventions were effective in increasing motivation and adherence to exercise; for example, LeGear et al [16] noted high participant enjoyment and engagement with Wii-based exergames, enhancing exercise adherence. Furthermore, Hoaas et al [12] reported long-term adherence to telerehabilitation, showing improved self-management and coping skills among patients with COPD.

QoL improvements were also observed, with several interventions enhancing emotional well-being and reducing anxiety; for example, Pancini et al [32] reported reduced anxiety and stress in participants using VR relaxation interventions, contributing to overall QoL. Such findings suggest that gamified approaches can not only support physical health but also promote psychological resilience and well-being in patients with COPD [8,13].

XSL•FO
RenderX

Table 4. Effectiveness of gamified chronic obstructive pulmonary disease (COPD) interventions.

Study	Effectiveness results	Behavioral outcomes	QoL[a] improvements
Tabak et al [8], 2014	Exacerbations: telehealth group=33 (median 2.0, IQR 1.0-3.0); hospitalizations: telehealth group=4 (median 5.5, IQR 4.8-6.3) d, CG[b]=5 (median 7.0, IQR 6.0-7.0) d; QoL (EQ-5D VAS[c]): telehealth group=72.3, CG=62.4; no statistically significant difference in clinical outcomes between groups	Improved self-management of exacerbations (86.4% diary adherence)	EQ-5D VAS score: telehealth group—from 64.7 (baseline) to 72.3 (3 mo); CG—from 65.0 (baseline) to 62.4 (3 mo); CCQ[d] score: telehealth group—from 2.0 (baseline) to 1.8 (3 mo)
Mazzoleni et al [14], 2014	6MWT[e]: EG[f]=+97.4 m, CG=+61.1 m (P=.03); TDI[g] score: EG=3.9, CG=2.2 (P<.001); SGRQ[h] score: EG=−10.8, CG=−12.7 (P=.66); significant improvement in 6MWT and dyspnea for EG compared to CG	Improved patient motivation and engagement in EG	SGRQ score improved
Kotrach et al [15], 2015	Exercise tolerance (6MWD[i]), heart rate, and oxygen saturation monitored; mean 6MWD was 306 (SD 81) m at baseline; preliminary results showed that participants could maintain exercise training after PR[j] using VGS[k]	Dyspnea and leg discomfort increased, indicating exertion during exercise	Not reported in the preliminary findings
Hoaas et al [12], 2016	Average adherence: 43.3% for daily diary, 56.2% for exercise training; no dropouts; long-term adherence despite motivational challenges	Participants reported better self-management and coping with COPD	Reported improved health and increased capacity for daily activities
LeGear et al [16], 2016	Energy expenditure: Wii group (mean 353.5, SD 134.1 J) vs treadmill group (mean 317.1, SD 105.2 J), mean difference 36.3 J (95% CI 31.4 to 104); heart rate: Wii group (mean 112.5, SD 13.2 bpm) vs treadmill group (mean 112.7, SD 10.2 bpm), mean difference −0.167 (95% CI −4.83 to 4.50); no significant difference in energy expenditure, heart rate, or perceived exertion between Wii and treadmill groups	Participants reported enjoyment and perceived feasibility of Wii exercises at home	Not specifically reported in this study
Liu et al [17], 2016	Patients with COPD walked 27.5 m less on GRAIL[l] vs overground 6MWT (P<.001); healthy older adults walked 23.6 m more on GRAIL (P<.001); GRAIL showed good reproducibility for both groups: ICC[m] of 0.80 for patients with COPD (95% CI 0.61 to 0.89) and 0.65 for healthy older adults (95% CI 0.05 to 0.86)	Improved reproducibility and patient engagement with virtual environment for patients with COPD	No QoL data reported
Bamidis et al [18], 2017	Efficacy of PSAs[n], e-cigarette interventions, and neurofeedback protocols; expected positive impacts on reducing smoking among groups considered high risk	Increased adherence to smoking cessation interventions using gamification and ICT[o]	Expected improvements in smoking-related morbidity and mortality rates
Burkow et al [19], 2018	Increase in physical activity from 2.9 to 5.9 sessions per wk during the program; 77% adherence to group exercises	Positive impact on motivation to engage in physical activity	Improved well-being and mood reported
De Las Heras et al [20], 2018	Positive perception of AR[p] glasses, particularly ease of use and exercise guidance; patients saw value in the AR glasses for telerehabilitation, although some found them heavy	Motivation to use AR glasses for physical exercise and rehabilitation	NR[q]
Parent et al [21], 2018	Peak minute ventilation (36.8 L/min in squatting game) and peak METs[r] (4.4 in squatting game); high-intensity games met exercise guidelines; Borg scores for leg exertion (13-14)	High perceived enjoyment and willingness to engage in home-based rehabilitation	QoL not directly measured
Rutkowski et al [22], 2019	Improved physical fitness as measured by the SFT[s]; significant within-group improvements (P<.05) in SFT (sit and reach test: from 0.0 to 0.7, 6MWT: from 494.9 to 469.9)	VR[t] group showed enhanced motivation and adherence	NR
Sutanto et al [23], 2019	6MWD, dyspnea (TDI), and health-related QoL (SGRQ); 6MWD improved significantly (EG—from 376.6 to 420.0 m; P<.001; CG—from 410.7 to 477.5 m; P<.001), without any difference between groups	NR	Significant SGRQ score reduction in both groups (EG—from 57.7 to 30.6; P<.05; CG—from 54.1 to 29.4; P<.05), without any difference between groups

XSL•FO
RenderX

Study	Effectiveness results	Behavioral outcomes	QoL[a] improvements
Jung et al [13], 2020	Improved compliance, physical health (mobility and flexibility), and psychological well-being; significant improvement in patients' physical function, along with reduced anxiety and depression	Increased confidence and motivation to exercise	Improved self-reported health-related QoL
Rutkowski et al [11], 2020	Significant improvement in SFT (arm curl, chair stand, and 6MWT; $P<.05$); ET[u]+VR superior to ET (eg, 6MWT: ET+VR=+39.11 m, ET=+16.24 m; $P<.05$)	Enhanced motivation and adherence in VR-based exercises	NR
Tu et al [24], 2020	Feasibility of smart in-home breathing training with RSA-BT[v]; real-time biofeedback effectively guided breathing patterns	Improved engagement with breathing exercises due to immersive VR	NR
Rutkowski et al [9], 2021	Reduction in emotional tension ($P<.001$), external stress ($P<.001$), depression ($P<.001$), and anxiety ($P<.001$); VR group showed significant stress, anxiety, and depression reduction compared to CG	Increased mood and emotional balance through immersive therapy	Statistically significant improvements in psychological well-being
Simmich et al [25], 2021	Use of the game (58.6% of d logged), daily steps, and MVPA[w]; 9 min/d increase in MVPA (EG) and 2% decrease in steps (EG) vs 13% decrease (CG)	Positive correlation between game use and steps; weak correlation with MVPA	No significant improvements reported
Simmich et al [26], 2021	Perceptions of wearables and AVGs[x] as tools for rehabilitation; participants found wearable trackers useful for quantifying activity, setting goals, and tracking improvements over time	AVGs were seen as fun and motivating for physical activity, but some participants felt that they were too difficult or not beneficial	No specific tools used to measure QoL, but general health benefits of physical activity were discussed
Baxter et al [27], 2022	Comparable inspiration durations between QUT Inspire (mean 7.3, SD 2.0 s) and Triflo II (mean 7.5, SD 2.3 s; $P=.79$); no significant differences in usability or performance between the app and the clinical device	Some users preferred app due to less perceived inspiratory effort	NR
Oberschmidt et al [28], 2022	Key patient values identified: independence, personal guidance, trust, and regularity; exergames supported values such as independence and challenge but hindered personal guidance and social interaction	Independence valued but personal guidance needed when using exergames	NR
Finkelstein et al [29], 2023	High usability and user acceptance (mean SUS[y] score: 95.8); 89% of the participants successfully completed the first task, and 100% completed tasks 2 and 3 without prompts	High interest in using VR for patient empowerment and PR education	NR
Gabriel et al [30], 2023	High usability scores (SUS score: 95.8/100); successful completion of PR tasks by all participants with minimal guidance	Increased willingness to engage with home-based PR through VR	NR
Gabriel et al [31], 2023	High acceptability and usability of the VR-based system; increased motivation and engagement due to the novel, immersive approach	Positive feedback on ease of use and enjoyment of the exercises	NR
Pancini et al [32], 2023	Reduction in anxiety, depression, and stress; increased relaxation and emotional well-being; expected to improve emotional well-being (based on prior research with similar methods)	Participants expected to experience increased emotional resilience	Expected improvements in emotional and psychological well-being
Pardos et al [33], 2023	Development of personalized recommendations based on health data; early results show potential for increased adherence to care plans using personalized recommendations	Expected improvement in health-related behavior through gamification	NR
Colombo et al [34], 2024	Adherence rate of 85.71%; mean 6MWT distance improved to 520.50 (SD 69.24) m; significant improvements in exercise capacity ($P<.05$)	Increased motivation to exercise through VR	NR
Jin et al [35], 2024	Significant improvements in 6MWD and Brief-BESTest[z] at 3 mo after the intervention ($P<.001$); EG maintained higher endurance and balance for 12 mo	Enhanced exercise tolerance and balance function; motivation sustained for 3 mo	Significant balance and exercise tolerance improvement
Kizmaz et al [36], 2024	Sit-to-stand test: significant improvement in PR+VR group ($P<.001$); COPD assessment test: significant reduction ($P<.001$), VR+PR group had greater improvements	Increased motivation and adherence to exercise reported in VR+PR group	Greater improvement in daily activities (London Chest Activity of Daily Living) in PR+VR group ($P<.001$)

XSL•FO
RenderX

Study	Effectiveness results	Behavioral outcomes	QoL[a] improvements
McAnirlin et al [10], 2024	Psychological well-being, heart rate, respiratory rate, and oxygen saturation; positive changes in well-being and presence; no cybersickness reported	Participants experienced positive emotional responses, reflective of nostalgic memories	Reported feelings of autonomy, positive emotions linked to memories, and restorative effects

[a]QoL: quality of life.
[b]CG: control group.
[c]VAS: visual analog scale.
[d]CCQ: Clinical COPD Questionnaire.
[e]6MWT: 6-minute walk test.
[f]EG: experimental group.
[g]TDI: transition dyspnea index.
[h]SGRQ: St George's Respiratory Questionnaire.
[i]6MWD: 6-minute walk distance.
[j]PR: pulmonary rehabilitation.
[k]VGS: virtual game systems.
[l]GRAIL: Gait Real-time Analysis Interactive Lab.
[m]ICC: intraclass correlation coefficient.
[n]PSA: public service announcement.
[o]ICT: information and communication technology.
[p]AR: augmented reality.
[q]NR: not reported.
[r]MET: metabolic equivalent of task.
[s]SFT: Senior Fitness Test.
[t]VR: virtual reality.
[u]ET: exercise training.
[v]RSA-BT: respiratory sinus arrhythmia biofeedback-based breathing training.
[w]MVPA: moderate to vigorous physical activity.
[x]AVG: active video game.
[y]SUS: System Usability Scale.
[z]Brief-BESTest: Brief Balance Evaluation Systems Test.

Engagement, Satisfaction, and Usability

The reviewed studies demonstrated strong engagement and satisfaction among patients with COPD using gamified interventions, with many participants expressing enjoyment and adherence to the programs; for instance, studies involving VR and exergames, such as the one by Rutkowski et al [9], showed high engagement due to immersive environments and interactive features, leading to consistent participation. Satisfaction metrics varied, with studies such as the one by Baxter et al [27] noting that visual rewards and timers within apps were especially motivating.

Adaptations to improve usability included simplified interfaces and controls, particularly for older adults with limited technological experience [29,30]. However, technical challenges were present, such as synchronization issues in Fitbit devices [25] and discomfort from equipment such as VR headsets [31]. Cultural adaptations were minimal but occasionally tailored to suit specific populations, as seen in the studies by Pardos et al [33] and Colombo et al [34], which adapted VR for older adults with COPD.

Table 5 provides detailed insights into engagement, satisfaction, and technical adjustments across the reviewed studies.

XSL•FO
RenderX

Table 5. Engagement, satisfaction, and usability characteristics of gamified interventions.

Study	Engagement	Satisfaction	Cultural adaptation	Technical adaptation	Challenges
Tabak et al [8], 2014	Web portal use: 86.4% of d; exercise adherence: 21%; activity coach used for 299 d (132 d monitoring and 167 d feedback)	Satisfaction (CSQ-8[a]): telehealth group—26.4, CG[b]—30.4 (out of 32)	NR[c]	Wearable and web-based portal; no adaptation for other platforms	Technical issues with activity coach (eg, cycling accuracy); low exercise adherence (21%)
Mazzoleni et al [14], 2014	7 additional Wii Fit sessions for EG[d]; all completed	Satisfaction: EG—42.4, CG—43.9 (out of 49)	NR	None beyond Wii Fit system	Initial difficulty with balance board; exclusion of patients with motor limitations
Kotrach et al [15], 2015	All participants adhered to the VGS[e] training	NR	NR	None beyond training sessions	Language barriers and patient ability to use VGS
Hoaas et al [12], 2016	On average, 3 diary entries per wk and 1.7 training sessions per wk	Increased self-efficacy and emotional safety; participants experienced health benefits	NRNR	iPad and treadmill used to adapt exercise training to home settings	Some technical difficulties with videoconferencing
LeGear et al [16], 2016	90% enjoyed Wii intervention, and 80% agreed that it could be used at home	NR	NR	None beyond standard Wii setup	Some participants required supervision for safe use
Liu et al [17], 2016	75% of patients with COPD[f] and 90% of healthy older adults improved in second GRAIL[g] test	NR	NR	No major technical adaptation beyond GRAIL VR[h] setup	Complex setup required; difficulty for patients using self-paced treadmill
Bamidis et al [18], 2017	Various engagement tools: achievements and self-reported progress via mobile apps	NR	Tailored to socioeconomic and cultural contexts of various countries	Use of ICT[i], mobile apps, SMS text messaging, and gamification across different platforms	Interoperability and customization for different health care systems; potential digital divide
Burkow et al [19], 2018	Peer monitoring and virtual group updates drove engagement	High acceptance; improved adherence to exercise routines; group motivation	NR	Tablet optimized with all other apps disabled	Minor technical issues (weather widget and activity sensor)
De Las Heras et al [20], 2018	High engagement; 12 out of 13 patients appreciated the AR[j] glasses	Suggestions for improvement: adjustable screen, brightness, and head fixation	NR	Adjustments to AR glasses design and usability proposed by patients	Issues with head fixation during movement; brightness control
Parent et al [21], 2018	91% of the participants reached high-intensity levels in squatting exercises	Reported enjoyment, motivation for home use, and exercise tolerance	NR	None beyond Kinect customization	Participants experienced some discomfort in using new technology
Rutkowski et al [22], 2019	High adherence to both standard and virtual rehabilitation programs	Significant improvement in exercise tolerance	NR	Basic Kinect setup for stationary use; no advanced technical customizations	Minor technical issues with Kinect system
Sutanto et al [23], 2019	High adherence to the Wii Fit program	NR	Conducted in an Indonesian context but no specific cultural adaptations noted	Wii Fit program customized to the local setting; no major technical challenges	Limited intensity tracking, high cost of the Wii Fit program
Jung et al [13], 2020	High engagement due to enjoyment and immersive aspects	Improved QoL[k], patient satisfaction, and engagement	NR	Feedback on improving headset weight and app functionality	Minor technical glitches; request for more customizable exercise levels
Rutkowski et al [11], 2020	High adherence (95% participation rate)	NR	NR	None beyond basic setup with Kinect	None significant; minor technical adjustments needed

XSL•FO
RenderX

Study	Engagement	Satisfaction	Cultural adaptation	Technical adaptation	Challenges
Tu et al [24], 2020	High engagement in demonstration sessions; real-time feedback kept users on track	User feedback on usability, engagement, and real-time performance improvements	NR	Used lightweight algorithms and readily available devices for home use	Some technical refinements (eg, headset comfort and sound effects) suggested by users
Rutkowski et al [9], 2021	High engagement in VR group with full participation over the 2 wk	NR	NR	Use of VR TierOne device; simple immersion setup	NR
Simmich et al [25], 2021	Participants' interest in wearables increased with social interaction and family involvement; challenges in long-term adherence were noted	Barriers and motivators for using wearables and AVGs[l]	NR	NR	Participants struggled with technological complexity and preferred more straightforward options
Simmich et al [26], 2021	High adherence to Fitbit (84.3% of d); moderate GEQ[m] score of 30.4	Engagement metrics (IMI[n] and GEQ); adherence to Fitbit	NR	No notifications, limiting engagement	Bluetooth synchronization issues with Fitbit
Baxter et al [27], 2022	High satisfaction with visual rewards; 75% found the timer motivating	User satisfaction with app's usability, responsiveness, and animations	NR	Distance measurement for inspiratory detection needed improvement	App required further technical refinement to improve microphone sensitivity
Oberschmidt et al [28], 2022	Exergames promoted challenge and seeing results, motivating participants	NR	NR	Issues with camera accuracy during exercise detection	Technical errors with exercise detection and loud notifications disrupted patient comfort
Finkelstein et al [29], 2023	Positive feedback for visual feedback, ease of navigation, and VR app structure	NR	NR	Simplified controls and interface for older adults with limited computer skills	Minor difficulties in finding and starting the app initially
Gabriel et al [30], 2023	NR	High satisfaction with visual feedback and educational content (mean posttask scores: 4.74-4.89 [out of 5])	NR	Simplified interface and navigation for older adults with limited technological experience	Minor difficulties in initial navigation and setup
Gabriel et al [31], 2023	High engagement; increased focus during exercises; minimal distractions	Improved motivation, focus on exercise content, and engagement	NR	Simplified interface for older adults with limited technological skills	Difficulty with headset weight, loading screens, and initial app navigation
Pancini et al [32], 2023	High engagement anticipated due to immersive VR and personalized savoring exercises	Enhanced positive emotions and psychological well-being	NR	Simplified interface to ensure ease of use for older adult patients	NR
Pardos et al [33], 2023	Scoring system with credits aimed at enhancing patient engagement	NR	NR	Data from smartwatches and Bluetooth devices integrated for monitoring	Further development required to expand recommendation domains
Colombo et al [34], 2024	86.85% attendance rate	High user engagement (mean Short Flow State Scale score 4.40, SD 0.36); fatigue and dyspnea improvements	Focused on older Italian patients with COPD	Use of semi-immersive VR to suit hospital settings	Issues with scaling workload increments
Jin et al [35], 2024	82.5% adherence in the intervention group	NR	Tailored to older Chinese patients with COPD	Visual feedback and game variety catered to balance and respiratory issues	Unclear measurement of exercise intensity

XSL•FO
RenderX

Study	Engagement	Satisfaction	Cultural adaptation	Technical adaptation	Challenges
Kizmaz et al [36], 2024	VR+PR[o] group had significantly longer pedaling time (508.44 s vs 357.56 s; *P*=.007)	NR	NR	Real-world footage of cycling in a forest used to enhance ecological realism	One patient could not continue due to dizziness related to VR use
McAnirlin et al [10], 2024	Participants cocreated their own VR experiences, leading to high engagement and satisfaction	NR	Customized to individual preferences and memories	Personalized VR experiences were created using 360-degree videos	Customization required multiple visits and effort to personalize scenes

[a]CSQ-8: Client Satisfaction Questionnaire-8.
[b]CG: control group.
[c]NR: not reported.
[d]EG: experimental group.
[e]VGS: virtual game systems.
[f]COPD: chronic obstructive pulmonary disease.
[g]GRAIL: Gait Real-time Analysis Interactive Lab.
[h]VR: virtual reality.
[i]ICT: information and communication technology.
[j]AR: augmented reality.
[k]QoL: quality of life.
[l]AVG: active video game.
[m]GEQ: Game Engagement Questionnaire.
[n]IMI: Intrinsic Motivation Inventory.
[o]PR: pulmonary rehabilitation.

Risk of Bias and Limitations

The reviewed studies presented varying levels of bias and limitations. For the RCTs, additional considerations included randomization procedures and blinding, while the non-RCT studies were evaluated based on their specific study design characteristics. Many of the studies (28/29, 97%), particularly early-stage trials, did not formally assess the risk of bias, often resulting in moderate risk due to factors such as small sample sizes and a lack of long-term follow-up [8,14]. Only a few studies (1/29, 3%), such as the one by Rutkowski et al [9], used formal tools and structured randomization to minimize bias, resulting in a lower risk profile.

Common limitations included a reliance on small sample sizes and subjective data, particularly in studies using self-reported outcomes, which may introduce reporting bias [25]. Technical challenges, such as issues with device accuracy and system usability, also affected study outcomes, as seen in the studies by Tabak et al [8] and Oberschmidt et al [28]. Furthermore, most of the studies (13/29, 45%) lacked long-term follow-up, limiting their ability to evaluate sustained effects of the interventions.

Table 6 provides a detailed breakdown of each study's risk-of-bias assessment, findings, and limitations.

XSL•FO
RenderX

Table 6. Risk of bias and study limitations.

Study	Risk-of-bias assessment	Risk-of-bias findings	Limitations
Tabak et al [8], 2014	NR[a]	Moderate risk (selection and attrition biases)	Small sample, high dropout (86% in CG[b]), and low exercise adherence (21%)
Mazzoleni et al [14], 2014	NR	Moderate risk (small sample size and lack of blinding)	Small sample, short duration, and lack of long-term follow-up
Kotrach et al [15], 2015	NR	Moderate risk (small sample size and preliminary data)	Small sample size, lack of long-term follow-up, and exclusion due to language barriers
Hoaas et al [12], 2016	NR	Moderate risk (small sample size)	Small sample size and seasonal effects on adherence
LeGear et al [16], 2016	NR	Moderate risk (small sample size and short duration)	Small sample size, lack of long-term follow-up, and supervision needed for safe exercise
Liu et al [17], 2016	NR	Moderate risk (lack of randomization and single-site study)	Small sample size, lack of long-term follow-up, monocentric, and limited applicability to patients classified as GOLD[c] stage 4
Bamidis et al [18], 2017	Not assessed in early stages of project	Risk of bias expected due to self-reported data and social desirability	No long-term results yet; potential for socioeconomic and geographic disparities in outcomes
Burkow et al [19], 2018	NR	Low generalizability due to small sample size and bias from prior rehabilitation experience	Small sample size, self-reported activity data, and lack of a CG
De Las Heras et al [20], 2018	NR	Moderate risk (small sample size and subjective feedback)	Small sample size and lack of long-term follow-up; only Nordic countries involved
Parent et al [21], 2018	NR	Moderate risk (small sample size; single session)	Short study duration, small sample size, and lack of long-term follow-up
Rutkowski et al [22], 2019	NR	Moderate risk (short intervention duration)	Short duration, lack of long-term follow-up, and no blinding of participants
Sutanto et al [23], 2019	NR	Moderate risk (small sample size)	Small sample size, unblinded study, and lack of intensity monitoring for the Wii exercises
Jung et al [13], 2020	NR	Moderate risk (small sample size and acknowledged limitations)	Small sample size and limited generalizability
Rutkowski et al [11], 2020	NR	Low risk (structured randomization and CG)	Short duration (2 wk); only patients classified as GOLD stages 2 and 3 included
Tu et al [24], 2020	NR (demonstration phase)	NR	Small-scale demonstration, short duration, and lack of long-term data
Rutkowski et al [9], 2021	Assessor-blinded RCT[d] with controlled randomization (low risk of bias)	Low risk (structured randomization and CG)	Short duration; only hospital based
Simmich et al [25], 2021	NR	Low risk (small sample size and self-reported data)	Lack of generalizability due to the small sample size and limited geographic representation
Simmich et al [26], 2021	NR	Possible bias due to the involvement of the EG[e] in co-design	Small sample size, short trial duration, Fitbit issues, and lack of notifications in the game
Baxter et al [27], 2022	NR	Low risk (randomization and crossover design)	Small sample, short session duration, and lack of clinical participants
Oberschmidt et al [28], 2022	NR	Some dropouts due to exacerbation but not directly related to intervention	Small sample size, dropouts after 12 wk, and occasional technical issues
Finkelstein et al [29], 2023	NR	Low risk (all participants completed the tasks without significant issues)	Small sample size; lack of long-term follow-up
Gabriel et al [30], 2023	NR	Low risk (comprehensive task completion by all participants)	Small sample size, no CG, and lack of long-term follow-up
Gabriel et al [31], 2023	NR	Low risk (most participants completed the tasks easily)	Small sample size, short duration, and lack of long-term follow-up

XSL•FO
RenderX

Study	Risk-of-bias assessment	Risk-of-bias findings	Limitations
Pancini et al [32], 2023	NR (planned study)	NR (study pending implementation)	Small sample size; short intervention duration
Pardos et al [33], 2023	NR (pilot)	NR	Exclusion of factors such as nutrition, smoking, and drinking; system still in development
Colombo et al [34], 2024	NR	Low risk (high adherence and positive outcomes)	Small sample size; no CG
Jin et al [35], 2024	NR	Low risk (group similarity and controlled environment)	Lack of long-term adherence tracking; reliance on self-report
Kizmaz et al [36], 2024	NR	Likely low risk, given the blinded evaluator and randomized design	No objective assessment of cybersickness or patient satisfaction; no third group for comparison with usual care
McAnirlin et al [10], 2024	NR	NR	Small sample size, no CG, and exploratory design

[a]NR: not reported.

[b]CG: control group.

[c]GOLD: Global Initiative for Chronic Obstructive Lung Disease.

[d]RCT: randomized controlled trial.

[e]EG: experimental group.

Discussion

Synthesis of Findings

Effectiveness in COPD Management

The reviewed studies highlighted the positive impact of gamified interventions on physical outcomes, QoL, and patient engagement in COPD management; for instance, interventions incorporating VR or exergames, such as those by Mazzoleni et al [14] and Jung et al [13], showed improved exercise tolerance with significant gains in the 6-minute walk test. Similarly, gamified tools often led to enhanced QoL, as patients reported reduced anxiety and a sense of achievement due to positive reinforcement mechanisms [9,32]. Such outcomes align with broader trends noted in the literature, where eHealth interventions had a positive effect on exercise endurance and QoL in patients with COPD [5]. Through interactive and motivational features, gamified interventions provide both physiological and psychological benefits, which seem crucial for managing COPD symptoms more effectively.

Engagement and Adherence

Gamified interventions demonstrated high engagement and adherence due to interactive features such as real-time feedback and immersive environments. Studies using VR or gamified apps, such as the ones by Tu et al [24] and Hoaas et al [12], reported consistently high engagement rates, with patients maintaining adherence due to motivational tools such as rewards and feedback loops [27,29]. These methods seem to support sustained behavioral changes necessary for chronic disease management [19]. Similar findings were echoed in reviews where gamification substantially enhanced adherence in COPD and other chronic disease management contexts [2,6]. Consequently, integrating gamified elements into COPD management is effective in maintaining long-term adherence, critical for preventing exacerbations.

Comparative Analysis With Nongamified Interventions

When comparing gamified and nongamified approaches, the evidence suggests that gamification provides superior patient engagement and adherence outcomes; for example, Chang et al [5] found that while telemonitoring improved exercise endurance, VR and game-based interventions fostered higher satisfaction. Liu et al [2] further emphasized that VR-based gamification was particularly beneficial in promoting exercise adherence, especially among patients with COPD who find traditional exercises repetitive. In addition, Selles et al [1] found that gamified components significantly increased physical activity in patients with chronic disease, outperforming conventional approaches. Similarly, Obrero-Gaitán et al [3] noted that VR gamified interventions enhanced motivation and enjoyment, making COPD management more accessible and engaging. These findings underscore that gamified strategies may provide a more effective, patient-centered approach to COPD care.

Main Findings in RCT Studies

RCTs are effective in reducing bias; controlling for external variables; and providing more precise and broadly applicable results, which is especially important in intervention studies of complex diseases such as COPD. To highlight the results from higher-quality studies, we further discuss the main findings of the RCTs (11/29, 38%) included in this review. Several of the RCTs (7/11, 64%) reported significant improvements in physical fitness and exercise capacity, including 6-minute walk distance, exercise tolerance, and performance on the Senior Fitness Test. Specifically, experimental groups showed greater improvements in these areas than control groups, indicating enhanced physical endurance and exercise capacity [11,14,22,23,26,35,36]. Another finding is that, compared to control groups, intervention groups demonstrated significant improvements in emotional well-being and mental health indicators, such as anxiety, depression, and stress [9,32]. However, there were also RCTs (1/29, 3%) that

XSL•FO
RenderX

found no statistically significant difference in clinical outcomes between the groups [8].

Strengths and Limitations

Strengths

Gamified interventions offer unique strengths in COPD management, particularly in enhancing patient engagement and supporting home-based or self-managed care. By incorporating features such as real-time feedback, interactive environments, and personalized adjustments, gamified tools address some of the limitations of traditional, clinic-based programs [8,13]. These interventions leverage technology to empower patients to monitor their symptoms, engage in rehabilitation activities, and adhere to exercise routines from home, providing significant flexibility and independence, which are highly valued in chronic care [6]. In addition, gamified tools have shown positive psychological effects, with many participants reporting improvements in mood, motivation, and self-efficacy [9,32]. This approach aligns well with the growing shift toward patient-centered care, emphasizing autonomy and self-management in chronic disease contexts such as COPD [3].

Limitations

Despite their strengths, the reviewed studies revealed several methodological challenges that limit the generalizability and consistency of the findings. One primary limitation was the heterogeneity across study designs, intervention types, and technology platforms, making it difficult to draw consistent conclusions about intervention efficacy; for example, while some of the studies (15/29, 52%) used VR-based environments, others (14/29, 48%) used simple exergames or mobile apps, leading to varied engagement and adherence outcomes [14,17]. Limited sample sizes and short intervention durations were also frequent issues, with some of the studies (15/29, 52%) having <20 participants or lasting only a few weeks [12,15]. Technical issues such as usability challenges and the need for advanced equipment can create barriers for older adults with limited technological familiarity [1,2]. Furthermore, a formal risk-of-bias assessment using validated scales, such as the Physiotherapy Evidence Database scale, was not conducted in this review; while we critically appraised the included studies based on study design, sample size, and methodological rigor, the absence of a standardized quality-scoring system may limit the comparability of study quality across different methodologies. Finally, while the inclusion of the China National Knowledge Infrastructure database allowed us to capture studies relevant to the Chinese population, it might limit the reproducibility of the work for researchers without access to this database. These factors contribute to biases and limitations in existing literature, indicating a need for larger, longer-term trials to assess gamified interventions effectively.

Potential of Gamified Interventions in COPD Management

Digital health technologies and gamification hold considerable potential as transformative approaches for managing chronic respiratory diseases such as COPD. By making rehabilitation more interactive and accessible, these interventions can support sustained patient engagement between medical presentations for exacerbations and beyond traditional clinical settings, addressing key challenges in long-term COPD care [5]. The ability of gamified interventions to integrate with real-time monitoring and provide personalized feedback aligns well with the goals of self-managed, home-based COPD care, making it easier for patients to monitor their symptoms and manage exacerbations effectively [24,29]. In addition, gamification encourages continuous engagement, which is crucial for adherence to exercise routines in COPD rehabilitation [6]. As research advances, digital health and gamification could become essential tools in the COPD management toolkit, offering a scalable, patient-centered approach to chronic respiratory care.

Conclusions

This systematic review highlights the positive impact of gamified interventions on COPD management, demonstrating improvements in physical outcomes, patient engagement, and QoL. Various studies have shown that gamified tools—ranging from exergames to VR-based PR—enhance patients' adherence to exercise routines and promote self-management [8,9]. In particular, features such as real-time feedback, interactive guidance, and personalized adjustments make these interventions highly engaging and effective, aligning with the chronic, self-managed nature of COPD care. However, the studies also reveal challenges, such as technological barriers and limitations in sample size and duration, suggesting that more robust trials are needed to confirm the efficacy of gamified interventions in COPD management.

The integration of gamified interventions into COPD care presents significant opportunities for patients, health care providers, and policy makers. Gamified interventions empower patients with interactive and flexible tools for managing COPD symptoms, promoting independence and engagement in their care. These tools offer a novel approach to supporting patient adherence to home-based rehabilitation, reducing the reliance on in-clinic visits. Policy makers can leverage this approach to address the rising costs of COPD management by supporting digital health policies that encourage the development and use of gamified COPD interventions. Such policies could also promote accessibility to these tools across various populations, making COPD care more equitable and scalable.

Future research should focus on addressing the identified limitations of many of the studies in this review by conducting larger, long-term RCTs to assess the sustained impact of gamified interventions on COPD outcomes. Developing culturally relevant and adapted gamified tools can enhance the global applicability of these interventions. Collaborating with gaming companies could facilitate the integration of advanced game mechanics and elements, making interventions more engaging and effective. Furthermore, integrating personalized gamification approaches that cater to individual patient needs and preferences into health care could enhance adherence. Finally, exploring the long-term effects of gamified interventions will be essential for establishing their place in the chronic management of COPD and similar conditions.

XSL•FO
RenderX

Acknowledgments

This study was partly funded by the New Zealand–China Tripartite Partnership Fund 2024 (3730686). The research team greatly appreciates the funding support obtained.

Data Availability

All data generated or analyzed during this study are included in this published paper and its supplementary information files.

Authors' Contributions

TY, JC, and CB conceptualized the study. JC, QH, and MP were responsible for methodology. JC and LL were responsible for formal analysis. YC, ZL, and HIL were responsible for investigation. QH and LL curated the data. LL and MP were responsible for visualization. TY and CB were responsible for supervision. JC was responsible for project administration. TY and CB were responsible for funding acquisition. JC and MP wrote the original draft. CB, TY, and HIL reviewed and edited the manuscript. All authors have read and approved the version to be published.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Detailed search strategies for each database.
[DOCX File , 17 KB - games_v13i1e69510_app1.docx]

Multimedia Appendix 2
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[PDF File (Adobe PDF File), 137 KB - games_v13i1e69510_app2.pdf]

Multimedia Appendix 3
Template for the data extraction and analysis form.
[DOCX File , 17 KB - games_v13i1e69510_app3.docx]

Multimedia Appendix 4
Details of data extraction and analysis of each reviewed study.
[DOCX File , 149 KB - games_v13i1e69510_app4.docx]

References

1. Selles WL, Santos EC, Romero BD, Lunardi AC. Effectiveness of gamified exercise programs on the level of physical activity in adults with chronic diseases: a systematic review. Disabil Rehabil 2024 Dec;46(26):6231-6239. [doi: 10.1080/09638288.2024.2323614] [Medline: 38444153]
2. Liu YX, Du QF, Jiang YL. The effect of virtual reality technology in exercise and lung function of patients with chronic obstructive pulmonary disease: a systematic review and meta-analysis. Worldviews Evid Based Nurs 2024 Jun 31;21(3):307-317. [doi: 10.1111/wvn.12698] [Medline: 38297408]
3. Obrero-Gaitán E, Chau-Cubero CY, Lomas-Vega R, Osuna-Pérez MC, García-López H, Cortés-Pérez I. Effectiveness of virtual reality-based therapy in pulmonary rehabilitation of chronic obstructive pulmonary disease. A systematic review with meta-analysis. Heart Lung 2024 May;65:1-10 [FREE Full text] [doi: 10.1016/j.hrtlng.2024.01.011] [Medline: 38330853]
4. Schmidt-Kraepelin M, Thiebes S, Tran MC, Sunyaev A. What's in the game? Developing a taxonomy of gamification concepts for health apps. In: Proceedings of the 51st Hawaii International Conference on System Sciences. 2018 Presented at: HICSS '18; January 3-6, 2018; Waikoloa Village, HI p. 1217-1226 URL: https://scholarspace.manoa.hawaii.edu/server/api/core/bitstreams/93fa9001-cd42-4379-b4d7-607c2d223d40/content
5. Chang H, Zhou J, Chen YD, Wang XH, Wang ZW. Comparative effectiveness of eHealth interventions on the exercise endurance and quality of life of patients with COPD: a systematic review and network meta-analysis. J Clin Nurs 2024 Aug 27;33(9):3711-3720. [doi: 10.1111/jocn.17225] [Medline: 38937908]
6. Dalko K, Elsuson HA, Kalter I, Zilezinski M, Hofstetter S, Stoevesandt D, et al. Virtual reality applications for the implementation of domestic respiratory rehabilitation programs for patients with long COVID and post-COVID condition: scoping review. JMIR Serious Games 2024 May 31;12:e52309 [FREE Full text] [doi: 10.2196/52309] [Medline: 38819890]

XSL•FO
RenderX

7. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
8. Tabak M, Brusse-Keizer M, van der Valk P, Hermens H, Vollenbroek-Hutten M. A telehealth program for self-management of COPD exacerbations and promotion of an active lifestyle: a pilot randomized controlled trial. Int J Chron Obstruct Pulmon Dis 2014 Sep;9:935-944 [FREE Full text] [doi: 10.2147/COPD.S60179] [Medline: 25246781]
9. Rutkowski S, Szczegielniak J, Szczepańska-Gieracha J. Evaluation of the efficacy of immersive virtual reality therapy as a method supporting pulmonary rehabilitation: a randomized controlled trial. J Clin Med 2021 Jan 18;10(2):352 [FREE Full text] [doi: 10.3390/jcm10020352] [Medline: 33477733]
10. McAnirlin O, Browning MH, Fasolino T, Okamoto K, Sharaievska I, Thrift J, et al. Co-creating and delivering personalized, nature-based VR experiences: proof-of-concept study with four U.S. adults living with severe COPD. Wellb Space Soc 2024 Dec;7:100212. [doi: 10.1016/j.wss.2024.100212]
11. Rutkowski S, Rutkowska A, Kiper P, Jastrzebski D, Racheniuk H, Turolla A, et al. Virtual reality rehabilitation in patients with chronic obstructive pulmonary disease: a randomized controlled trial. Int J Chron Obstruct Pulmon Dis 2020;15:117-124 [FREE Full text] [doi: 10.2147/COPD.S223592] [Medline: 32021150]
12. Hoaas H, Andreassen HK, Lien LA, Hjalmarsen A, Zanaboni P. Adherence and factors affecting satisfaction in long-term telerehabilitation for patients with chronic obstructive pulmonary disease: a mixed methods study. BMC Med Inform Decis Mak 2016 Feb 25;16:26 [FREE Full text] [doi: 10.1186/s12911-016-0264-9] [Medline: 26911326]
13. Jung T, Moorhouse N, Shi X, Amin MF. A virtual reality-supported intervention for pulmonary rehabilitation of patients with chronic obstructive pulmonary disease: mixed methods study. J Med Internet Res 2020 Jul 07;22(7):e14178 [FREE Full text] [doi: 10.2196/14178] [Medline: 32673224]
14. Mazzoleni S, Montagnani G, Vagheggini G, Buono L, Moretti F, Dario P, et al. Interactive videogame as rehabilitation tool of patients with chronic respiratory diseases: preliminary results of a feasibility study. Respir Med 2014 Oct;108(10):1516-1524 [FREE Full text] [doi: 10.1016/j.rmed.2014.07.004] [Medline: 25087837]
15. Kotrach H, Dajczman E, Tremblay G, Baltzan M, Wardini R, Levitz S, et al. A pilot study using virtual game system to maintain adherence to home-based exercise following pulmonary rehabilitation in chronic obstructive pulmonary disease. Chest 2015 Oct;148(4):709A. [doi: 10.1378/chest.2257078]
16. LeGear T, LeGear M, Preradovic D, Wilson G, Kirkham A, Camp PG. Does a Nintendo Wii exercise program provide similar exercise demands as a traditional pulmonary rehabilitation program in adults with COPD? Clin Respir J 2016 May;10(3):303-310. [doi: 10.1111/crj.12216] [Medline: 25351282]
17. Liu WY, Meijer K, Delbressine JM, Willems PJ, Franssen FM, Wouters EF, et al. Reproducibility and validity of the 6-minute walk test using the gait real-time analysis interactive lab in patients with COPD and healthy elderly. PLoS One 2016 Sep 8;11(9):e0162444 [FREE Full text] [doi: 10.1371/journal.pone.0162444] [Medline: 27607426]
18. Bamidis PD, Paraskevopoulos E, Konstantinidis E, Spachos D, Billis A. Multimodal e-Health services for smoking cessation and public health: the SmokeFreeBrain project approach. Stud Health Technol Inform 2017;245:5-9. [Medline: 29295041]
19. Burkow TM, Vognild LK, Johnsen E, Bratvold A, Risberg MJ. Promoting exercise training and physical activity in daily life: a feasibility study of a virtual group intervention for behaviour change in COPD. BMC Med Inform Decis Mak 2018 Dec 18;18(1):136 [FREE Full text] [doi: 10.1186/s12911-018-0721-8] [Medline: 30563507]
20. De Las Heras JC, Tulppo M, Kiviniemi A, Hilberg O, Løkke A, Ekholm S, et al. Augmented reality glasses as a new tele-rehabilitation tool for home use: patients' perception and expectations. Disabil Rehabil Assist Technol 2022 May;17(4):480-486. [doi: 10.1080/17483107.2020.1800111] [Medline: 32750254]
21. Parent AA, Gosselin-Boucher V, Houle-Peloquin M, Poirier C, Comtois AS. Pilot project: physiologic responses to a high-intensity active video game with COPD patients-Tools for home rehabilitation. Clin Respir J 2018 May;12(5):1927-1936. [doi: 10.1111/crj.12760] [Medline: 29316273]
22. Rutkowski S, Rutkowska A, Jastrzębski D, Racheniuk H, Pawełczyk W, Szczegielniak J. Effect of virtual reality-based rehabilitation on physical fitness in patients with chronic obstructive pulmonary disease. J Hum Kinet 2019 Oct;69:149-157 [FREE Full text] [doi: 10.2478/hukin-2019-0022] [Medline: 31666897]
23. Sutanto YS, Makhabah DN, Aphridasari J, Doewes M, Suradi, Ambrosino N. Videogame assisted exercise training in patients with chronic obstructive pulmonary disease: a preliminary study. Pulmonology 2019;25(5):275-282 [FREE Full text] [doi: 10.1016/j.pulmoe.2019.03.007] [Medline: 31076287]
24. Tu L, Hao T, Bi C, Xing G. BreathCoach: a smart in-home breathing training system with bio-feedback via VR game. Smart Health 2020 May;16:100090. [doi: 10.1016/j.smhl.2019.100090]
25. Simmich J, Mandrusiak A, Smith ST, Hartley N, Russell TG. A co-designed active video game for physical activity promotion in people with chronic obstructive pulmonary disease: pilot trial. JMIR Serious Games 2021 Jan 27;9(1):e23069 [FREE Full text] [doi: 10.2196/23069] [Medline: 33502321]
26. Simmich J, Mandrusiak A, Russell T, Smith S, Hartley N. Perspectives of older adults with chronic disease on the use of wearable technology and video games for physical activity. Digit Health 2021 May 30;7:20552076211019900 [FREE Full text] [doi: 10.1177/20552076211019900] [Medline: 34104468]

XSL•FO
RenderX

27. Baxter CA, Carroll JA, Keogh B, Vandelanotte C. Virtual respiratory therapy delivered through a smartphone app: a mixed-methods randomised usability study. BMJ Open Respir Res 2022 Jun 27;9(1):e001221 [FREE Full text] [doi: 10.1136/bmjresp-2022-001221] [Medline: 35760497]
28. Oberschmidt K, Broekhuis M, Grünloh C. Patient values associated with an exergame supporting COPD treatment. Stud Health Technol Inform 2022 May 25;294:730-734. [doi: 10.3233/SHTI220573] [Medline: 35612193]
29. Finkelstein J, Parvanova I, Huo X. Feasibility of a virtual reality app to promote pulmonary rehabilitation. Stud Health Technol Inform 2023 May 18;302:458-462. [doi: 10.3233/SHTI230172] [Medline: 37203716]
30. Gabriel AS, Tsai TY, Xhakli T, Finkelstein J. Mixed-methods assessment of a virtual reality-based system for pulmonary rehabilitation. Stud Health Technol Inform 2023 Oct 20;309:245-249. [doi: 10.3233/SHTI230789] [Medline: 37869851]
31. Gabriel AS, Tsai TY, Xhakli T, Finkelstein J. Patient perceptions of a virtual reality-based system for pulmonary rehabilitation: a qualitative analysis. Stud Health Technol Inform 2023 Jun 29;305:406-409. [doi: 10.3233/SHTI230517] [Medline: 37387051]
32. Pancini E, Villani D, Riva G. oVeRcomING COPD: virtual reality and savoring to promote the well-being of patients with chronic obstructive pulmonary disease. Cyberpsychol Behav Soc Netw 2023 Jan 01;26(1):65-67. [doi: 10.1089/cyber.2022.29265.ceu] [Medline: 36595364]
33. Pardos A, Gallos P, Menychtas A, Panagopoulos C, Maglogiannis I. Enriching remote monitoring and care platforms with personalized recommendations to enhance gamification and coaching. Stud Health Technol Inform 2023 May 18;302:332-336. [doi: 10.3233/SHTI230129] [Medline: 37203673]
34. Colombo V, Mondellini M, Fumagalli A, Aliverti A, Sacco M. A virtual reality-based endurance training program for COPD patients: acceptability and user experience. Disabil Rehabil Assist Technol 2024 May 05;19(4):1590-1599. [doi: 10.1080/17483107.2023.2219699] [Medline: 37272556]
35. Jin XL, Jin MN, Zhang BL, Niu ME, Han YX, Qian JL. The association of conventional therapy associated with somatosensory interactive game enhances the effects of early pulmonary rehabilitation for patients with acute exacerbation of chronic obstructive pulmonary disease: a randomized controlled trial. Games Health J 2025 Apr 29;14(2):127-135. [doi: 10.1089/g4h.2023.0095] [Medline: 39207252]
36. Kizmaz E, Telli Atalay O, Çetin N, Uğurlu E. Virtual reality for COPD exacerbation: a randomized controlled trial. Respir Med 2024;230:107696. [doi: 10.1016/j.rmed.2024.107696] [Medline: 38857811]

Abbreviations

COPD: chronic obstructive pulmonary disease
PR: pulmonary rehabilitation
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QoL: quality of life
RCT: randomized controlled trial
VR: virtual reality

Edited by A Coristine; submitted 02.12.24; peer-reviewed by G Cangelosi, W Selles; comments to author 20.01.25; revised version received 20.02.25; accepted 18.04.25; published 30.05.25.

Please cite as:
Chen J, Yang T, He Q, Pang M, Cao Y, Liu Z, Li L, Liu HI, Bullen C
The Impact of Gamified Interventions on the Management of Chronic Obstructive Pulmonary Disease: Systematic Literature Review
JMIR Serious Games 2025;13:e69510
URL: https://games.jmir.org/2025/1/e69510
doi:10.2196/69510
PMID:

©Jinsong Chen, Tingzhong Yang, Qilian He, Mingli Pang, Ying Cao, Zheng Liu, Linfei Li, Hsing-I Liu, Christopher Bullen. Originally published in JMIR Serious Games (https://games.jmir.org), 30.05.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Technology-Assisted Motor-Cognitive Training Among Older Adults: Rapid Systematic Review of Randomized Controlled Trials

Yaqin Li[1], MS; Yaqian Liu[1], MS; Angela YM Leung[1], Prof Dr; Jed Montayre[1], Prof Dr

School of Nursing, Hong Kong Polytechnic University, Hong Kong, China (Hong Kong)

Corresponding Author:
Jed Montayre, Prof Dr
School of Nursing
Hong Kong Polytechnic University
11 Yuk Choi Road, Hung Hom, Kowloon
Hong Kong,
China (Hong Kong)
Phone: 852 2766 6393
Fax: 852 2766 6393
Email: jed-ray.montayre@polyu.edu.hk

Abstract

Background: Age-related physiological changes in older adults involve a rapid decline in motor exercise ability; some older adults may also experience difficulties in maintaining focus, memory loss, and a decline in reaction time, which consequently impair their ability to perform dual tasks. Motor-cognitive training (MCT) refers to a blend of motor activity and cognitive training that occurs simultaneously and can assist older adults in enhancing their physical function, cognitive abilities, and dual-task performance. In recent years, the use of technology for delivering MCT has become increasingly popular in research. This has been achieved through various technologies that simplify MCT for older adults.

Objective: This study aimed to systematically examine the feasibility and effectiveness studies on technology-assisted MCT among older adults.

Methods: This rapid review was conducted following the updated PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 standards, and the Synthesis Without Meta-analysis (SWiM) in systematic reviews reporting guideline. Four databases were searched, including CINAHL, Embase, PubMed, and Scopus, from January 2013 to March 2025. Search strategies were constructed based on three main topics: (1) older adults, (2) MCT, and (3) technology. Inclusion criteria followed the population, intervention, comparator, outcome, and study design framework as follows: older adults (population); technology-assisted MCT (intervention); standard treatment control, active control, partial intervention control, placebo control, and dose-response control (comparator); various measures of physical, cognitive, and dual-task performance (outcome); and randomized controlled trials (RCTs) and pilot RCTs (study design). The Cochrane Risk of Bias Tool was applied for quality appraisal of the included studies. The feasibility of the included studies was assessed using completion rates and attrition rates. Descriptive statistics were used to describe the demographic and clinical characteristics of the groups, while narrative methods were used to categorize and synthesize their effectiveness.

Results: In total, 20 studies were included, comprising 16 RCTs and 4 pilot RCTs, most of which were conducted within a 6-week period. Each session typically lasted between 10 and 30 minutes and was held 2 to 3 times per week. Feasibility analysis showed that technology-assisted MCT was generally feasible. While the workload was high, the perceived usability was also high, with a considerable amount of positive feedback and very few reported adverse events. The types of MCT varied in terms of components, duration, and frequency. The majority of studies (18/20, 90%) demonstrated statistically significant improvements in physical, cognitive, and dual-task performance because of technology-assisted MCT.

Conclusions: The feasibility of technology-assisted MCT among older adults was high regardless of the perceived high workload, and most studies showed statistical effectiveness in improving physical, cognitive, and dual-task performance.

Trial Registration: Open Science Foundation (OSF) Registries 10.17605/OSF.IO/5SRCQ; https://osf.io/5srcq

(JMIR Serious Games 2025;13:e67250) doi:10.2196/67250

XSL•FO
RenderX

KEYWORDS

technology-assisted; motor-cognitive; performance; older adults; randomized controlled trials

Introduction

Background

The United Nations projects that the number of people aged ≥65 years will reach 1.6 billion by 2050, accounting for 16% of the total population [1]. Older adults will experience age-related decline regarding their physical, cognitive, and dual-task reserves and functions [2]. For example, physically, most of them will experience osteoporosis, connective tissue problems [3,4], and muscle fiber loss [5], contributing to dysfunction [6,7]. Cognitively, processing speed demonstrates the earliest decline [8], along with a serious decline in learning ability, attention, visual-spatial capability, and working memory [9-11]. These factors collectively decrease their physical and motor reserves and functions, while cognitive decline will further lead to issues such as memory loss, difficulty maintaining focus, and ultimately slower thinking [12,13]. "Dual tasking" refers to performing 2 tasks simultaneously, typically combining cognitive and motor tasks of varying complexity, such as doing simple arithmetic like addition or subtraction while walking [14,15]. Dual tasks are common in daily life and demand the simultaneous processing of multiple pieces of information and tasks, placing greater strain on physical and cognitive reserves and functions [16,17]. Older adults not only experience a simultaneous reduction in both motor and cognitive reserves and functions, resulting in a gradual or sharp decline in their ability to perform dual tasks, but also reductions in both processing speed, as mentioned earlier, and reaction time [18], which further impairs their ability to perform dual tasks. In the meantime, limited opportunities to perform dual tasks in daily life conversely result in the deterioration of physical, cognitive, and dual-task reserves and functions [19]. These changes result in decreased physical, cognitive, and dual-task reserve and functions in older adults [5-8], and they ultimately impact their activities of daily living (ADLs) [20].

Motor-cognitive training (MCT) is the most typical type of dual task, which combines motor activity (eg, aerobic, strength, and endurance training) with cognitive training (eg, memory and attention training) [21,22]. This approach significantly improves physical, cognitive, and dual-task performance in various populations [23-25]. For instance, a systematic review and meta-analysis of 11 randomized controlled trials (RCTs) with 322 participants showed that MCT improved gait, motor symptoms, and balance in patients with Parkinson disease (PD) [26]. Another review of 13 studies with 584 patients found MCT to be beneficial for balance in patients with multiple sclerosis [27]. In addition, MCT offers synergistic benefits by combining motor and cognitive training, enhancing brain adaptations and cognition [28]. A review of 21 studies with 2221 participants showed small-to-medium improvements in cognitive functions in older adults with cognitive impairment [23]. Synthesized evidence on MCT showed that it improves dual-task performance in patients with PD [29] and positively impacts cognitive and physical functions in individuals with dementia or mild cognitive impairment [30].

In recent years, as technology is becoming increasingly integrated into our daily lives, its use for delivering MCT has also become increasingly popular in research [31]. Technology-assisted intervention involves incorporating technology—such as digital devices, tele-technology, monitoring, and assistive technology—into clinical interventions to assess, monitor, educate, and enhance the effectiveness of an intervention [32]. Technology can address challenges in traditional MCT by refining technological innovations in response to specific needs and creating interactive scenarios [31]. The use of technology can provide participants with standardized and uniform instructions on their interventions, ensuring consistency of MCT [33,34]. Technology can make MCT more feasible, enjoyable, and relaxing for older adults [35-38], providing timely assistance and feedback [39-42], and increasing the likelihood of voluntary participation and long-term adherence [43-46]. Technology enhances the motor activity component of MCT by offering premade demonstrations and real-time environments, reducing the need for trainers and reserved spaces [47,48]. Sensors and interactive technologies can monitor movements in real time, ensuring proper posture and technique [49-51], and record exercise processes for future improvement [52,53]. In addition, technology can expand cognitive training beyond basic tasks, offering diverse and engaging exercises through virtual reality (VR) and exergames, which simulate complex cognitive scenarios [54,55]. Despite these advantages, few reviews have investigated the feasibility and effectiveness of technology-assisted MCT among older adults.

Objectives

In this study was a rapid systematic review to identify various types of technology-assisted MCT and examine the literature in terms of the feasibility and effectiveness of technology-assisted MCT in older adults.

Methods

Overview

This rapid review was conducted in accordance with the updated Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 standards [56], and the Synthesis Without Meta-analysis (SWiM) in systematic reviews reporting guidelines [57]. The PRISMA and SWiM checklists can be found in Multimedia Appendices 1 and 2, respectively. In addition, 2 other guidelines were used: "Rapid Reviews to Strengthen Health Policy and Systems: A Practical Guide" from the Alliance for Health Policy and Systems Research, World Health Organization (WHO) [58], and "Conducting a Rapid Review for Quick Turnaround Knowledge Synthesis" [59,60]. These methodologies were integrated into the process from the search strategy to the presentation of results. This review was registered on the Open Science Framework platform [61]. Details of registration are provided in Multimedia Appendix 3.

XSL•FO
RenderX

Search Methods

We conducted a rapid systematic search for relevant studies using 4 databases, including CINAHL, Embase, PubMed, and Scopus, focusing on literature from the past decade (from January 2013 to March 2025) [58,60]. To ensure comprehensiveness, we also reached out to the authors of eligible studies without full texts, authors of conference abstracts, and authors of important studies. Our search terms centered on three main topics: (1) older adults, (2) MCT, and (3) technology, by using Medical Subject Headings (MeSH), terms Emtree terms, free-text terms, combinations of these terms, and Boolean operators (eg, and, or, and not). For example, terms like *aged* (MeSH, Emtree), *senior**, *elderly* (Emtree) for older adults; *DT*, *concurrent-task*, *cognitive-motor* for MCT; and *technology* (MeSH, Emtree), *virtual reality* (MeSH, Emtree), *exergaming* (MeSH) for technology were applied. Details of our search strategies can be found in Multimedia Appendix 4.

Inclusion and Exclusion Criteria

Published papers were included in this review according to the inclusion criteria following the population, intervention, comparator, outcome, and study design (PICOS) framework [62]. The exclusion criteria are also provided in the subsequent sections. More details of the inclusion and exclusion criteria for the included studies in this review are in Textbox 1.

Textbox 1. The inclusion and exclusion criteria of the studies included in this review.

Inclusion criteria

- Older adults aged >55 years
- Motor-cognitive training (MCT) delivered using any technology as a whole or part of the intervention
- Standard treatment control, active control, partial intervention control, placebo control, and dose-response control
- A variety of measures related to the physical, cognitive, and dual-task performance
- Randomized controlled trials (RCTs) and pilot RCTs

Exclusion criteria

- Adults aged <55 years
- Non-MCT or MCT not delivered using any technology as a whole or part of the intervention
- Other measures not focusing on physical, cognitive, and dual-task performance
- Other non-RCT studies (eg, pre-post studies)

Inclusion Criteria

Details on inclusion criteria following the PICOS framework are as follows:

For population, we included studies that reported that their participants were older adults, without restrictions based on sociodemographic characteristics such as gender, medical background (clinical or nonclinical population), residence, ethnicity, educational background, or occupational status.

For interventions, technology-assisted MCT was defined as MCT delivered using any technology (including VR, exergames, phones, computers, tele-technology, and other types of technology). We also included studies that used technology-assisted MCT as part of the overall intervention.

For comparators or control, we included studies that compared technology-assisted MCT to treatment. This could be standard treatment control, active control, partial intervention control, placebo control, or dose-response control.

For main outcomes, indicators included a variety of measures related to the physical, cognitive, and dual-task performance of the participants. These measures involved the use of scales, instruments, and sensors to measure physiological metrics of the patient and other tools.

For study design, RCTs and pilot RCTs were included. RCTs use an experimental research design that minimizes bias by randomly assigning participants to intervention and control groups to evaluate the effect of a specific intervention on study outcomes [63]. In contrast, a pilot RCT is a smaller-scale version of an RCT, typically conducted before a formal RCT study, to assess the feasibility of the intervention, the validity of experimental procedures, and the operationalization of data collection, as well as to test the study design [64]. RCTs were included in this review because they provide robust evidence, are better able to minimize bias, and offer high internal validity. Pilot RCTs were also considered because the research question addressed by this rapid systematic review concerns a relatively new intervention that has not been widely studied. Pilot RCTs offer preliminary evidence on feasibility, aligned with the review's objectives, and provide a reasonable foundation for informing larger RCTs.

Exclusion Criteria

Exclusion criteria were studies of adults aged <55 years; non-MCT or MCT not delivered using any technology as a whole or part of the intervention; other measures not focusing on physical, cognitive, and dual-task performance; and other non-RCT studies (eg, pre-post studies).

Screening and Selection

All articles retrieved from the literature search were imported into EndNote (Clarivate), where duplicates were automatically removed. Two researchers (Y Li and Y Liu) managed and

XSL•FO
RenderX

selected the literature, following this screening process: first, titles and abstracts were independently screened by the researchers. Full texts were obtained if at least 1 reviewer believed an article met the inclusion criteria. Subsequently, 2 reviewers independently verified the eligibility of references through full-text screening. If disagreements on inclusion arose and could not be resolved through discussion, a third reviewer (JM) was consulted. Reasons for exclusion were documented. The general process of literature search and study selection was described using a flowchart following the updated PRISMA 2020 guidelines [56].

Quality Appraisal

Two researchers independently reviewed each study for methodological quality and rigor. They used the Cochrane Risk of Bias Tool (version 2.0) [65] for RCTs. In cases of disagreement, a third researcher (JM) was consulted.

Data Abstraction

Data were independently extracted by 2 researchers using a standardized form, and all processes involved in obtaining and validating data from the original authors were documented. The form included the following information: (1) characteristics of the included papers, such as author, country, year of publication, information regarding participants (sample size, age, and gender), and experimental design; (2) intervention details, including study duration, type of intervention, intervention components (physical and cognitive part), dosages, duration of each session, frequency, intensity, type of control, the content of control, technological device applied in the studies, the form of delivery (online, offline, mixed); (3) information on feasibility, including overall recruitment, retention, attainment, and dropout rates; (4) measures of participants' physical, cognitive, and dual-task performance. Disagreements during the data extraction process were resolved through discussions with a third-party researcher (JM). All recommendations were incorporated once a consensus was reached. When study data were published in multiple articles, the article with the most detailed information or the largest sample size was selected.

Data Analysis

The feasibility of the included studies in this review was assessed using completion rates and attrition rates for calculations. In addition, the most-used scales to assess the feasibility of the intervention systems were presented in this review, along with statistical analysis of scores obtained in each study. We reviewed feedback on each intervention from participants in each study and presented several narratives from participants. The frequency of adverse events associated with the interventions was also calculated, along with specific descriptions of those events.

In this study, a rapid systematic review was conducted due to the significant variation among the included studies. These variations encompassed differences in the intervention components (motor and cognitive), settings, the total duration, the duration of each session, frequency, the technologies applied, the time points for outcome measurement, and the outcome indicators. Descriptive statistics were used to describe the demographic and clinical characteristics of the groups, while narrative methods were used to categorize and synthesize the effectiveness.

Quality Evaluation of the Configuration of the Intervention Components (Motor and Cognitive)

This review analyzed the quality of the configuration for each component, including motor activity and cognitive training parts, and evaluated them according to previous standards or guidelines, to determine if they meet the training standards. The criteria for the quality assessment of each component compiled in this review simply represent the criteria for conducting quality evaluations in this review, to facilitate the following analysis of the effects of all included studies. The aspects and criteria for each inspection were as follows. First, the motor activity aspect: components were considered effective if the study stated that the intervention design was based on exercise guidelines, specific research findings (eg, previous studies or expert opinions), or if the design—including the components, session duration, frequency, and total volume—met WHO recommendations for exercise for older adults [66]. In such cases, interventions were marked with a checkmark. Second, the cognitive training aspect: interventions were considered effective if the study indicated that the intervention design was based on guidelines related to cognitive interventions, or specific research findings (eg, previous studies or expert opinions). In these cases, interventions were marked with a checkmark.

Evaluation of Effects

In this review, we carefully examined each included study for statistical significance across the following 3 outcome metrics: physical, cognitive, or dual task–related measures. If a study demonstrated statistical significance for a particular outcome (any related outcome indicators in physical, cognitive, or dual-task performance), it was considered valid in that area (eg, physical, cognitive, or dual-task performance). In these cases, if there are some outcome indicators with statistical significance, they were marked with a checkmark.

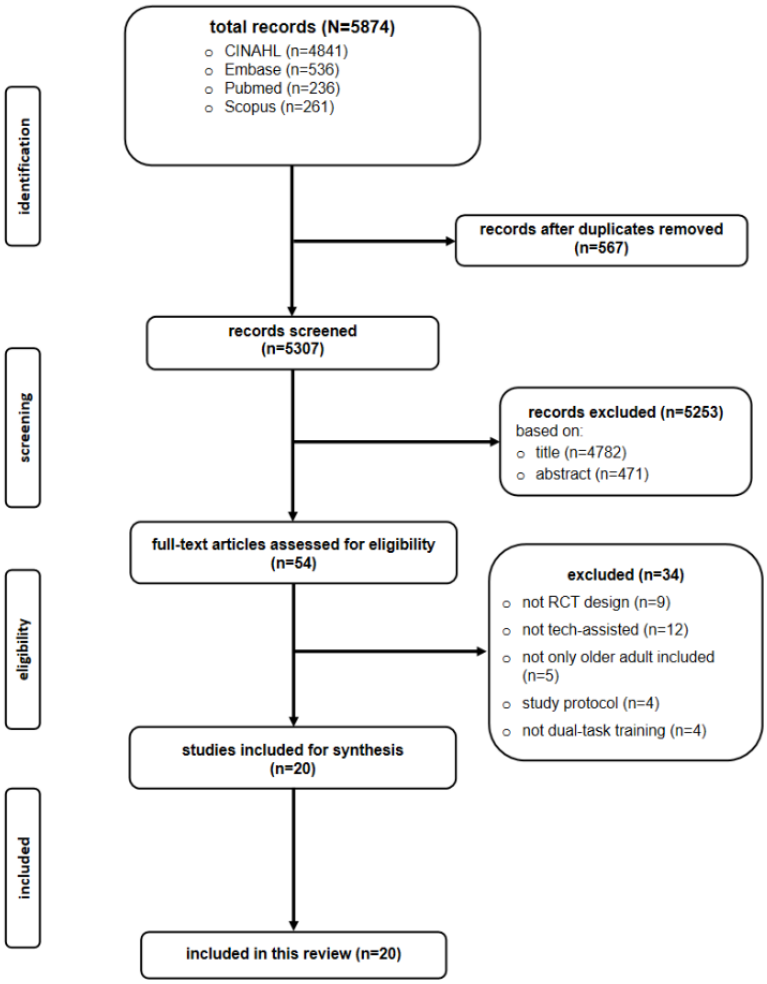
Results

Overview

The literature search yielded 5874 potentially relevant studies. After removing 567 duplicates using EndNote, 5307 were screened by 2 independent researchers. After a thorough evaluation of titles and abstracts, 5253 studies were excluded, of which 4782 were excluded based on their titles, and an additional 471 were removed after reviewing their abstracts, as they did not meet the inclusion criteria of this review. Following this preliminary screening, we conducted a full-text review of the remaining 54 articles, of which 34 were further excluded due to specific reasons outlined in the PRISMA flowchart in Figure 1. Consequently, 20 articles met our inclusion criteria and were included.

XSL•FO
RenderX

Figure 1. The PRISMA flowchart. RCT: randomized controlled trial.



Quality Appraisal

Of the 20 RCTs, 8 (40%) had high risk, 10 (50%) had some risk, and only 2 (10%) had low risk. This indicates that most of the trials had some level of risk in their design, resulting in an overall moderate quality. To be specific, there were high levels of risks in 10% (2/20) of studies (studies 8 and 20; the number of studies is consistent with the values presented the tables) and some risks in 45% (9/20) of studies (studies 1, 2, 4, 7, 9, 10, 14, 17, and 18) regarding randomization [67-86]. Regarding deviations from the intended intervention, 10% (2/20) of studies (studies 16 and 18) had a high level of risk, and 30% (6/20) of studies (studies 5, 11, 13, 14, 19, and 20) had some risk. For the measurement of outcomes, 25% (5/20) of studies (studies 5, 8, 13, 17, and 19) showed high risk. Finally, in the selection of reported results, 50% (10/20) of studies (studies 2, 3, 4, 7, 8, 14, 15, 16, 17, and 19) had some risk. The details of quality appraisals of these 20 RCT studies are presented in Figures 2 and 3 and Table 1.

XSL•FO
RenderX

Figure 2. The quality assessment of the included randomized controlled trials (RCTs; n=15; intention-to-treat) as percentages.

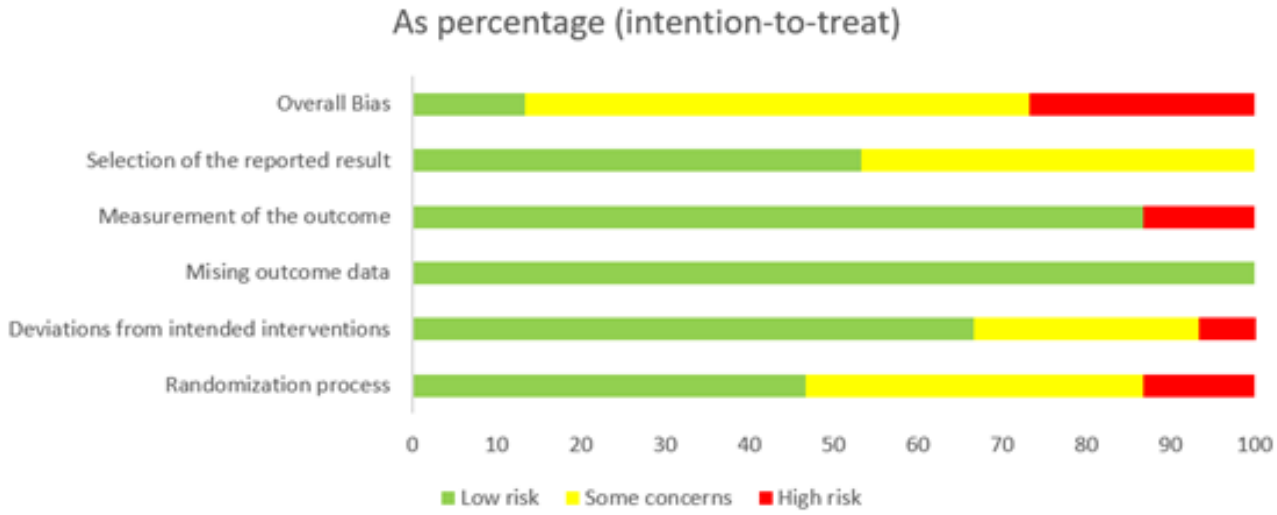
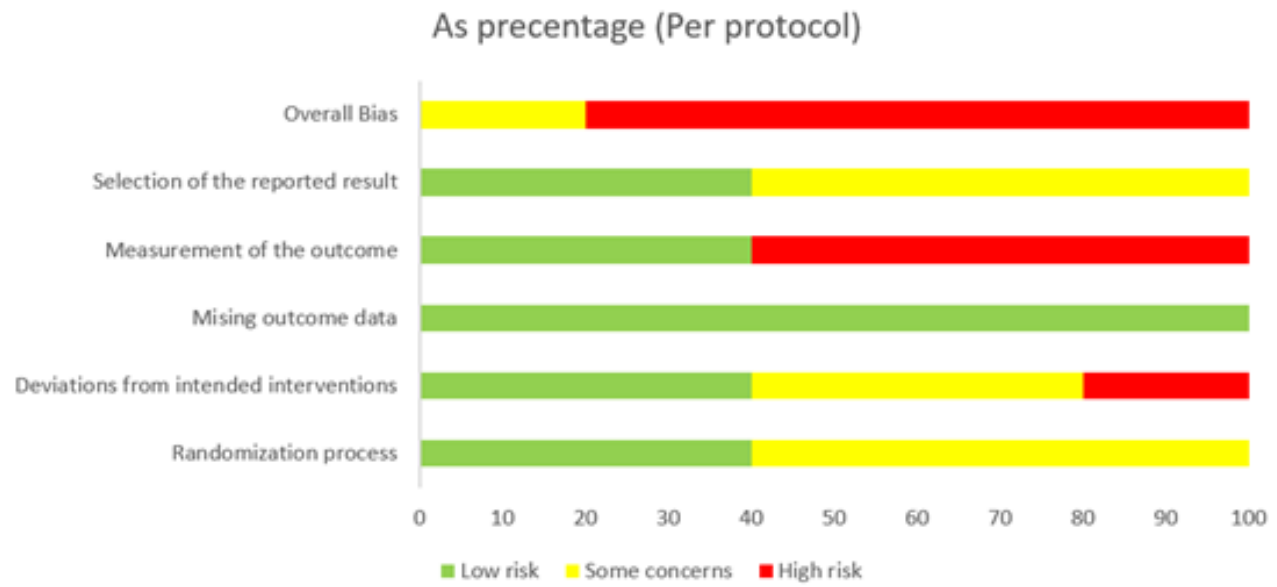


Figure 3. The quality assessment of the included randomized controlled trials (RCTs; n=5; per protocol) as percentages.



XSL•FO
RenderX

Table 1. Risk of bias for included randomized controlled trials (n=20).

Number	Study	Randomization	Deviations from intended intervention	Missing outcome data	Measurement of outcome	Selection of reported result	Overall
1	Menengi et al [67], 2022	Some concerns	Low	Low	Low	Low	Some concerns
2	Jäggi et al [68], 2023	Some concerns	Low	Low	Low	Some concerns	Some concerns
3	Kwan et al [69], 2021	Low	Low	Low	Low	Some concerns	Some concerns
4	Altorfer et al [70], 2021	Some concerns	Low	Low	Low	Some concerns	Some concerns
5	Manser et al [71], 2023	Low	Some concerns	Low	High	Low	High
6	Fishbein et al [72], 2019	Low	Low	Low	Low	Low	Low
7	Kannan et al [73], 2019	Some concerns	Low	Low	Low	Some concerns	Some concerns
8	Uematsu et al [74], 2023	High	Low	Low	High	Some concerns	High
9	Liao et al [75], 2019	Some concerns	Low	Low	Low	Low	Some concerns
10	Schoene et al [76], 2015	Some concerns	Low	Low	Low	Low	Some concerns
11	Forte et al [77], 2023	Low	Some concerns	Low	Low	Low	Some concerns
12	Pelosin et al [78], 2021	Low	Low	Low	Low	Low	Low
13	Eggenberger et al [79], 2015	Low	Some concerns	Low	High	Low	High
14	Delbroek et al [80], 2017	Some concerns	Some concerns	Low	Low	Some concerns	Some concerns
15	Hagovska and Nagyova[81], 2017	Low	Low	Low	Low	Some concerns	Some concerns
16	Park et al [82], 2020	Low	High	Low	Low	Some concerns	High
17	Villa-Sánchez et al [83], 2023	Some concerns	Low	Low	High	Some concerns	High
18	Buele et al [84], 2024	Some concerns	High	Low	Low	Low	High
19	Zak et al [85], 2024	Low	Some concerns	Low	High	Some concerns	High
20	Kwan et al [86], 2024	High	Some concerns	Low	Low	Low	High

Data Characteristics of Included Studies

A total of 20 RCTs were included in this review, of which only 4 (20%) were pilot RCTs (studies 1, 2, 4, and 5); the remaining 16 (80%) studies were RCTs. Only 5% (1/20) of study focused solely on feasibility (study 17), while 20% (4/20) of studies examined both feasibility and effects (studies 2, 3, 4, and 5). The remaining 75% (15/20) of studies exclusively investigated the effectiveness of their research subject (eg, improvement of gait performance, balance performance, and fall prevention). These papers were published between 2015 and 2024, with a notable uptick in 2023 (6/20, 30%), and were published in 13 different countries, namely, Switzerland (4/20, 20%), Italy (3/20, 15%), China (3/20, 15%), and other countries (10/20, 50%), including the United States, Korea, Ecuador, Poland, Turkey, Israel, Japan, Australia, Belgium, and Slovak Republic. These works were disseminated across 17 distinct journals, with the highest number of publications found in *Frontiers in Aging*

XSL•FO
RenderX

Neuroscience (4/20,20%). In total, 10% (2/20) of studies applied co-design (studies 5 and 20) while others reported no co-design elements, and only 5% (1/20) of study reported study fidelity (study 20). Co-design is a design methodology that emphasizes collaboration among multiple parties, particularly stakeholders such as end users, researchers, and others [87]. It encourages active participation throughout the design process by communication, creative input, sharing insights, and testing new ideas, to ensure the outcome better aligns with the needs, desires, and real-world contexts of the stakeholders (eg, development of a new exergame). Half of the included studies (studies 3, 6-9, 12-15, and 18; 10/20, 50%; all RCTs) did not perform sample size calculations. More details are provided in the summary table in Multimedia Appendix 5.

This review analyzed data from 20 studies, encompassing 1197 participants. The population size in these studies ranged from a minimum of 16 to a maximum of 293 participants, with an average of 59.9 (SD 4075.5) participants per study. Although all the included studies stated they would focus on older adults, and most participants were aged ≥60 years, only 1 (5%) study included individuals aged ≥50 years in their study (study 7).

Study Designs of Included Studies

The following are some of the basic characteristics of the included studies.

Inclusion and Exclusion Criteria of Participants in the Included Studies

Most of the included studies stipulated that participants should be older adults; however, the age of participants stipulated in these studies ranged from >50 years (studies 2, 4, and 5), to ≥70 years (studies 10, 13, and 19). The participants' physical conditions differed, encompassing healthy and nondependent community-dwelling older adults (studies 3, 8, 10, 11, 13, and 17), individuals with mild cognitive impairment (studies 9, 16, 18, and 20), those diagnosed with idiopathic PD (studies 2 and 12), stroke patients (studies 6 and 7), hospitalized older adults (studies 4 and 14), and patients with Alzheimer (study 1).

At baseline, many studies screened participants using cognitive assessment tools, such as the Brief Mental State Examination (studies 1, 2, 4, 16, 17, and 19), or the Montreal Cognitive Assessment (studies 3, 9, 14, 18, 20). In addition, regarding physical abilities, standing or walking for a specified duration was another criterion. These included standing for 3 minutes (studies 2 and 4), standing unassisted or assisted for 5 minutes (studies 7 and 12), and standing for 10 minutes (study 5). Additional criteria encompassed walking with or without a walker (study 10), walking unassisted or assisted for 10 meters (studies 6, 9, and 14), and walking with or without assistance for 20 meters (study 13).

Independence in performing daily activities was also a common requirement (studies 5, 8, 9, 10, 11, and 16). Furthermore, participants were expected to possess the ability to provide informed consent (studies 2, 4, 13, 16, and 18), and adequate communication skills (studies 1, 7, and 16). Finally, some studies also had criteria related to a stable medication regimen or a history of specific treatments (studies 1, 2, and 12) and if the participant had a history of falls (studies 8 and 12).

Exclusion Criteria Applied in the Studies

The most-mentioned exclusion criteria among included studies were cognitive impairment, such as dementia (studies 3, 7, 9, 10, 12-15, 18, and 20), or significant cognitive deficits in potential participants. In addition, the inability to comprehend instructions or effectively communicate (often due to language barriers or sensory impairments), often led to exclusion in some cases (studies 1-7, 12, 14-16, and 18-20). Moreover, individuals with unstable or acute medical conditions, including major mental illnesses, were also typically ineligible (studies 1, 2, 4, 5, 9, 10, 12, 15, 16, 18, and 19). Finally, participants who had recently undergone major surgery or had acute orthopedic conditions were generally excluded (study 14). The history of falls also served as an exclusion criterion in certain studies (study 6).

Technologies Applied in the Studies and Configuration of the Intervention Components (Motor and Cognitive) of MCT

In this review, we reviewed and confirmed that all the included studies used simultaneous MCT, except the study by Buele et al [84]. Notably, we found that MCT showed significant variation in the configuration of components (motor activity and cognitive training), as well as in the duration and frequency of each session of each component. Among the included types of MCT, some were more focused on motor activity interventions involving simple cognitive tasks such as performing addition and subtraction while engaging in specific-intensity, multicomponent, structured motor exercise (eg, aerobic, strength, and balance training). Others were more cognitively focused, structured within a broad cognitive exercise framework, where participants performed simple gross motor activities such as moving their legs back and forth. Also, there were no golden-standard guidelines on how to structure MCT interventions to simultaneously improve physical, cognitive, and dual-task performance. More details are provided in Multimedia Appendix 5.

In Table 2, we briefly describe the components of the motor activities and cognitive training interventions for each MCT, noting whether technology was used in these 2 components, and specifying the technologies used in either 1 part or the overall MCT. We also report whether either intervention component (physical and cognitive) was developed based on ADLs (eg, dressing, eating, or mobility), or instrumental ADLs (IADLs; eg, shopping, housekeeping, or laundry). In Table 2, for the motor activity part, studies 2, 4, and 5 were developed based on ADLs, while study 16 was based on IADLs. Studies 3, 6, 12, 15, 18, and 20 did not use technology for motor activities. Among these, studies 3, 6, and 12 used treadmills, whereas study 15 focused on stamina training. Regarding cognitive training, studies 2, 4, 5, and 15 were based on ADLs, whereas studies 3, 9, 16, 18, and 20 were based on IADLs. Only study 7 did not incorporate technology. The following technologies frequently mentioned among the included studies, including the Dividat Senso (3/20, 15%), the Wii Fit Games (2/20, 10%), exergames besides Wii Fit Games (5/20, 25%), VR noted in various forms (8/20, 40%), and 2 other technologies in 2 (10%) studies. In this study, there were 7 active controls

XSL•FO
RenderX

(studies 3, 9, 11, 15, 16, 18, and 19), 5 (25%) studies used standard treatment control as a control group (studies 1, 2, 4, 5, and 20), 4 placebo controls (studies 8, 10, 14, and 17), 3 partial intervention control (studies 6, 7, and 13), and 1 dose-response control (study 12). In addition, regarding the implementation of MCT, 12 (60%) out of 20 studies reported their intervention settings. Specifically, 4 (20%) out of 20 studies (studies 3, 18, 19, and 20) in community settings, 3 (15%) studies (studies 1, 5, and 10) implemented interventions at home, 3 (15%) studies (studies 2, 4, and 7) in clinical rehabilitation settings, 1 (5%) study (study 15) in outpatient settings, and 1 (5%) study (study 14) in a residential care center. More details are provided in Table 2.

XSL•FO
RenderX

Table 2. Intervention components and how technologies assisted the interventions.

Number	Study	Physical	Cognitive	How to cooperate	Control group	Settings
1	Menengi et al [67], 2022	Chair-based exercises	Cognitive training facilitated through a computer or tablet	Online supervision (Zoom)	Standard treatment control	Home-based
2	Jäggi et al [68], 2023	Regular rehabilitation and exercises included in the Dividat Senso[a]	Exercises included in the Dividat Senso[a]	The Dividat Senso	Standard treatment control	In an inpatient rehabilitation setting
3	Kwan et al [69], 2021	Cycling on an ergometer[b]	Cognitive training on VR[c]	VR platform	Active control	In community
4	Altorfer et al [70], 2021	Regular rehabilitation therapy and exercises included in the Dividat Senso[a]	Exercises included in the Dividat Senso[a]	The Dividat Senso	Standard treatment control	In a geriatric inpatient rehabilitation setting
5	Manser et al [71], 2023	Exercise based on the brain-IT system (heart rate variability-induced resonance respiration) in combination with the exercises included in the Dividat Senso[a]	Combination of cognitive training[a] based on the brain-IT system and the Dividat Senso	The brain-IT system and the Dividat Senso	Standard treatment control	At home
6	Fishbein et al [72], 2019	Treadmill[b]	The SeeMee system	The SeeMee system	Partial intervention control	—[d]
7	Kannan et al [73], 2019	Body movement, balance games	Addition, subtraction, and multiplication[b]	Wii Fit, balance board	Partial intervention control	In clinical stroke rehabilitation settings
8	Uematsu et al [74], 2023	Balance training	Cognitive training included in Wii Fit Games	Wii Fit	Placebo control	—
9	Liao et al [75], 2019	The physical regimen comprised resistance, aerobic, and balance exercises aligned with the American College of Sports Medicine standards for older adults	Activities like reciting poems while walking, naming flowers and animals while navigating obstacles, and solving math problems during resistance exercises[c]	VR	Active control	—
10	Schoene et al [76], 2015	4 specific stepping games, balance training	Stepping training targeted cognitive functions of fall prevention in older adults	Electronic pedals, computer interface, and television screen	Placebo control	At home
11	Forte et al [77], 2023	Gross motor coordination (upper and lower body movements designed to enhance both static and dynamic balance)	Stimulus-response cognitive tasks generated by an device (Witty-SEM)	Witty-SEM	Active control	—
12	Pelosin et al [78], 2021	Using a Kinect camera to capture participants' foot movements on a treadmill, the system integrated these movements into a computer-generated virtual environment displayed on a screen, and participants were required to avoid virtual obstacles[b]	Various cognitive domains, such as executive functions, attention, working memory, and visual processing	A TT+VR system (V-TIME), a Kinect camera	Dose-response control	—

XSL•FO
RenderX

Number	Study	Physical	Cognitive	How to cooperate	Control group	Settings
13	Eggenberger et al [79], 2015	The training was designed following current physical fitness and fall prevention recommendations for older adults, focusing on aerobic endurance exercises	Cognitive training	A VR video game, DANCE	Partial intervention control	—
14	Delbroek et al [80], 2017	Physical training in 9 exercises designed to improve balance, weight-bearing, memory, attention, and dual-tasking abilities	Cognitive training in 9 exercises designed to improve balance, weight-bearing, memory, attention, and dual-tasking abilities	The BioRescue	Placebo control	In a residential care center
15	Hagovska and Nagyova [81], 2017	Physical training[b]	Cognifit[a]	Cognifit	Active control	In an outpatient psychiatric clinic of the Highly Specialized Geriatric Institute
16	Park et al [82], 2020	A device with software using VR for activities such as driving, bathing, cooking, and shopping[c]	This setup facilitated training in attention, memory, problem-solving, and executive functions[c]	The MOTOcog system	Active control	—
17	Villa-Sánchez et al [83], 2023	Walking on carpet	Counting backward	Sensing carpet	Placebo control	—
18	Buele et al [84], 2024	Balance exercises (eg, walking in a straight line, squats, brisk walking, going up and down stairs, and social dancing individually and in pairs)[b]	Collective social interaction activities, and an immersive VR-based system that simulates a task of searching for ingredients in a kitchen cupboard[c]	VR	Active control	Community
19	Zak et al [85], 2024	Walking, walking with putting a ball between the hands, and walking with tossing a ball	Cognitive exercises, talking, adding up numbers, subtracting numbers, repeating phrases, reciting a word chain, and identifying objects	VR	Active control	Community
20	Kwan et al [86], 2024	Biking[b]	The VR game contains 8 different themes: orientation, finding a bus stop, reporting lost items, finding a supermarket, grocery shopping, cooking, finding a travel hot spot, and bird watching[c]	VR	Standard treatment control	Community

[a]Interventions developed based on activities of daily living.

[b]Nontechnologically supported components.

[c]Interventions developed based on instrumental activities of daily living.

[d]Not available.

Duration of Total Experiments and Each Session and Intervention Frequency

The intervention period varied from 2 weeks to 6 months, with many studies lasting 6 weeks (6/20, 30%), 2 to 4 weeks (4/20, 20%), 8 weeks (3/20, 15%), and 12 weeks (3/20, 15%). Each intervention session lasted between 10 and 30 minutes (12/20, 60%), followed by 30 minutes (4/20, 20%), 40 to 60 minutes (3/20, 15%), the other studies did not specify the exact times for the interventions. Regarding frequency, most interventions occurred 2 to 3 days per week (8/20, 40%), followed by 4 to 5 days per week (4/20, 20%). More details can be found in Multimedia Appendix 5.

XSL•FO
RenderX

Feasibility Analysis

Feasibility, Acceptability, and Adherence

Regarding feasibility, acceptability, and adherence, most included studies have shown high completion and low attrition rates (studies 1, 2, 4, 6-8, 10, and 13-17), indicating high participant engagement and intervention suitability.

The NASA Task Load Index and the System Usability Scale

The NASA Task Load Index (TLX) is a widely used subjective multidimensional assessment tool that rates perceived workload to evaluate the effectiveness or other aspects of the performance of a task, system, or team. The scale has a total score of 100, with workload score values of 0 to 9 as low, 10 to 29 as moderate, 30 to 49 as slightly high, 50 to 79 as high, and 80 to 100 as very high [88]. Two articles in this review applied the NASA-TLX (studies 2 and 4), with the highest NASA-TLX mean score of 56.2 (study 2) and the lowest of 45.5 (study 4), indicating high workloads.

The System Usability Scale (SUS) is a commonly used tool to analyze the perceived usability of a system, product, or service, scored on a scale of 0 to 100, where the higher the score, the better the usability [89]. Total 3 (15%) of the 20 included studies measured SUS (studies 2, 4, and 5), with mean SUS scores ranging from 71.7 to 83.6, providing better information on overall participant satisfaction and user-perceived usability of the system.

Positive Feedback

Positive feedback from participants and medical staff (studies 1 and 14) included the following: “the caregivers stated that they agreed 100% with the expressions-my patient was satisfied with the online exercise treatment” and “I was satisfied with the online exercise treatment” from study 1 and “Interviews with participants from the intervention group showed that they found the program useful for their concentration, memory, and balance, according to the results of the IMI, which resulted in high compliance. They scored the program as very interesting and pleasant to do and perceived their performance of the different exercises as good to very good” from study 14. This demonstrated acceptance and satisfaction with the interventions. These similarities suggest that the interventions were generally well-received, feasible to implement, and safe for participants from different study backgrounds.

Adverse Events

The results from the included studies consistently highlight the absence of significant adverse events throughout the study period across studies as follows: most of the studies reported no dropouts or adverse events linked to the exercise treatment (studies 1, 2, 4, and 6-17). Among these studies, 20% (4/20) of minor adverse events were noted; however, all evidence suggests that there is no direct link between these adverse events and the intervention. They were “minor technical issues, such as video sound and connection problems, were reported but did not impede the sessions, and caregivers expressed high levels of satisfaction” (study 1); “most participants did not experience symptoms of VR sickness, while two individuals in the control group withdrew due to moderate joint and muscle pain” (study 3); and “specifically falls resulting in bruises, were noted in the intervention group, importantly, these events were unrelated to the ‘Brain-IT’ training” (study 5). Overall, the prevalence of adverse events in these included studies underscores the safety and minimal adverse effects associated with the interventions examined in the review.

Quality Evaluation of the Configuration of the Intervention Components (Motor and Cognitive)

On the basis of the quality evaluation criteria mentioned in the Quality Evaluations of the Configuration of the Intervention Components (Motor and Cognitive) subsection in the Methods section, 6 of 20 (30%) studies met the standards for motor activity design (studies 5, 6, 9, 11, 13, and 19). In total, 12 of 20 (60%) studies were qualified for cognitive training design (studies 2-5, 10-12, 14-16, 19, and 20). However, only 3 of 20 (15%) studies qualified in both aspects (studies 5, 11, and 19). Details are provided in Table 3.

XSL•FO
RenderX

Table 3. The quality assessment of intervention components.

Number	Study	Physical	Cognitive
1	Menengi et al [67], 2022		
2	Jäggi et al [68], 2023		✓
3	Kwan et al [69], 2021		✓
4	Altorfer et al [70], 2021		✓
5	Manser et al [71], 2023	✓	✓
6	Fishbein et al [72], 2019	✓	
7	Kannan et al [73], 2019		
8	Uematsu et al [74], 2023		
9	Liao et al [75], 2019	✓	
10	Schoene et al [76], 2015		✓
11	Forte et al [77], 2023	✓	✓
12	Pelosin et al [78], 2021		✓
13	Eggenberger et al [79], 2015	✓	
14	Delbroek et al [80], 2017		✓
15	Hagovska and Nagyova [81], 2017		✓
16	Park et al [82], 2020		✓
17	Villa-Sánchez et al [83], 2023		
18	Buele et al [84], 2024		
19	Zak et al [85], 2024	✓	✓
20	Kwan et al [86], 2024		✓

Effects of Interventions and Study Characteristics

Except for 10% (2/20) studies (studies 5 and 17), nearly all studies reported the effects of the intervention, including in physical, cognitive, and dual-task performance, in at least 1 or 2 aspects. The physical and cognitive quality evaluations in Table 4 are derived from the assessments presented in Table 3. More details are provided in Table 4, which is a summary of whether there were any statistically significant outcome indicators in each 3 domains (“yes,” “no,” or “was not tested in this domain”), and Multimedia Appendix 6, which provides the details of statistically significant outcome indicators in the 3 domains.

XSL•FO
RenderX

Table 4. The assessment of outcome indicators.

Number	Study	Physical (quality)[a]	Physical (results)[b]	Cognitive (quality)[a]	Cognitive (results)[b]	Dual-task (results)[c]
1	Menengi et al [67], 2022	X[d]	✓[d]	X[e]	✓[e]	—
2	Jäggi et al [68], 2023[f]	X[d]	✓[d]	✓	✓	X
3	Kwan et al [69], 2021[g]	X[d]	✓[d]	✓	✓	—
4	Altorfer et al [70], 2021[f]	X[h]	X[h]	✓	✓	✓
5	Manser et al [71], 2023[f]	✓[i]	X[i]	✓[j]	X[j]	—
6	Fishbein et al [72], 2019	✓[k]	✓[k]	X[l]	—	—
7	Kannan et al [73], 2019	X[d]	✓[d]	X[e]	✓[e]	X
8	Uematsu et al [74], 2023	X[d]	✓[d]	X[l]	—	—
9	Liao et al [75], 2019[g]	✓[k]	✓[k]	X[e]	✓[e]	✓
10	Schoene et al [76], 2015	X[h]	X[h]	✓	✓	—
11	Forte et al [77], 2023	✓[k]	✓[k]	✓	✓	—
12	Pelosin et al [78], 2021	X[h]	X[h]	✓	✓	—
13	Eggenberger et al [79], 2015	✓[k]	✓[k]	X[l]	—	✓
14	Delbroek et al [80], 2017	X[h]	X[h]	✓	✓	X
15	Hagovska and Nagyova [81], 2017[f]	X[d]	✓[d]	✓	✓	—
16	Park et al [82], 2020[g]	X[h]	—[h]	✓	✓	—
17	Villa-Sánchez et al [83], 2023	X[h]	X[h]	X[l]	X[l]	X
18	Buele et al [84], 2024	X[h]	—[h]	X[e]	✓[e]	—
19	Zak et al [85], 2024	✓[k]	✓[k]	✓[j]	X[j]	—
20	Kwan et al [86], 2024	X[d]	✓[d]	✓	✓	—

[a]In the physical (quality) and cognitive (quality) columns, "✓" means met the criteria, "X" means did not meet the criteria, and "—" means not measured.

[b]In the physical (results) and cognitive (results) columns, "✓" means met statistical significance, "X" means did not meet statistical significances, and "—" mean not measured.

[c]In the dual-task (results) column, "✓" means effective in dual-task performance, "X" means ineffective in dual-task performance, and "—" mean not measured.

[d]Consistency of design and results in the motor activity designation did not meet the previous training standards, guidelines for motor activity design, but had motor effects.

[e]Consistency of design and results in the cognitive training designation did not meet the previous training standards, guidelines of cognitive design, but had cognitive effects.

[f]Interventions developed based on activities of daily living.

[g]Interventions developed based on instrumental activities of daily living.

[h]Consistency of design and results in the motor activity designation did not meet the previous training standards, guidelines of motor design, and was without motor effects.

[i]Consistency of design and results in the motor activity designation met the previous training standards, guidelines of motor activity design, but without motor effects.

[j]Consistency of design and results in the cognitive training designation met the previous training standards, guidelines of cognitive training design, but without cognitive effects.

[k]Consistency of design and results in the motor activity designation met the previous training standards, guidelines of motor activity design, and received statistically significant outcomes.

[l]Consistency of design and results in the cognitive training designation did not meet the previous training standards, guidelines of cognitive design, and was without cognitive effects.

XSL•FO
RenderX

Effects on Physical Performance

This part reported physical performance across multiple studies that showed effectiveness (studies 1-3, 6-9, 11, 13, 15, 19, and 20). Among the physical function-related outcome indicators demonstrating statistical significance, the most commonly applied indicators, including the 5 times sit to stand test (studies 1, 2, and 11), walking speed or time (studies 6, 11, and 13), the timed up and go test (studies 1 and 11), ADL (studies 1 and 15), the Short Physical Performance Battery (studies 2 and 13), gait performance (studies 9 and 13), and fall-related metrics (studies 10 and 13).

Effects on Cognitive Performance

Results in cognitive function measurements were also reported as effective in some studies (studies 1-4, 7, 9-12, 14-16, 18, and 20). Among the cognitive outcome indicators demonstrating statistical significance, the most common ones included the Montreal Cognitive Assessment (studies 3, 11, 16, 18, and 20), the trail making test parts A and B (studies 9, 15, and 16), the go or no-go test (studies 2 and 4), and visuospatial ability (studies 10 and 12).

Effectiveness on Dual-Task Performance

Dual-task performance was assessed in only 35% (7/20) of the studies (studies 2, 4, 7, 9, 13, 14, and 17), while only 15% (3/20) of them (studies 4, 9, and 13) showed the effectiveness of the intervention in dual-task performance. Of these 7 studies, there were 2 studies (studies 2 and 4) that were pilot RCTs, while the others were full RCTs.

Discussion

General Characteristics of the Included Studies

This is the first systematic review specifically assessing the feasibility and effectiveness of technology-assisted MCT in older adults with various physical conditions. Upon review, there were only 20 RCT studies of technology-assisted MCT interventions. Although interest has been growing yearly, and an uptick in 2023 was captured (6/20, 30%), the number of studies remains relatively sparse. Most studies included in this review used exergame technology to deliver MCT (including systems such as Dividat Senso, Wii Fit Games, and other exergames; 10/20, 50%), and this was closely followed by the use of VR technology (8/20, 40%). Only 5% (1/20) of studies used remote intervention design, and nearly half (10/20, 50%) focused on fall prevention, walking performance, balance, or gait performance. Moreover, co-design was not widely applied (2/20, 10%), despite its importance for boosting user engagement, system usability, user experience [90], and potentially increasing the effectiveness of the training [91,92]. Finally, the sample sizes were generally small, for both the total number of participants and the average number per study. In the meantime, proper sample size estimation methods were only applied in half (10/20, 50%) of the included studies. The average number of participants was <60 [93] and was not only typically seen in all pilot RCTs (studies 1, 2, 4, and 5), but most of the RCTs (studies 3, 6-9, 11, 14, and 16-18).

There are some positive aspects to highlight. This review includes a relatively representative sample involving both healthy and older adults with chronic conditions. Participants were selected based on their physical and cognitive functioning to ensure a proper intervention basis for MCT (eg, sufficient physical, cognitive functioning, communication skills, while excluding participants with severe cognitive or mobility impairments, sensory deficits, acute illnesses, unstable medical conditions, recent major surgeries, or a significant history of falls). The most common total experiment duration was 6 weeks, each session typically lasted 10 to 30 minutes, and they took place 2 to 3 times per week. In experiments, shorter durations with higher frequency effectively avoid participant fatigue, which may also have the potential to impact performance and response [94], reduce attrition, maintain compliance [95,96], and lower the overall cost of the experiment. In addition, the primary outcome of this review demonstrated excellent feasibility, acceptability, and adherence. Although the perceived workload was average to high, the perceived usability of the systems was also high. Furthermore, most studies received positive feedback with few negative events. Overall, the data suggest that the feasibility of this technology-assisted MCT intervention is high, with minimal prevalence of adverse events.

Relationship Between the Varied Effects and the Different Configurations (Motor and Cognitive Components) of the Technology-Assisted MCT

As mentioned in the introduction, MCT has proven effective in improving physical, cognitive, and dual-task performance among older adults [97]. However, the effectiveness of MCT varied across studies. For example, it varied in exercise effectiveness among older adults [98], gait and balance effectiveness varied among older adults with cognitive impairments [23], cognitive function effectiveness varied among individuals with various clinical conditions [99], and balance-oriented motor activity, cognitive training, and dual-task performance enhancement through performance-related interventions varied among normal older adults [100]. To achieve these benefits, training interventions should have increased difficulty, appropriate intensity, sufficient duration, task specificity, and variable task prioritization [100]. In addition, after carefully examining and cross-checking the effects from each included study of their MCT delivered via various technologies (eg, exergames, VR, and other technologies), we found no significant results. On this basis, we can see that the differences in effects may be blamed on the various configurations of MCT, including motor and cognitive components.

In this review, all the studies included were interpretable except for study 5, which showed inconsistent results. This may be attributed to the small sample size in this study (16/20, 80%) among all the included studies, the older aged participants (mean age 79.9 years in the intervention group and 73.7 years in the control group), and the occurrence of adverse events (3 slight adverse events happened in the intervention group, such as falls in homes with bruises, but no more serious injuries) weakened the effectiveness of the interventions. For control groups, there were 7 active controls, 25% (5/20) of studies used standard treatment as a control, 4 used placebo controls, 3 used partial

XSL•FO
RenderX

intervention controls, and 1 used a dose-response control. Except for studies that included standard treatment as a control or a placebo control, the remaining studies with other kinds of control groups can be compared to noninferior experiment designs. As a result, there may be an ability to account for the greater effectiveness in the experimental group than the control group.

This review found that physical designs that follow the recommended requirements (eg, for volume, duration, frequency) for motor activity training based on previous research and WHO guidelines [14,66] form a basis to receive exercise benefits. Cognitive gain, on the other hand, depends not only on physical design, but also the cognitive design and whether the participant has an associated cognitive impairment. Moreover, it is possible that physical designs meeting the above criteria are a basis for obtaining beneficial performance in dual tasks. Regarding the design of motor activity interventions, 5 of 20 (25%) studies, (partial intervention control: studies 6 and 13; active control: studies 9, 11, and 19) met the criteria for motor activity, and all these 5 physical measures showed statistical significance. For cognitive design, studies 6 and 13 did not meet the criteria and did not assess cognitive performance. The cognitive intervention in study 11 met the criteria and showed statistically significant cognitive outcomes. Study 9 did not meet the criteria for cognitive training but demonstrated statistical significance in cognitive outcome measures. This may be attributed to the achieved level of physical activity in study 9, which was based on neuroplasticity theory [101], positively influencing cognitive function. Conversely, study 19 met the criteria for cognitive training, but did not show significant cognitive improvements. Upon reviewing study 19, which focused on healthy older adults whose score on the Brief Mental State Examination was >23, its absence of cognitive impairment potentially explains the lack of statistical significance. Finally, studies 6, 11, and 19 did not measure dual-task effects, while studies 9 and 13 demonstrated statistical significance.

All adults should undertake regular physical activity (a strong recommendation with moderate-certainty evidence) according to the WHO guidelines [66]. In this review, 7 studies (standard treatment control: studies 1, 2, and 20; active control: studies 3 and 15; partial intervention control: study 7; placebo control: study 8) did not meet the criteria for motor activity yet showed physical statistical significance. This discrepancy, along with the WHO recommendations for older adults' motor activity, may suggest that even though their motor activity did not meet recommended standards, participants still gained exercise benefits. In addition, this finding also reveals that this rule is adaptive to technology-assisted MCT. Among these 8 studies, only 4 (studies 2, 3, 15, and 20) met the criteria for cognitive intervention. However, all studies except for study 8, which did not measure cognitive outcomes, showed significant results. These results also can be explained by the neuroplasticity theory that motor activity can enhance participants' cognitive performance [101]. In addition, the WHO recommendations for cognitive intervention for older adults with or without cognitive decline [14] support the use of cognitive interventions to prevent cognitive deterioration. Overall, the cognitive intervention designs in these studies yielded statistically significant results, even though only half of these studies (4/20, 20%) met the criteria for cognitive intervention design. Of the 7 studies, only studies 2 and 7 assessed dual-task performance, but neither showed statistical significance.

Although motor activity prompts positive and statistically significant effects as advocated by the WHO, the exact tipping point for achieving physical gains remains unclear. In this review, 7 studies (standard treatment control: study 4; placebo control: studies 10, 14, and 17; active control: studies 16 and 18; dose-response control: study 12) did not meet the criteria for motor activity and did not show any statistically significant physical effects (study 16 did not measure physical effects). More research will be needed to determine the type of motor activity, frequency, session duration, and total volume to start to achieve physical benefits. For cognitive training, 5 studies (studies 4, 10, 12, 14, and 16) met the criteria and showed effects, while study 18, despite it not meeting the criteria, showed statistically significant cognitive gains. Compared to the 5 studies (studies 6, 9, 11, 13, and 19) and 7 studies (studies 1, 2, 3, 7,8, 15, and 20) mentioned earlier, this demonstrates that, in technology-assisted MCT, even if the motor activity does not meet the criteria and does not result in effects, the cognitive component, if met, can still achieve a significant cognitive effect (studies 4, 10, 12, 14, and 16). In addition, these 5 studies (studies 4, 10, 12, 14, and 16) also support the WHO recommendation that "cognitive interventions should be provided to older adults, regardless of their cognitive status, to decrease their deterioration in cognitive functions" [14].

Relationship Between the Varied Dual-Task Effects and Whether the Motor and Cognitive Components in MCT Were Constructed Based on the ADL and IADL Frameworks

Finally, this review included 4 studies with at least one part of the intervention (motor or cognitive) designed based on an ADL framework (studies 2, 4, 5, and 15), and 5 studies with at least one part of the intervention (motor or cognitive) designed based on an IADL framework (studies 3, 9, 16, 18, and 20). Among these, studies 3, 5, 15, 16, 18, and 20 did not measure dual-task performance. Study 2 measured dual-task performance but found no statistical significance, while studies 4 and 9 measured dual-task performance and reported statistical significance. Examining the motor and cognitive design of studies 4 and 9, we found that although the motor activity in study 4 was substandard and the cognitive design in study 9 was substandard, the intervention design in study 4 followed the ADL framework for both motor and cognitive parts, while study 9 followed the IADL design for the cognitive part. On the basis of these findings and the small sample sizes in this review, we may tentatively hypothesize that it may be possible to improve participants' dual-task performance or even achieve statistical significance if the intervention constructs the motor and cognitive components of MCT based on the ADL and IADL frameworks.

Conclusions

This rapid systematic review provides an up-to-date synthesis of evidence for technology-assisted MCT in older adults,

XSL•FO
RenderX

addressing a significant knowledge gap in the field. The review found that there is little research on technology-assisted MCT; however, it has been relatively successful in participant inclusion and exclusion for this type of intervention. The review also outlines the most used experimental setups in the included studies. Preliminary results indicate that technology-assisted MCT is highly feasible, with almost no reported adverse events, suggesting it merits further research and replication. In addition, this review analyzed and evaluated the composition and quality of motor activity, cognitive training components, the techniques used, and the effectiveness of these interventions. Most of the included studies showed statistically significant effectiveness. We thoroughly examined the contextual settings of each study (eg, the techniques used and the populations studied) to assess their research outcomes. However, we did not observe any significant differences in the effects of the various techniques applied in these included studies. This review suggests that motor activity and cognitive training that meet the criteria yield statistical significance. However, the exact tipping point for achieving physical or cognitive gains remains unclear. Moreover, given the relatively limited sample sizes in this review, it might be tentatively suggested that to potentially enhance participants' dual-task performance or even achieve statistical significance, researchers can try to construct the intervention of the motor and cognitive components of MCT based on the ADL and IADL frameworks.

Recommendation

More research will be needed to determine types of motor and cognitive components, frequency, session duration, total volume, and type of technology to start to achieve physical and cognitive benefits from technology-assisted MCT.

Acknowledgments

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Authors' Contributions

The authors contributed to the research and manuscript in the following ways: Y Li, AYML, and JM conceptualized the study. Y Li and JM developed the methodology. Y Li and Y Liu curated the data, were responsible for the software, and analyzed the data. Y Li, AYML, and JM interpreted the results. Y Li drafted the original manuscript. Y Li and JM contributed to the review and editing. AYML and JM supervised the project.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA 2020 checklist.
[DOCX File , 34 KB - games_v13i1e67250_app1.docx]

Multimedia Appendix 2
Synthesis Without Meta-analysis (SWiM) in systematic reviews checklist.
[DOCX File , 21 KB - games_v13i1e67250_app2.docx]

Multimedia Appendix 3
Open Science Framework Registries technology-assisted motor-cognitive training among older adults; a rapid systematic review of randomized controlled trials.
[PDF File (Adobe PDF File), 413 KB - games_v13i1e67250_app3.pdf]

Multimedia Appendix 4
Search strategy (updated version in March 2025).
[DOCX File , 26 KB - games_v13i1e67250_app4.docx]

Multimedia Appendix 5
Summary table.
[DOCX File , 92 KB - games_v13i1e67250_app5.docx]

Multimedia Appendix 6

XSL•FO
RenderX

Details on statistically significant outcome indicators in each of 3 domains.
[DOCX File , 21 KB - games_v13i1e67250_app6.docx]

References

1. Global issues: population. United Nations. URL: https://www.un.org/en/global-issues/population [accessed 2025-05-14]
2. Hertzog C, Kramer AF, Wilson RS, Lindenberger U. Enrichment effects on adult cognitive development: can the functional capacity of older adults be preserved and enhanced? Psychol Sci Public Interest 2008 Oct;9(1):1-65. [doi: 10.1111/j.1539-6053.2009.01034.x] [Medline: 26162004]
3. Zanker J, Duque G. Osteoporosis in older persons: old and new players. J Am Geriatr Soc 2019 Apr;67(4):831-840. [doi: 10.1111/jgs.15716] [Medline: 30570741]
4. Barnsley J, Buckland G, Chan PE, Ong A, Ramos AS, Baxter M, et al. Pathophysiology and treatment of osteoporosis: challenges for clinical practice in older people. Aging Clin Exp Res 2021 Apr;33(4):759-773 [FREE Full text] [doi: 10.1007/s40520-021-01817-y] [Medline: 33742387]
5. Faulkner JA, Larkin LM, Claflin DR, Brooks SV. Age-related changes in the structure and function of skeletal muscles. Clin Exp Pharmacol Physiol 2007 Nov;34(11):1091-1096 [FREE Full text] [doi: 10.1111/j.1440-1681.2007.04752.x] [Medline: 17880359]
6. Tieland M, Trouwborst I, Clark BC. Skeletal muscle performance and ageing. J Cachexia Sarcopenia Muscle 2018 Mar;9(1):3-19 [FREE Full text] [doi: 10.1002/jcsm.12238] [Medline: 29151281]
7. Mitchell WK, Williams J, Atherton P, Larvin M, Lund J, Narici M. Sarcopenia, dynapenia, and the impact of advancing age on human skeletal muscle size and strength; a quantitative review. Front Physiol 2012;3:260 [FREE Full text] [doi: 10.3389/fphys.2012.00260] [Medline: 22934016]
8. Hughes ML, Agrigoroaei S, Jeon M, Bruzzese M, Lachman ME. Change in cognitive performance from midlife into old age: findings from the Midlife in the United States (MIDUS) study. J Int Neuropsychol Soc 2018 Sep;24(8):805-820 [FREE Full text] [doi: 10.1017/S1355617718000425] [Medline: 30019663]
9. Piefke M, Onur Ö, Fink GR. Aging-related changes of neural mechanisms underlying visual-spatial working memory. Neurobiol Aging 2012 Jul;33(7):1284-1297. [doi: 10.1016/j.neurobiolaging.2010.10.014] [Medline: 21130531]
10. Novotný JS, Gonzalez-Rivas JP, Medina-Inojosa JR, Lopez-Jimenez F, Geda YE, Stokin GB. Investigating cognition in midlife. Alzheimers Dement (N Y) 2021;7(1):e12234 [FREE Full text] [doi: 10.1002/trc2.12234] [Medline: 35005209]
11. Inzitari M, Newman AB, Yaffe K, Boudreau R, de Rekeneire N, Shorr R, et al. Gait speed predicts decline in attention and psychomotor speed in older adults: the health aging and body composition study. Neuroepidemiology 2007;29(3-4):156-162 [FREE Full text] [doi: 10.1159/000111577] [Medline: 18042999]
12. Murman DL. The impact of age on cognition. Semin Hear 2015 Aug;36(3):111-121 [FREE Full text] [doi: 10.1055/s-0035-1555115] [Medline: 27516712]
13. Glisky EL, Riddle DR. Changes in cognitive function in human aging. In: Brain Aging: Models, Methods, and Mechanisms. Boca Raton, FL: CRC Press; 2007.
14. Risk reduction of cognitive decline and dementia: WHO guidelines. World Health Organization. 2019. URL: https://www.who.int/publications/i/item/risk-reduction-of-cognitive-decline-and-dementia [accessed 2025-05-14]
15. Tait JL, Duckham RL, Milte CM, Main LC, Daly RM. Influence of sequential vs. simultaneous dual-task exercise training on cognitive function in older adults. Front Aging Neurosci 2017;9:368 [FREE Full text] [doi: 10.3389/fnagi.2017.00368] [Medline: 29163146]
16. Liebherr M, Schubert P, Schiebener J, Kersten S, Haas CT. Dual-tasking and aging—about multiple perspectives and possible implementations in interventions for the elderly. Cogent Psychol 2016;3(1):1261440. [doi: 10.1080/23311908.2016.1261440]
17. Mack M, Stojan R, Bock O, Voelcker-Rehage C. The association of executive functions and physical fitness with cognitive-motor multitasking in a street crossing scenario. Sci Rep 2023 Jan 13;13(1):697 [FREE Full text] [doi: 10.1038/s41598-022-26438-x] [Medline: 36639402]
18. Schmidt GJ, Boechat YE, van Duinkerken E, Schmidt JJ, Moreira TB, Nicaretta DH, et al. Detection of cognitive dysfunction in elderly with a low educational level using a reaction-time attention task. J Alzheimers Dis 2020;78(3):1197-1205. [doi: 10.3233/JAD-200881] [Medline: 33136095]
19. Abo M, Hamaguchi T. Effectiveness of a dual-task intervention involving exercise and vocalized cognitive tasks. J Clin Med 2024 May 17;13(10):2962 [FREE Full text] [doi: 10.3390/jcm13102962] [Medline: 38792503]
20. Brustio PR, Magistro D, Zecca M, Liubicich ME, Rabaglietti E. Fear of falling and activities of daily living function: mediation effect of dual-task ability. Aging Ment Health 2018 Jun;22(6):856-861 [FREE Full text] [doi: 10.1080/13607863.2017.1318257] [Medline: 28485621]
21. Ewolds H, Broeker L, de Oliveira RF, Raab M, Künzell S. Ways to improve multitasking: effects of predictability after single- and dual-task training. J Cogn 2021 Jan 07;4(1):4 [FREE Full text] [doi: 10.5334/joc.142] [Medline: 33506170]
22. Strobach T, Wendt M, Janczyk M. Editorial: multitasking: executive functioning in dual-task and task switching situations. Front Psychol 2018;9:108 [FREE Full text] [doi: 10.3389/fpsyg.2018.00108] [Medline: 29497390]

XSL•FO
RenderX

23. Ali N, Tian H, Thabane L, Ma J, Wu H, Zhong Q, et al. The effects of dual-task training on cognitive and physical functions in older adults with cognitive impairment; a systematic review and meta-analysis. J Prev Alzheimers Dis 2022;9(2):359-370. [doi: 10.14283/jpad.2022.16] [Medline: 35543010]
24. He Y, Yang L, Zhou J, Yao L, Pang MY. Dual-task training effects on motor and cognitive functional abilities in individuals with stroke: a systematic review. Clin Rehabil 2018 Jul;32(7):865-877. [doi: 10.1177/0269215518758482] [Medline: 29473480]
25. Malik J, Stemplewski R, Maciaszek J. The effect of juggling as dual-task activity on human neuroplasticity: a systematic review. Int J Environ Res Public Health 2022 Jun 09;19(12):7102 [FREE Full text] [doi: 10.3390/ijerph19127102] [Medline: 35742356]
26. Li Z, Wang T, Liu H, Jiang Y, Wang Z, Zhuang J. Dual-task training on gait, motor symptoms, and balance in patients with Parkinson's disease: a systematic review and meta-analysis. Clin Rehabil 2020 Nov;34(11):1355-1367. [doi: 10.1177/0269215520941142] [Medline: 32660265]
27. Martino Cinnera A, Bisirri A, Leone E, Morone G, Gaeta A. Effect of dual-task training on balance in patients with multiple sclerosis: a systematic review and meta-analysis. Clin Rehabil 2021 Oct;35(10):1399-1412. [doi: 10.1177/02692155211010372] [Medline: 33874763]
28. Manser P, de Bruin ED. Making the best out of IT: design and development of exergames for older adults with mild neurocognitive disorder - a methodological paper. Front Aging Neurosci 2021;13:734012 [FREE Full text] [doi: 10.3389/fnagi.2021.734012] [Medline: 34955806]
29. Johansson H, Folkerts AK, Hammarström I, Kalbe E, Leavy B. Effects of motor-cognitive training on dual-task performance in people with Parkinson's disease: a systematic review and meta-analysis. J Neurol 2023 Jun;270(6):2890-2907 [FREE Full text] [doi: 10.1007/s00415-023-11610-8] [Medline: 36820916]
30. Yu D, Li X, He S, Zhu H, Lam FM, Pang MY. The effect of dual-task training on cognitive ability, physical function, and dual-task performance in people with dementia or mild cognitive impairment: a systematic review and meta-analysis. Clin Rehabil 2024 Apr;38(4):443-456. [doi: 10.1177/02692155231216621] [Medline: 38009067]
31. Tuena C, Borghesi F, Bruni F, Cavedoni S, Maestri S, Riva G, et al. Technology-assisted cognitive motor dual-task rehabilitation in chronic age-related conditions: systematic review. J Med Internet Res 2023 May 22;25:e44484 [FREE Full text] [doi: 10.2196/44484] [Medline: 37213200]
32. Chong MS, Sit JW, Karthikesu K, Chair SY. Effectiveness of technology-assisted cardiac rehabilitation: a systematic review and meta-analysis. Int J Nurs Stud 2021 Dec;124:104087. [doi: 10.1016/j.ijnurstu.2021.104087] [Medline: 34562846]
33. Peska DN, Lewis KO. Uniform instruction using web-based, asynchronous technology in a geographically distributed clinical clerkship: analysis of osteopathic medical student participation and satisfaction. J Am Osteopath Assoc 2010 Mar;110(3):135-142. [Medline: 20386022]
34. DiCerbo KE, Shute V, Kim YJ. The future of assessment in technology-rich environments: psychometric considerations. In: Learning, Design, and Technology: An International Compendium of Theory, Research, Practice, and Policy. Cham, Switzerland: Springer; 2023.
35. Brivio E, Serino S, Negro Cousa E, Zini A, Riva G, De Leo G. Virtual reality and 360° panorama technology: a media comparison to study changes in sense of presence, anxiety, and positive emotions. Virtual Real 2020 Jul 09;25(2):303-311. [doi: 10.1007/s10055-020-00453-7]
36. Furszyfer Del Rio DD, Sovacool BK, Martiskainen M. Controllable, frightening, or fun? Exploring the gendered dynamics of smart home technology preferences in the United Kingdom. Energy Res Soc Sci 2021 Jul;77:102105. [doi: 10.1016/j.erss.2021.102105]
37. Ong TL, Ruppert MM, Akbar M, Rashidi P, Ozrazgat-Baslanti T, Bihorac A, et al. Improving the intensive care patient experience with virtual reality-a feasibility study. Crit Care Explor 2020 Jun;2(6):e0122 [FREE Full text] [doi: 10.1097/CCE.0000000000000122] [Medline: 32695991]
38. Ponsignon F, Derbaix M. The impact of interactive technologies on the social experience: an empirical study in a cultural tourism context. Tour Manag Perspect 2020 Jul;35:100723. [doi: 10.1016/j.tmp.2020.100723]
39. Hurling R, Fairley BW, Dias MB. Internet-based exercise intervention systems: are more interactive designs better? Psychol Health 2006 Dec;21(6):757-772. [doi: 10.1080/14768320600603257]
40. Ouyang F, Zheng L, Jiao P. Artificial intelligence in online higher education: a systematic review of empirical research from 2011 to 2020. Educ Inf Technol 2022 Feb 26;27(6):7893-7925. [doi: 10.1007/s10639-022-10925-9]
41. Sallam M, Salim NA, Barakat M, Al-Tammemi AB. ChatGPT applications in medical, dental, pharmacy, and public health education: a descriptive study highlighting the advantages and limitations. Narra J 2023 Apr;3(1):e103 [FREE Full text] [doi: 10.52225/narra.v3i1.103] [Medline: 38450035]
42. Zhai X, Chu X, Chai CS, Jong MS, Istenic A, Spector M, et al. A review of artificial intelligence (AI) in education from 2010 to 2020. Complexity 2021 Apr 20;2021. [doi: 10.1155/2021/8812542]
43. Bennell K, Nelligan RK, Schwartz S, Kasza J, Kimp A, Crofts SJ, et al. Behavior change text messages for home exercise adherence in knee osteoarthritis: randomized trial. J Med Internet Res 2020 Sep 28;22(9):e21749 [FREE Full text] [doi: 10.2196/21749] [Medline: 32985994]

XSL•FO
RenderX

44. Canan CE, Waselewski ME, Waldman AL, Reynolds G, Flickinger TE, Cohn WF, et al. Long term impact of PositiveLinks: clinic-deployed mobile technology to improve engagement with HIV care. PLoS One 2020;15(1):e0226870 [FREE Full text] [doi: 10.1371/journal.pone.0226870] [Medline: 31905209]
45. Collado-Mateo D, Lavín-Pérez AM, Peñacoba C, Del Coso J, Leyton-Román M, Luque-Casado A, et al. Key factors associated with adherence to physical exercise in patients with chronic diseases and older adults: an umbrella review. Int J Environ Res Public Health 2021 Mar 19;18(4):2023 [FREE Full text] [doi: 10.3390/ijerph18042023] [Medline: 33669679]
46. Pudkasam S, Feehan J, Talevski J, Vingrys K, Polman R, Chinlumprasert N, et al. Motivational strategies to improve adherence to physical activity in breast cancer survivors: a systematic review and meta-analysis. Maturitas 2021 Oct;152:32-47. [doi: 10.1016/j.maturitas.2021.06.008] [Medline: 34674806]
47. Ollermark K. Designing for body awareness: exploring the interaction between screen and participant in pre-recorded online workouts. Malmö University. 2022. URL: https://www.diva-portal.org/smash/get/diva2:1676394/FULLTEXT02 [accessed 2025-05-14]
48. Woessner MN, Tacey A, Levinger-Limor A, Parker AG, Levinger P, Levinger I. The evolution of technology and physical inactivity: the good, the bad, and the way forward. Front Public Health 2021;9:655491 [FREE Full text] [doi: 10.3389/fpubh.2021.655491] [Medline: 34123989]
49. Gao L, Zhang G, Yu B, Qiao Z, Wang J. Wearable human motion posture capture and medical health monitoring based on wireless sensor networks. Measurement 2020 Dec;166:108252. [doi: 10.1016/j.measurement.2020.108252]
50. Rana M, Mittal V. Wearable sensors for real-time kinematics analysis in sports: a review. IEEE Sensors J 2021 Jan 15;21(2):1187-1207. [doi: 10.1109/jsen.2020.3019016]
51. Roggio F, Ravalli S, Maugeri G, Bianco A, Palma A, Di Rosa M, et al. Technological advancements in the analysis of human motion and posture management through digital devices. World J Orthop 2021 Jul 18;12(7):467-484 [FREE Full text] [doi: 10.5312/wjo.v12.i7.467] [Medline: 34354935]
52. Minardi HA, Ritter S. Recording skills practice on videotape can enhance learning - a comparative study between nurse lecturers and nursing students. J Adv Nurs 1999 Jun;29(6):1318-1325. [doi: 10.1046/j.1365-2648.1999.01017.x] [Medline: 10354225]
53. Yang K, Perez M, Hubert N, Hossu G, Perrenot C, Hubert J. Effectiveness of an integrated video recording and replaying system in robotic surgical training. Ann Surg 2017 Mar;265(3):521-526. [doi: 10.1097/SLA.0000000000001699] [Medline: 28169927]
54. Chen EH, Bailey DH. Dual-task studies of working memory and arithmetic performance: a meta-analysis. J Exp Psychol Learn Mem Cogn 2021 Mar;47(2):220-233. [doi: 10.1037/xlm0000822] [Medline: 32011157]
55. Liederman J. Subtraction in addition to addition: dual task performance improves when tasks are presented to separate hemispheres. J Clin Exp Neuropsychol 1986 Oct;8(5):486-502. [doi: 10.1080/01688638608405172] [Medline: 3805249]
56. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Syst Rev 2021 Mar 29;10(1):89 [FREE Full text] [doi: 10.1186/s13643-021-01626-4] [Medline: 33781348]
57. Campbell M, McKenzie JE, Sowden A, Katikireddi SV, Brennan SE, Ellis S, et al. Synthesis without meta-analysis (SWiM) in systematic reviews: reporting guideline. BMJ 2020 Jan 16;368:l6890 [FREE Full text] [doi: 10.1136/bmj.l6890] [Medline: 31948937]
58. Tricco AC, Langlois EV, Straus SE. Rapid reviews to strengthen health policy and systems: a practical guide. World Health Organization. 2017. URL: https://iris.who.int/bitstream/handle/10665/258698/9789241512763-eng.pdf [accessed 2025-05-14]
59. Hartling L, Guise JM, Hempel S, Featherstone R, Mitchell MD, Motu'apuaka ML, et al. Fit for purpose: perspectives on rapid reviews from end-user interviews. Syst Rev 2017 Mar 17;6(1):32 [FREE Full text] [doi: 10.1186/s13643-017-0425-7] [Medline: 28212677]
60. Kazi MR, Chowdhury N, Chowdhury MM, Turin TC. Conducting a rapid review for quick turnaround knowledge synthesis. Health Prim Care 2021;5. [doi: 10.15761/HPC.1000213]
61. Technology-assisted motor-cognitive training among older adults: a rapid systematic review. OSF Registries. 2024 Jul 31. URL: https://osf.io/5srcq [accessed 2025-05-14]
62. Amir-Behghadami M, Janati A. Population, Intervention, Comparison, Outcomes and Study (PICOS) design as a framework to formulate eligibility criteria in systematic reviews. Emerg Med J 2020 Jun;37(6):387. [doi: 10.1136/emermed-2020-209567] [Medline: 32253195]
63. Zabor EC, Kaizer AM, Hobbs BP. Randomized controlled trials. Chest 2020 Jul;158(1S):S79-S87 [FREE Full text] [doi: 10.1016/j.chest.2020.03.013] [Medline: 32658656]
64. Feeley N, Cossette S. Pilot studies for randomized clinical trials. In: Henly SJ, editor. Routledge International Handbook of Advanced Quantitative Methods in Nursing Research. London, UK: Routledge; 2015:199-212.
65. Tikka C, Verbeek J, Ijaz S, Hoving JL, Boschman J, Hulshof C, et al. Quality of reporting and risk of bias: a review of randomised trials in occupational health. Occup Environ Med 2021 Sep;78(9):691-696 [FREE Full text] [doi: 10.1136/oemed-2020-107038] [Medline: 34162718]
66. WHO Guidelines Approved by the Guidelines Review Committee. Geneva, Switzerland: World Health Organization; 2020.

XSL•FO
RenderX

67. Menengi Ç KN, Yeldan İ, Çınar N, Şahiner T. Effectiveness of motor-cognitive dual-task exercise via telerehabilitation in Alzheimer's disease: an online pilot randomized controlled study. Clin Neurol Neurosurg 2022 Dec;223:107501. [doi: 10.1016/j.clineuro.2022.107501] [Medline: 36368169]
68. Jäggi S, Wachter A, Adcock M, de Bruin ED, Möller JC, Marks D, et al. Feasibility and effects of cognitive-motor exergames on fall risk factors in typical and atypical Parkinson's inpatients: a randomized controlled pilot study. Eur J Med Res 2023 Jan 16;28(1):30 [FREE Full text] [doi: 10.1186/s40001-022-00963-x] [Medline: 36647177]
69. Kwan RY, Liu JY, Fong KN, Qin J, Leung PK, Sin OS, et al. Feasibility and effects of virtual reality motor-cognitive training in community-dwelling older people with cognitive frailty: pilot randomized controlled trial. JMIR Serious Games 2021 Aug 06;9(3):e28400 [FREE Full text] [doi: 10.2196/28400] [Medline: 34383662]
70. Altorfer P, Adcock M, de Bruin ED, Graf F, Giannouli E. Feasibility of cognitive-motor exergames in geriatric inpatient rehabilitation: a pilot randomized controlled study. Front Aging Neurosci 2021;13:739948 [FREE Full text] [doi: 10.3389/fnagi.2021.739948] [Medline: 34912206]
71. Manser P, Poikonen H, de Bruin ED. Feasibility, usability, and acceptance of "Brain-IT"-a newly developed exergame-based training concept for the secondary prevention of mild neurocognitive disorder: a pilot randomized controlled trial. Front Aging Neurosci 2023;15:1163388 [FREE Full text] [doi: 10.3389/fnagi.2023.1163388] [Medline: 37810620]
72. Fishbein P, Hutzler Y, Ratmansky M, Treger I, Dunsky A. A preliminary study of dual-task training using virtual reality: influence on walking and balance in chronic poststroke survivors. J Stroke Cerebrovasc Dis 2019 Nov;28(11):104343. [doi: 10.1016/j.jstrokecerebrovasdis.2019.104343] [Medline: 31495668]
73. Kannan L, Vora J, Bhatt T, Hughes SL. Cognitive-motor exergaming for reducing fall risk in people with chronic stroke: a randomized controlled trial. NeuroRehabilitation 2019;44(4):493-510. [doi: 10.3233/NRE-182683] [Medline: 31256084]
74. Uematsu A, Tsuchiya K, Fukushima H, Hortobágyi T. Effects of motor-cognitive dual-task standing balance exergaming training on healthy older adults' standing balance and walking performance. Games Health J 2023 Aug;12(4):302-309. [doi: 10.1089/g4h.2022.0203] [Medline: 36944151]
75. Liao YY, Chen IH, Lin YJ, Chen Y, Hsu WC. Effects of virtual reality-based physical and cognitive training on executive function and dual-task gait performance in older adults with mild cognitive impairment: a randomized control trial. Front Aging Neurosci 2019;11:162 [FREE Full text] [doi: 10.3389/fnagi.2019.00162] [Medline: 31379553]
76. Schoene D, Valenzuela T, Toson B, Delbaere K, Severino C, Garcia J, et al. Interactive cognitive-motor step training improves cognitive risk factors of falling in older adults - a randomized controlled trial. PLoS One 2015;10(12):e0145161 [FREE Full text] [doi: 10.1371/journal.pone.0145161] [Medline: 26673919]
77. Forte R, Trentin C, Tocci N, Lucia S, Aydin M, Di Russo F. Motor-cognitive exercise with variability of practice and feedback improves functional ability and cognition in older individuals. Aging Clin Exp Res 2023 Nov;35(11):2797-2806. [doi: 10.1007/s40520-023-02568-8] [Medline: 37853304]
78. Pelosin E, Ponte C, Putzolu M, Lagravinese G, Hausdorff JM, Nieuwboer A, et al. Motor-cognitive treadmill training with virtual reality in Parkinson's disease: the effect of training duration. Front Aging Neurosci 2021;13:753381 [FREE Full text] [doi: 10.3389/fnagi.2021.753381] [Medline: 35069171]
79. Eggenberger P, Theill N, Holenstein S, Schumacher V, de Bruin ED. Multicomponent physical exercise with simultaneous cognitive training to enhance dual-task walking of older adults: a secondary analysis of a 6-month randomized controlled trial with 1-year follow-up. Clin Interv Aging 2015;10:1711-1732 [FREE Full text] [doi: 10.2147/CIA.S91997] [Medline: 26604719]
80. Delbroek T, Vermeylen W, Spildooren J. The effect of cognitive-motor dual task training with the biorescue force platform on cognition, balance and dual task performance in institutionalized older adults: a randomized controlled trial. J Phys Ther Sci 2017 Jul;29(7):1137-1143 [FREE Full text] [doi: 10.1589/jpts.29.1137] [Medline: 28744033]
81. Hagovska M, Nagyova I. The transfer of skills from cognitive and physical training to activities of daily living: a randomised controlled study. Eur J Ageing 2017 Jun;14(2):133-142 [FREE Full text] [doi: 10.1007/s10433-016-0395-y] [Medline: 28804398]
82. Park JS, Jung YJ, Lee G. Virtual reality-based cognitive-motor rehabilitation in older adults with mild cognitive impairment: a randomized controlled study on motivation and cognitive function. Healthcare (Basel) 2020 Sep 11;8(3):335 [FREE Full text] [doi: 10.3390/healthcare8030335] [Medline: 32932997]
83. Villa-Sánchez B, Gandolfi M, Emadi Andani M, Valè N, Rossettini G, Polesana F, et al. Placebo effect on gait: a way to reduce the dual-task cost in older adults. Exp Brain Res 2023 Jun;241(6):1501-1511. [doi: 10.1007/s00221-023-06620-x] [Medline: 37085646]
84. Buele J, Avilés-Castillo F, Del-Valle-Soto C, Varela-Aldás J, Palacios-Navarro G. Effects of a dual intervention (motor and virtual reality-based cognitive) on cognition in patients with mild cognitive impairment: a single-blind, randomized controlled trial. J Neuroeng Rehabil 2024 Aug 01;21(1):130 [FREE Full text] [doi: 10.1186/s12984-024-01422-w] [Medline: 39090664]
85. Zak M, Sikorski T, Michalska A, Sztandera P, Szczepanowska-Wolowiec B, Brola W, et al. The effects of physiotherapy programmes, aided by virtual reality solutions, on balance in older women: a randomised controlled trial. J Clin Med 2024 Oct 28;13(21):6462 [FREE Full text] [doi: 10.3390/jcm13216462] [Medline: 39518601]

XSL•FO
RenderX

86. Kwan RY, Liu J, Sin OS, Fong KN, Qin J, Wong JC, et al. Effects of virtual reality motor-cognitive training for older people with cognitive frailty: multicentered randomized controlled trial. J Med Internet Res 2024 Sep 11;26:e57809 [FREE Full text] [doi: 10.2196/57809] [Medline: 39259959]
87. Eisenstein C. Implementing and evaluating co-design: a step-by-step toolkit. New Philanthropy Capital. 2019. URL: https://policycommons.net/artifacts/1572973/implementing-and-evaluating-co-design-a-step-by-step-toolkit-contents/2262752/ [accessed 2025-05-14]
88. Prabaswari AD, Mahfudhi MICCC. Comparative analysis of mental workloads for disruption technicians and new installation technicians using the NASA-TLX method (case study: PT Telkom Akses Kandatel Sleman). AIP Conf Proc 2023:050017c [FREE Full text] [doi: 10.1063/5.0107051]
89. Grier RA, Bangor A, Kortum P, Peres SC. The System Usability Scale: beyond standard usability testing. Proc Hum Factors Ergon Soc Annu Meet 2013 Sep 30;57(1):187-191. [doi: 10.1177/1541931213571042]
90. Bird M, McGillion M, Chambers EM, Dix J, Fajardo CJ, Gilmour M, et al. A generative co-design framework for healthcare innovation: development and application of an end-user engagement framework. Res Involv Engagem 2021 Mar 01;7(1):12 [FREE Full text] [doi: 10.1186/s40900-021-00252-7] [Medline: 33648588]
91. Maher LM, Hayward B, Hayward P, Walsh C. Increasing sustainability in co-design projects: a qualitative evaluation of a co-design programme in New Zealand. Patient Exp J 2017 Jul 26;4(2):44-52. [doi: 10.35680/2372-0247.1150]
92. Tay BS, Cox DN, Brinkworth GD, Davis A, Edney SM, Gwilt I, et al. Co-design practices in diet and nutrition research: an integrative review. Nutrients 2021 Oct 14;13(10):3593 [FREE Full text] [doi: 10.3390/nu13103593] [Medline: 34684593]
93. Shanyinde M, Pickering RM, Weatherall M. Questions asked and answered in pilot and feasibility randomized controlled trials. BMC Med Res Methodol 2011 Aug 16;11:117 [FREE Full text] [doi: 10.1186/1471-2288-11-117] [Medline: 21846349]
94. Boksem MA, Meijman TF, Lorist MM. Effects of mental fatigue on attention: an ERP study. Brain Res Cogn Brain Res 2005 Sep;25(1):107-116. [doi: 10.1016/j.cogbrainres.2005.04.011] [Medline: 15913965]
95. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
96. Gustavson K, von Soest T, Karevold E, Røysamb E. Attrition and generalizability in longitudinal studies: findings from a 15-year population-based study and a Monte Carlo simulation study. BMC Public Health 2012 Oct 29;12:918 [FREE Full text] [doi: 10.1186/1471-2458-12-918] [Medline: 23107281]
97. Šlosar L, Voelcker-Rehage C, Paravlić AH, Abazovic E, de Bruin ED, Marusic U. Combining physical and virtual worlds for motor-cognitive training interventions: position paper with guidelines on technology classification in movement-related research. Front Psychol 2022;13:1009052 [FREE Full text] [doi: 10.3389/fpsyg.2022.1009052] [Medline: 36591050]
98. Varela-Vásquez LA, Minobes-Molina E, Jerez-Roig J. Dual-task exercises in older adults: a structured review of current literature. J Frailty Sarcopenia Falls 2020 Jun;5(2):31-37 [FREE Full text] [doi: 10.22540/JFSF-05-031] [Medline: 32510028]
99. Pereira Oliva HN, Mansur Machado FS, Rodrigues VD, Leão LL, Monteiro-Júnior RS. The effect of dual-task training on cognition of people with different clinical conditions: an overview of systematic reviews. IBRO Rep 2020 Dec;9:24-31 [FREE Full text] [doi: 10.1016/j.ibror.2020.06.005] [Medline: 33336101]
100. Wollesen B, Voelcker-Rehage C. Training effects on motor–cognitive dual-task performance in older adults. Eur Rev Aging Phys Act 2013 Feb 24;11(1):5-24. [doi: 10.1007/s11556-013-0122-z]
101. Hötting K, Röder B. Beneficial effects of physical exercise on neuroplasticity and cognition. Neurosci Biobehav Rev 2013 Nov;37(9 Pt B):2243-2257. [doi: 10.1016/j.neubiorev.2013.04.005] [Medline: 23623982]

Abbreviations

ADL: activities of daily living
IADL: instrumental activities of daily living
MCT: motor-cognitive training
MeSH: Medical Subject Headings
NASA-TLX: NASA Task Load Index
PD: Parkinson disease
PICOS: population, intervention, comparator, outcome, and study design
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
SUS: System Usability Scale
SWiM: Synthesis Without Meta-analysis
VR: virtual reality
WHO: World Health Organization

XSL•FO
RenderX

Edited by A Coristine; submitted 06.10.24; peer-reviewed by M Takemi, M Gasmi, F Bruni; comments to author 16.12.24; revised version received 25.02.25; accepted 18.04.25; published 03.06.25.

Please cite as:
Li Y, Liu Y, Leung AYM, Montayre J
Technology-Assisted Motor-Cognitive Training Among Older Adults: Rapid Systematic Review of Randomized Controlled Trials
JMIR Serious Games 2025;13:e67250
URL: https://games.jmir.org/2025/1/e67250
doi:10.2196/67250
PMID:

©Yaqin Li, Yaqian Liu, Angela YM Leung, Jed Montayre. Originally published in JMIR Serious Games (https://games.jmir.org), 03.06.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Serious Games for Preventing Musculoskeletal Disorders in Occupational Settings: Scoping Review

Thomas Rivière[1], MS; Thibaud Hulin[2], PhD; Farzan Sasangohar[3], PhD; Mohsen Zare[1], PhD

[1]ELLIADD (UR 4661), UTBM, Université Marie et Louis Pasteur, Belfort, France

[2]ELLIADD (UR 4661), Université Marie et Louis Pasteur, Montbéliard, France

[3]Department of Industrial and Systems Engineering, Texas A&M University, College Station, TX, United States

Corresponding Author:
Thomas Rivière, MS
ELLIADD (UR 4661)
UTBM
Université Marie et Louis Pasteur
Campus Montbeliard
Belfort, 90100 Cedex
France
Phone: 33 384583320
Email: thomas.riviere@utbm.fr

Abstract

Background: Musculoskeletal disorders (MSDs) are a significant health concern in the workplace, and while ergonomic interventions are commonly used, their long-term effectiveness is often questioned. Serious games (SGs), designed to go beyond entertainment, have emerged as a promising tool that may address some of the limitations of traditional interventions, such as the need for sustained impact and greater worker engagement.

Objective: This review aims to identify and analyze the key characteristics of SGs—including design, gameplay, and expected outcomes—that have been developed for the prevention or mitigation of work-related MSDs. Additionally, it explores the documented effects of SG implementation, assessing their potential contribution to MSD prevention and intervention strategies.

Methods: A scoping review was conducted across 6 scientific databases (APA PsycInfo, Web of Science, Science Direct, MEDLINE, IEEE Xplore, and Google Scholar) to identify relevant studies published up to 2025. The selection process involved a multistep screening, including title and abstract review, followed by full-text assessment by 2 independent reviewers. Studies included were original research articles in English addressing MSD prevention and mitigation. Exclusions applied to studies on nonwork-related MSDs, limited content, duplicates, or repurposed entertainment games or gamification solutions. Data extraction was performed using a standardized form to capture key study characteristics. A 2-level analysis was applied: descriptive analysis, categorizing studies based on study characteristics and primary focus (design, evaluation, or both), and content-based analysis, examining game design, gameplay, expected outcomes, and evaluation methods to provide a structured synthesis of findings.

Results: The initial search identified 2700 records, with 15 studies meeting the inclusion criteria. These studies explored diverse applications of SGs for MSD prevention, focusing either on game design and development or on educational impact assessment. Notably, only 2 studies comprehensively addressed both design methodology and educational evaluation. Findings revealed considerable variability in design approaches, technological platforms, gameplay mechanics, and expected outcomes. Additionally, the literature exhibited significant inconsistencies in evaluating SG effectiveness, with methodological limitations affecting comparability. While some studies targeted rehabilitation or occupational health and safety, only a few explicitly focused on MSD prevention, with a predominant emphasis on physical risk factors, whereas psychosocial and organizational aspects remained largely underexplored.

Conclusions: This review highlights the need for standardized protocols and criteria for the design and evaluation of SGs to enable further synthesis and impact measurement. The integration of MSD prevention into SGs remains limited and is often approached indirectly through related themes such as workplace safety or rehabilitation. Future research should focus on developing and validating more comprehensive SG-based interventions and exploring their potential as effective tools in occupational health. The findings indicate a substantial gap in empirical evidence regarding the effectiveness of SGs for MSD prevention, largely due to the disparity in experimental approaches.

XSL•FO
RenderX

(JMIR Serious Games 2025;13:e66913) doi:10.2196/66913

KEYWORDS

musculoskeletal disorders; serious games; occupational health; prevention; exergaming; video games

Introduction

Shifting MSD Prevention Toward Interactive Solutions

Work-related musculoskeletal disorders (MSDs) are a major public health issue and the leading cause of occupational illness in Europe, with incidence rates increasing [1,2]. The European Agency for Safety and Health at Work reports that MSDs account for 60% of work-related health issues in the European Union [1]. In 2019, more than 100 million European workers were affected, with back pain (43%) and neck or shoulder discomfort (41%) being the most common [1,3]. The prevalence is particularly high among sectors such as construction, water supply, agriculture, forestry, and fishing [1]. Data from the French social health insurance system estimates the economic burden of MSDs to be approximately 2 billion euros (US $2.35 billion) [4], contributing to significant labor force loss and financial strain. Intensified work conditions, task monotony, shortened cycles, and high psychological demands contribute to this trend [5,6]. Given the multiple risk factors, it is crucial to implement effective prevention strategies [7].

To address this growing concern, various preventive strategies have been developed. Effective interventions require a comprehensive approach that integrates technical solutions, organizational changes, and worker training [8-10]. Operator training is a widely adopted preventive measure aimed at increasing awareness, promoting behavior change, and supporting optimal work practices [11].

Numerous industries integrate MSD prevention programs into their operational frameworks, often including operator training. However, the implementation of such training varies across sectors [12] and faces challenges related to engagement, practical application, and knowledge retention [13]. Studies suggest that traditional training alone is insufficient for achieving sustained behavior change because of inconsistent implementation and a lack of standardization [7]. Moreover, the effectiveness of these programs remains uncertain, as many interventions are not systematically documented, and randomized controlled trials often fail to demonstrate a significant impact [11]. For instance, Hogan et al [14] found that manual material handling training did not significantly reduce MSD incidence, while Yu et al [15] concluded that training alone is insufficient for preventing MSDs. Even when training interventions show some efficacy [16], their long-term impact remains questionable [17].

While training programs and awareness campaigns are essential, they may not always ensure the long-term adoption of preventive behaviors, necessitating a paradigm shift. This shift involves using technologies to create immersive and interactive learning environments that are better aligned with workers' realities. Among these technologies, serious games (SGs) offer a complementary approach to diversify prevention strategies and enhance existing interventions. Rather than replacing conventional methods, SGs capitalize on interactivity and engagement to reinforce key messages and promote behavioral change [18,19].

Serious Games for Occupational Settings

To fully understand the role of SGs in occupational training, it is important to define their key characteristics. An SG is a computer application designed to integrate the characterizing goals [20] such as prevention, education, communication, or behavior change with the engaging features of video games [21]. SGs enhance engagement, sustain motivation, and create an environment where users can experiment, fail, and learn without real-world consequences [22-24]. This aspect is particularly relevant for MSD prevention, as it allows workers to explore risk situations and test alternative solutions in a controlled, risk-free space. Additionally, SGs offer flexibility in training schedules, allowing workers to engage in learning during their leisure time, which may contribute to higher adoption rates and sustained motivation [25-27].

Beyond their general benefits, SGs can also be tailored to meet the specific needs of different workplace environments. They can simulate realistic work settings in a way that personalizes the experience and may lead to enhanced performance outcomes [28]. These features are not inherent to SGs but can be incorporated depending on the target audience and intended outcomes. In MSD prevention, SGs can help workers recognize risk factors, understand ergonomic principles, and apply preventive strategies in their daily activities.

Research Gap and Objectives

Despite their potential, the effectiveness of SGs in occupational settings remains uncertain. The difficulty in demonstrating their impact is largely due to methodological limitations that undermine the reliability of findings [29]. Material and software constraints also present challenges, potentially affecting implementation and adoption. Additionally, a misalignment between game mechanics and intended outcomes may diminish the educational and preventive value [30]. While gamification can produce notable cognitive outcomes, its impact on motivation and behavior is inconsistent. In particular, extrinsic motivation mechanisms can sometimes overshadow intrinsic engagement, reducing overall effectiveness [31].

Given these uncertainties, a scoping review of the literature is essential to understand how SGs have been designed and evaluated for MSD prevention. As a preliminary step to investigate this gap, this study aims to identify the key characteristics of SGs—including design, gameplay, and expected outcomes—used to address MSD prevention. Additionally, it explores whether any observable effects related to SG implementation have been documented.

XSL•FO
RenderX

Methods

Study Design

This study follows a scoping review approach, adhering to the methodological framework of Arksey and O'Malley [32] and further refined by Levac et al [33]. The findings are reported in accordance with the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist (Multimedia Appendix 1) [34].

Search Strategy

A systematic literature search was conducted in APA PsycInfo, Web of Science, Science Direct, MEDLINE via PubMed, and IEEE Xplore. Additionally, Google Scholar was searched, and ResearchRabbit was used to identify further relevant studies. The search covered all studies published up to 2025. The search strategy combined terms related to MSD prevention and SGs: ("musculoskeletal disorders" OR "musculoskeletal injury" OR "musculoskeletal disease" OR "MSD" OR "repetitive strain injury" OR "cumulative trauma disorder" OR "occupational health" OR "occupational safety" OR "occupational prevention" OR "occupational disorder" OR "ergonomic" OR "prevention" OR "risk reduction" OR "workplace intervention") AND ("serious game" OR "game-based" OR "educational game" OR "digital game" OR "learning game" OR "prevention game" OR "digital educational game" OR "video game" OR "gamification" OR "game" OR "gaming").

Search queries were tailored to each database's syntax (eg, title or abstract in MEDLINE, topics in Web of Science, and abstract or full text in IEEE Xplore). Detailed search queries are provided in Multimedia Appendix 2.

Eligibility Criteria and Study Selection

To ensure the selection of relevant studies, specific inclusion and exclusion criteria were established based on the scope of this review. These criteria are outlined in Textbox 1.

Studies on occupational safety management were included if they addressed MSD prevention.

Textbox 1. Eligibility criteria.

Inclusion criteria

- Written in English and published as original research articles (excluding review articles and meta-analyses).
- Addressed musculoskeletal disorder (MSD) prevention, rehabilitation, addressing work-related MSDs, or occupational safety.
- Considered at least one aspect of MSD prevention, including biomechanical, organizational, or psychosocial factors.
- Investigated serious games explicitly designed for MSD-related prevention or training.

Exclusion criteria

- Focused on musculoskeletal conditions unrelated to work (eg, sports injuries, nonwork-related rehabilitation, and general physiology).
- Had limited content (eg, posters and short conference abstracts).
- Were duplicate publications from the same project (only the most comprehensive and recent version was retained).
- Examined entertainment video games repurposed for training or gamification solutions (ie, game mechanics applied in nongame contexts [35]) as these do not meet the strict definition of serious games.

Study Selection Process

A rigorous multistep selection process was implemented to ensure methodological robustness and transparency (Multimedia Appendix 3).

Title and Abstract Screening

One reviewer (TR) initially screened all retrieved records (n=2700) to exclude studies that did not meet the inclusion criteria. The remaining studies were then compiled into a structured database, documenting key bibliographic details for systematic evaluation.

Full-Text Screening and Eligibility Assessment

Following title and abstract screening, 2 independent reviewers (TR and MZ) conducted a full-text assessment of the remaining studies to determine their eligibility. In cases of disagreement, a third reviewer provided arbitration to reach a consensus. To enhance efficiency and ensure consistency in decision-making, Rayyan software [36] was used for study selection.

Final Study Selection

All eligible studies underwent a final review before inclusion in the analysis. The overall selection process is presented as a flow diagram in the Results section.

Data Charting Process and Analysis

A standardized data-charting form was used to systematically extract relevant study characteristics, including publication details (year and country, study design, and target populations), MSD approach (biomechanical, psychosocial, or organizational), type and description of the serious game (utilitarian objective, gameplay, and expected outcomes), and experimental methodology (assessment measures design, experimental methodology, and platforms and control devices of games; Multimedia Appendix 3).

To systematically examine the focus and contributions of the included studies, a 2-level analytical approach was used. The first level focused on general study characteristics, including publication year, discipline, study design, and target populations.

XSL•FO
RenderX

It also categorized studies based on their primary objective. This level of descriptive analysis is outlined in Textbox 2.

The second level of analysis examined the core components of SG development and evaluation through 5 key parameters. To establish these parameters, we analyzed commonalities and divergences across studies regarding SG development, deployment, and assessment. Thematic coding revealed 5 recurring dimensions listed in Textbox 3.

Textbox 2. Descriptive analysis.

- Design-focused studies: Research primarily investigating game development, technologies, and methodologies.
- Evaluation-focused studies: Research assessing the effectiveness, usability, and engagement of serious games in achieving intended outcomes [26].
- Combined design and evaluation studies: Studies integrating both game development and assessment of its impact.

Textbox 3. Content analysis.

1. Design methodologies: These include collaborative, iterative, and participatory approaches, with a focus on technology-centered development.
2. Platforms and control devices: Games can be implemented on various platforms, including personal computers, consoles, mobile devices, and virtual reality systems. Control devices may involve haptic controllers, motion capture systems, and keyboards and mice.
3. Gameplay: Games vary across different categories such as simulations, exergames (games designed for physical exercise [37]), and management games. This feature encompasses mini-games, narrative elements, feedback systems, customization options, points and scoring, competition, and audio components.
4. Expected outcomes: The anticipated results include improvements in rehabilitation, heightened occupational health and safety awareness, musculoskeletal disorder prevention, user engagement, usability, user experience, and educational impact.
5. Evaluation methods: Evaluation methods assessed engagement, usability, user experience, and educational impact.

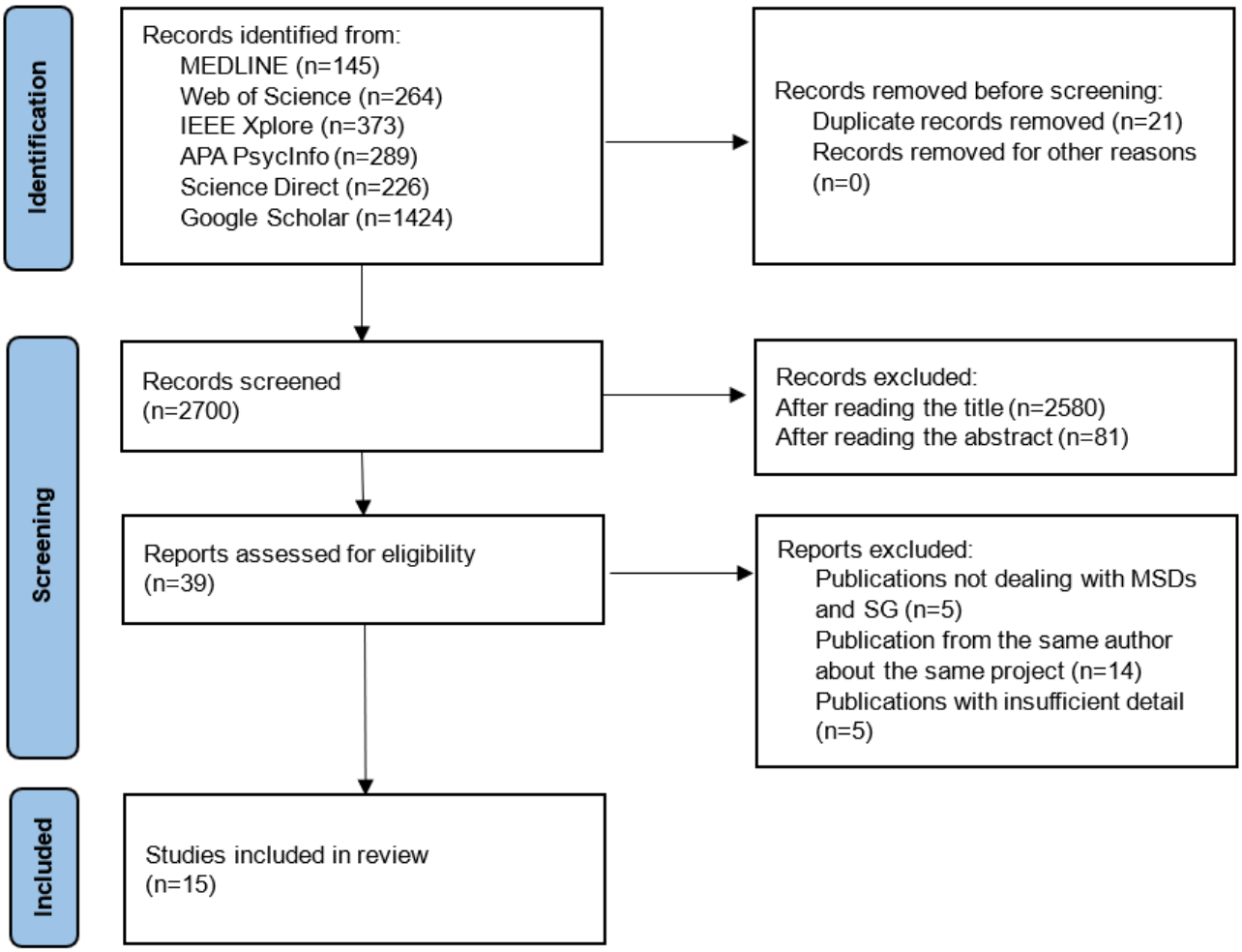
Results

Study Selection

A total of 2700 records were initially retrieved. After title screening, irrelevant and duplicate records were removed, leaving 120 studies for further evaluation. Abstract screening reduced the selection to 39 eligible studies, and after full-text assessment, 15 studies met the predefined inclusion criteria. The data selection procedure is shown in Figure 1 using a PRISMA-ScR flow diagram.

XSL•FO
RenderX

Figure 1. The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram of the selection process of the studies included in this review. MSD: musculoskeletal disorder; SG: serious games.



Characteristics of the Included Studies

Based on the first-level analysis, the included studies were published between 2013 and 2025. Geographically, studies were carried out in Europe (n=8), Asia (n=2), North America (n=2), South America (n=1), and Oceania (n=1). One study has a dual continental origin, the result of a collaboration between Colombian and Canadian researchers.

The studies addressed 3 primary application areas: rehabilitation SGs addressed work-related MSDs (n=4) designed for patients with MSDs [38-41], occupational health and safety games (n=5) aimed at enhancing risk management and injury prevention in workplace settings [42-46], and MSD prevention games (n=6), targeting physical risk factors and ergonomic training to mitigate occupational MSDs [47-52]. The games developed by Jansen-Kosterink et al [38] and France and Thomas [40], while not explicitly designed for occupational settings, were retained in our review due to their focus on chronic low back pain—one of the most prevalent forms of work-related MSDs. The target populations varied across studies: people with MSDs (n=4), health care professionals (n=1), industrial workers (n=3), including automotive industry workers (n=2) and watchmakers (n=1), sedentary office workers (n=2), high school students (n=1), children aged between 8 and 14 years (n=1), and unspecified target population (n=3). Across all studies, SGs were proposed as an alternative to traditional training methods, using their interactive and engaging nature to enhance learning and behavior change.

Additionally, the reviewed studies were further classified based on their primary focus on game design, evaluation, or both (Multimedia Appendix 4 [38-52]). Design-oriented studies (n=8) primarily investigated game development methodologies, with limited focus on evaluation. In most cases, evaluation was restricted to usability testing conducted at the final stage of game development [39,41,44-46,48,49,51]. These studies did not assess the long-term behavioral impact or MSD symptom reduction. Evaluation-oriented studies (n=5) conducted detailed assessments of SG usability, engagement, and effectiveness, but provided limited insights into game design methodologies [38,40,42,50,52]. Only 2 studies covering both design and evaluation comprehensively examined both development methodologies and games' impact on behavioral change [43,47].

Content Analysis of Game Design and Its Evaluation

Design Methodology

Among the design methods identified in the reviewed studies, collaborative approaches are predominant, with multiple stakeholders participating in various stages of game development. The collaborative design approach is prominently characterized by the involvement of various stakeholders,

XSL•FO
RenderX

including end users, ergonomists, and game developers. This approach aligns with the principles of design-based research, as outlined by Sanchez and Monod-Ansaldi [53]. Design-based research emphasizes a systematic yet flexible methodology that seeks to enhance educational practices through iterative analysis, design, development, and implementation, driven by collaboration among researchers and practitioners within real-world contexts.

Three articles provided very limited details regarding their design methodology, concentrating almost exclusively on the technical characteristics of the game [39,41,51]. In these cases, the perspective appears to be more oriented toward demonstrating technological feasibility than toward adopting a participatory approach that incorporates the needs of end users. While such approaches may serve as exploratory or experimental objectives, they do not provide clear insight into how learning objectives, user needs, engagement strategies, and evaluation criteria were defined and integrated into the game, thereby limiting the understanding of whether the intended educational or behavioral outcomes were achieved.

Although not all studies explicitly detail their design methodologies, a common practice observed is the involvement of users in varying degrees and stages of the design process with an iterative approach. Several studies [39,43,45,49] have documented user involvement in the early stages of the development process. For example, Lanzotti et al [43] engaged users in shaping various design aspects of the game, such as character perspective, avatar visualization, and graphical detail. This user feedback, gathered from a sample of approximately 10 participants, guided design decisions. The optimal configuration was determined through a 10-point evaluation process. On the other hand, Rodrigues et al [48] and Idriss et al [39] included users primarily during the testing phase to evaluate whether the serious game met its basic functional requirements. However, their focus did not extend to aspects such as user experience or broader usability considerations.

The methods used to engage users in the design phase also exhibit notable diversity, with some studies using virtual simulations and structured feedback mechanisms. For example, Sisto et al [49] used a virtual simulation of decontextualized activities within the game that are conceptually similar to real-world work tasks. Ergonomics experts and game developers collaborated closely to define these game scenarios, using insights drawn from field data related to actual operational activities to ensure that the simulated activities effectively represent the types of tasks encountered in real work environments. Rapp et al [45] similarly integrated users into their game development process through a combination of semistructured interviews and web-based questionnaires. The data obtained helped specify the educational objectives of the game and refine the initial design elements.

While most included studies incorporate user input at specific stages, 3 studies integrate stakeholders throughout the entire development process, offering a more comprehensive co-design approach. Pietrafesa et al [44] used an advanced collaborative design approach that used both qualitative and quantitative methods. Users were engaged actively across various phases of the game's development, including through surveys, focus groups, questionnaires, and discussions. User testing was conducted toward the end of the development cycle. Particularly, the authors sought input from both primary users and developers involved in the design process. Similarly, Kuipers et al [47] and Rebelo and Filgueiras [46] adopted an iterative design method grounded in collaboration between prospective users and experts. This method involved distinct stages: defining requirements, setting objectives and selecting technology, constructing prototypes, deploying the prototypes, evaluating them with target users, and conducting a comprehensive review by an expert panel.

Overall, the sophistication of the design methodologies appears to correlate with the effectiveness of the developed games. Games developed through advanced co-design methodologies often exhibit greater sophistication and functionality. Many researchers emphasize the importance of multidisciplinary and participatory collaboration in the development of SGs [53-55]. Hulin [56] underscores the specifics of the co-design approach within a research-action framework, introducing a novel method for analyzing design processes through stratigraphic modeling of experiences. Additionally, involving end users as consultants and testers—focusing their feedback on game dynamics—has been shown to enhance the game impact [57,58].

Platforms and Control Devices

To understand the technological landscape of SGs for MSD prevention, we examined the platforms and control mechanisms used in the reviewed studies. Among the reviewed studies, a variety of platforms and control mechanisms for SGs are evident. Four games were designed for personal computers (PCs) and use keyboard and mouse controls [42-45], while others apply haptic controllers [48] and cameras [41]. Three games were developed for smartphones and tablets [42,44,50]. Several games used motion-capture technologies such as Microsoft Kinect [39,47,48,52], inertial measurement unit (IMU) [51], Leap Motion [49], or similar tools like Vive Tracker [40]. One study used a more advanced motion capture system combining a reflective-marker suit, infrared cameras, and surface electromyography electrodes [38]. In some cases, the games used virtual reality (VR) devices [40,49] alongside motion-capture technologies, with the latter often serving as measurement tools. It is noteworthy that all the reviewed SGs were digital in format [59].

Sisto et al [49] used a combination of motion-capture technologies (Microsoft Kinect and Leap Motion) within a VR game. Their approach aimed to capture subtle hand movements with Leap Motion while using Microsoft Kinect to assess the positions of various body parts, including the back, arms, and neck. Similarly, Stranick and Lopez [52] used a VR headset in order to immerse the user in the game environment, leveraging Microsoft Kinect to adopt the expected posture for stretching exercises. Rodrigues et al [48] used multiple gaming peripherals, incorporating gestural control via Microsoft Kinect, and haptic devices. Their game architecture featured segments that are played on a PC with haptic controllers, while other segments used Kinect to detect when players needed breaks and to suggest exercises.

XSL•FO
RenderX

Some games, such as those developed by Greuter and Tepe [42] and Pietrafesa et al [44], privileged minimal hardware, designed specifically for tablets, smartphones, and PCs. The motivation for choosing minimal hardware requirements in game development often revolved around increasing accessibility, particularly for classroom settings or enabling students to use the game on their smartphones beyond class hours. However, the rationale behind specific sophisticated technology choices, such as VR, or haptic controllers, is rarely detailed in the reviewed studies. For example, Rodrigues et al [48] used haptic devices in their game, but the study lacks a clear justification for this choice, missing a clear explanation of the benefits and limitations of using haptic technology.

Games designed for rehabilitation, or the acquisition of physical skills, predominantly used motion capture technologies [38-40,47-49]. The choice of technology in these cases is closely tied to its functional utility within the game's context. However, the implementation of such technologies can introduce specific constraints. For example, in Valdivia et al [51], although the use of an IMU allows for faster sampling and more accurate capture, it is invasive and requires the expertise of specialists. Similarly, Stranick and Lopez [52] and Sisto et al [49] reported notable technical limitations associated with the use of VR technology. In Stranick and Lopez [52], the Microsoft Kinect sometimes failed to capture player movements in time. In Sisto et al [49], players had to keep their hands within their field of vision to avoid losing in-game items, and the VR system occasionally demonstrated suboptimal fluidity, requiring players to repeat certain actions. Additionally, the setup of the VR system was time-consuming and required the expertise of a VR specialist to address potential system bugs and ensure the game's proper functioning.

Gameplay

To clearly define the scope of gameplay in the reviewed studies, it is necessary to distinguish between its core components and related elements such as narrative scenarios. Here, “gameplay” is defined as the interactive and functional aspects of a game, encompassing its rules, mechanics, objectives, and the overall player experience. In contrast, a “scenario” refers to the narrative or storyline that provides the contextual framework for the game and structures the sequence of events during gameplay [59]. A single gameplay framework can support multiple scenarios.

The reviewed studies reveal a wide range of formats, from very simple structures to more elaborate ones. A first category of gameplay includes short and simple mini-games, based on basic interactions and often limited to a single task or objective [41,51,52]. A second category consists of relatively simple games that incorporate a progression or performance tracking system, allowing for a sense of continuity in the player experience [43,49,50]. A third configuration involves games composed of multiple successive mini-games, designed to diversify situations and gaming experiences [38-40,48]. The last category of gameplay consists of a more complex structure, combining various mechanics, levels of games, and integrating more advanced progression systems [42,44,45,47]. This diversity in gameplay reflects a variety of pedagogical intentions, as well as technical and contextual constraints specific to each project.

Several of the developed games show similarities in their underlying concepts and rules, while their distinctions appear at the level of narrative elements, artistic attributes, and control mechanisms. For instance, France and Thomas [40] and Jansen-Kosterink et al [38] both crafted memory-based games characterized by identical rule sets but differentiated by graphical elements and control mechanisms. Moreover, we observed games of similar genres, especially within the risk management category, within those designed for PCs [42,44,45].

In some PC-based games, player interaction extends beyond individual actions to managing environments and nonplayer characters (NPCs), adding a strategic component to gameplay. For example, in Pietrafesa et al [44], players are responsible for maintaining a safe work environment by overseeing NPCs' tasks and protecting them from workplace hazards. The player's success is tied to the productivity of the NPCs, which is affected by the implemented safety measures. Greuter and Tepe [42] require players to identify and address hazardous conditions affecting NPCs' construction work. Players mitigate these risks by providing necessary equipment and adjusting the work environment, enabling NPCs to continue their tasks. In Rapp et al [45], a multiplayer game simulates restaurant management, where players oversee various aspects such as personnel recruitment, budget management, and employee welfare. The game includes a debriefing phase aimed at enhancing learning outcomes, although details on the gameplay mechanics were not extensively provided in the study.

Some studies incorporated features designed to enhance player motivation and engagement, such as competition and collaboration [42-44,47]. Lanzotti et al [43] considered incorporating a shared scoreboard to introduce competition as a motivational element. Pietrafesa et al [44] added a competitive component by having 2 groups of students compete against each other, though the study does not provide detailed information on the conditions of this competition. Greuter and Tepe [42] and Kuiper et al [47] organized weekly group gaming sessions to facilitate testing, but this collaborative aspect was not further examined in the study.

Another critical aspect of gameplay complexity is the extent to which games integrate dynamic feedback and progression systems to reinforce learning objectives. Stranick and Lopez [52] incorporate elements such as achievements and ranking in their game, which are clear feedback and progression mechanisms for the player. Sisto et al [49] integrate task completion time and a posture risk assessment score as core gameplay elements. These scores are calculated using an algorithm that factors in exposure time and joint angles within the simulated environment. At the end of each game level, players can view their risk scores for awkward postures alongside their time performance scores, enabling them to refine their strategies for task execution within the game. In contrast, the game by Kuipers et al [47] is notably more sophisticated, incorporating elements of customization, scoring, and progression. Players are encouraged to explore the in-game world by visiting specific locations that trigger mini-games, which allow them to personalize the virtual environment. The primary objective of this game is to induce behavioral changes, motivating players to experiment with new behaviors while

XSL•FO
RenderX

confronting challenging tasks and objectives. Throughout the gameplay, players receive precise and tailored feedback to ensure compliance with safety and in-game rules. Both the games developed by Kuipers et al [47] and Sisto et al [49] incorporate in-game experiences and dashboards with data visualization to support players in understanding and interpreting the information presented during gameplay.

Interestingly, despite the emphasis on interactivity, none of the studies, except for Greuter and Tepe [42], report considering audio elements and music in the game design which could influence player immersion and experience. In their study, participants provided negative feedback about the sound environment, which the authors hypothesized was due to all 20 games being played simultaneously in the same room, creating an overwhelming auditory experience. The authors suggest that the use of headphones might have mitigated this issue, although it could have also impeded player collaboration. Table 1 summarizes the technologies, gameplay, and expected outcomes used in the reviewed studies.

XSL•FO
RenderX

Table 1. List of studies according to platforms and control devices, gameplay, and expected outcomes of the game.

	Platforms and control devices	Gameplay	Expected outcomes
Jansen-Kosterink et al [38], 2013	Motion capture system (via suit, sensors, and infrared cameras)	Puzzle game and exergame: (1) control an avatar's movement on an obstacle course by walking pace, (2) control avatar's climb progress by raising arms, and (3) memorize and replicate a neck stretching movement sequence	Physically train patients with chronic neck and back pain through gameplay
Idriss et al [39], 2017	Microsoft Kinect	Exergame: (1) kick a ball after positioning correctly and (2) move an object between two tables	Make rehabilitation more engaging and fun for patients with MSDs[a]
France and Thomas [40], 2018	VR[b] and motion capture system (Vive Tracker)	Puzzle game and exergame: (1) memorize a sequence of icons and replicate it by selecting icons through gestures and movements, (2) catch fish with a certain gesture and avoid shark attacks by ducking, and (3) dodge or catch balls thrown by characters (dodgeball)	Reduce fear and perception of pain and increase physical activity in patients with chronic back pain
Husna et al [41], 2025	PC[c] and webcam	Exergame: making hand gestures to get a character out of a maze	Accompanying the rehabilitation work of people with MSD
Greuter and Tepe [42], 2013	iPad and PC (mouse)	Management game: identify hidden risks in a scene and propose corrective measures via an icon bar to allow NPCs[d] to work on building a tower	Engage learners in OHS[e] training and teach danger management procedures
Lanzotti et al [43], 2019	PC game (keyboard and mouse)	Simulation game: solve an action sequence by selecting the right tool and PPE[f] and select the necessary actions to control a control panel	Train industrial operators in safety measures and practices for specific, personalized, and hazardous activities
Pietrafesa et al [44], 2021	Web-based game accessible on PC, tablet, and mobile	Management game: manage a healthy work environment by organizing work tasks and protecting characters via PPE	Raise awareness among high school students in vocational schools about health and safety at work
Rapp et al [45], 2019	PC	Management game: a multiplayer simulation game where players must manage a restaurant as executive managers	Provide understanding of occupational health and promote employee culture among the participants
Rebelo and Filgueiras [46], 2012	*Not documented*	*Not documented*	Raising young people's awareness of occupational health
Kuipers et al [47], 2016	Microsoft Kinect	Exergame: (1) catch sheep based on their proximity to the player and (2) maintain an orchard by harvesting fruits and watering trees	Instruct health care staff on lifting and transfer techniques
Rodrigues et al [48], 2018	Kinect, haptic controllers, and PC game	Exergame: (1) adapt the workspace by interacting with environmental elements to recreate a healthy environment for characters, (2) complete stretching exercises by moving puzzle pieces, and (3) replicate and hold the posture of a silhouette displayed on screen	Raise awareness of proper office work postures and stretching exercises
Sisto et al [49], 2018	VR and motion capture system (Kinect and Leap Motion)	Puzzle and simulation game: place gears in a specific arrangement (puzzle)	Train operators to find and practice the best postures and movements in their work activities
Intipanya et al [50], 2025	Smartphone with camera	Exergame: perform head stretching movements to control aircraft	Preventing cervical pain
Valdivia et al [51], 2017	Kinect and IMU[g]	Exergame: perform bending and extension movements to “paint” a room	Promote stretching exercises for office workers
Stranick and Lopez [52], 2022	Oculus VR and Kinect	Exergame: successively reproduce postures to cross obstacles	Promote stretching exercises for workers in industry

[a]MSDs: musculoskeletal disorders.

[b]VR: virtual reality.

[c]PC: personal computer.

[d]NPCs: nonplayer characters.

[e]OHS: occupational health and safety.

[f]PPE: personal protective equipment.

[g]IMU: inertial measurement unit.

XSL•FO
RenderX

Expected Outcomes

Defining the expected outcomes and learning objectives of SGs is crucial, as they specify the skills, knowledge, and competencies that developers aim for users to acquire. For SGs designed for the rehabilitation of patients with work-related MSDs [38-41], the scenarios are structured around movements and postures critical for recovery or training. The expected outcome is to engage patients in performing rehabilitation exercises through play scenarios and to help them overcome fear of injury through interactive participation. An example of a rehabilitation-focused SG is the VR-based intervention developed by France and Thomas [40], designed to gradually improve lumbar flexion. While the game may not directly influence psychological factors for behavior change, it serves as a distraction from pain, reinforcing movement by increasing expectations of lumbar flexion. This approach helps individuals cope with movement-related fear, promoting the understanding that lumbar flexion is not inherently harmful, particularly for those with chronic low back pain and pain-related fear.

The learning outcomes of the games addressing occupational health and safety [42-45] revolve around raising awareness and facilitating learning. These studies typically focus on the risks of work-related accidents, particularly safety procedures and equipment. For instance, the game developed by Greuter and Tepe [42] aligns with occupational health and safety training objectives, aiming to educate users on hazard identification, control measures, communication processes, and reporting procedures. However, the study did not assess whether playing the game improves students' ability to identify occupational hazards. Moreover, Rapp et al [45] expanded this approach by integrating broader workplace learning objectives, aiming to enhance leadership, communication, and safety culture. For example, they aim to enhance knowledge by involving employees in operational decision-making, develop skills by allowing players to explore various actions with feedback, and promote awareness and attitude shifts toward workplace health and safety by emphasizing its importance for both employee well-being and organizational success. However, the authors did not provide any results demonstrating the achievement of these objectives through their game.

Finally, a distinct category of SGs focuses on MSD prevention by training professionals to adopt proper postures and ergonomic behaviors [46-52]. Rodrigues et al [48] developed 3 complementary mini-games: one is designed to raise awareness about the risks associated with static desk postures, while the other two encourage players to engage in regular stretching exercises. Kuipers et al [47] and Sisto et al [49] designed games aimed at promoting behavioral changes, providing tailored feedback to support the effective adoption of new ergonomic practices. Within this category, three studies [50-52] specifically used exergames to promote stretching and movement exercises. Intipanya et al [50] created a smartphone-based exergame where head movements control an aircraft, targeting cervical pain prevention. Valdivia et al [51] used Kinect and IMU sensors to prompt office workers to perform bending and extension movements through a room painting task. Stranick and Lopez [52] designed a VR-based exergame where players reproduce postures to overcome obstacles, promoting stretching exercises among industrial workers. The exergames reviewed primarily focus on the correct execution of physical exercises, with short-term learning objectives centered on performance. Players' cognitive involvement remains limited: few games encourage understanding the purpose of movements, reflecting on posture, or transferring learning beyond the game context. Moreover, game mechanics often emphasize repetition or movement accuracy without providing elements that help players understand their relevance or their connection to MSD prevention. The game included in Kuipers et al [47], however, goes further by promoting behavioral change among health care workers, particularly in patient handling practices.

Evaluation Methods

According to Gris and Bengston [60], the assessment of game-based learning effectiveness revolves around the key components including engagement, user experience, usability, and educational impact. The evaluation uses a combination of quantitative and qualitative measures for assessing educational impact, with qualitative methods being more commonly used for evaluating engagement, user experience, and usability [60]. In medical literature, another prevalent approach is to assess a game's ability to facilitate skill transfer [61]. The following section examines how different included studies evaluated various components of game effectiveness.

Engagement, Usability, and User Experience

The assessment of engagement, usability, and user experience is evident in 10 of the studies reviewed [38,39,41-43,46-48,51,52], although these evaluations exhibit variations in the questionnaires, variables, and methodologies used. Among these studies, some focused more specifically on user experience, using distinct evaluation methods and tools. Greuter and Tepe [42] evaluated participants' user experience by a questionnaire across various dimensions, including fun, engagement, success, control, motivation, feedback, usability, and difficulty [62]. Additionally, semistructured interviews were conducted following the test sessions to gather further insights. While the authors reported satisfactory levels of motivation, engagement, and learning among participants, with most students indicating increased motivation as they progressed through the game, several limitations in the study's design should be acknowledged. These include a small sample size, the subjective nature of the evaluation (which relied solely on self-reported questionnaires from a single group), and the absence of a longitudinal follow-up period.

Similarly, Jansen-Kosenterink et al [38] adopted a multifaceted approach to evaluate user experience, incorporating standardized and customized assessment tools. In their study, user experience was evaluated using multiple metrics: usability was measured with the System Usability Scale (SUS) [63], satisfaction was gauged through 2 specific questions, motivation was assessed using 2 additional questions, and gaming experience was evaluated with the Core Elements of Gaming Experience Questionnaire [64]. All 10 participants in the study reported a positive user experience, indicating that the game effectively distracted them from pain and enhanced their focus on gameplay. This study addressed the longitudinal aspects by tracking participants over a 4-week period of gameplay.

XSL•FO
RenderX

However, the study's limitations include the absence of a control group and a relatively small sample size. Similarly, Husna et al [41] used the SUS and the User Experience Questionnaire to assess their game, although the choice of these tools is questionable given the simplicity of the prototype, in which a character is guided through a maze using hand gestures captured via video. The minimal graphics and interactive complexity of the system raise concerns about the actual relevance of evaluating usability and user experience in this context.

Some studies focused more on usability assessment, with Lanzotti et al [43] evaluating their SG based on effectiveness, efficiency, and satisfaction. Effectiveness and efficiency were measured using quantitative metrics, including interaction errors, incorrect movements, gameplay duration, and task completion rates, while satisfaction was assessed through a subjective questionnaire. The usability index was calculated as the average of these 3 factors. The study involved 2 groups of 6 operators: the case group, which used an SG for training on mould press safety procedures, and the control group, which completed a written test with a similar structure. Although the test group reported higher satisfaction, their overall usability index was lower, particularly in correctly sequencing commands for safe machine operation. Interpretation of the results was challenging due to the small sample size and the lack of information on participants' baseline knowledge levels. The authors suggested that the written test may have assessed theoretical knowledge, whereas the SG required practical application, potentially leading to discrepancies. Additionally, no weighting was applied to usability factors in calculating the index, and the questionnaire lacked details on reliability and validity. Furthermore, the study did not specify how variables such as error frequency were quantified, particularly for the control group, limiting the clarity of the findings.

In addition to evaluating usability, some studies also explored user preferences and subjective satisfaction. Rodrigues et al [48] assessed user preferences among 20 participants, divided into 2 age groups, focusing on 5 key aspects: game relevance, ease of use, effectiveness of stretching exercises, perceived health (before and after the session), and overall satisfaction. Participants rated each aspect on a 3-point scale (low, medium, and very good). Most participants rated the game as “very good” or “medium” across the 5 categories. However, fewer than 40% rated the effectiveness of stretching exercises as “very good,” and only about half gave the highest rating for overall satisfaction. This suggests that while participants found the game engaging, its perceived effectiveness in promoting stretching exercises was more variable. A key limitation of this study is the lack of details on the validity, reliability, and development of the assessment questionnaire. The absence of such information raises concerns about the robustness of the evaluation method, as it is unclear whether the tool effectively captured user perceptions and game impact.

Other studies further investigated usability and acceptability through comparative analysis between healthy participants and those with MSD symptoms. In that respect, Idriss et al [39] conducted an evaluation of the usability and acceptability of a serious game that featured 2 distinct scenarios: football and object manipulation. Feedback was collected through a questionnaire. The study also involved a comparison of game scores between 2 groups: healthy participants and those with MSD symptoms. The results indicated that healthy participants achieved significantly higher scores than those with MSD symptoms. Participants were asked to rate various parameters, including the clarity of objectives, level of difficulty, and the attractiveness of the game environment, using a 5-point Likert scale. Both groups found the games to be motivational, attractive, and challenging. However, the study did not include a comparative analysis of the usability results between the 2 groups. Additionally, the questionnaire used for evaluation had not been validated for reliability or validity, raising concerns about whether it adequately captured all relevant aspects of usability, engagement, and playability in the games.

Finally, two studies focused on the assessment of engagement. Valdivia et al [51] mainly assessed exergame’s engagement skills through the Game Engagement Questionnaire, which measures dimensions such as immersion, focus, and sense of control. This evaluation followed two brief 1-minute exercise sessions in which participants performed trunk flexion and extension movements using 2 versions of the exergame—one based on the Kinect V2 sensor and the other using an IMU. Despite the use of a standardized questionnaire, such short-duration exposure with a small subject raises questions about the ability of the sessions to foster meaningful engagement or allow users to fully grasp the potential of the system. While most studies relied on self-reported measures, Kuipers et al [47] introduced a different approach by quantifying engagement through voluntary re-engagement metrics. They used a quantitative measure for evaluating player engagement, where the number of game sessions voluntarily conducted by players over a given period served as an indicator of their engagement. This approach is analogous to the voluntary reengagement observed in Sanchez et al [65], where players' self-initiated participation in the task was used as a measure of motivation. During the experimentation phase, participants were given free time in which they could choose to replay the game at their discretion. The decision of a participant to replay the game independently was interpreted as a sign of motivation. However, Sanchez et al [65] emphasized that this measure is more closely associated with intrinsic motivation rather than engagement. This distinction between the two concepts should be considered in the interpretation of results.

In summary, in line with Gris and Bengston [60], the assessment of engagement, usability, and user experience in SGs predominantly relies on qualitative measures. Moreover, the evaluation tools used in these studies are generally nonstandardized. Among the studies reviewed that assess these concepts, only three [38,41,51] used a standardized questionnaire. Furthermore, there are notable discrepancies in the concepts and variables measured across different studies. The concept of usability, for instance, is variably defined depending on the study. For example, in the study of Lanzotti et al [43], usability encompasses a range of quantitative and qualitative measures, focusing on the dimensions of effectiveness, efficiency, and satisfaction.

XSL•FO
RenderX

Educational Impact of Serious Games

Assessing the educational impact of SGs requires robust evaluation methodologies to determine their effectiveness in training and behavior change. However, many studies fail to implement comprehensive assessment processes, relying instead on limited or inconsistent evaluation methods. Among the studies included in the analysis, many lack a comprehensive process for evaluating the impact of developed SGs on training or achieving targeted outcomes. Three reviewed studies [44,45,49] do not provide any evaluation of the game's impact on learning, while others rely solely on postgame questionnaires without including a control group or had significant limitations. For instance, Lanzotti et al [43] assessed the efficacy of game-based learning by comparing in-game metrics with responses from a control group to a multiple-choice questionnaire addressing the same simulated process. However, the level of training received by the control group was not clearly defined, and the results indicated that the control group outperformed the experimental group, suggesting the serious game's effectiveness was limited. Similarly, Greuter and Tepe [42] evaluated the game's effectiveness through a written questionnaire on covered concepts. Since the game aimed to enhance hazard identification skills, effectiveness was measured through a written test requiring participants to identify risks from photographs. Educational impacts were assessed by comparing pretest and posttest questionnaire results of the experimental group. Although the authors mentioned the presence of a control group, results between the two groups were not provided, and the overall study design lacked robustness due to the use of a validated questionnaire and postsession data only.

However, 3 studies implemented game evaluations through prolonged test sessions involving multiple groups and clear, comparable measurement variables. Kuipers et al [47] measured the evolution of participants' game scores based on the number and duration of sessions over a period of 3 weeks. They observed that the more players played, the higher the scores increased. The increase in scores thus reflects the effectiveness of the game: players autonomously exhibit the desired behavior. The authors posited that performing in-game movements mirrors actual workplace lifting techniques, thereby increasing game scores leading to training outcomes aligning with desired behavioral changes in lifting techniques. Nevertheless, further validation is required through longitudinal studies assessing behavioral changes in real workplace settings.

Similarly, Jansen-Kosterink et al [38] evaluated learning outcomes by analyzing motor skill improvements and clinical indicators before and after gameplay sessions. The game was administered to a cohort of patients afflicted with chronic neck and back pain, who engaged with the game over a 4-week duration. Among the parameters measured, including walking velocity, reach height, cervical range of motion, pain intensity, and pain disability, only cervical range of motion exhibited a significant increase over the testing period. Although the authors asserted that patients generally exhibited faster walking speeds, increased reach, and enhanced neck mobility after 4 weeks of gameplay, these results did not reach statistical significance, likely due to the limited sample size and the short duration of the testing period.

France and Thomas [40] conducted a 9-week clinical trial to evaluate their VR-based rehabilitation game for pain management and functional recovery. They assessed patient-reported disability, pain levels, spinal condition, and pain expectations, with follow-up monitoring over 48 weeks. Results showed that the experimental group experienced greater pain and disability reduction, along with decreased pain expectancy and improved lumbar flexion compared to the control group. These findings suggest that the VR game may help individuals with low back pain and movement-related fear engage in repeated sessions, promoting graded spinal motion.

Discussion

Challenges and Perspectives for Serious Games in MSD Prevention

The objective of this review was to identify and analyze the key characteristics of SGs developed for work-related MSD prevention and mitigation while exploring their potential impact. However, the lack of empirical evidence and methodological inconsistencies across studies make it challenging to conclusively determine the effectiveness of SGs in reducing MSDs. One major challenge is the inconsistency of experimental protocols, particularly due to variability in measurement methods across studies. While certain studies, particularly in the medical field, rely on detailed and robust experimental protocols (randomized controlled trials), others often suffer from methodological limitations such as small sample sizes and weak study designs lacking control groups. This is in line with the findings of Steiner et al [66] in their scoping review focusing on gamification in shoulder rehabilitation, highlighting significant methodological limitations across studies, including small sample sizes, a lack of control groups, and a lack of robust experimental designs. Such limitations are also identified in previous studies [67,68] and appear to be a key factor contributing to the inconsistency of results we documented across SG studies. Ideally, a comparison of 3 conditions—training or learning sessions, gaming sessions, and combined training or learning and gaming sessions—would have provided valuable insights, yet such comparisons were often unfeasible in the studies included.

Beyond experimental limitations, the diversity of research objectives and methodological frameworks further complicates the synthesis of convergent knowledge related to SGs developed for occupational health objectives. Understanding the educational impact of these games necessitates a thorough contextualization of the existing games [69]. The variability in the characteristics of SGs—including aspects such as design, gameplay, pedagogical objectives, target population, and context of use—further complicates this endeavor. A review of SGs in health professions education [70] found similar limitations and challenges in drawing conclusions about SGs' efficacy.

Additionally, the complexity of evaluating SGs is exacerbated by the difficulty in quantifying learning outcomes and behavioral changes over time, particularly outside the realm of exergames,

XSL•FO
RenderX

where quantitative and physiological data cannot be used for measurement. A systematic review focusing on digital SGs in nursing education highlighted the scarcity of data on sustained behavioral outcomes and the need for more rigorous and longitudinal research designs [71]. Another review highlighted the importance of integrating behavioral change theories into game design to effectively measure and achieve desired health outcomes [72].

Moreover, none of the studies reviewed have considered users' level of experience with video games or their gaming culture. Previous studies on behavior change in persuasive health games found that participants' engagement and outcomes were significantly affected by their familiarity with gaming and concluded that users' gaming backgrounds should be considered to tailor games [73]. Silva et al [74] highlighted instances where participants easily navigated game challenges due to their regular engagement with video games.

Another challenge lies in the simplistic design of many SGs, which often fail to balance pedagogical objectives with engaging and immersive gameplay elements. As seen in studies such as Pietrafesa et al [44], game scenarios do not replicate pedagogical objectives effectively. As a result, these games often lack the critical ludic quality essential for maintaining player engagement [75]. A systematic review by Ambros-Antemate et al [76] showed the absence of iterative design and testing phases resulting in games that may not effectively meet user needs or learning objectives. Garske et al [77] examined SGs for prosthetic training and identified similar design shortcomings—such as the lack of user-centered design, oversimplified mechanics, and misalignment with learning objectives—that align with the results of our study. The success of a serious game depends on effectively balancing ludic elements with characterizing goals [78]. When this balance is disrupted, players may rapidly lose interest and motivation. Consequently, even SGs that are content-rich and technically sound may fail to achieve their intended outcomes if the ludic aspects are not sufficiently developed to maintain long-term user engagement [79]. Previous research has shown that the degree to which users find a game to be enjoyable is related to their motivation levels, leading to better engagement and learning outcomes [80].

To enhance educational impact, several researchers [45,68,81] argue for the necessity of debriefing in SGs, highlighting its potential to enhance understanding of game content, clarify misconceptions, and internalize important lessons. This reflective process not only facilitates immediate learning but also contributes to long-term retention of knowledge and skills acquired during gameplay. For example, although not directly addressing MSDs, Michelet et al [82] found that participants who engaged in a structured debriefing session postgameplay demonstrated better knowledge retention and skill acquisition compared to those who did not. However, debriefing remains largely underexplored in the reviewed SGs on MSD prevention.

Moreover, the emphasis of existing research remains primarily on rehabilitation and occupational safety, with limited focus on the direct prevention of MSDs. A limited number of studies [47-52] specifically address the topic of MSD prevention, which primarily focused on physical ergonomics or manual material handling. These results underscore the critical need for further research to explore additional facets of MSD prevention, such as organizational and psychosocial factors, within the design framework of SGs.

Despite the variability in objectives, design methodologies, technologies, and contexts, which complicates the synthesis of results, this literature review underscores several key challenges that future research on the design and evaluation of SGs for the prevention of MSDs must address. A critical area for improvement is the detailed documentation of user involvement in the design process, as this engagement is vital for promoting both the acceptance and effectiveness of these games. The debriefing phases and evaluations of engagement and usability must be rigorously formalized to ensure more consistent and reliable outcomes. Future studies must overcome existing methodological limitations by developing standardized evaluations to ensure the empirical validation of serious game interventions. It is also essential to broaden the scope of research beyond rehabilitation and safety to encompass the prevention of MSDs. This expansion should include a focus on the organizational and psychosocial dimensions of MSDs, using SGs to develop more comprehensive and effective interventions.

Therefore, we recommend adopting design methodologies that integrate multifactorial MSD prevention models, while ensuring that technological implementations remain practical and adaptable to real-world industrial settings. Strengthening evaluation frameworks, particularly in measuring long-term effectiveness, is also crucial for establishing SGs as viable intervention tools in occupational health.

Limitations

This study has several limitations. First, the final selection includes a relatively small number of studies, primarily due to the specialized nature of the research topic and the selection methodology, which inherently constrained the available literature. Second, grey literature, including unpublished studies and industry-based reports, was excluded, potentially omitting relevant work on SGs for MSD prevention.

Furthermore, to maintain a balance between comprehensiveness and relevance, search queries were adapted to the specific constraints of each database. In some cases, broad terms such as “game” and “occupational” were limited to titles or combined with more specific qualifiers (eg, “health” and “disease”) to reduce the retrieval of irrelevant results. While necessary, these refinements may have inadvertently restricted the scope of the search.

To mitigate these limitations, 2 independent authors (TR and MZ) conducted the search in several stages of study, and in addition to querying major scientific databases, we also searched Google Scholar and ResearchRabbit to ensure the inclusion of all relevant studies.

Conclusions

This review underscores that SGs for MSD prevention developed through collaborative and iterative design processes, particularly those involving end users and domain experts, tend

XSL•FO
RenderX

to exhibit greater pedagogical and functional robustness, whereas limited stakeholder involvement often results in narrower and less effective outcomes. The technological landscape of these SGs is diverse, ranging from basic PCs to advanced VR and motion-capture systems; however, most studies do not provide clear justifications for their technology choices. Learning objectives are frequently limited to task execution, with minimal emphasis on reflective learning or transfer beyond the game. Moreover, the expected outcomes predominantly address physical ergonomics, often neglecting psychosocial and organizational contributors to MSD risk. The evaluation of SG effectiveness remains inconsistent, with many studies lacking methodological rigor or long-term assessment, and while some report gains in knowledge or behavior, the evidence base remains limited and fragmented, especially regarding sustained health improvements.

Acknowledgments

The authors are grateful to the University of Technology of Belfort and Montbéliard for the provision of computing resources and data resources. The authors would also like to thank the ELLIADD laboratory and, more specifically, the center of research ERCOS, for their support and interest in this project. Generative AI (ChatGPT) was used to assist with text refinement and language clarity. However, all intellectual contributions, critical analyses, and interpretations were conducted by the authors. The final manuscript was thoroughly reviewed and edited by the authors to ensure accuracy and originality.

Authors' Contributions

TR conceptualized the study, managed data curation, conducted formal analysis, developed the methodology, wrote a part of the original draft, and reviewed the draft. TH conceptualized the study, wrote a part of the original draft, and reviewed the draft. FS wrote a part of the original draft and reviewed the draft. MZ conceptualized the study, managed data curation, conducted formal analysis, supervised the research group, wrote a part of the original draft, and reviewed the draft.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA-ScR checklist.
[PDF File (Adobe PDF File), 551 KB - games_v13i1e66913_app1.pdf]

Multimedia Appendix 2
Search queries.
[DOCX File , 23 KB - games_v13i1e66913_app2.docx]

Multimedia Appendix 3
Detailed screening.
[XLSX File (Microsoft Excel File), 57 KB - games_v13i1e66913_app3.xlsx]

Multimedia Appendix 4
Studies classification.
[DOCX File , 37 KB - games_v13i1e66913_app4.docx]

References

1. de Kok J, Vroonhof P, Snijders J, Roullis G, Clarke M, Peereboom K, et al. Work-related MSDs: prevalence, costs and demographics in the EU. European Agency for Safety and Health at Work. 2019. URL: https://osha.europa.eu/en/publications/summary-msds-facts-and-figures-overview-prevalence-costs-and-demographics-msds-europe [accessed 2025-09-20]
2. Roveshti MM, Pouya AB, Pirposhteh EA, Khedri B, Khajehnasiri F, Poursadeqiyan M. Work-related musculoskeletal disorders and related risk factors among bakers: a systematic review. Work 2024;77(2):463-476. [doi: 10.3233/WOR-220165] [Medline: 37718825]
3. Yarmohammadi H, Niksima SH, Yarmohammadi S, Khammar A, Marioryad H, Poursadeqiyan M. Evaluating the prevalence of musculoskeletal disorders in drivers systematic review and meta-analysis. J Health Saf Work 2019;9(3):221-230.
4. Rapport annuel 2018 de l'Assurance Maladie - Risques professionnels. l'Assurance Maladie. 2019 Dec 1. URL: https://www.assurance-maladie.ameli.fr/etudes-et-donnees/2018-rapport-annuel-assurance-maladie-risques-professionnels [accessed 2025-02-08]

XSL•FO
RenderX

5. Roquelaure Y. Troubles musculo-squelettiques et facteurs psychosociaux au travail. European Trade Union Institute. 2018. URL: https://www.etui.org/fr/publications/rapports/troubles-musculo-squelettiques-et-facteurs-psychosociaux-au-travail [accessed 2022-02-09]
6. Hassanzadeh P, Ghahramani A, Mohebbi I. An assessment of association between macro-ergonomics status and employees' prevalence of musculoskeletal disorders and job stress outcomes in Urmia educational and medical hospitals. Iran J Ergon 2020;7(4):40-51. [doi: 10.30699/jergon.7.4.40]
7. Nambiema A, Bodin J, Stock S, Aublet-Cuvelier A, Descatha A, Evanoff B, et al. Proportion and number of upper-extremity musculoskeletal disorders attributable to the combined effect of biomechanical and psychosocial risk factors in a working population. Int J Environ Res Public Health 2021;18(8):3858 [FREE Full text] [doi: 10.3390/ijerph18083858] [Medline: 33916971]
8. van de Wijdeven B, Visser B, Daams J, Kuijer PP. A first step towards a framework for interventions for individual working practice to prevent work-related musculoskeletal disorders: a scoping review. BMC Musculoskelet Disord 2023;24(1):87 [FREE Full text] [doi: 10.1186/s12891-023-06155-w] [Medline: 36726094]
9. Winkel J, Westgaard RH. Development and implementation of interventions managing work-related musculoskeletal disorders: inadequacy of prevalent research framework and future opportunities. Scand J Work Environ Health 2019;45(3):316-317 [FREE Full text] [doi: 10.5271/sjweh.3815] [Medline: 31041996]
10. Hemati K, Darbandi Z, Kabir-Mokamelkhah E, Poursadeghiyan M, Ghasemi MS, Mohseni-Ezhiye M, et al. Ergonomic intervention to reduce musculoskeletal disorders among flour factory workers. Work 2020;67(3):611-618. [doi: 10.3233/WOR-203275] [Medline: 32986645]
11. Rodrigues Ferreira Faisting AL, de Oliveira Sato T. Effectiveness of ergonomic training to reduce physical demands and musculoskeletal symptoms: an overview of systematic reviews. Int J Ind Ergon 2019;74(7):102845. [doi: 10.1016/j.ergon.2019.102845]
12. Yazdani A, Wells R. Barriers for implementation of successful change to prevent musculoskeletal disorders and how to systematically address them. Appl Ergon 2018;73:122-140 [FREE Full text] [doi: 10.1016/j.apergo.2018.05.004] [Medline: 30098627]
13. Van Eerd D, Munhall C, Irvin E, Rempel D, Brewer S, van der Beek AJ, et al. Effectiveness of workplace interventions in the prevention of upper extremity musculoskeletal disorders and symptoms: an update of the evidence. Occup Environ Med 2016;73(1):62-70 [FREE Full text] [doi: 10.1136/oemed-2015-102992] [Medline: 26552695]
14. Hogan DA, Greiner BA, O'Sullivan L. The effect of manual handling training on achieving training transfer, employee's behaviour change and subsequent reduction of work-related musculoskeletal disorders: a systematic review. Ergonomics 2014;57(1):93-107. [doi: 10.1080/00140139.2013.862307] [Medline: 24387742]
15. Yu W, Yu ITS, Wang X, Li Z, Wan S, Qiu H, et al. Effectiveness of participatory training for prevention of musculoskeletal disorders: a randomized controlled trial. Int Arch Occup Environ Health 2013;86(4):431-440. [doi: 10.1007/s00420-012-0775-3] [Medline: 22544420]
16. Esmaeili R, Shakerian M, Esmaeili SV, Jalali M, Pouya AB, Karimi A. A multicomponent quasi-experimental ergonomic interventional study: long-term parallel four-groups interventions. BMC Musculoskelet Disord 2023;24(1):107 [FREE Full text] [doi: 10.1186/s12891-023-06220-4] [Medline: 36759793]
17. Daltroy LH, Iversen MD, Larson MG, Lew R, Wright E, Ryan J, et al. A controlled trial of an educational program to prevent low back injuries. N Engl J Med 1997;337(5):322-328. [doi: 10.1056/NEJM199707313370507] [Medline: 9233870]
18. Cheong YG, Khaled R, Grappiolo C, Campos J, Martinho C, Ingram G, et al. A computational approach towards conflict resolution for serious games. In: Association for Computing Machinery. New York, NY: Presented at: 6th International Conference on Foundations of Digital Games, FDG 2011; June 29 - July 1 2011, Bordeaux, France; 2011. [doi: 10.1145/2159365.2159368]
19. Din SU, Baig MZ, Khan MK. Serious games: an updated systematic literature review. ArXiv Preprint posted online on May 31, 2023. [doi: 10.48550/arXiv.2306.03098]
20. Dörner R, Göbel S, Effelsberg W, Wiemeyer J, editors. Serious Games: Foundations, Concepts and Practice. Cham, Switzerland: Springer International Publishing; 2016.
21. Alvarez J, Djaouti D. Introduction Au Serious Game - Serious Game: An Introduction. 1ère édition. Paris: Questions Théoriques; 2010.
22. Xu Y, Lau Y, Cheng LJ, Lau ST. Learning experiences of game-based educational intervention in nursing students: a systematic mixed-studies review. Nurse Educ Today 2021;107:105139. [doi: 10.1016/j.nedt.2021.105139] [Medline: 34563963]
23. Hamari J, Shernoff DJ, Rowe E, Coller B, Asbell-Clarke J, Edwards T. Challenging games help students learn: an empirical study on engagement, flow and immersion in game-based learning. Comput Hum Behav 2016;54:170-179. [doi: 10.1016/j.chb.2015.07.045]
24. Larson K. Serious games and gamification in the corporate training environment: a literature review. TechTrends 2019;64(2):319-328. [doi: 10.1007/s11528-019-00446-7]
25. Kwon J, Lee Y. Serious games for the job training of persons with developmental disabilities. Comput Educ 2016;95:328-339. [doi: 10.1016/j.compedu.2016.02.001]

XSL•FO
RenderX

26. Vigoroso L, Caffaro F, Micheletti Cremasco M, Cavallo E. Innovating occupational safety training: a scoping review on digital games and possible applications in agriculture. Int J Environ Res Public Health 2021;18(4):1868 [FREE Full text] [doi: 10.3390/ijerph18041868] [Medline: 33671867]
27. Martínez-Miranda J, Espinosa-Curiel IE. Serious games supporting the prevention and treatment of alcohol and drug consumption in youth: scoping review. JMIR Serious Games 2022;10(3):e39086 [FREE Full text] [doi: 10.2196/39086] [Medline: 36006694]
28. Sajjadi P, Ewais A, De Troyer O. Individualization in serious games: a systematic review of the literature on the aspects of the players to adapt to. Entertain Comput 2022;41:100468. [doi: 10.1016/j.entcom.2021.100468]
29. Gorbanev I, Agudelo-Londoño S, González RA, Cortes A, Pomares A, Delgadillo V, et al. A systematic review of serious games in medical education: quality of evidence and pedagogical strategy. Med Educ Online 2018;23(1):1438718 [FREE Full text] [doi: 10.1080/10872981.2018.1438718] [Medline: 29457760]
30. Lavigne M. Les faiblesses ludiques et pédagogiques des serious games. In: Nume´rique et e´ducation: dispositifs, jeux, enjeux, hors jeux. France: Universitaires de Lorraine; 2016:71-82.
31. Sailer M, Homner L. The gamification of learning: a meta-analysis. Educ Psychol Rev 2019;32(1):77-112. [doi: 10.1007/s10648-019-09498-w]
32. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol Theory Pract 2005;8(1):19-32. [doi: 10.1080/1364557032000119616]
33. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci 2010;5:69 [FREE Full text] [doi: 10.1186/1748-5908-5-69] [Medline: 20854677]
34. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Med 2009;6(7):e1000100 [FREE Full text] [doi: 10.1371/journal.pmed.1000100] [Medline: 19621070]
35. Deterding S, Dixon D, Khaled R, Nacke L. From game design elements to gamefulness: defining "gamification". New York, NY: Association for Computing Machinery; 2011 Presented at: MindTrek '11: Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments; September 28-30, 2011; Tampere, Finland p. 9-15. [doi: 10.1145/2181037.2181040]
36. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan – a web and mobile app for systematic reviews. Syst Rev 2016;5(1):210 [FREE Full text] [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
37. Choi SD, Guo L, Kang D, Xiong S. Exergame technology and interactive interventions for elderly fall prevention: a systematic literature review. Appl Ergon 2017;65:570-581. [doi: 10.1016/j.apergo.2016.10.013] [Medline: 27825723]
38. Jansen-Kosterink SM, Huis In 't Veld RMHA, Schönauer C, Kaufmann H, Hermens HJ, Vollenbroek-Hutten MMR. A serious exergame for patients suffering from chronic musculoskeletal back and neck pain: a pilot study. Games Health J 2013;2(5):299-307 [FREE Full text] [doi: 10.1089/g4h.2013.0043] [Medline: 24761327]
39. Idriss M, Tannous H, Istrate D, Perrochon A, Salle J, Ho Ba Tho M, et al. Rehabilitation-oriented serious game development and evaluation guidelines for musculoskeletal disorders. JMIR Serious Games 2017;5(3):e14 [FREE Full text] [doi: 10.2196/games.7284] [Medline: 28676468]
40. France CR, Thomas JS. Virtual immersive gaming to optimize recovery (VIGOR) in low back pain: a phase II randomized controlled trial. Contemp Clin Trials 2018;69:83-91 [FREE Full text] [doi: 10.1016/j.cct.2018.05.001] [Medline: 29730393]
41. Husna R, Brata KC, Anggraini IT, Funabiki N, Rahmadani AA, Fan CP. An investigation of hand gestures for controlling video games in a rehabilitation exergame system. Computers 2025;14(1):25. [doi: 10.3390/computers14010025]
42. Greuter S, Tepe S. Tampre, Finland: Digital Games Research Association; 2013 Presented at: DiGRA 2013 Conference; August 26-29, 2013; Atlanta, GA. [doi: 10.26503/dl.v2013i1.680]
43. Lanzotti A, Vanacore A, Tarallo A, Nathan-Roberts D, Coccorese D, Minopoli V, et al. Interactive tools for safety 4.0: virtual ergonomics and serious games in real working contexts. Ergonomics 2019;63(3):324-333. [doi: 10.1080/00140139.2019.1683603] [Medline: 31648616]
44. Pietrafesa E, Bentivenga R, Lalli P, Capelli C, Farina G, Stabile S. Becoming safe: a serious game for occupational safety and health training in a WBL Italian experience. Cham, Switzerland: Springer; 2021 Presented at: 10th International Conference on Methodologies and Intelligent Systems for Technology Enhanced Learning; June 17-19, 2020; L'Aquila, Italy p. 264-271. [doi: 10.1007/978-3-030-52287-2_27]
45. Rapp J, Rose J, Narciss S, Kapp F. How to set the game characteristics, design the instructional content and the didactical setting for a serious game for health prevention in the workplace. Cham, Switzerland: Springer; 2019 Presented at: 7th International Conference on Games and Learning Alliance, GALA 2018; December 5-7, 2018; Palermo, Italy p. 210-220. [doi: 10.1007/978-3-030-11548-7_20]
46. Rebelo F, Filgueiras EV. Ergoshow: a user-centred design game to make children aware of ergonomics and occupational safety and health. Theor Issues Ergon Sci 2012;13(1):4-17. [doi: 10.1080/1463922x.2010.525761]
47. Kuipers DA, Wartena BO, Dijkstra BH, Terlouw G, van T Veer JTB, van Dijk HW, et al. iLift: a health behavior change support system for lifting and transfer techniques to prevent lower-back injuries in healthcare. Int J Med Inform 2016;96:11-23. [doi: 10.1016/j.ijmedinf.2015.12.006] [Medline: 26797571]

XSL•FO
RenderX

48. Rodrigues MAF, Serpa YR, Macedo DV, Sousa ES. A serious game to practice stretches and exercises for a correct and healthy posture. Entertain Comput 2018;28:78-88. [doi: 10.1016/j.entcom.2017.11.002]
49. Sisto M, Zare M, Ouerhani N, Bolinhas C, Divernois M, Mignot B, et al. Virtual reality serious game for musculoskeletal disorder prevention. Cham, Switzerland: Springer; 2018 Presented at: Augmented Reality, Virtual Reality, and Computer Graphics, AVR 2018; June 24-27, 2018; Otranto, Italy p. 43-59. [doi: 10.1007/978-3-319-95282-6_4]
50. Intipanya N, Sihawong R, Janwantanakul P. Effects of a smartphone game to facilitate active neck movements on the incidence of neck pain among office workers: a 6-month cluster-randomized controlled trial. Musculoskelet Sci Pract 2025;75:103243. [doi: 10.1016/j.msksp.2024.103243] [Medline: 39647258]
51. Valdivia S, Blanco R, Uribe A, Penuela L, Rojas D, Kapralos B. A spinal column exergame for occupational health purposes. Cham, Switzerland: Springer; 2017 Presented at: 6th International Conference on Games and Learning Alliance, GALA 2017; December 5-7, 2017; Lisbon, Portugal p. 83-92. [doi: 10.1007/978-3-319-71940-5_8]
52. Stranick T, Lopez C. Adaptive virtual reality exergame: promoting physical activity among workers. J Comput Inf Sci Eng 2022;22(3):031002. [doi: 10.1115/1.4053002]
53. Sanchez E, Monod-Ansaldi R. Recherche collaborative orientée par la conception. Educ Didact 2015;9(2):73-94. [doi: 10.4000/educationdidactique.2288]
54. Jaccard D, Suppan L, Bielser F. Contribution of the co.LAB framework to the collaborative design of serious games: mixed methods validation study. JMIR Serious Games 2021;9(4):e33144 [FREE Full text] [doi: 10.2196/33144] [Medline: 34822336]
55. Vermeulen M. Une approche meta-design des learning games pour développer leur usage [PhD thesis]. Paris, France: Sorbonne Université; 2018. URL: https://tel.archives-ouvertes.fr/tel-01871048 [accessed 2025-09-23]
56. Hulin T. Co-design et scénarisation interactive d'une expérience ludique. Revue de recherches francophones en sciences de l'information et de la communication 2021;10:hal-03185449 [FREE Full text]
57. Verschueren S, Buffel C, Vander Stichele G. Developing theory-driven, evidence-based serious games for health: framework based on research community insights. JMIR Serious Games 2019;7(2):e11565 [FREE Full text] [doi: 10.2196/11565] [Medline: 31045496]
58. DeSmet A, Van Cleemput K, Bastiaensens S, Poels K, Vandebosch H, Malliet S, et al. Bridging behavior science and gaming theory: using the intervention mapping protocol to design a serious game against cyberbullying. Comput Hum Behav 2016;56:337-351 [FREE Full text] [doi: 10.1016/j.chb.2015.11.039]
59. Alvarez J. Du jeu vidéo au serious game: approches culturelle, pragmatique et formelle [PhD Thesis]. Toulouse, France: Université Toulouse; 2007. URL: https://hal.archives-ouvertes.fr/tel-01240683 [accessed 2025-09-23]
60. Gris G, Bengtson C. Assessment measures in game-based learning research: a systematic review. Int J Serious Games 2021;8(1):3-26. [doi: 10.17083/ijsg.v8i1.383]
61. Kato PM. Evaluating efficacy and validating games for health. Games Health J 2012;1(1):74-76. [doi: 10.1089/g4h.2012.1017] [Medline: 26196436]
62. Ijsselsteijn W, De Kort Y, Poels K, Jurgelionis A, Bellotti F. Characterising and measuring user experiences in digital games. 2007 Presented at: International Conference on Advances in Computer Entertainment Technology, ACE 2007; June 13-15, 2007; Salzburg, Austria.
63. Brooke J. SUS: a quick and dirty usability scale. In: Jordan PW, Thomas B, Weerdmeester BA, McClelland IL, editors. Usability Evaluation in Industry. London: CRC Press; 1996:189-194.
64. Calvillo-Gámez EH, Cairns P, Cox AL. Assessing the core elements of the gaming experience. In: Bernhaupt R, editor. Game User Experience Evaluation. Human–Computer Interaction Series. Cham, Switzerland: Springer; 2015:37-62.
65. Sanchez DR, Nelson T, Kraiger K, Weiner E, Lu Y, Schnall J. Defining motivation in video game‐based training: exploring the differences between measures of motivation. Int J Train Dev 2021;26(1):1-28. [doi: 10.1111/ijtd.12233]
66. Steiner B, Elgert L, Saalfeld B, Wolf K. Gamification in rehabilitation of patients with musculoskeletal diseases of the shoulder: scoping review. JMIR Serious Games 2020;8(3):e19914 [FREE Full text] [doi: 10.2196/19914] [Medline: 32840488]
67. Girard C, Ecalle J, Magnan A. Serious games as new educational tools: how effective are they? A meta‐analysis of recent studies. J Comput Assist Learn 2012;29(3):207-219. [doi: 10.1111/j.1365-2729.2012.00489.x]
68. Sanchez É. Le Paradoxe du Marionnettiste - Jeu et Apprentissage. France: Octarès Édition; 2022.
69. Ricciardi F, De Paolis LT. A comprehensive review of serious games in health professions. Int J Comput Games Technol 2014;2014(108):787968. [doi: 10.1155/2014/787968]
70. Gentry SV, Gauthier A, L'Estrade Ehrstrom B, Wortley D, Lilienthal A, Tudor Car L, et al. Serious gaming and gamification education in health professions: systematic review. J Med Internet Res 2019;21(3):e12994 [FREE Full text] [doi: 10.2196/12994] [Medline: 30920375]
71. Lee M, Shin S, Lee M, Hong E. Educational outcomes of digital serious games in nursing education: a systematic review and meta-analysis of randomized controlled trials. BMC Med Educ 2024;24(1):1458 [FREE Full text] [doi: 10.1186/s12909-024-06464-1] [Medline: 39696231]
72. Damaševičius R, Maskeliūnas R, Blažauskas T. Serious games and gamification in healthcare: a meta-review. Information 2023;14(2):105. [doi: 10.3390/info14020105]

XSL•FO
RenderX

73. Gómez-León MI. Serious games to support emotional regulation strategies in educational intervention programs with children and adolescents: systematic review and meta-analysis. Heliyon 2025;11(4):e42712 [FREE Full text] [doi: 10.1016/j.heliyon.2025.e42712] [Medline: 40034297]
74. Silva JF, Almeida JE, Rossetti RJF, Coelho AL. A serious game for EVAcuation training. New York, NY: IEEE; 2013 Presented at: 2013 IEEE 2nd International Conference on Serious Games and Applications for Health (SeGAH); May 2-3, 2013; Vilamoura, Portugal. [doi: 10.1109/segah.2013.6665302]
75. Kocher M. Ludic didactics: for an inspired, motivating and playful education. Cham, Switzerland: Springer International Publishing; 2022 Presented at: 11th International Conference on Games and Learning Alliance, GALA 2022; November 30 - December 2, 2022; Tampere, Finland p. 193-201. [doi: 10.1007/978-3-031-22124-8_19]
76. Ambros-Antemate JF, Beristain-Colorado MDP, Vargas-Treviño M, Gutiérrez-Gutiérrez J, Hernández-Cruz PA, Gallegos-Velasco IB, et al. Software engineering frameworks used for serious games development in physical rehabilitation: systematic review. JMIR Serious Games 2021;9(4):e25831 [FREE Full text] [doi: 10.2196/25831] [Medline: 34762052]
77. Garske CA, Dyson M, Dupan S, Morgan G, Nazarpour K. Serious games are not serious enough for myoelectric prosthetics. JMIR Serious Games 2021;9(4):e28079 [FREE Full text] [doi: 10.2196/28079] [Medline: 34747715]
78. Caserman P, Hoffmann K, Müller P, Schaub M, Straßburg K, Wiemeyer J, et al. Quality criteria for serious games: serious part, game part, and balance. JMIR Serious Games 2020;8(3):e19037 [FREE Full text] [doi: 10.2196/19037] [Medline: 32706669]
79. Gillani F, Inayat I, Val de Carvalho C. Impact of serious game design factors and problem based pedagogy on learning outcome. 2022 Presented at: 2022 International Conference on Frontiers of Information Technology (FIT); December 12-13, 2022; Islamabad, Pakistan p. 314-319. [doi: 10.1109/fit57066.2022.00064]
80. Franzwa C, Tang Y, Johnson A, Bielefeldt T. Balancing fun and learning in a serious game design. Int J Game Based Learn 2014;4(4):37-57. [doi: 10.4018/ijgbl.2014100103]
81. Brom C, Preuss M, Klement D. Are educational computer micro-games engaging and effective for knowledge acquisition at high-schools? A quasi-experimental study. Comput Educ 2011;57(3):1971-1988. [doi: 10.1016/j.compedu.2011.04.007]
82. Michelet D, Barre J, Truchot J, Piot M, Cabon P, Tesniere A. Effect of computer debriefing on acquisition and retention of learning after screen-based simulation of neonatal resuscitation: randomized controlled trial. JMIR Serious Games 2020;8(3):e18633 [FREE Full text] [doi: 10.2196/18633] [Medline: 32780021]

Abbreviations

IMU: inertial measurement unit
MSD: musculoskeletal disorder
NPC: nonplayer character
PC: personal computer
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
SG: serious game
SUS: System Usability Scale
VR: virtual reality

Edited by A Coristine; submitted 30.09.24; peer-reviewed by M Poursadeqiyan, J Moreno-Chaparro; comments to author 07.11.24; revised version received 07.02.25; accepted 20.05.25; published 21.10.25.

Please cite as:
Rivière T, Hulin T, Sasangohar F, Zare M
Serious Games for Preventing Musculoskeletal Disorders in Occupational Settings: Scoping Review
JMIR Serious Games 2025;13:e66913
URL: https://games.jmir.org/2025/1/e66913
doi:10.2196/66913
PMID:

©Thomas Rivière, Thibaud Hulin, Farzan Sasangohar, Mohsen Zare. Originally published in JMIR Serious Games (https://games.jmir.org), 21.10.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Mobile Exergames to Reduce Sedentary Time in Patients With Heart Failure: What Do Health Care Professionals Expect?

Leonie Klompstra[1], PhD; Anna Strömberg[1], PhD; Antoni Bayes-Genis[2], PhD; Maria Boldo[2], MD; Beatriz González[2], RN; Nuria Santaularia[3], PhD; Tiny Jaarsma[1], PhD

[1]Department of Health, Medicine and Caring Sciences, Linköping University, Linköping, Sweden

[2]Rehabilitation Service and Heart Failure Unit, Hospital Universitari Germans Trias i Pujol, Badalona, Spain

[3]Althaia, Xarxa Assistencial Universitària de Manresa, Manresa, Spain

Corresponding Author:
Leonie Klompstra, PhD
Department of Health, Medicine and Caring Sciences, Linköping University, Linköping, Sweden

Abstract

Health care professionals emphasized that while exergaming holds promise as a motivational approach to supporting physical activity among patients with heart failure, its success depends on thoughtful integration into existing care pathways, personalization to patient needs, and addressing technological barriers.

Trial Registration: ClinicalTrials.gov NCT05641662; https://clinicaltrials.gov/study/NCT05641662

(JMIR Serious Games 2025;13:e69126) doi:10.2196/69126

KEYWORDS

exergaming; heart failure; implementation; sedentary; serious games

Introduction

Exergaming, which combines exercise and video gaming, has emerged as a promising intervention for decreasing older adults' sedentary time [1]. Sedentary behavior is a significant concern in patients with heart failure (HF). Regular physical activity (PA) improves quality of life and physical function and reduces hospitalizations [2]. However, despite these benefits, many patients struggle to engage in sufficient PA due to various barriers, including physical limitations, lack of motivation, and psychological factors [2,3].

Exergaming can encourage PA among patients with cardiovascular disease [4]. In a randomized controlled trial (RCT) [5], we tested a commercial exergame platform on a television (Nintendo Wii). Participants expressed a desire to engage in exergaming outdoors and have the exergame tailored to their PA levels [6]. Based on these insights, we developed a mobile phone exergame [7,8] tailored to patients' exercise capacity, and a new RCT [9] involving patients with HF is ongoing. Health care professionals' (HCPs') expectations regarding mobile exergaming are important, as they play a key role in recommending and supervising physical activities for patients with HF and for mobile exergame implementation in clinical practice. Given the lack of existing literature on HCPs' expectations, this study's purpose was to explore HCPs' expectations regarding mobile exergaming as an approach to decreasing sedentary time in patients with HF.

Methods

Ethical Considerations

Ethical approval was not required for this study, as no sensitive or personally identifiable data were collected, in accordance with institutional guidelines. Verbal informed consent was obtained from all participants. Interview data were anonymized during transcription, no compensation was provided, and all procedures complied with institutional and GDPR (General Data Protection Regulation) guidelines [10-12].

Study Design

A qualitative study with semistructured interviews was performed between September and October 2023 with 17 HCPs (9 cardiologists, 2 rehabilitation physicians, 5 nurses, and 1 physiotherapist) who were involved in the care of patients with HF at the cardiology departments of Althaia, Xarxa Assistencial Universitària de Manresa (Manresa, Spain) and Hospital Universitari Germans Trias i Pujol (Barcelona, Spain).

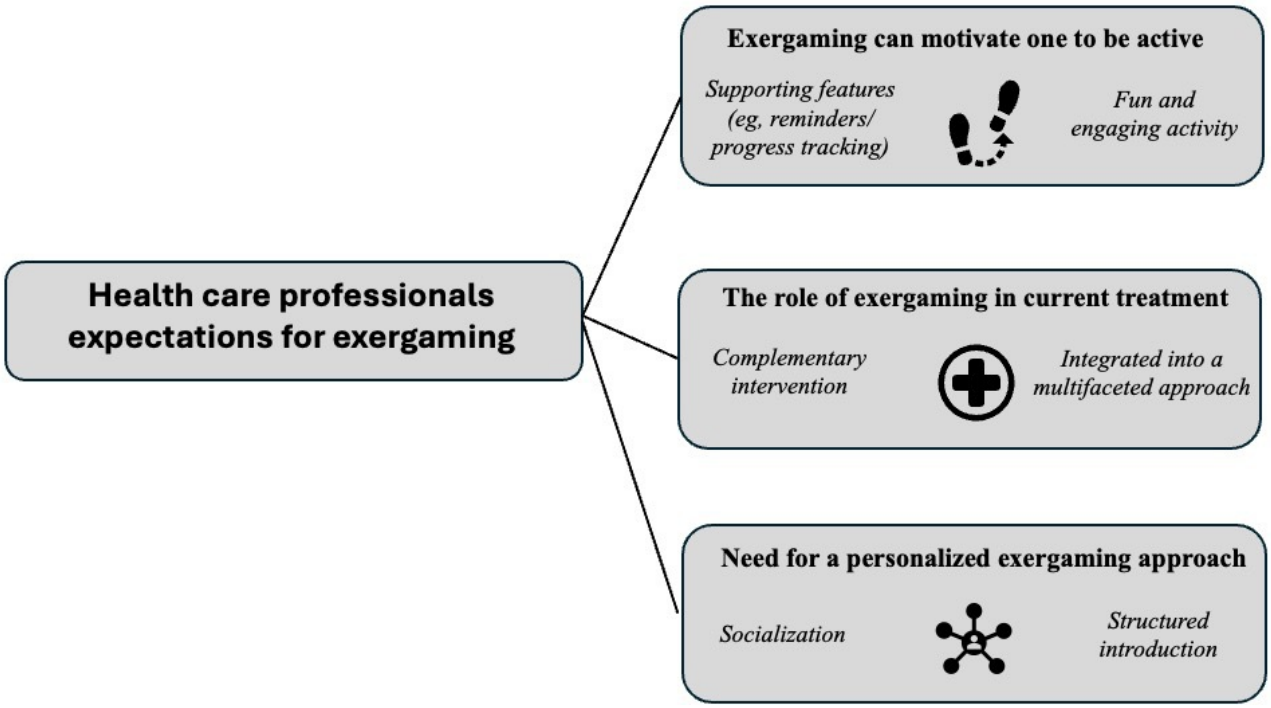
For semistructured interviews on HCPs' experiences with sex differences in exercise among patients with HF, HCPs were asked one additional question to explore HCPs' expectations regarding mobile exergaming as an approach to decreasing sedentary time in patients with HF (Multimedia Appendix 1). The interviews were conducted by a medical student (supervised by 2 experienced researchers [TJ and LK]), recorded, transcribed verbatim, and analyzed via thematic analysis [13].

XSL•FO
RenderX

Results

Three key themes on HCPs' expectations regarding mobile exergaming for patients with HF were found (Figure 1).

Figure 1. Themes (with strategies in italics) describing expectations that health care professionals had regarding exergaming as an approach to decreasing sedentary time in patients with heart failure.



Exergaming Can Motivate One to Be Active

One of the most expressed expectations was that exergaming may motivate patients with HF to engage in regular PA. The interactive and gamified nature of exergaming was seen as appealing, as it can transform PA from a chore into a fun and engaging activity. HCPs also highlighted mobile exergaming's potential to support regular PA through reminders and progress tracking. They believed that such features could help patients maintain a consistent PA routine, which would ultimately lead to reduced sedentary time.

Adding but Not Replacing

HCPs viewed exergaming as a complementary intervention rather than as a replacement for traditional PA advice or cardiac rehabilitation, acknowledging that while exergaming could offer benefits, it should not be seen as a stand-alone solution. Instead, it should be integrated into a broader, multifaceted approach to managing HF, potentially enhancing the effects of standard rehabilitation.

Need for a Personalized Exergaming Approach

HCPs expressed concerns about the technological challenges associated with exergaming, noting that some patients, particularly older adults, may lack the necessary skills or confidence to use exergaming technology effectively. This can be a barrier to exergaming's successful implementation. HCPs emphasized the importance of familiarizing patients with exergaming technology to ensure its successful adoption. Furthermore, socialization (eg, involving family members in the exergaming process) was seen as an important facilitator of engagement. Opinions diverged with regard to the influence of patients' sex on the ease of adopting exergaming. Some HCPs believed that men might find exergaming easier due to a perceived greater familiarity with video games or technology. Conversely, others suggested that women might have an advantage, particularly in using mobile phones and apps, which are often integral to mobile exergaming.

Discussion

Per HCPs, while exergaming has the potential to motivate patients to become more physically active, it should be used in conjunction with other interventions (eg, cardiac rehabilitation). This study also highlighted potential challenges related to technology use. HCPs expressed concerns about the digital divide, particularly for older patients who may be less familiar with exergaming technology, as recently described in a systematic review on older adults' experiences with exergaming [14]. There is a need for adequate familiarization and for tailoring exergaming experiences to individual patient characteristics.

Acknowledgments

We would like to thank Malin Olsson, for performing the interviews, and all the participants in this study.

XSL•FO
RenderX

Funding

This work is supported by The Swedish Research Council (2020 - 01109), the Swedish Heart and Lung Foundation (20200190 and 20240456), the Kamprad Foundation (20210074), and the Research Council in South East Sweden (FORSS-940933 and FORSS-995300).

Data Availability

Due to the nature of the data and study materials, data cannot be made freely available. However, by contacting the corresponding author, procedures for sharing data, analytic methods, and study materials for reproducing the results or replicating the procedure can be arranged, following Swedish legislation.

Authors' Contributions

Conceptualization: LK, TJ
Formal analysis: LK, TJ
Funding acquisition: TJ
Investigation: LK, TJ
Methodology: LK, TJ
Project administration: LK, TJ
Resources: ABG, MB, BG, NS
Supervision: LK, TJ
Visualization: LK
Writing – original draft: LK
Writing – review & editing: LK, AS, ABG, MB, BG, NS, TJ

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview guide.
[DOCX File, 15 KB - games_v13i1e69126_app1.docx]

References

1. Ismail NA, Hashim HA, Ahmad Yusof H. Physical activity and exergames among older adults: a scoping review. Games Health J 2022 Feb;11(1):1-17. [doi: 10.1089/g4h.2021.0104] [Medline: 34851732]
2. Pelliccia A, Sharma S, Gati S, et al. 2020 ESC Guidelines on sports cardiology and exercise in patients with cardiovascular disease. Eur Heart J 2020 Jan 1;42(1):17-96. [doi: 10.1093/eurheartj/ehaa605] [Medline: 32860412]
3. Resurrección DM, Moreno-Peral P, Gómez-Herranz M, et al. Factors associated with non-participation in and dropout from cardiac rehabilitation programmes: a systematic review of prospective cohort studies. Eur J Cardiovasc Nurs 2019 Jan;18(1):38-47. [doi: 10.1177/1474515118783157] [Medline: 29909641]
4. Blasco-Peris C, Fuertes-Kenneally L, Vetrovsky T, Sarabia JM, Climent-Paya V, Manresa-Rocamora A. Effects of exergaming in patients with cardiovascular disease compared to conventional cardiac rehabilitation: a systematic review and meta-analysis. Int J Environ Res Public Health 2022 Mar 15;19(6):3492. [doi: 10.3390/ijerph19063492] [Medline: 35329177]
5. Jaarsma T, Klompstra L, Ben Gal T, et al. Effects of exergaming on exercise capacity in patients with heart failure: results of an international multicentre randomized controlled trial. Eur J Heart Fail 2021 Jan;23(1):114-124. [doi: 10.1002/ejhf.1754] [Medline: 32167657]
6. Klompstra L, Jaarsma T, Mårtensson J, Strömberg A. Exergaming through the eyes of patients with heart failure: a qualitative content analysis study. Games Health J 2017 Jun;6(3):152-158. [doi: 10.1089/g4h.2016.0087] [Medline: 28422516]
7. Berglund A, Klompstra L, Orädd H, et al. The rationale behind the design decisions in an augmented reality mobile eHealth exergame to increase physical activity for inactive older people with heart failure. JMIR Serious Games 2024 Aug 21;12:e50066. [doi: 10.2196/50066] [Medline: 39185820]
8. Berglund A, Jaarsma T, Orädd H, et al. The application of a serious game framework to design and develop an exergame for patients with heart failure. JMIR Form Res 2024 Aug 7;8:e50063. [doi: 10.2196/50063] [Medline: 39110976]
9. Jaarsma T. Effects of exergaming to reduce sedentary time in inactive patients with heart failure (Heart-eXg). ClinicalTrials.gov. URL: https://clinicaltrials.gov/study/NCT05641662 [accessed 2025-08-06]
10. Comitè D'ètica De La Investigació Amb Medicaments [Webpage in Catalan]. Hospital Universitari Germans Trias I Pujol. URL: http://www.ceicgermanstrias.cat/ [accessed 2025-09-09]

XSL•FO
RenderX

11. Plataforma de serveis de suport [Webpage in Catalan]. La Unio – Associacio D'entitats Sanitaries I Socials. URL: https://www.uch.cat/fundacio-unio/serveis/plataforma-de-serveis-de-suport.html [accessed 2025-09-09]
12. Voigt P, Von dem Bussche A. The EU General Data Protection Regulation (GDPR): A Practical Guide, 1st edition: Springer International Publishing; 2017. [doi: 10.1007/978-3-319-57959-7]
13. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
14. Rytterström P, Strömberg A, Jaarsma T, Klompstra L. Exergaming to increase physical activity in older adults: feasibility and practical implications. Curr Heart Fail Rep 2024 Aug;21(4):439-459. [doi: 10.1007/s11897-024-00675-9] [Medline: 39023808]

Abbreviations:

GDPR : General Data Protection Regulation
HCP: health care professional
HF: heart failure
PA: physical activity
RCT: randomized controlled trial

Edited by A Coristine; submitted 22.11.24; peer-reviewed by O Chiala, R Hervás; revised version received 09.07.25; accepted 09.07.25; published 26.11.25.

Please cite as:
Klompstra L, Strömberg A, Bayes-Genis A, Boldo M, González B, Santaularia N, Jaarsma T
Mobile Exergames to Reduce Sedentary Time in Patients With Heart Failure: What Do Health Care Professionals Expect?
JMIR Serious Games 2025;13:e69126
URL: https://games.jmir.org/2025/1/e69126
doi:10.2196/69126

© Leonie Klompstra, Anna Strömberg, Antoni Bayes-Genis, Maria Boldo, Beatriz González, Nuria Santaularia, Tiny Jaarsma. Originally published in JMIR Serious Games (https://games.jmir.org), 26.11.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org

XSL•FO
RenderX